Te Whare Wānanga o Otāgo

NEW ZEALAND

Genome-Wide DNA Methylation in Polycystic Kidney Disease

Sarah Bowden

A thesis submitted in partial fulfilment of the requirements
for the Degree of Master of Science

University of Otago
Dunedin, New Zealand
April 2019

Abstract

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is a heritable renal disease that causes the enlargement of kidneys due to the bilateral development of fluid-filled cysts. This results in end-stage kidney disease in adults and a reduced life expectancy.

While it is known that a mutation within a PKD-causing gene is required for the development of ADPKD, the underlying mechanisms causing cystogenesis and allowing the progression of disease are not well understood.

As a result of this poor understanding there are few treatment options for patients with ADPKD, therefore a large proportion of patients will progress to end-stage renal disease for which they will need dialysis or renal transplantation.

Epigenetic modifications including DNA methylation are known to be altered in neoplasia, for which there are now several FDA-approved therapeutic drugs. As there are many similarities between ADPKD and neoplasia, we postulate that like tumour tissue, ADPKD tissue contains differentially methylated regions that may be exploited for future therapeutic discovery.

To investigate this, we have performed reduced representation bisulfite sequencing (RRBS) on four ADPKD kidney tissue samples, and three non-ADPKD kidney tissue samples. In this analysis we confirm that there are 13 regions in the genome with differential methylation, and there is a global trend of hypomethylation in ADPKD. Furthermore, the 3' end of the PKD-associated gene *PKD1* shows increased methylation associated with increased mRNA expression.

To investigate whether DNA methylation changes are universally changed in ADPKD cysts, we performed RRBS on a further eight ADPKD samples, each from unique cysts from a single ADPKD patient. In this analysis there were differential methylation patterns in each cyst, however these changes were not consistent between cysts.

These data show trends in global methylation in ADPKD not previously reported, and methylation changes within the genes *NAGLU* and *GET4* concomitant with gene expression, which require further investigation to identify their role in ADPKD.

Keywords: Autosomal Dominant Polycystic Kidney Disease, ADPKD, DNA Methylation, RRBS, miRNA

Abbreviations

ADPKD	Autosomal Dominant Polycystic Kidney Disease
ANOVA	Analysis of Variance
ARPKD	Autosomal Recessive Polycystic Kidney Disease
cAMP	Cyclic adenosine 3′,5′-monophosphate
cDNA	Complementary DNA
CKD	Chronic Kidney Disease
CpG	Cytosine-Guanine dinucleotide
Cq	Quantitation Cycle
DMEM	Dulbecco's Modified Eagle Medium
DMF	Differentially Methylated Fragment
DNMT	DNA Methyltransferase
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
ER	Endoplasmic Reticulum
ESRD	End Stage Renal Disease
EtOH	Ethanol

FDA	Food and Drug Administration
FDR	False Discovery Rate
FFPE	Formalin-Fixed Paraffin-Embedded
HAT	Histone Acetyltransferase
HDAC	Histone Deacetylase
LOH	Loss of Heterozygosity
LTA	Lipoteichoic acid
MCKD	Medullary Cystic Kidney Disease
MDCK	Madin-Darby Canine Kidney
miRNA	Micro RNA
mRNA	Messenger RNA
mTOR	Mammalian Target of Rapamycin
NBF	Neutral Buffered Formalin
NF-κB	Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells
NPHP	Nephronopthisis
OFG	Otago Genomics Facility
PC1	Polycystin 1
PC2	Polycystin 2

PCC	Pearson's Correlation Coefficient
PCR	Polymerase Chain Reaction
PKD	Polycystic Kidney Disease
PLD	Polycystic Liver Disease
qPCR	Quantitative Polymerase Chain Reaction
RC	R3277C
RRBS	Reduced Representation Bisulfite Sequencing
SV40	Simian Vacuolating Virus 40
TAE	Tris-Acetate-EDTA
TRPP2	Transient Receptor Protein 2
TSA	Trichostatin A
TSS	Transcription Start Site
VPA	Valproic Acid
Wnt	Wingless/Integrated
WT 9-7	Wild Type Recombinant Adeno-SV40 Virus 9-7 Cell Line
WT 9-12	Wild Type Recombinant Adeno-SV40 Virus 9-12 Cell Line

Acknowledgements

First and foremost, I would like to acknowledge and thank my supervisors Prof Mike Eccles, Dr Cherie Stayner and Dr Aniruddha Chatterjee for their guidance through this project. As I discovered during this process, scientific research is unpredictable and has lots of ups and downs, and I am grateful for your guidance through the many challenges I faced.

I would also like to thank the Eccles lab group and Department of Pathology for their assistance with protocols and analysis, particularly Jackie Ludgate for her expertise and aid with the library preparation, and Dr Euan Rodger for his time assisting with the data analysis. I also must acknowledge my predecessor Michael Bates, for generating the preliminary data in this project and helping me as I started out as a baby scientist in the lab.

Thank you also to my friends and family, for putting up with years of rants about PCR optimisation and DNA methylation. You've all encouraged my weird brand of nerd, and I blame you all for this. Most importantly, my parents deserve credit for getting me to this point, and continually supporting me along the way. Thank you for everything you do!

Table of Contents

Table of Tables

Table of Figures

Chapter 1: Introduction

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is the most common heritable renal disease in humans. Characterised by the development of large, fluid-filled cysts throughout both kidneys, this disease has an estimated prevalence of 3 in 10,000 [1]. As cysts continue to develop and grow throughout a patient's lifetime, the kidney will ultimately enlarge from less than 200 g in a healthy person to upwards of 1.5 kg in some ADPKD patients. The amount of cystic growth (both by number and volume) varies between patients, from patients who develop only a few cysts and have adequate renal function in late adulthood [2], to patients with cysts visible as early as *in utero,* and who experience severe and rapidly progressive disease [3].

As a consequence of the growth of these cysts, renal function is impaired. It is assumed that either the destruction of the renal parenchyma [4], the fibrosis of interstitial kidney tissue [5], the disruption of the renal architecture [6, 7], or a combination of any of these precludes the regular function of nephrons [8]. Patients present with symptoms that reflect the decline of renal function, including flank pain, hypertension and various urinary complications. Interestingly, ADPKD patients experience several cardiac-related symptoms, and patients often develop hypertension ten years earlier than the average person (even when they exhibit adequate renal function). Common extrarenal symptoms include cysts in other organs such as the liver, pancreas and intestines. The prevalence of intracranial aneurysms is also higher in ADPKD patients than the general population [9, 10]. Many of these extrarenal symptoms are caused by connective tissue defects [11].

Associated with significant morbidity, the progression of ADPKD will ultimately inhibit renal function to the point of end-stage renal disease (ESRD), at which point patients may require dialysis or renal transplantation. An estimated 50% of ADPKD patients will reach ESRD by age 60 [12, 13]; in New Zealand ADPKD is the fourth most common reason for dialysis [14]. Despite the significant burden ADPKD can play on the medical system, there is currently no cure for the disease, and patients are treated with a limited number of therapeutics that show moderate improvement in some patients. One of the currently available therapeutic options is the vasopressin V2-receptor antagonist tolvaptan. Tolvaptan has been approved for the treatment of ADPKD in several countries, as it attenuates the rate of cyst growth by half, delaying the onset of ESRD. However, this drug has undesirable side effects including resulting in abnormal liver function tests, which results in poor patient tolerance [15]. Additionally, the efficacy of tolvaptan has been deemed insufficient for approval from the US Food and Drug Administration (FDA) [16]. As a result of these factors, ADPKD patients have a reduced life expectancy than that of the general population [12, 17].

1.1 The genetic background of ADPKD

ADPKD is genetically heterogeneous – a germline mutation within a PKD-causing gene is a prerequisite for patients to develop this disease. *PKD1* is the most prevalent disease-causing gene for ADPKD, with approximately 85% of patients having mutations within this gene. It was previously assumed *PKD2* accounted for the remaining 15% of patients [18]; however, the sequencing of patients with atypical ADPKD has led to the discovery of patients with mutations in genes including *GANAB* [19] and *DNAJB11* [20], which are now considered to be PKD-causing genes for a small proportion of patients. Approximately 10% of patients have no family history of ADPKD, and 25% of these have proven *de novo* mutations [21].

There is no mutation 'hot-spot' within these genes; over 1,800 mutations have been identified as contributing to the pathogenicity in ADPKD. These are highly variable, spread across each gene, and are typically unique within families. A database of ADPKD mutations within *PKD1* and *PKD2* has been established (pkdb.mayo.edu) [22], with frameshift mutations accounting for the largest percentage of predicted pathogenic mutations recorded to date. The pattern of inheritance in ADPKD follows that of autosomal dominance, as most patients are heterozygotes for the pathogenic mutation they carry.

Patients with a mutation within *PKD1* are classed as type I ADPKD patients, while patients with a mutation within *PKD2* are type II ADPKD patients. *PKD1* mutations affect males and females equally, with a median onset of ESRD (or death) at 53 years. *PKD2* mutations affect males and females differently, however, as males have an earlier median onset ESRD than females (67 and 71 years, respectively) [18]. There is insufficient data on the epidemiology of other pathogenic genes in ADPKD due to their recent characterisation and low prevalence.

Features of the *PKD1* gene (16p13.3) make it likely to be more susceptible to somatic mutation. It has a high GC content [23], which makes it more vulnerable to slippage errors during replication [24], and the presence of a long polypyrimidine tract within intron 21, which may predispose the gene to mutation [25]. *PKD1* spans almost 50 kb of the genome, but it has an open reading frame (ORF) of 14 kb, with most of this being untranslated (Fig. 1.1) [26].

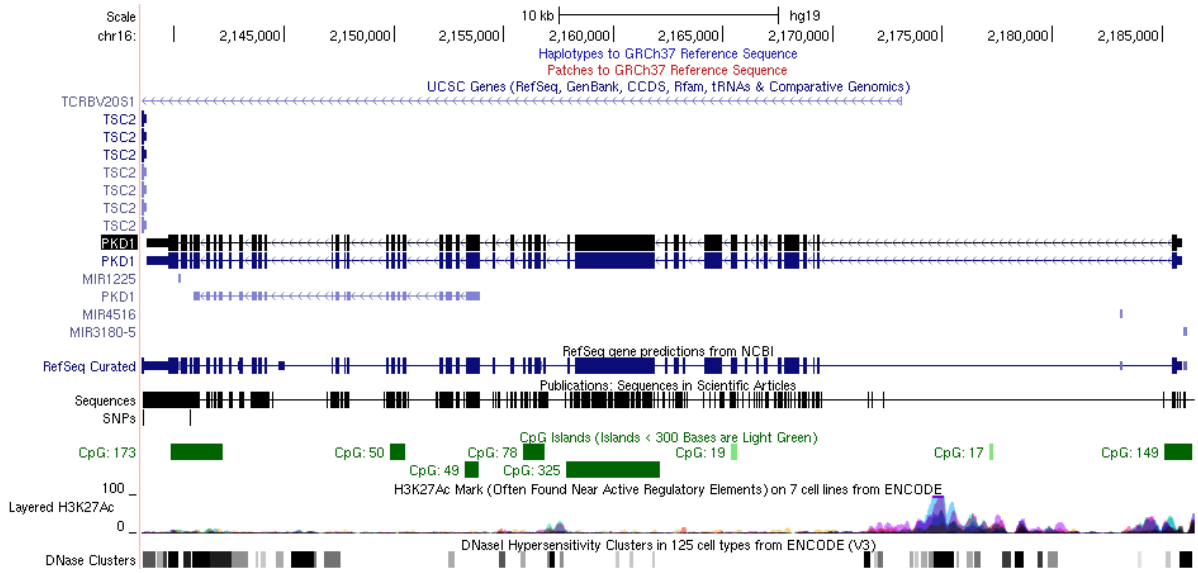


Figure 1.1: Features of the ADPKD-associated gene *PKD1*
PKD1 is 50 kb in length and runs in the 3' 5' direction. Regulatory elements within the gene include multiple CpG islands, an area of H3K27Ac within intron 1, DNase clusters associated with the CpG islands and H3K27Ac region. There are also three miRNAs within the gene and promoter: miR-1225, miR-4516 and miR-3180-5. The 3' end of TSC2 is directly adjacent to the 3' end of PKD1. Image generated on the UCSC genome browser with the GRCh37/hg19 genome annotation. http:/genome.ucsc.edu [27, 28].

Alternative splicing of *Pkd1* is observed in mouse and rat models [29, 30], however analysis of PKD1 is complicated due to the multiple pseudogenes, so it is unclear whether alternate splicing is present or plays a role in PKD1 expression. There are also six pseudogenes ~13 Mb upstream of PKD1 which share a 97-99% homology with sequences from exons 1-33 [31, 32]. These may indicate a propensity for this region of the genome to experience somatic mutation; in cancer somatically acquired pseudogenes mirror features seen in retrotransposons [33].

Conversely, *PKD2* (4q21) is a single copy gene. It spans 68 kb of the genome, with an ORF ~5 kb [34]. Multiple alternative splicing transcripts have been identified in *PKD2*, including some which are not present in polysomal fractions, indicating that perhaps they play a role in the expression of other *PKD2* transcripts [35].

For the purposes of this research, patients with assumed or confirmed type I ADPKD are investigated, and thus the discussion into ADPKD predominantly features type I ADPKD characteristics.

1.2 *PKD* genes and Polycystins

PKD1 encodes the protein polycystin-1 (PC1), a 462 kD integral membrane protein with both an extracellular and an intracellular terminus (Fig. 1.2). The extracellular N-terminus, which makes up approximately 75% of the protein, contains various domains capable of binding to a range of molecules including collagen, fibronectin, proteins and lipids. Of particular interest are the cell wall integrity and stress response component homology domain (which may bind to carbohydrates) and a C-type lectin domain for which carbohydrate binding is calcium-dependent [31]. The intracellular C-terminus also interacts with lipids and proteins, although the function of this terminus appears to play a role in signalling, as this tail contains a G-protein-activation site and phosphorylation signalling sites. Overall, PC1 has a structure indicative of a receptor or adhesion molecule.

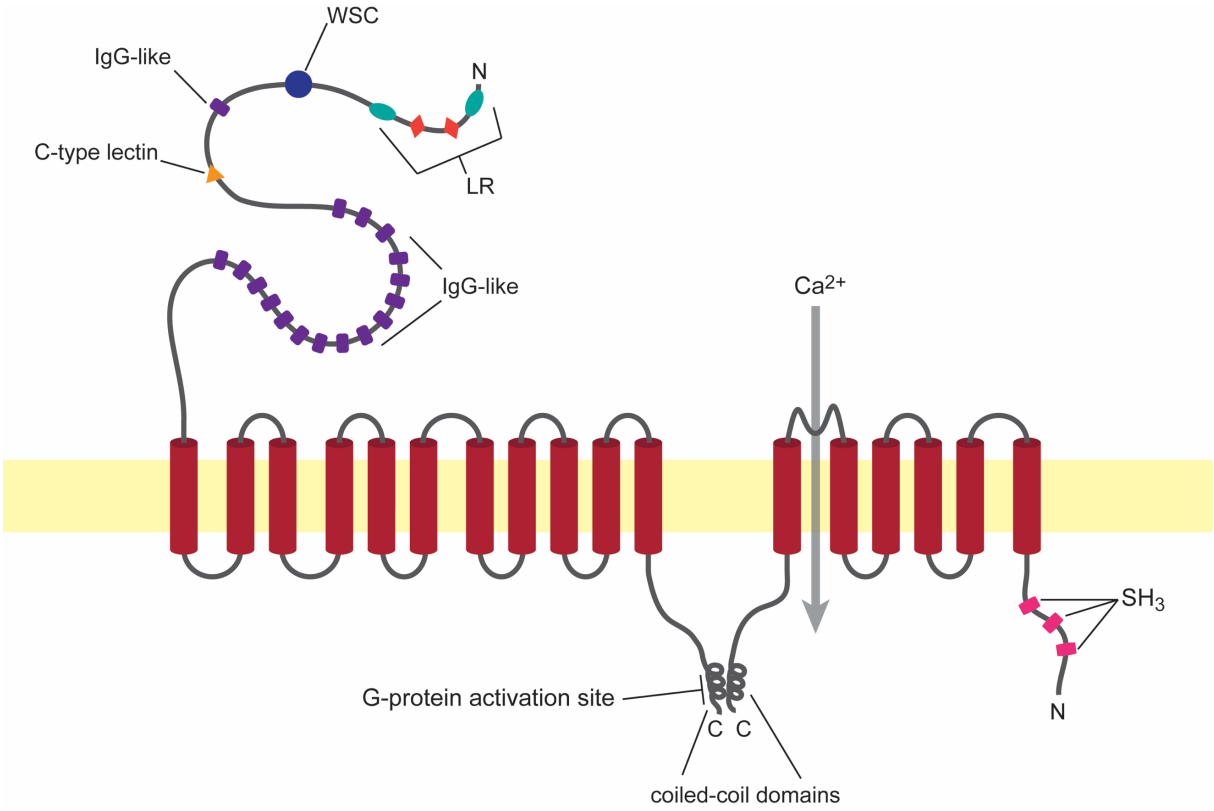


Figure 1.2: Structure of Polycystin 1 and Polycystin 2
Polycystin 1 (left) is a 462 kD protein, predominantly comprised of an extracellular N terminus. This terminus contains a cytosine-flanked leucine repeat sequence (LR), which is capable of binding collagen, fibronectin and laminin; a cell wall integrity and stress response component domain (WSC); PKD1-like IgG-like repeats (IgG-like) which can bind protein ligands; and a C-type lectin which binds carbohydrates in a calcium-dependent manner. The intracellular C terminus contains a G-protein activation site, and a coiled-coil domain which has been postulated to assist with binding to the coiled-coil domain of Polycystin 2. PC2 (right) is a smaller protein of 110 kD. The transmembrane domains share homology with the transmembrane domains of PC1, and the extracellular loop between domains 5 and 6 forms a pore (which allows calcium to flow into the cell). The intracellular N terminus of PC2 contains SH_3 domains which allow the binding of focal adhesion proteins such as tensin. Adapted from Wilson (2004) [5].

Polycystin-2 (PC2) is a 110 kD protein (Fig. 1.2). The transmembrane domain of this protein shares a homologous region with a section of the transmembrane domain within PC1, however the proteins have different functions. The six transmembrane domains of PC2 form a pore, which acts as a nonselective cation channel capable of transporting Ca^{2+} [36]. PC2 belongs to a subfamily of transient receptor potential channels.

It is believed that PC1 and PC2 form a polycystin complex together, which may have a mechanotransduction role in which the receptor molecules of PC1 influence the capacity of PC2 to transport Ca^{2+} [37].

1.3 Molecular characteristics of renal cysts in ADPKD

ADPKD cysts demonstrate variable thickening of the basement membranes as well as alterations in the matrix composition [5, 12, 38]. Epithelial cells derived from ADPKD are more adherent to type I and IV collagen [39, 40]. The relationship between the cells and the matrix may also be influenced by abnormal numbers of integrin receptors. In mice, it has been demonstrated that the over-expression of matrix and focal adhesion complex-associated proteins such as tensin results in the growth of a high number of cysts [41, 42].

A key process during development is regulating the polarity of cells through the location of specific proteins in order for correct renal function to occur [43]. Tissue from ADPKD has altered polarity, including the relocation of the Na^+/K^+-ATPase pump from the basolateral to the apical cell membrane [44]. This suggests the ability to regulate polarity is lost (or never activated) in cystic tissue. Additionally, the isoform present in the cystic cell is the fetal form $\alpha1\beta2$ rather than the mature isoform $\alpha1\beta1$, suggesting that the fetal isoform is never downregulated in mature cells. Furthermore, polarity of *PKD1* is also altered in ADPKD. While typically characterised as a membrane protein, *PKD1* has been found primarily in the cytoplasm of ADPKD epithelial cells [45].

The ADPKD epithelia respond abnormally to signalling molecules. In the ADPKD kidney, apoptosis is abnormally persistent, destroying the renal parenchyma, which allows cystic

epithelia to proliferate. In mice it has been demonstrated that inactivation of the apoptosis inhibitor *bcl-2* is sufficient to cause renal cysts.

Conversely, the proliferation of renal epithelia – which should cease at adulthood – continues in ADPKD. Epithelial cells originating from the ADPKD kidney have been proven to have an increased capacity for proliferation. ADPKD is also more susceptible to stimulation from the growth factor EGF, which is another cause of cystic proliferation. This can be demonstrated in transgenic mice overexpressing the receptor *erb-b2*, which develop focal dilation and proliferation of tubular epithelia [46].

EGF is secreted into the cystic lumen in quantities than can initiate proliferation. Combined with the abnormal location of EGF receptors in the luminal membrane, this mechanism contributes to the growth of cells. The inhibition of these receptors in mice reduces the number of cysts. Other growth factors, cytokines, lipid factors, ATP and cAMP all are present within the cyst fluid and contribute to the proliferative capabilities of the epithelia.

The ADPKD-causing genes *PKD1* and *PKD2* have many purported roles in ADPKD. One such role involves a signalling cascade which leads to cell proliferation, migration and differentiation by repressing transcription factors such as TCF/LEF, AP-1 and STAT1. These genes control other pathways such as *Wnt* signalling.

1.3.1 Characteristics of PKD1 in ADPKD

Although it is known that a germline mutation to either *PKD1* or *PKD2* is typically required for ADPKD, the direct relationship these genes have with the onset of disease is not clear. It may be expected that the germline mutation would result in the loss of function at this allele

(either due to truncation or misaligned protein folding), thereby causing a loss of polycystin dosage, or function, sufficient to cause cystogenesis.

Loss of polycystin dosage has been demonstrated in mice, where the knockout or reduction of *Pkd1* is sufficient to cause cystic disease, as well as extrarenal manifestations typical of APDKD (Table 1.1).

Table 1.1: Evidence demonstrates either an increase or decrease of *PKD1* and Polycystin 1 causes ADPKD
Mouse model section adapted from Happé and Peters (2014) [67].

Mouse models: Reduction of *Pkd1*	Knockout mice Conditional knockout mice Inducible knockout mice Hypomorphic mice	[47-51] [52-54] [49, 55] [56-59]
Mouse models: Increased expression of (human) *PKD1*		[60, 61]
Human tissue: Variable *PKD1* expression in human tissue		[62]
Human tissue: Increased expression of *PKD1*		[6, 63-66]

However, the loss of dosage model is contrary to observational evidence in human ADPKD tissue. Various analyses of *PKD1* expression in tissue derived from ADPKD cysts has shown evidence that *PKD1* expression is increased in cystic tissue (Table 1.1). These analyses are not conclusive evidence that ADPKD is caused by the upregulation of *PKD1,* as a significant proportion of cyst (10-20%) are deemed to have an absence of the protein in some studies [6, 63]. However, they do demonstrate that the relationship between *PKD1* and ADPKD is not as simple as loss-of-function animal models previously described.

1.4 Cystogenesis in ADPKD

Cysts initially arise from the proliferation of epithelial cells within a nephron tubule. The accepted model of cyst development is based upon focal proliferation (Fig. 1.3), where a single epithelial cell undergoes some form of genetic event, altering the function of the cell and initiating abnormal proliferation [4]. The resulting mass of cells then fills with fluid and will bud off from the nephron of origin to become an independent cyst. The monoclonal nature of ADPKD cysts has been demonstrated [68], supporting the concept that the origin of a cyst is likely from one epithelial cell.

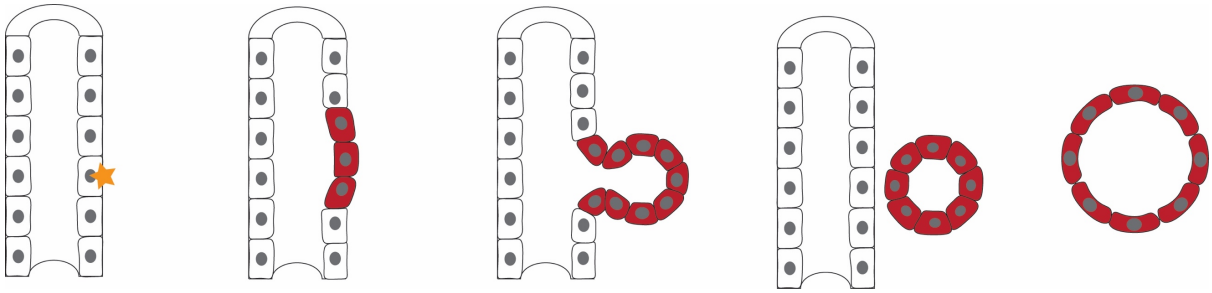


Figure 1.3: Focal proliferation as a model of cyst development
Focal proliferation is widely agreed upon as the mechanism by which cysts develop. In this model, a genetic event generates cystic cells (red), which proliferate and expand out of the renal tubule, until they bud off and expand as a single, independent cyst.

Cysts continue to grow, likely as a result of fluid intake and the continual cellular proliferation. Cyst growth as measured by total kidney volume is variable, and is calculated to increase at a rate of 1-10% per year [69, 70].

Although each kidney contains roughly one million nephrons [71] which all carry the germline PKD-causing mutation, less than 1% of these will develop into a cyst [72]. In this respect, ADPKD is dominant at the genetic level, but recessive at the cellular level. Therefore, it is not the germline mutation alone causing cystogenesis, but one or more additional molecular features initiating cystogenesis. The feature(s) would then need to occur hundreds of times in

each PKD kidney in order for the characteristic cystic kidney to develop. A number of hypotheses have been suggested to explain the initiation of cystogenesis in ADPKD.

1.4.1 Loss of heterozygosity

In 1992 it was first proposed that ADPKD was caused by the loss of the second 'wild-type' copy of *PKD1,* comparable to the repression of tumour suppressor genes in heritable cancers [73]. This mechanism is dubbed the "two-hit hypothesis", and there is experimental evidence to demonstrate that this does occur in some cysts from ADPKD patients. Multiple studies have identified loss of heterozygosity (LOH) in *PKD1,* in approximately 12-24% of selected ADPKD cysts, based upon microsatellite analysis. [68, 74, 75]. The true rate of LOH in ADPKD may be higher due to single nucleotide polymorphisms rather than the specific translocations tested in these studies, although it is unlikely to account for the development of all cysts.

In the LOH model, the rate at which patients acquire a second hit plays a role in determining the course of disease. This would also explain the phenotypic variability between different patients. However, for this model to account for all cyst formation, there must be extremely high rates of somatic mutation within *PKD1* and *PKD2*. In addition, while LOH is frequently cited as a theory to explain the mechanism underlying cystogenesis initiation, there are some experimental observations (such as increased gene expression, see Section 1.3.1) which LOH cannot sufficiently explain.

1.4.2 Incomplete penetrance

Papers published in the late 2000s report the screening of *PKD1* and *PKD2* in consanguineous ADPKD families that do not fit the normal paradigm of single dominantly inherited mutations in ADPKD.

A notable *PKD1* mutation R3277C (RC) was found and determined to be the pathogenic mutation in a family. The RC allele appears to show a "low potency" where heterozygous family members show minimal disease and develop only a few renal cysts. This is unlike typical *PKD1* mutations, where heterozygotes develop significant ADPKD. However, the homozygous family members with two copies of RC presented with severe ADPKD that progressed to ESRD and required transplantation [76]. In another family, the *in utero* onset of polycystic kidney disease was found to be the accumulation of the RC allele *in trans* with another, more severely pathogenic and completely penetrant *PKD1* mutation [77]. Therefore, the RC allele could also be described as incompletely penetrant, able to induce severe ADPKD in combination with another pathogenic mutation.

1.4.3 Gene dosage and networks

As the dosage of PC1 has been found to be both increased and decreased in ADPKD tissue and animal models (Table 1.1), it has been questioned whether the dosage itself influences cystogenesis, or whether the dysregulation of this molecule has downstream consequences in gene networks.

An early example of this is *PAX2*, which is known to play a role in kidney cell differentiation [78]. Transgenic ADPKD compound mutant mice carrying, in an addition to a homozygous *Pkd1* mutation, a heterozygous *Pax2* mutation, demonstrate less severe cyst formation [79],

inferring that perhaps reduction of *Pax2* dosage may be equivalent to inhibition of factors involved in the signalling pathways (such as cMet, NF-κB or Wnt) that form a regulatory cascade to activate *Pax2* expression in kidney tubules.

It is also been demonstrated that the protein kinase mTOR (mechanistic Target of Rapamycin) plays a role in ADPKD. mTOR plays a role in the coordination of cellular growth and metabolism, and is regulated by PC1 [80]. As a target of Rapamycin, this has been an investigated route for therapeutic treatment of ADPKD. While Rapamycin has lowered total kidney volume, this molecule, as a treatment for ADPKD, has been unsuccessful due to poor patient tolerance to side effects [16].

A recent study has demonstrated that the oncogene *Myc* is upregulated both in the absence and gain of polycystin-1 in transgenic mice [81]. Furthermore, the expression of PC1 was modulated by the overexpression of *Myc*, suggesting that there is a regulatory loop mechanism influencing pathogenesis.

1.5 Polycystic Kidney Disease is a ciliopathy

Cysts from ADPKD arise from the epithelial cells of the nephron. Renal epithelial cells have a primary (immotile) cilium protruding on the apical membrane, which can act as a mechanosensor to detect fluid flow within the lumen, in turn triggering an influx of intracellular Ca^{2+} (which has a role in cell signalling) [82].

ADPKD is considered a ciliopathy – a disorder of the cilia – due to evidence linking dysfunctional cilia with cystic kidney disease. In mice, inactivation of either the gene *Kif3a* or Polaris prevents the synthesis of cilia, subsequently causing polycystic kidney disease [83]. The

absence of cilia is not a feature in *Pkd1* mutant mice nor ADPKD patients, so inactivation of *Kif3a* is not a requirement for the disease but it does infer that the impairment of cilia is crucial in ADPKD.

PC1 and PC2 can both localise to the cilium [37, 84], where together they are thought to function as sensors of the extracellular environment (the renal tubular lumen) and initiate signalling cascades. Cultured renal epithelial cells from *Pkd1*– mutant embryonic mice do not respond to fluid flow sheer stress, while their wild-type counterparts do by increasing the intracellular Ca^{2+} concentration [85]. It is postulated that the large extracellular portion of PC1 acts as a sensor, which results in a conformational change to activate PC2.

Additionally, when wild-type renal epithelial cells were treated with an inhibitory antibody designed to target the intracellular portion of PC2, the cells continue to respond to fluid flow by increasing the concentration of intracellular Ca^{2+}. However, when the same cells are treated with an antibody targeting the extracellular portion of PC2, the cells no longer respond to fluid flow [37].

In addition to ADPKD, there are various forms of renal ciliopathies, including Autosomal Recessive Polycystic Kidney Disease (ARPKD). This disease includes subtypes such as Nephronophthisis (NPHP) and Medullary Cystic Kidney Disease (MCKD) (Table 1.2). ARPKD is an inherited, recessive disease caused by the gene *PKHD1*. Recessive PKD is much less common than the dominant form, and often results in fetal or neonatal death [86]. It is generally accepted that ARPKD cysts arise from the collecting duct of the nephron, which expand but stay attached to their respective tubule [5]. In ADPKD it is believed the cysts bud off from the nephron, and while there is some experimental evidence to suggest a high

proportion of ADPKD cysts contain collecting duct markers, ADPKD cysts are believed to arise from all segments of the nephron as well as the collecting ducts [87].

Table 1.2: A summary of the major renal ciliopathies

Disease	Genetic cause	Age when affected/diagnosed	Cyst location
ADPKD	*PKD1* *PKD2* *GANAB* *DNAJB11*	Primarily adults	Arise from all segments of the nephron and bud off into the parenchyma
ARPKD	*PKHD1*	Neonates	Derived from and remain connected to the collecting duct
	NPHP genes	Infants - Adolescents	Corticomedullary border
	MCKD1 *MCKD2*	Primarily adults	Within the medulla

NPHP, which is the primary cause of renal failure in children, is the result of recessive inheritance of one of a large number of identified *NPHP* genes, which affect adolescents, juveniles and infants. In this disease, cysts develop at the corticomedullary border of the kidney [88]. There are also multiple forms of MCKD, consequential of the dominant inheritance of *MCKD1* or *MCKD2* [89, 90]. In MCKD, cysts form exclusively within the medulla and result in the bilateral shrinkage of the kidneys. MCKD is typically milder than NPHP, and primarily affects adults.

The development of hepatic cysts can occur in ADPKD patients, which is usually indicative of co-morbid Polycystic Liver Disease (PLD). PLD can occur as an independent disease or alongside ADPKD or ARPKD, however hepatic cysts typically do not affect liver function. PLD alone is associated with the genes *PRKCSH* and *SEC63* [91, 92].

1.6 Similarities between ADPKD and neoplasia

ADPKD demonstrates similarities to a number of features associated with cancer, which is why it has been described as "neoplasia in disguise" [93]. Features of ADPKD align with eight of the ten hallmarks of cancer (Table 1.3) [94].

Table 1.3: Comparison of ADPKD and the hallmarks of cancer
Adapted from Seegar-Nukpezah et al. (2015) [95].

Hallmarks of cancer	ADPKD
Genomic instability and mutation	LOH has been proposed to initiate cystogenesis [68, 74, 75]; emerging data that epigenetic changes contribute to cystogenesis [96]. ADPKD lymphocytes demonstrate high levels of DNA damage [97].
Sustaining proliferative signalling	Elevation of epithelial proliferation in ADPKD cysts due to growth factors/cytokines/cAMP in cystic fluid.
Evading growth suppressors	Tumour suppressor signalling (such as p53) is dysregulated [98, 99].
Resisting cell death	Low levels of apoptosis in cysts (elevated in parenchyma) [100].
Tumour-promoting inflammation	Interstitial inflammation is observed in ADPKD [101]. This is thought to contribute to disease progression.
Inducing angiogenesis	Abnormal capillaries around renal cysts [102, 103].
Deregulating cellular energetics	High rate of aerobic glycolysis in ADPKD [104].
Avoiding immune destruction	Macrophages are thought to promote cyst growth through C3 induction [105].
Enabling replicative immortality	Replicative immortality does not occur in ADPKD – replicative crisis is delayed in *Pkd1*$^{-/-}$ mouse embryo fibroblasts [98].
Activating invasion and metastasis	Metastasis does not occur in ADKD.

Most notably, both diseases can be characterised by genomic instability and mutation. In cancer, this is either due to heritable phenotypes or the "chance acquisition" of a mutant genotype causing the subclonal population of cells to proliferate over neighbouring cells [94]. Emerging data now recognises that epigenetic modification regulating gene expression is also an aspect of instability in the genome. In ADPKD, the heritable phenotype originates within a

PKD-causing gene (such as *PKD1*). The chance acquisition of mutant genotypes is the LOH proposed to occur in more than 10% of clonal cyst populations.

Defects in the DNA-maintenance mechanisms within the genome have been documented in neoplasia [106], limiting the cell's ability to repair or destroy misfolded proteins and to prevent further genomic damage. Interestingly, one of the genes recently attributed to PKD pathogenesis, *GANAB*, is involved in the quality control complex of the endoplasmic reticulum.

Another similarity between neoplasia and ADPKD is that epithelial cells originating from renal cysts have sustained proliferative signalling, wherein molecules such as EGF, cAMP and cytokines contribute to cyst growth and are speculated to drive cystogenesis in an oncogenic fashion. The deregulation of apoptosis is also observed in both diseases, although in cancer this process is limited; in ADPKD apoptosis is increased in the parenchyma and marginally reduced in tubule epithelia.

One key hallmark of cancer that ADPKD does not possess is the invasion and metastasis of malignant cells throughout other organs of the body. While cysts are observed in other organs (such as within the liver and intestines), these do not proliferate and cause organ failure in the same manner that the primary cysts within the kidney do, and there is no evidence to suggest they originate from migration of renal cysts.

There are two fundamental characteristics of cystic cells in ADPKD which are not considered to be in the hallmarks of cancer: altered ciliary signalling, and the dysregulation of intracellular Ca^{2+} and cAMP.

Available treatments for ADPKD are limited due to both the incomplete knowledge of the disease, and little financial incentive for pharmaceutical companies to develop ADPKD-specific products [95]. Additionally, a number of therapies have failed due to insufficient patient responses during clinical trials or poor patient tolerance to side effects. The similarities between ADPKD and cancer means that there is the possibility of therapeutics currently developed for use in cancers to be repurposed for use in ADPKD, some of which have already been the subject of clinical trials in ADPKD [16].

1.7 Epigenetic and post-translational influences in ADPKD

In addition to germline or somatic mutations, gene and protein expression can be easily altered by epigenetic mechanisms. These are molecular changes that occur to the DNA molecule without changing the genetic sequence and are easily altered to change gene expression. Epigenetic factors are important for many reasons. During development, epigenetic mechanisms play a role in cellular growth and differentiation, and as cells mature, these epigenetic factors change to suit the role of a cell. Once the cell differentiates, epigenetic modifications are specific for every cell type.

Two key forms of epigenetic regulation involve histone modification (such as acetylation and methylation) and DNA methylation. These mechanisms act to modify the chromatin state and accessibility of DNA to transcription factors. A similar biological mechanism is the post-translational regulation of genes by microRNA. The most significant mechanisms of epigenetic control of gene expression in ADPKD are explored below.

1.7.1 microRNA

microRNA are a class of non-coding RNAs, approximately 20 nucleotides long. These sequences function as post-translational regulators to alter protein expression by cleaving mRNA (as part of a silencing complex containing Dicer). It has been estimated that one third of protein coding genes may be regulated by miRNA [107]; in cancer it has been documented that a large number of genes are modified by specific miRNA. Some notable pathways regulated by miRNA include the anti-apoptosis gene *BCL2* and the oncogene *MYC* [108].

miRNAs are transcribed from the genome as primary miRNA, which are then spliced with Drosha and Dicer enzymes to become mature miRNAs (Fig. 1.4). Mature miRNAs contain a "seed site", which is a sequence estimated to be from six to ten nucleotides long that has the ability to target complementary sequences in the genome. Due to the frequency of the complementary sequence in the genome, each miRNA has the ability to transcriptionally regulate a multitude of genes [109, 110].

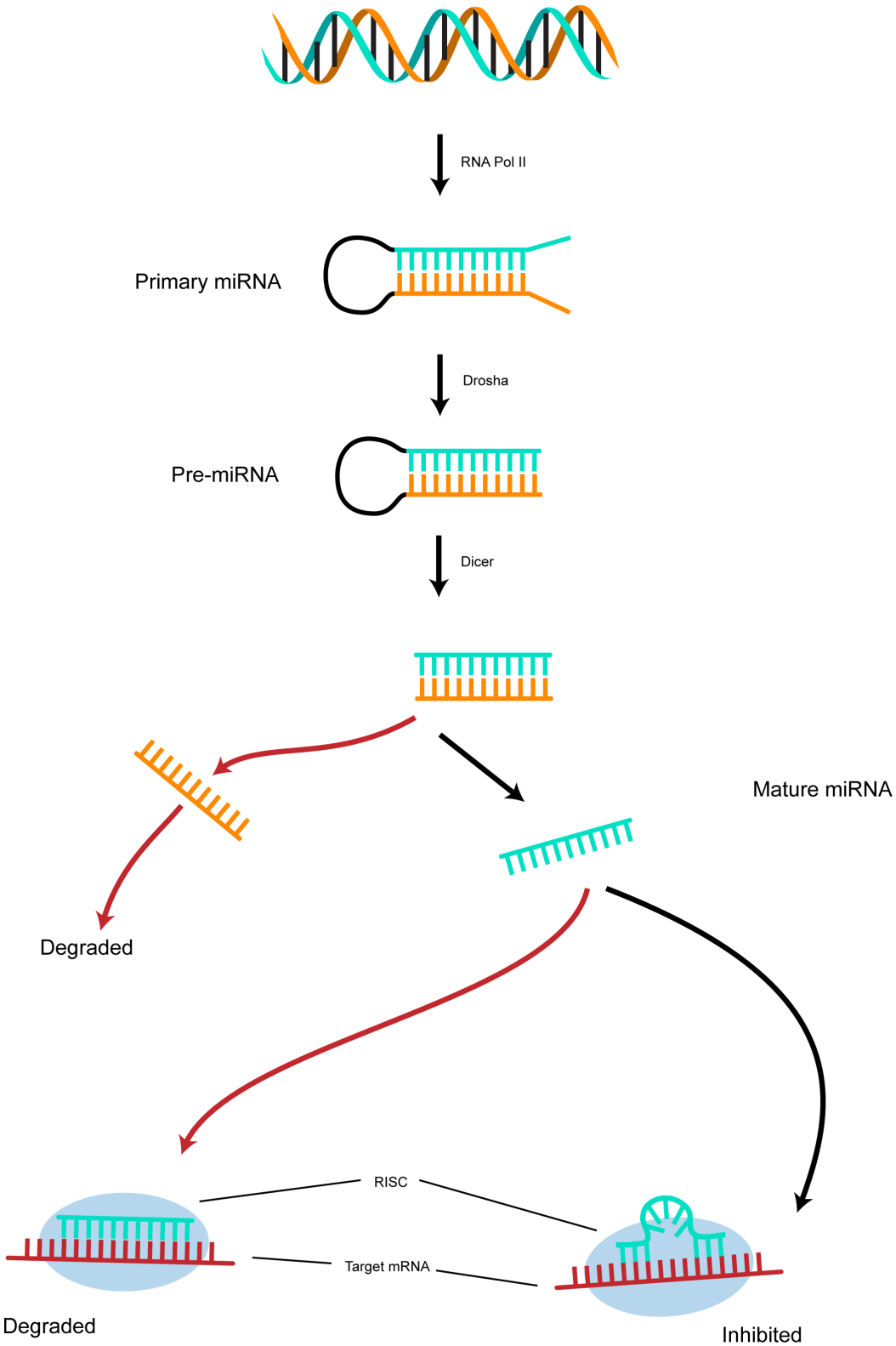


Figure 1.4: Maturation and mechanism of miRNA
miRNAs are transcribed from the genome with RNA Pol II. The primary miRNA is then modified by Drosha enzymes to become pre-miRNA. Pre-miRNA are further modified by Dicer. The mature miRNA can then form a complex with a RISC (RNA-induced silencing complex) to bind to a complementary sequence on a mRNA molecule where it can attenuate expression through two mechanisms: repression or degradation.

1.7.2 miRNA in ADPKD

Previous research in renal development has confirmed that miRNAs are required for embryonic development of the kidney, and several miRNAs have been implicated in glomerular and tubular diseases [111]. Screening of miRNA in tissue and plasma samples has identified a wide range of miRNA as being dysregulated in ADPKD. Current findings of miRNA dysregulation in ADPKD include the upregulation of miR-182-5p [112] and 199a-5p [113], as well as the downregulation of miR-192 and miR-194 due to hypermethylation [114]. It has also been demonstrated that miR-501-5p is upregulated in ADPKD, causing activation of the mTOR pathway through p53-mediated mechanisms [115].

Most recently, inhibition of the miR-17~92 cluster has been demonstrated to reduce cyst proliferation in ADPKD mice [116]. Further analysis of the six miRNAs within this cluster (miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1, miR-92a-1) provides evidence that modulation of miR-17 alone is sufficient to attenuate the growth of cysts [117]. Additionally, the anti-miR-17 treatment in this study inhibited mTOR signalling. As mTOR has previously been assessed as a target for therapies in ADPKD (see Section 1.4.3), inhibition of miRNA suggests an additional target for treatment of the disease.

1.7.3 Histone Modification

DNA is wound around histone octamers, which allow for the tight packaging and organisation of the DNA molecule in the nucleus. The histone proteins contain many tails, to which a large range of chemical modifications can occur. These modifications determine how closely the histones are packaged together, in turn facilitating how freely DNA can be transcribed. There are three main modifications to the histone tails: phosphorylation, methylation and acetylation [118].

Acetylated histones result in relaxed chromatin (euchromatin) which is associated with greater levels of gene transcription, while deacetylation results in heterochromatin, which is tightly packed and has reduced levels of gene transcription (Fig. 1.5).

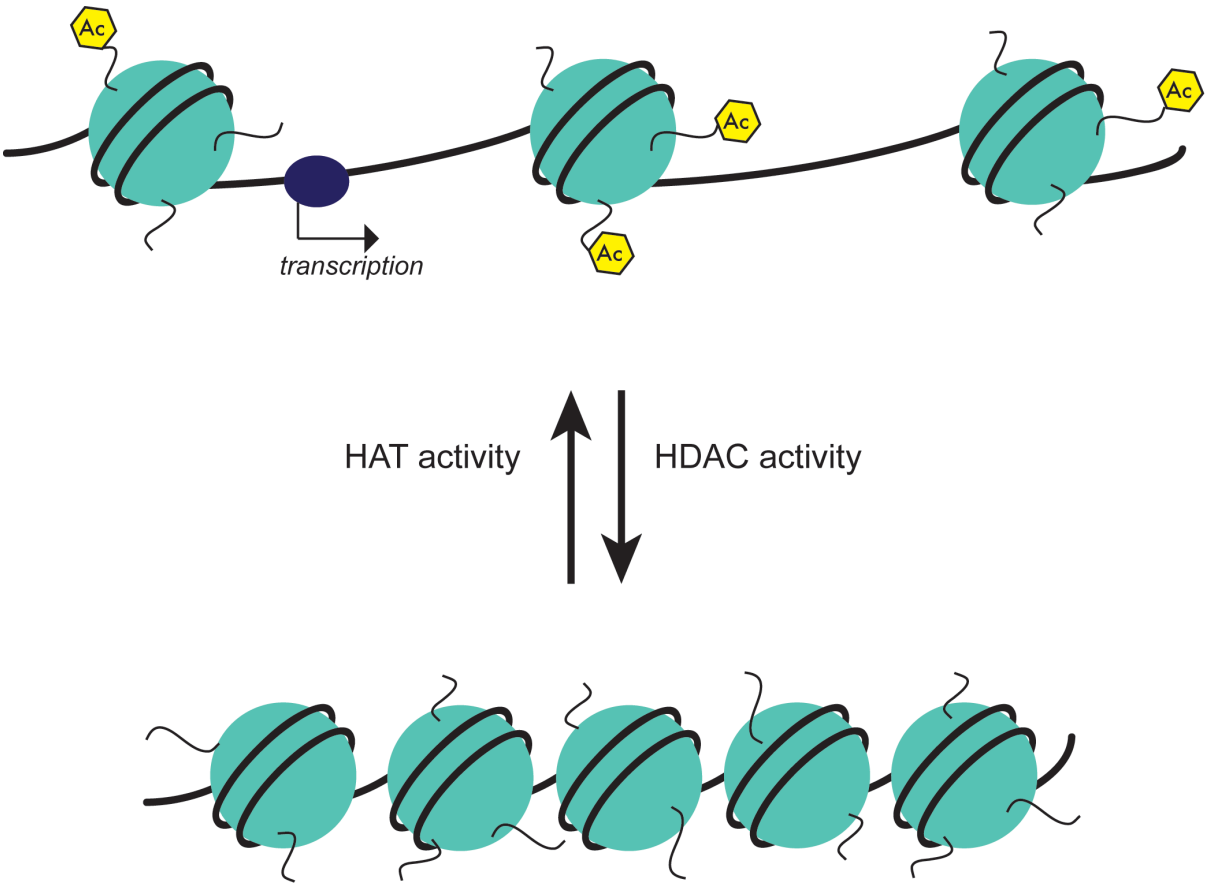


Figure 1.5: Histone acetylation regulates gene transcription
The modification of lysine residues attached to the histone octamers that DNA is wound around determines how tightly packaged the chromatin is. Euchromatin (upper) is associated with greater levels of gene transcription than heterochromatin (lower) which prevents transcriptional mechanisms from accessing the DNA. Chromatin packaging is modified by the attachment of acetyl group (Ac) to the lysine residues with histone acetyltransferases (HATs), or dissociation of acetyl group from the lysine residue with histone deacetylases (HDACs).

Histone acetylation is catalysed with histone acetyltransferases (HATs), while deacetylation requires histone deacetylases (HDACs). There are several classes of HATs and HDACs, based upon homologous sites and the target of their epigenetic affects [119].

1.7.4 Histone deacetylases in ADPKD

While multiple mechanisms of histone modification are able to occur, there is little evidence of histone phosphorylation and methylation in the context of ADPKD. The regulation of histones with HDACs, however, has been well documented [96]. Some of the studies implicating HDAC activity in ADKD include p53-induced repression of the *PKD1* promoter which has been shown to be attenuated with HDAC treatment [120], targeting of HDAC5 expression through polycystin-dependent fluid stress signalling [121], and HDAC6 inhibition preventing the release of Ca^{2+} from the endoplasmic reticulum, consequentially attenuating cyst growth through cAMP signalling [122].

Numerous studies have identified the activity of HDACs as a driver of neoplasia, due to the aberrant expression of HDACs in tumours [123], and the ability to use post-translational modification of histones with HDACs and HATs as biomarkers in human tumours [124]. As such, there are now several HDAC inhibitors that have been approved by the Food and Drug Administration (FDA) for the use in human cancer [125]. The class I HDAC inhibitor VPA, and class I and II HDAC inhibitor TSA have both been shown to reduced cyst formation and slow cyst growth in animal models [99, 120, 121, 126], yet these drugs are yet to show sufficient beneficial effects in human clinical trials and be approved by the FDA.

1.7.5 DNA methylation

DNA has the potential for nucleotides to be modified by the addition of a methyl group via DNA methyltransferase (DNMT) proteins. This typically occurs on cytosine residues which are immediately followed by guanine (CpGs), although DNA methylation can also occur at other dinucleotide combinations [127]. Mammalian DNA is methylated at 70-80% of all CpG sites

in the genome [128], as DNA methylation is required for the controlled suppression and expression of genes (Fig. 1.6).

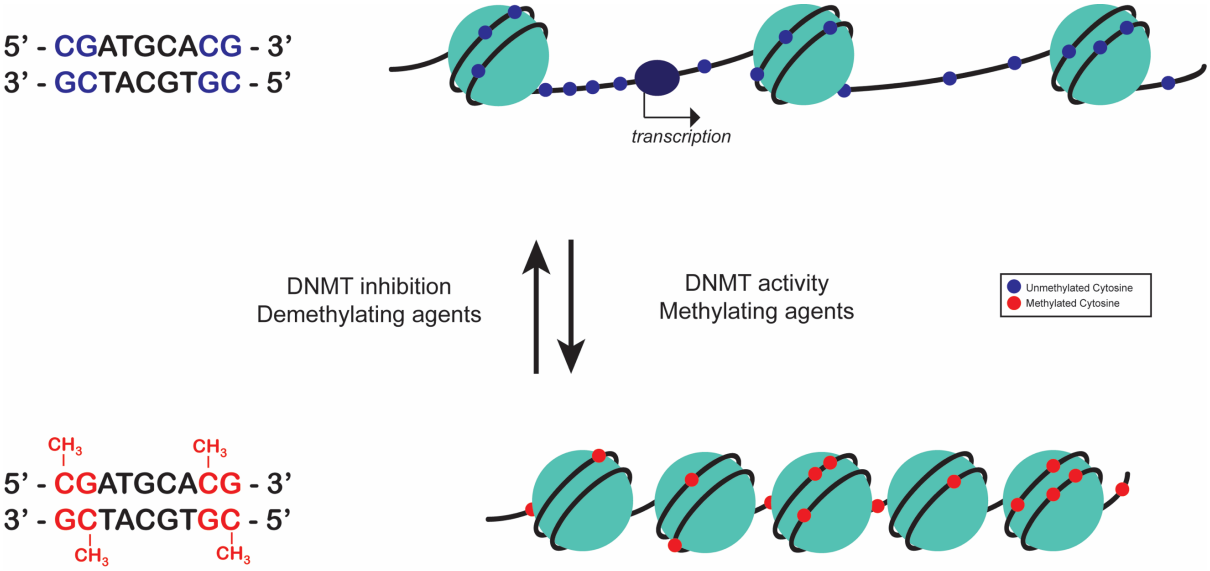


Figure 1.6: DNA methylation as an epigenetic mechanism
The addition of a methyl group to a CpG dinucleotide is created by DNMTs (or methylating agents). Unmethylated DNA can be facilitated by the inhibition of DNMTs, or demethylating agents. Hypomethylation of promoters and other enhancer elements allow the transcription of genes, while hypermethylation of promoters prevents the DNA machinery from accessing these regions to enable transcription. The inverse is seen in gene bodies, as hypermethylation at the 3' end of a gene is often correlated with increased gene expression, although an explanatory mechanism for this is not clear.

Clusters of CpGs are known as CpG islands, and these are often found in areas of high importance for gene expression, such as the promoter or enhancer regions of the genome [129]. Methylation at the promoter regions of the genome have historically been associated with the repression of a gene, by reducing the ability of transcription factors to bind [130]. One mechanism of transcription factor inhibition includes chromatin remodeling, suggesting that there is epigenetic cross-talk between DNA methylation and histone modification in the epigenetic control of genes [131].

Methylation of CpGs within the gene body is not as well understood as that of the promoter, but typically results in sustained or increased expression of a gene [132, 133]. A variety of mechanisms have been proposed to explain this phenomenon. These all postulate that the methylated regions contain genomic elements, either responsible for alternative splicing, containing transcription factors which when hypomethylated interfere with the host gene expression, or as residual epigenetic marks from embryonic stages of development [132].

DNA methylation is required at all stages of life, from differentiation of embryonic stem cells during development [134], to the maintenance of X inactivation [135] and specialization of cells [136]. The pathology of several diseases, most notably cancers, has been linked to dysregulation of DNA methylation both globally and locally [137]. As such, DNMT inhibitors have been developed as a therapeutic drug for certain cancers [138].

1.7.6 DNA methylation in ADPKD

A 2014 paper by Woo et al. [139] was the first to report global DNA methylation in ADPKD patients. This was conducted with pyrosequencing on the kidney tissue from three patients with ADPKD (and three from non-ADPKD tissue). The results from this analysis identified over 13,000 unique fragments in the genome which were differentially methylated; 91% of these were found to be hypermethylated. The exonic region of the genome was found to be the major target of DNA methylation changes in ADPKD, with 5.93-fold higher hypermethylation occurring at these regions.

This group also analysed DNA methylation in the *PKD1* gene body (exon 43), and demonstrated hypermethylation in ADPKD samples, negatively correlating with the expression of the gene. Demethylation of an immortalised cell line (WT 9-12) resulted in increased *PKD1*

mRNA, and treatment of the cyst-forming cell line MDCK with a DNMT inhibitor repressed the growth of cysts.

Inhibition of DNMTs has been developed as a therapeutic for various neoplasia [137]. It is theorized that if the hypermethylation of ADPKD was resulting in cystogenesis, the same pharmaceuticals could be used for the treatment and management of ADPKD patients.

1.8 Rationale for study

Preliminary data from a previous student [140] showed global hypomethylation in an analysis of ADPKD tissue from two patients. This was in contrast to the previously reported study which identified hypermethylation as a contributing factor to ADPKD [139]. As there are so few data currently available on DNA methylation in ADPKD, my first aim in this study was to analyse further ADPKD tissues, to add to the previous research performed in our laboratory on the methylation of tissue from ADPKD kidneys.

All previous work on the ADPKD methylome has been performed using whole tissue samples, which contain multiple ADPKD cysts. As cysts are thought to arise independently, one of my specific aims in this study was to identify whether changes seen in ADPKD whole tissue are representative of the ADPKD methylome, or whether cysts each contain their own unique methylation pattern.

1.9 Hypothesis

I hypothesise that there will be significant differences in the methylome of ADPKD tissue and non-ADPKD kidney. I further hypothesise that there will be multiple differentially methylated loci in the ADPKD tissue that may have a role in the pathogenesis of ADPKD.

1.10 Aims

- To use reduced representation bisulfite sequencing (RRBS) to identify key regions of differential methylation in a larger panel of ADPKD kidney tissue.
- To use RRBS to show the (representative) methylome of ADPKD cysts and determine whether the changes seen in whole tissue are seen consistently in individual cysts.
- To investigate the three miRNAs associated with CpG islands in the *PKD1* gene and determine if aberrant expression of these miRNAs is involved in ADPKD pathogenesis.

Chapter 2: Methods

2.1 Sources of nucleic acids for genetic analysis

DNA and RNA isolated for this research came from kidney tissue from ADPKD patients and healthy individuals, as well as ADPKD-derived cell lines. One patient has a confirmed germline *PKD1* mutation. The remining four patients are assumed to have *PKD1* mutations, based upon the age of patients at time of nephrectomy, however these patients have not been genotyped. Data about each sample such as age and sex can be found in Appendix A. The use of human tissue samples was approved by the Otago Human Ethics Committee H15/H110.

2.1.1 Kidney tissue dissection

Tissue segments were dissected from the kidneys of ADPKD patients immediately following renal transplantation at Dunedin Hospital. Multiple segments were cut from the cortex of the kidney (although this region is not clearly defined in ADPKD due to the extent of cyst growth, thus the location was approximated), in which each tissue section contained tissue from multiple cysts as well as blood cells, connective tissue, fibrous tissue and any remaining parenchyma. Tissue samples were cut from multiple regions in the kidney by grasping a segment of tissue with forceps and cutting it out with surgical scissors. Samples were immediately frozen in liquid nitrogen for storage at -80 ºC.

An additional kidney in this analysis was treated with neutral buffered formalin (NBF) for seven days prior to dissection. The selection of tissue segments was carried out in the same manner as above but were stored in 70% EtOH for 24 hours before embedding in a paraffin block (rather than freezing at -80 ºC).

Segments of non-ADPKD tissue were provided by the Christchurch Tissue Bank. These samples were collected from the cortex of the kidney during renal tumour surgery – dissected as far from the tumour as possible, and with histological testing to confirm the absence of malignant tissue – which were collected fresh and frozen at -80 ºC when delivered to our laboratory. The non-ADPKD tissue was used as the control group in this study, as by the time ADPKD patients undergo nephrectomy, the disease has developed to an extent where the kidneys contain no unaffected tissue that could be used as a direct comparison against cystic tissue.

2.1.2 Isolation of cyst wall tissue from an ADPKD kidney

In addition to segments of tissue, sections of cyst wall were also dissected from an ADPKD kidney. These samples were intended to only contain tissue from a single cyst wall, thereby providing a more precise picture of the methylation landscape in cystic tissue. These samples were excised from an ADPKD kidney that had been treated in NBF for 24 hours and rinsed with 70% EtOH to wash off any sources of contaminating DNA (such as blood cells) from the cyst wall. These samples were submerged in EtOH and stored at 4 ºC.

2.1.3 Immortalised cell line growth

The two cell lines used in this analysis were renal epithelial cells WT 9-7 (ATCC® CRL-2830™) and WT 9-12 (ATCC® CRL-2833™), each derived from an independent cyst of a female ADPKD patient. The cell lines have been characterised as having originated from a proximal tubule based upon CD13 antibody staining and LTA lectin binding (although the WT 9-12 cell line demonstrates distal and proximal characteristics). Mutation screening has shown that these

cells both carry the truncating *PKD1* mutation Q2556X – WT 9-7 is heterozygous for this mutation while WT 9-12 is homozygous [141].

Cells were grown on flasks coated with type I collagen (0.05 mg/mL) in DMEM media (Table 2.1). The cells were grown at 37 ºC with 5% CO_2 for ten days, during which time they went through one passage. Cells were lifted with Trypsin-EDTA (0.25%) and spun down at 300 g for five minutes. The supernatant was aspirated, and the pellets were stored at -20 ºC until required for DNA and RNA extraction.

Table 2.1: DMEM media for cell culture

Dulbecco's Modified Eagle Medium (DMEM)	89%
Penicillin-Streptomycin (10,000 U/mL)	1%
Fetal Calf Serum	10%

2.1.4 DNA isolation from cells and tissue

DNA was extracted from the cell line pellets using a QIAmp DNA Mini kit (Qiagen, cat. no. 51304). All elutions followed the manufacturer's instruction.

Approximately 25 mg of tissue was cut from the frozen kidney segments over dry ice with a clean razor blade. The ADPKD samples required further maceration with the blade to assist with the homogenization steps, as they are fibrous and difficult to breakdown in the following steps. The samples were then transferred to a 2 mL microcentrifuge tube to grind with a clean pestle, then homogenized in a QIAShredder column (Qiagen, cat. no. 79656) before continuing with the QIAmp DNA Mini kit protocol.

DNA was isolated from cyst wall samples, which had been treated with NBF, with the QIAmp DNA FFPE kit (Qiagen, cat. no. 56404), omitting the paraffin-melting steps. Incubation at 57ºC

was increased from one hour to overnight (approximately 16 hours) to ensure maximum lysing, and the ATE incubation was extended to five minutes to increase the final yield of DNA.

2.1.5 RNA isolation from cells and kidney tissue

RNA was extracted from the cell line pellets and fresh-frozen kidney tissue segments using a miRNeasy Mini kit (Qiagen, cat. no. 217004). As above, 25 mg of the tissue samples were cut over dry ice and macerated with a clean razor blade. The miRNeasy protocol was followed as per the manufacturer's instruction to isolate total RNA.

The segment of ADPKD kidney in paraffin was sliced into four sections 10 µm thick, and RNA was extracted using the miRNeasy FFPE kit (Qiagen, cat. no. 217504). Deparaffinisation was performed with xylene, and the manufacturer's instructions were followed after this point.

2.2 Creation of reduced representation bisulfite sequencing libraries

The methylome of samples in this research was generated using a method known as reduced representation bisulfite sequencing (RRBS), wherein a fraction of the genome enriched with CpG sites is amplified and sequenced to provide an overall picture of the methylation across the whole genome.

This protocol is comprised of six key steps (summarised in Fig. 2.1). It is based upon published protocols [142], although it has been modified to allow a lower quantity of input DNA. Key processes of the protocol are detailed in the following sections. Prior to the development of this protocol, five whole tissue samples were generated from 2.5 µg of DNA by a previous student [140], and the cell line libraries were generated from 2.5 µg of DNA with the assistance of Jackie Ludgate.

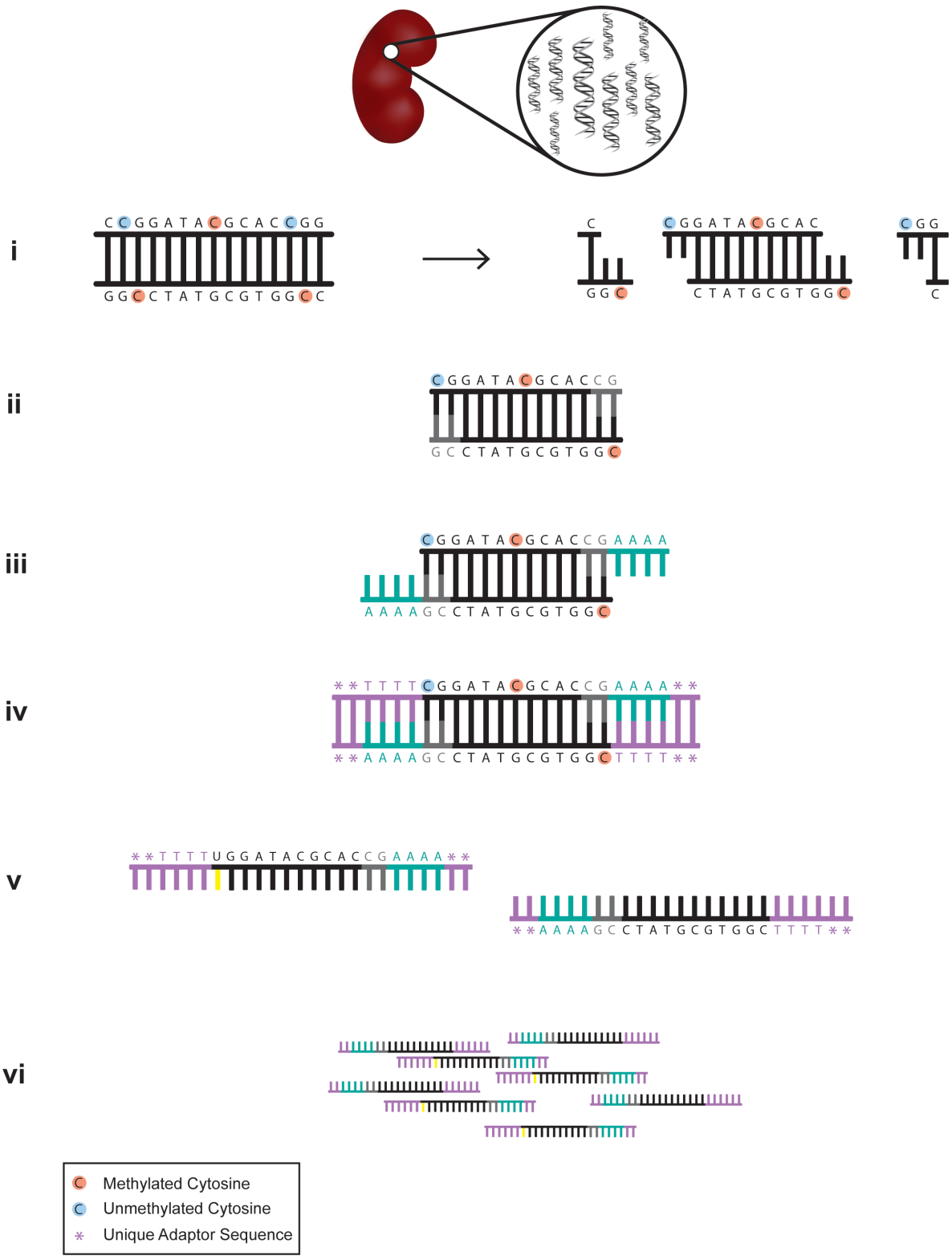


Figure 2.1: Summary of the RRBS protocol
RRBS libraries are generated from DNA isolated from kidney tissue. **(i)** The DNA is fragmented at CpG sites using an MspI digest. **(ii)** Fragments are then modified to repair sticky ends (grey), **(iii)** an A-tail (green) is added to the 3' end of each fragment, **(iv)** and a sequencing adaptor (purple) ligated to the A-tail. **(v)** The library is then bisulfite converted, creating single stranded fragments with uracil (yellow) in place of methylated cytosines. **(vi)** The library is then amplified and is ready to be sequenced.

2.2.1 MspI digest

An MspI restriction enzyme digest was set up for each sample (Table 2.2) and incubated at 37 ºC overnight. Low-bind tubes were used for this and all subsequent steps performed in the thermal cycler to minimise the loss of DNA due to sample-tube binding.

Table 2.2: MspI restriction enzyme digest

DNA sample	~0.5 µg
NEB MspI (20 U/µL)	8.0 µL
10x NEB Buffer 4	3.0 µL
MQ H_2O	To 30.0 µL
Total	**30.0 µL**

Once digested, libraries were purified with the QIAQuick PCR Purification kit (Qiagen, cat. no. 28104) and eluted into 60.0 µL.

2.2.2 Fragment repair

Following the digest and purification, samples were incubated with End Repair Mix 2 according to Table 2.3 at 30 ºC for 30 minutes.

Table 2.3: Fragment repair for one RRBS library

MspI-digested DNA sample	60.0 µL
End Repair Mix 2	40.0 µL
Total	**100.0 µL**

This sample was then purified with the MinElute PCR Purification kit (Qiagen, cat. no. 28006) and eluted into 17.5 µL.

A 30A overhang was formed at the 3’ end of each fragment using A-Tailing Mix (Table 2.4). Libraries were incubated at 37 ºC for 30 minutes, 70 ºC for five minutes and 4 ºC for five minutes.

Table 2.4: A-Tailing

DNA sample	17.5 µL
A-Tailing Mix	12.5 µL
Total	**30.0 µL**

To the A-Tailed sample, adaptors with a 30T overhang were bound to the fragments in each library, set up as in Table 2.5 in a preheated thermal cycler at 30 ºC for ten minutes. Unique adaptor sequences were selected for each library, based on the published sequencing guidelines for multiplexing [143], which allows for multiple libraries to be pooled together and sequenced across the same lanes.

Table 2.5: Adaptor ligation

DNA sample	60.0 µL
RSB Buffer	2.5 µL
DNA Adaptor	2.5 µL
Ligation Mix 2	2.5 µL
Total	**37.5 µL**

Immediately following ligation, 5 µL of Stop Ligase Mix was added to each sample, and a final MinElute PCR Purification was carried out.

2.2.3 Bisulfite conversion

To identify the methylation status of the DNA, samples underwent bisulfite conversion. Although two different kits were used (EZ DNA Methylation-Direct, Zymo, cat. no. D5020 and MethylCode Bisulfite Conversion, Invitrogen, cat. no. MECOV50), the reagents and protocols used are the same.

The protocol for each kit was followed, each time using freshly prepared CT reagent, as this is only stable for a short period of time, and subject to degradation when exposed to UV light. Once converted, the remaining library preparation steps were carried out within 24 hours, as single-stranded bisulfite converted libraries are not as stable as double stranded genomic DNA.

2.2.4 Semi-quantitative PCR

A PCR reaction was set up for each library (Table 2.6) and split in two tubes. Each sample underwent thermal cycling at 15 and 20 cycles (Table 2.7).

Table 2.6: PCR master mix

10x *PfuTurbo Cx* Reaction Buffer	2.5 µL
2.5mM dNTP Stock (3.3 mM final)	3.0 µL
TruSeq PCR Primer Cocktail	3.0 µL
DNA Sample	3.0 µL
PfuTurbo Cx HotStart DNA Polymerase	1.2 µL
MQ H_2O	12.3 µL
Total	**25.0µL**

Table 2.7: Thermal cycling conditions for library amplification
With the semi-quantitative analysis, $n = 15$ or $n = 20$ cycles; in large scale amplification, n = the number determined by semi-quantitative PCR analysis.

	95 ºC	2 minutes	
Denaturation	95 ºC	30 seconds	
Annealing	65 ºC	30 seconds	x *n* cycles
Extension	72 ºC	45 seconds	
	72 ºC	7 minutes	
	4 ºC	hold	

A 3% (w/v) NuSieve gel was prepared by mixing NuSieve™ GTG™ agarose (Lonza, cat. no. 50081) in a glass bottle with freshly prepared 0.5 X TAE buffer, heating in small increments in a microwave at 50% power until the powder is fully dissolved. The gel was then poured into a tray with a 12-toothed comb and left to set for an hour. Once set it was placed in the electrophoresis tank, submerged but not covered in freshly prepared 0.5 X TAE buffer. 7 µL of 25 bp DNA ladder (Invitrogen, cat. no. 10597011) was loaded into the first well, and PCR products were loaded in pairs with 3 µL Xylene Cyanol (XC) loading dye. The gel was run at 50 V for 5 minutes, then paused as additional 0.5 X TAE buffer was added to the tank to submerge the gel and the gel was run for a final 85 minutes. This method of dry-loading the

NuSieve™ gel prevents PCR products from rising out of the well (which can happen due to remaining EtOH in the sample).

Once the gel was run for a total of 90 minutes, the gel was visualised under UV in the BioRad GelDoc. The resulting image (see Appendix B) was used to judge the number of cycles required to successfully amplify the remaining library, by comparing the brightness of bands between 150 and 325 bp. These decisions were supported with advice from senior researchers in the laboratory who have experience with this protocol.

2.2.5 Large scale amplification

PCR master mix was prepared (Table 2.8) and amplified under cycling conditions determined by the semi-quantitative PCR. For each library, the DNA and master mix solution was pipetted across an 8-tube strip of 0.2 mL microcentrifuge tubes in volumes of 12 µL, replicating the surface area of the semi-quantitative PCR. Once amplified, the PCR products were pooled from the 8-tube strip into a 2 mL microcentrifuge tube and purified with a MinElute PCR Purification kit.

Table 2.8: Large scale PCR master mix

10x *PfuTurbo Cx* Reaction Buffer	10.0 µL
2.5mM dNTP Stock (3.3 mM final)	12.0 µL
TruSeq PCR Primer Cocktail	12.0 µL
DNA Sample	12.0 µL
PfuTurbo Cx HotStart DNA Polymerase	4.8 µL
MQ H_2O	49.2 µL
Total	**100.0 µL**

2.2.6 Size selection

In order to isolate the RRBS library from the adaptors and any DNA fragments remaining in the sample, libraries were run through another 3% (w/v) NuSieve gel as prepared above. Fresh 0.5 X TAE buffer was added to the gel electrophoresis tank to reach but not cover the gel. 7 µL

of 25 bp DNA ladder was loaded into the first well, and libraries (18 µL of DNA and 2 µL of XC loading dye) were loaded into every second well, with a maximum of four libraries on any gel. The gel was run at 50V for a few minutes, and once the samples were seen to move into the gel, and tank was paused and 0.5 X TAE buffer was added to fully cover the gel. The gel was then run at 50 V for a total of 90 minutes.

Once this was completed, the gel was carefully removed from the tank and placed on a clean, sterile surface (a T75 cell culture flask was used). The first lane containing the ladder was cut from the gel with a clean razor blade and transferred to the GelDoc to visualise and measure the DNA ladder without exposing the RRBS libraries to the UV lamp. Using a ruler laid alongside the lane, the distance to the 150 and 325 bp bands was measured (Fig. 2.2).

Lanes containing the library were cut from the gel with a clean razor blade. Perpendicular cuts were made at the 150 bp and 325 bp points, using the ruler laid alongside as a reference. Each cut was made with a new razor blade, to eliminate the likelihood of contamination from the larger DNA fragments and smaller adaptor dimers on the blade surface.

The section of the gel containing the library was then placed into a 2 mL LoBind microcentrifuge tube (Eppendorf, cat. no. 022431048) to perform a MinElute Gel Extraction (Qiagen, cat. no. 28604); sections weighing over 400 mg were split into two tubes. The remaining agarose was pieced together and imaged under UV in the GelDoc to confirm the process had correctly removed the section containing DNA library (Appendix B).

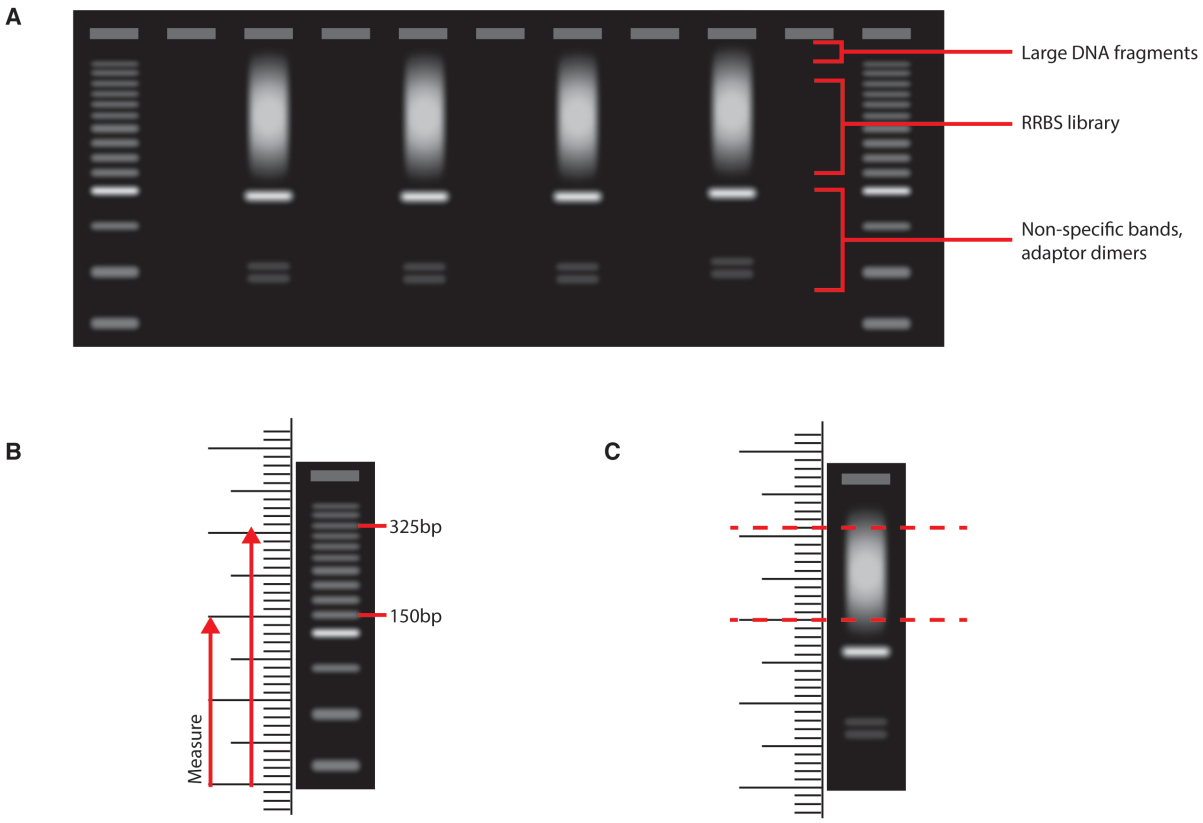


Figure 2.2: Illustration of size selection and gel extraction
A) Expected RRBS library following large-scale library amplification. **B)** The lane containing a 1 kb ladder is cut from the gel and imaged under UV; distance to the 150 and 325bp bands is measured on a ruler. **C)** Lane containing RRBS library is cut from gel. The gel is cut at the distances measured from the ladder to excise solely the section of the gel containing RRBS library.

2.2.7 Library preparation

Isolated libraries were quantified with the Qubit 2.0 Fluorometer (Life Technologies) with the dsDNA HS Assay (Invitrogen, cat. no. Q32851), before being sent to Otago Genomics Facility (OGF) to be run through the 2100 Bioanalyzer.

Several libraries required an additional round of size selection, when initial bioanalyzer traces indicated adaptor contamination. For this, AMPure XP Beads (New England BioLabs, cat. no. E6270) were used, following the manufacturer's protocol to remove fragments under 100 bp.

Bead size selection does not lose as much of the DNA library as an additional gel extraction. These samples were run through the bioanalyzer again before continuing with sequencing.

Libraries were diluted in Qiagen elution buffer 10 nM in 10 µL solutions for high-throughput sequencing on the Illumina HiSeq 2500 with single-ended 100 bp reads. Some libraries had a yield too low to prepare a 10 nM solution from, and so smaller quantities of DNA were submitted (as listed in Appendix C).

2.3 Analysis of sequencing data

The libraries produced in Section 2.2 were mapped to the human genome. The resulting alignments were used alongside additional RRBS libraries from ADPKD and non-ADPKD tissue previously generated in our laboratory, to analyse the methylome of ADPKD. Analysis was performed with DMAP programs [144] in command line and with RStudio functions. An outline of the protocol is detailed below, step-by-step examples of the programs are in Appendix D.

2.3.1 Preparation of sequencing alignments

Each file was trimmed using the program cleanadaptors (v 1.35), which trimmed the 5' and 3' ends of each fragment to remove residual adaptor sequence from being included in the subsequent analyses. Sequence quality was checked with fastqc (v 0.11.5), and the resulting file was opened in a web browser to check for features such as overrepresented sequences. Overrepresented sequences indicate that adaptor sequences are still present in the library and the file should be further trimmed with more stringent criteria. The switches -t and -T instruct the program how far to cut off the 5' and 3' end of the fragment, respectively, and -x rejects fragments which are less than 4 bp in length following trimming.

As the samples had been sequenced across multiple lanes, each sample had multiple files (i.e. three for each individual ADPKD cyst) which were merged together to create a single file per library.

The fastq files were mapped to the human genome (GRCh37/hg19 construct) using bismark (v 0.16.1), which generates a bam file.

A non-ADPKD reference genome was created from all three non-ADPKD libraries using samtools (v 1.3), which could be used to compare against all ADPKD cyst libraries.

2.3.2 Analysis of RRBS data

Text files of the methylation data were generated using diffmeth (v 1.75). In all cases, the following switches were used: -F 2 (requires every fragment to contain two or more CpGs) and -t 10 (each fragment must contain ten or more reads). Additionally, all fragments used in analyses were to be between 40 and 220 bp in length.

Lists of all RRBS fragments in a sample were generated with the switch -L. ANOVA analysis between two groups (-R and -S) were performed with the switch -B.

The files generated from diffmeth analysis were annotated with identgeneloc (v 0.15), which gave information about the location of the nearest protein-coding gene for each fragment.

2.3.3 Post-processing of RRBS data

Once generated by DMAP programs, data was further analysed in RStudio.

The lists of all RRBS fragments in each sample were compiled into one data frame, and the merge function was used to create a data frame of fragments common between all nineteen libraries. Pearson's correlation coefficient (PCC) was calculated using this data frame to assess the similarities between samples.

Global methylome analysis was performed with a Wilcoxon rank sum test for each library. Hyper- and hypomethylation trends were calculated between the median methylation values of the libraries. Violin plots were generated, which visually demonstrated trends across the whole genome.

To identify the DMFs in ADPKD tissue, the *p* values calculated by ANOVA analysis in diffmeth were adjusted using False Discovery Rate (FDR) to account for multiple testing. The difference between non-ADPKD and ADPKD methylation was calculated and fragments with an absolute difference greater than 10% with an FDR-adjusted p value < 0.05 were identified as DMFs.

Two methods of clustering were applied to the data sets. The first was unsupervised hierarchical clustering, which was performed on the common analysed fragments of all 19 RRBS libraries.

The second clustering technical was applied to only the common analysed fragments from the eight ADPKD cysts. This approach used the program ConsensusClusterPlus [145] to estimate the number of classes in an unsupervised dataset.

2.4 Gene expression

The expression of ADPKD-associated gene *PKD1* was investigated in ADPKD and non-ADPKD tissue sources through qPCR. Additional qPCR was performed on the same samples for several of the DMF-associated genes (cDNA and analysis performed by me, qPCR of DMF genes performed by Michael Bates).

2.4.1 cDNA generation

cDNA was generated from total RNA extractions using High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, cat. no. 4368814), following the protocol without RNase inhibitor. 10 µL of total RNA (10 ng) from each sample was added to 10 µL of master mix (set up as in Table 2.9). The samples were incubated at 25 ºC for ten minute, 37 ºC for two hours and 85ºC for five minutes.

Table 2.9: cDNA Reverse Transcription master mix

10X RT Buffer	2.0 µL
25X dNTP Mix (100 mM)	0.8 µL
10X RT Random Primers	2.0 µL
MultiScribe™ Reverse Transcriptase	1.0 µL
MQ H_2O	4.2 µL
Total	**10.0 µL**

2.4.2 Reference gene selection

Four genes – *B2M, EEF1A1, RPL27* and *SRP14* – were selected as potential reference genes, as these have been used in previous studies of gene expression in ADPKD. Primers were prepared to 10 ng/µL (sequences in Appendix E) for qPCR with SYBR® *Premix Ex Taq*™ (Tli RNase H Plus) (Takara Bio, cat. no. RR420A). Reactions were performed in triplicate on a LightCycler480 system, the set up for a single reaction is in Table 2.10.

Table 2.10: qPCR reaction set up

SYBR® *Premix Ex Taq*™	5.0 µL
PCR Forward Primer (10 µM)	0.2 µL
PCR Reverse Primer (10 µM)	0.2 µL
cDNA sample	1.0 µL
MQ H_2O	2.6 µL
Total	**10.0 µL**

The most suitable genes were required to show consistent expression across all the tissue samples, which was measured with the algorithm geNorm and reported as "M" and "CV" values in qbase+ [146, 147].

2.4.3 Quantification of *PKD1* expression

To calculate the relative expression of *PKD1* in the tissue samples, qPCR was carried out in a 96-well plate. Each sample underwent qPCR for two reference gene primer pairs and a primer pair that targets *PKD1* (sequences in Supplementary Table E.1), all in triplicate and set up as per Section 2.4.2 on the LightCycler480. The Cq values were imported into qbase+, where the expression of *PKD1* in each sample was normalised to geometric mean of the relative quantities of the two reference genes [146, 147].

2.4.4 Statistical significance

In order to determine the statistical significance in *PKD1* expression between ADPKD and non-ADPKD samples, the *p* value was calculated by way of a non-parametric t-test. The Mann-Whitney U-Test was used between the two groups in Prism.

2.4.5 miRNA expression

The expression of miRNA within the *PKD1* sequence was carried out using the TaqMan® systems, which has pre-designed probes to target mature miRNA in a sample. This protocol

includes the TaqMan® Advanced miRNA cDNA Synthesis (Applied Biosystems, cat. no. A28007), followed by the Advanced miRNA Assay (Applied Biosystems, cat. no. A25576).

In this protocol, cDNA is produced from four consecutive reactions as outlined in the Advanced miRNA Assays User Guide. 10 ng of total RNA was used for cDNA synthesis. The amplification of miRNA for qPCR analysis was performed on three target miRNA (Assay ID 478639_mir, 479661_mir, 478303_mir) and the reference gene miR-191-5p (Assay ID 477952_mir) with the TaqMan® Fast Advanced Master Mix as recommended by the manufacturer.

miRNA data was analysed on qbase+ by normalising the Cq expression of the genes to the reference gene, and by performing the Mann-Whitney U-Test between sample groups (as in Section 2.4.4).

Chapter 3: Analysis of the ADPKD kidney methylome

Reduced representation bisulfite sequencing (RRBS) libraries were created from two ADPKD samples as per the methodology in Section 2.2. These sequencing data were analysed with RRBS data from an additional two ADPKD and three non-ADPKD tissue samples (prepared by a previous Master's student in our laboratory [140]).

The purpose of this was to create data representative of the methylome of ADPKD and non-ADPKD kidneys to elucidate the differences between the two groups and demonstrate key differences that occur in ADPKD. The methylome analyses presented in this chapter have since been published [148] and can be read in Appendix F.

3.1 Source material

3.1.1 Kidney tissue

Samples from one of the ADPKD kidneys used in this analysis (ADPKD D) were collected immediately following nephrectomy. This kidney weighed 1,164 g, and morphology was difficult to distinguish due to the presence of numerous cysts throughout the kidney (Fig. 3.1). It was noted that the centre of the kidney appeared to contain a large volume of adipose tissue, approximately where the medulla may be located.

Multiple sections of tissue were collected from this kidney, and two sections were used in this analysis to provide biological and technical replicates (eventually to be called samples ADPKD D I, II and III). The second sample of ADPKD tissue prepared in this study (ADPKD 05) was donated in 2005 and had been stored at -80 ºC. All samples used in this analysis are summarised in Table 3.1.

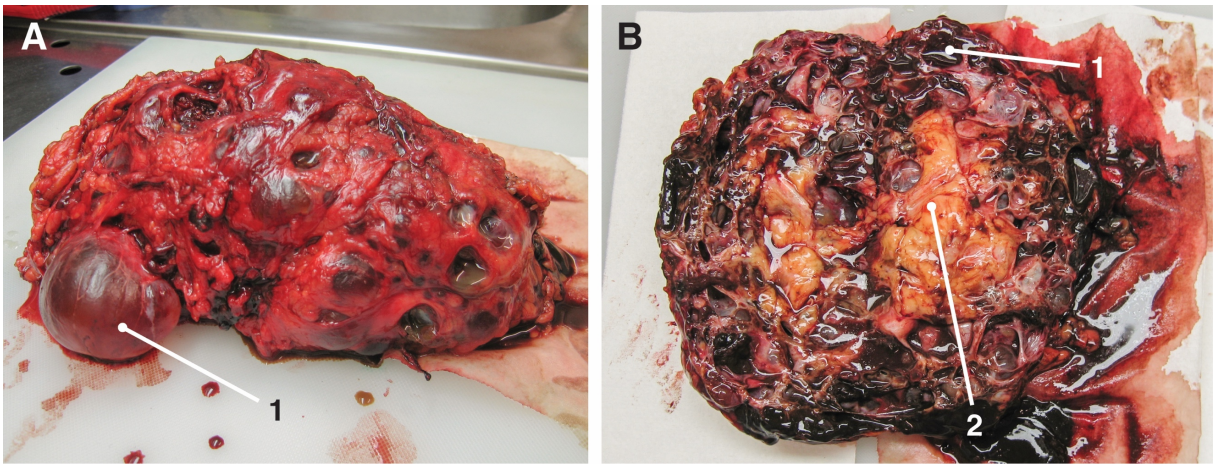


Figure 3.1: ADPKD kidney following nephrectomy
The left kidney from patient ADPKD D (1,164 g) was dissected for collection of whole tissue ADPKD samples. **A)** Large fluid-filled cysts characteristic of ADPKD are throughout the entire kidney. **B)** The kidney, opened along the coronal plane, is filled with cysts. A large deposit of what appears to be adipose tissue is found within the centre. (1 = ADPKD cyst, 2 = adipose deposit).

Table 3.1: Sample summary

Tissue was sourced from three non-ADPKD patients (blue) and six ADPKD patients (red). Of the ADPKD samples, biological replicates were generated from repeat analyses of patient ADPKD D, and one pair of technical replicates was prepared from this same patient. Cell lines originally isolated from the same patient were also biological replicates. qPCR analyses were performed with tissue samples from three non-ADPKD patients and five ADPKD patients (whole tissue only). Patient ADPKD W tissue was only used for qPCR (not RRBS).

Patient	Tissue type	Biological and technical replicate origin		Sample ID	RRBS Input	qPCR Input
Non-ADPKD E	Fresh-frozen			Non-ADPKD E	2.5 µg	10 ng
Non-ADPKD G	Fresh-frozen			Non-ADPKD G	2.5 µg	10 ng
Non-ADPKD H	Fresh-frozen			Non-ADPKD H	2.5 µg	10 ng
ADPKD 05	Fresh-frozen			ADPKD 05	0.5 µg	10 ng
ADPKD 07	Fresh-frozen			ADPKD 07	2.5 µg	10 ng
ADPKD 08	Fresh-frozen			ADPKD 08	2.5 µg	10 ng
ADPKD D	Fresh-frozen	Left kidney	Tissue Sample 1	ADPKD D I	0.5 µg	10 ng
			Tissue Sample 2	ADPKD D II	0.5 µg	
				ADPKD D III	0.5 µg	
	Formalin	Right kidney		Cyst 1	0.5 µg	-
				Cyst 2	0.5 µg	-
				Cyst 3	0.5 µg	-
				Cyst 4	0.5 µg	-
				Cyst 5	0.5 µg	-
				Cyst 6	0.5 µg	-
				Cyst 7	0.5 µg	-
				Cyst 8	0.5 µg	-
ADPKD W	Formalin			ADPKD W	-	10 ng
WT 9	Cell line	Proximal tubule		WT 9-7	2.5 µg	-
		Proximal and distal tubule		WT 9-12	2.5 µg	-

3.1.2 Generation of RRBS libraries

RRBS libraries were generated from small sections of kidney tissue (approximately 25 mg). It was noted that one section of ADPKD D tissue (which would generate sample ADPKD D I) was distinctly redder than the second ADPKD D sample, which appeared to be more fibrotic (and generated samples ADPKD D II and III, which were technical replicates of each other). Libraries were generated from 0.5 µg of DNA using the modified RRBS methodology in Section 2.2.

3.1.3 Assessing the quality of data procured for this analysis

Mapping and bisulfite conversion rates were calculated for the processed RRBS files to ensure they were of sufficient quality (Table 3.2). While all libraries had bisulfite conversion > 98%, only the ADPKD D sample with the highest coverage and mapping efficiency (ADPKD D III) was selected for further analysis in this section of the research project and will be referred to as 'ADPKD D' in the remainder of this chapter.

Table 3.2: Coverage and mapping efficiency of generated RRBS libraries
Data on the sequencing alignment and efficiency generated as part of the Bismark alignment.

	Sequences analysed in total	No. alignments with unique best hit	Mapping efficiency	Bisulfite conversion
ADPKD D I	9.95E+06	4.75E+06	47.70%	> 98%
ADPKD D II	1.43E+07	7.26E+06	50.80%	> 98%
ADPKD D III	1.42E+07	7.70E+06	54.20%	> 98%
ADPKD 05	1.86E+07	1.01E+07	54.20%	> 98%

The sequences underwent quality control by FastQC analysis, which analyses each sequenced fragment and assigns a quality score based upon the ability to identify nucleobases at each position in a sequence. This is represented by a box and whisker plot (Fig. 3.2) which plots the median and interquartile range (IQR) of all analysed sequences at each position in the read. It

is expected that good quality sequences should have an IQR scoring above 28, which estimates accuracy of base-calling at 99.9%. Additionally, sequenced libraries with IQR scores falling below 20 (accuracy of base-calling ≤ 90%) should be trimmed to eliminate poor quality reads.

The quality analysis of the RRBS libraries for ADPKD 05 and ADPKD D had sufficient quality scores and thus were suitable for further analyses.

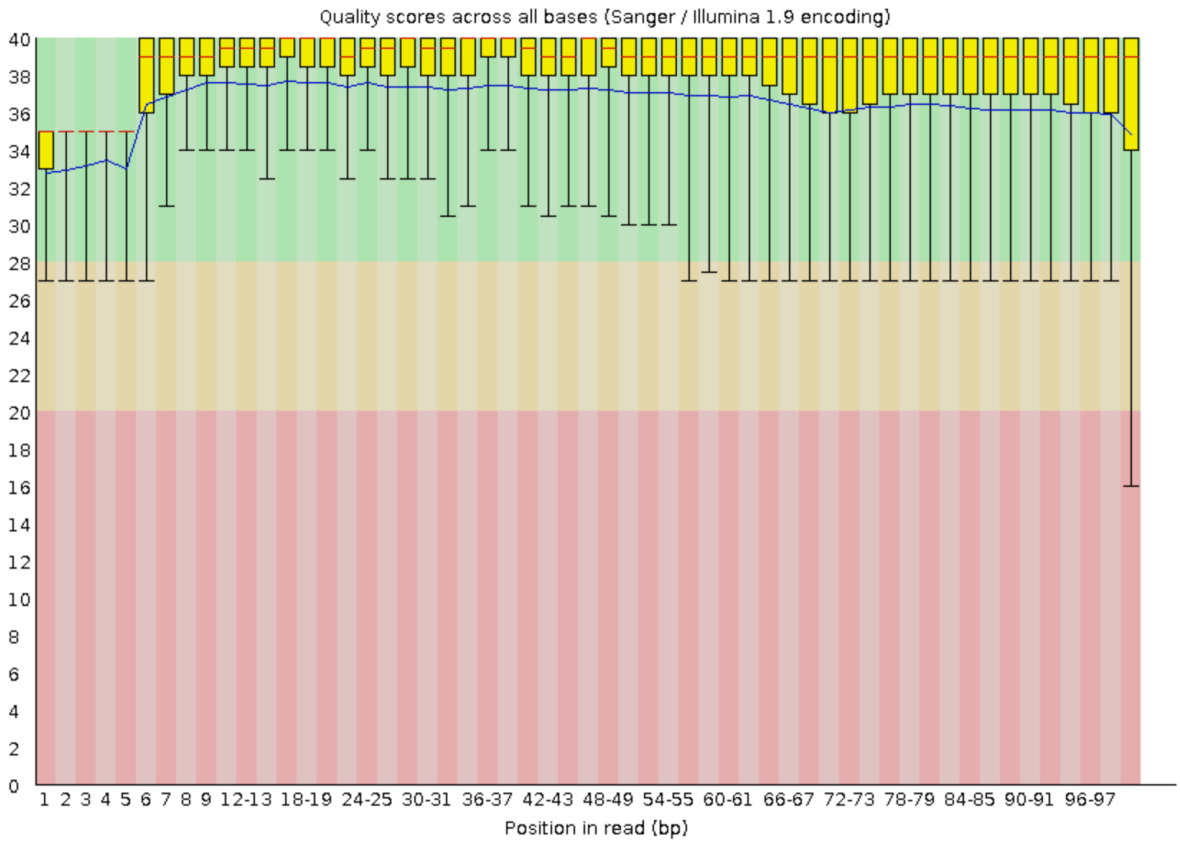


Figure 3.2: Quality score of RRBS library ADPKD 05
The quality of RRBS fragments across the sequenced lengths is represented by a box and whisker plot. The above sample, ADPKD 05, is representative of the sequence quality observed in all of the RRBS libraries generated in this project. It is typically recommended that when the interquartile range is above 28 (green), the sequencing data is of good quality, and then it drops below 20 (red) the sequences should be trimmed at this position in the read to eliminate poor quality.

3.1.4 Validation of the RRBS protocol for assessing the methylome of kidney tissue

Pearson's correlation coefficient (PCC) values were calculated between each pair in the RRBS analysis (Table 3.3). These measure how far the observed data lie from the line of perfect concordance. Therefore, the variation between libraries can be measured as 1 PCC. These data validate the accuracy of the RRBS protocol, and also provide insight into the relationship between methylomes of kidney tissue in different groups (Table 3.4). A definition of correlation for highly similar objects has been described as: < 0.90 as poor, 0.90-0.95 as moderate, 0.95-0.99 as substantial and > 0.99 as almost perfect concordance [149].

Table 3.3: Pearson's correlation coefficients between pairs of RRBS libraries
Libraries 05, 07, 08, D I, D II and D III are ADPKD samples, while E, G and H are non-ADPKD tissue. All pairs $p < 2.2E\text{-}16$.

	05	07	08	D I	D II	D III	E	G
07	0.989							
08	0.990	0.990						
D I	0.969	0.976	0.965					
D II	0.982	0.982	0.981	0.973				
D III	0.982	0.982	0.981	0.972	0.993			
E	0.972	0.976	0.977	0.953	0.982	0.982		
G	0.967	0.971	0.972	0.945	0.976	0.977	0.989	
H	0.977	0.982	0.984	0.955	0.980	0.980	0.987	0.985

Technical replicates (ADPKD D II and III) were the most closely related (PCC = 0.993) in this group, which supports the validity of the RRBS methodology to accurately replicate the methylome of kidney tissue as the concordance is "almost perfect".

Table 3.4: Mean PCC between experimental variables
Mean PCC values calculated from data available in Table 3.3.

	Mean PCC
Technical Replicates Samples prepared from the same DNA extraction	0.993
Biological Replicates Samples from the same patient; different DNA extractions	0.972
Unrelated ADPKD Samples from unrelated ADPKD patient libraries	0.982
ADPKD v Non-ADPKD Comparisons between ADPKD and non-ADPKD libraries	0.973
Unrelated Non-ADPKD Samples from unrelated non-ADPKD libraries	0.987

3.2 Methylation in kidney tissue

3.2.1 The global methylome in kidney tissue

There were 345,711 fragments shared between ADPKD ($n = 4$) and non-ADPKD ($n = 3$) groups which satisfied the criteria of containing at least two CpGs and a read depth of at least ten at each fragment site. The coverage from each individual RRBS library is in Table 3.5. These fragments also had to be present in at least two of the ADPKD and two of the non-ADPKD libraries. This group of fragments was termed the analysed fragments, which covered 1,791,585 CpG sites.

Table 3.5: Coverage of RRBS libraries
Fragments were required contain two or more CpGs with a read depth of 10 or higher, and to be in at least two non-ADPKD and two ADPKD samples in order to be included in analysed fragments pool.

	Fragments analysed	CpGs covered
Non-ADPKD E	342,174	1,772,557
Non-ADPKD G	313,924	1,619,926
Non-ADPKD H	342,028	1,768,358
ADPKD 05	250,407	1,353,018
ADPKD 07	341,981	1,757,350
ADPKD 08	340,723	1,762,144
ADPKD D	170,346	957,481
Total analysis	345,711	1,791,585

Total methylation within each of the whole tissue libraries is illustrated in Fig. 3.3. It is noted that the ADPKD D library has a significantly lower coverage and median methylation. The comparison between the ADPKD and non-ADPKD groups is plotted in Fig. 3.4.

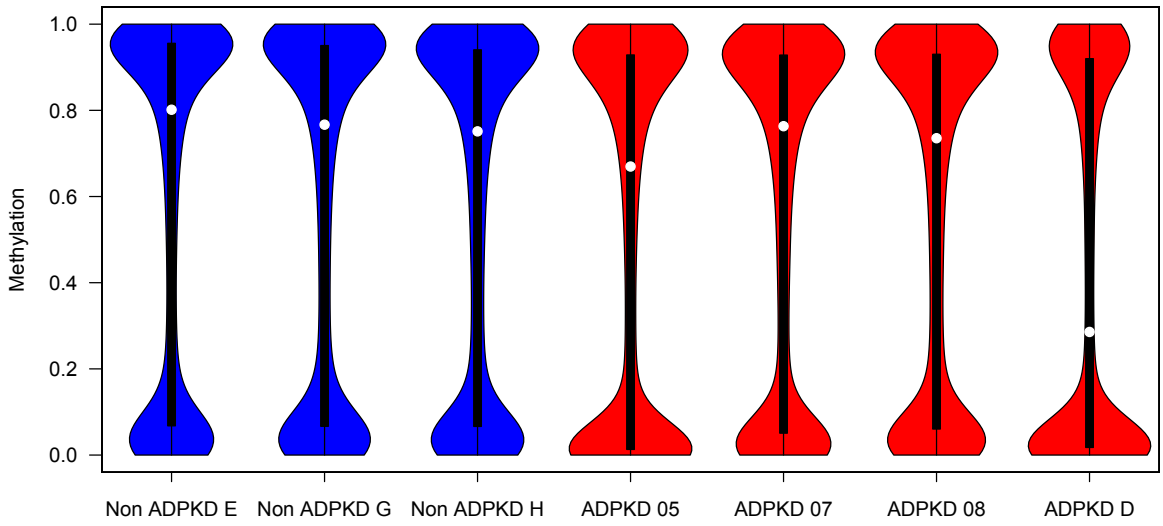


Figure 3.3: Global methylome of RRBS libraries
The total sequenced methylome of each whole tissue RRBS library (*n* listed in Table 3.5). Median values (represented by the black bar on each violin plot) are on average lower in the ADPKD (red) samples compared to non-ADPKD (blue). The methylation values from the non-ADPKD (blue) and ADPKD (red) libraries are plotted where the median value is represented by the white dot. The distribution of methylation values are represented by the shape and area of the violin plot. Sample ADPKD D has less coverage of highly methylated fragments, as the kernel density in the area between 0.8 and 1.0 methylation is not as wide as the other samples.

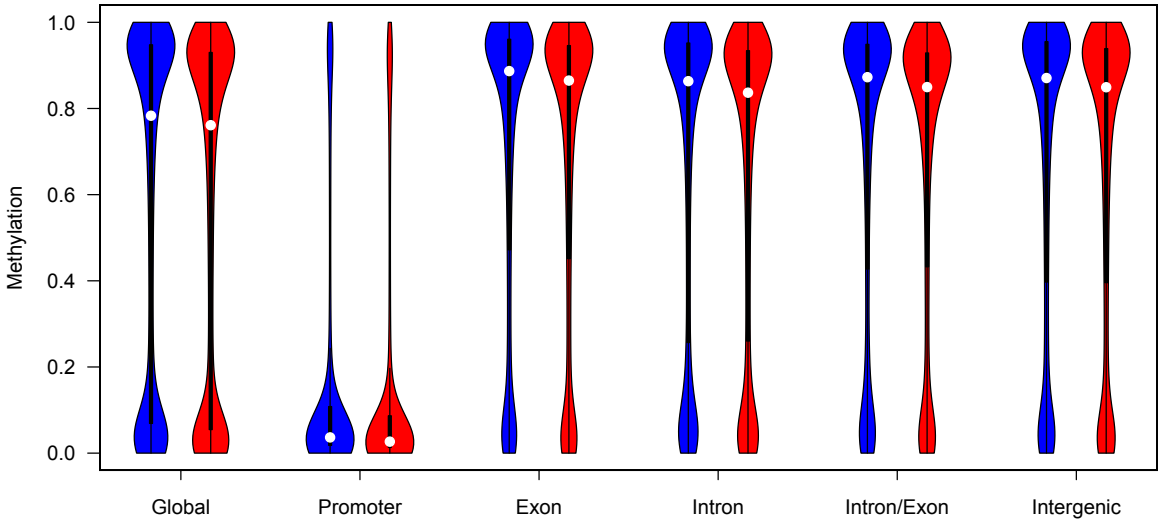


Figure 3.4: Total methylation within genomic elements of ADPKD and non-ADPKD tissue
The methylation beta values extend from 0 (no methylation in a fragment) to 1 (all CpGs are methylated in a fragment). The methylation values from the non-ADPKD (blue) and ADPKD (red) libraries are plotted where the median value is represented by the white dot. The distribution of methylation values are represented by the shape and area of the violin plot. *n*(Genome-Wide) = 345,711).

The median methylation value of the ADPKD tissue group was 0.56, whilst the non-ADPKD tissue group had a median methylation score of 0.58. This demonstrates a small but statistically significant trend of hypomethylation in ADPKD tissue ($p < 2.2E-06$, Wilcoxon rank sum test). Because sample ADPKD D had a significantly lower median, the global median of the other three ADPKD libraries was plotted to check that this library had not created a bias in the analysis. The global median value of methylation of libraries ADPKD 05, 07 and 08 was also 0.56 (Supplementary Fig. G.1).

The violin plot of genomic elements (Fig. 3.4) indicates a small amount of global hypomethylation in the intron, exon and intergenic fragments, as well as the same trend of global hypomethylation seen in Fig. 3.3 (Wilcoxon rank sum test $p < 2.2E-16$, Table 3.6).

Table 3.6: Significance of differences in methylation seen in ADPKD tissue
Median difference is calculated as methylation of non-ADPKD tissue subtracted from ADPKD tissue; negative values indicate hypomethylation of ADPKD. The significance was assessed with a Wilcoxon rank sum test (W).

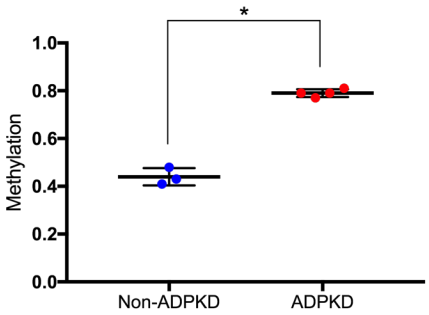
	Median difference	W	*p*
Genome-wide	-0.0214	6.3755E+10	< 2.2E-16
Promoter	-0.0102	2.9147E+9	< 2.2E-16
Intron	-0.0217	9.0363E+9	< 2.2E-16
Exon	-0.0317	5.9062E+8	< 2.2E-16
Intron/Exon	-0.0029	1.3295E+8	< 2.2E-16
Intergenic	-0.0199	9.293E+09	< 2.2E-16

3.2.2 Differentially methylated fragments in ADPKD tissue

Of the 345,711 analysed fragments in the whole tissue analysis, a total of 13 differentially methylated fragments (DMFs) were identified (Fig. 3.5). These were distinguished using the criteria set out in Chapter 2 (difference in methylation between ADPKD and non-ADPKD ≥ 10% and an FDR-adjusted p value < 0.05).

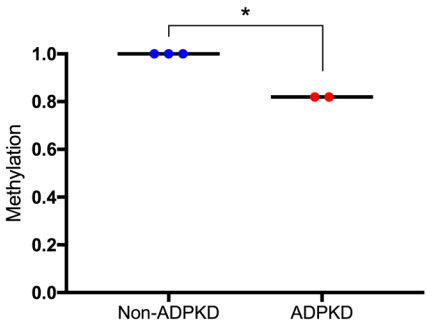
2:771624 - 771728

Upstream of *TMEM18*



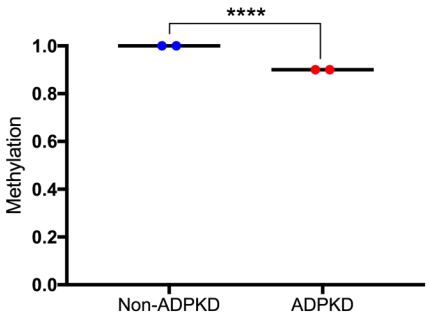
4:793929 - 794143

Within *CPLX1*



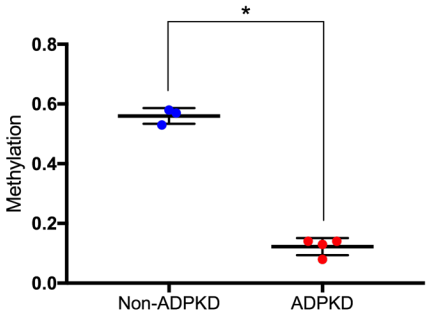
7:6495310 - 6495378

Within *KDELR2*



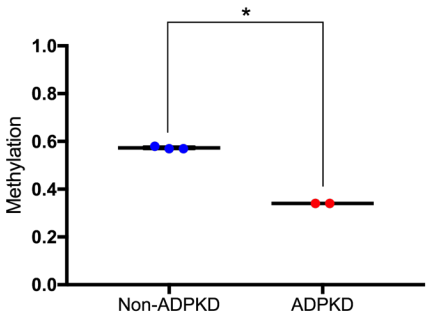
9:130768316 - 130768366

Upstream of *FAM102A*



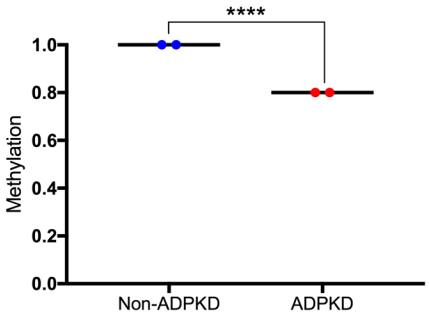
17:40683608 - 40683711

Within promoter of *NAGLU*



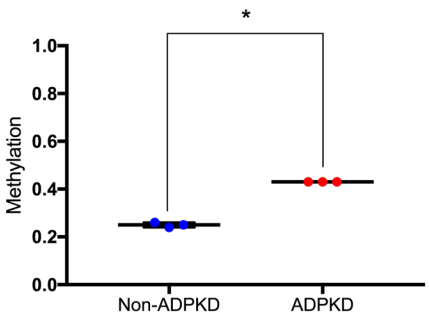
18:10271529 - 10271568

Upstream of *APCDD1*



22:37499386 - 37499523

Within *TMPRSS6*



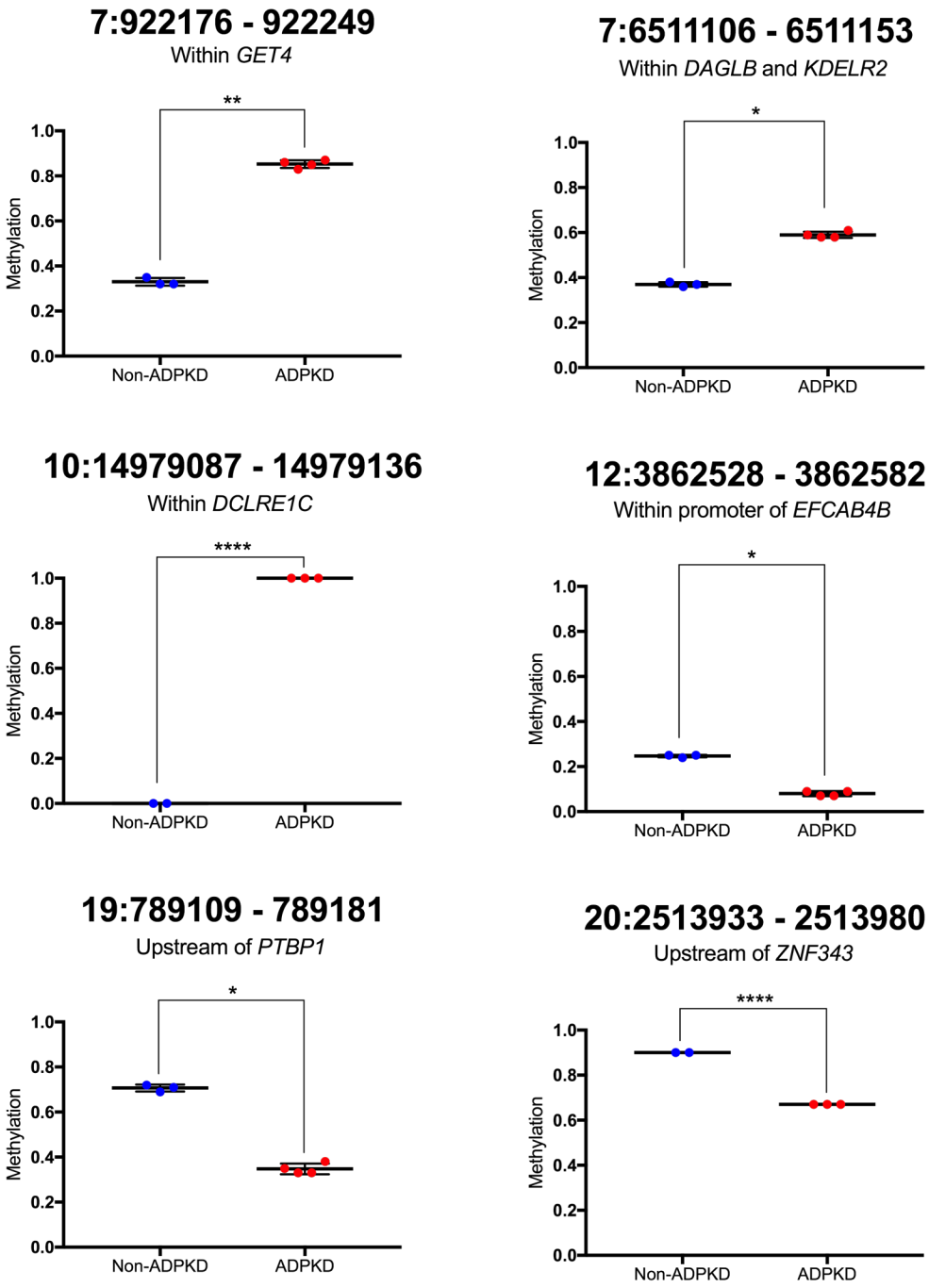


Figure 3.5: Methylation of non-ADKPD and ADPKD tissue at DMFs
DMFs (pages 58, 59) were calculated from ANOVA analysis and have a difference in methylation > 10% and FDR-adjusted p value < 0.05. Bars show mean value ± SD, FDR-adjusted p values: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

There was an enrichment of DMFs within intergenic regions and intron regions (Fig. 3.6), although due to the low number of total DMFs this limits the significance of any observed trends.

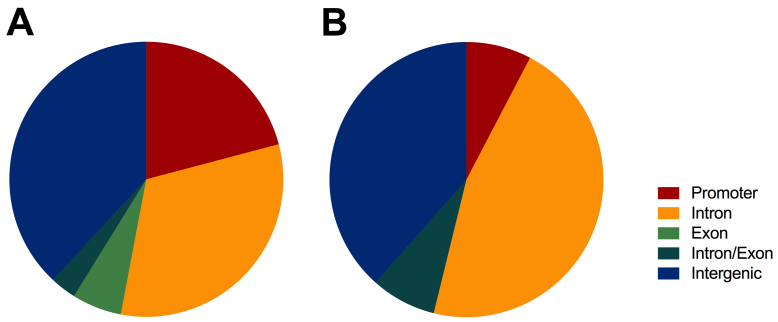


Figure 3.6: Distribution of genomic elements in DMFs
A) Distribution of elements in analysed fragments (*n* = 345,711). **B)** Distribution of elements in identified DMFs (*n* = 13).

Eight of these DMFs demonstrated hypomethylation in ADPKD, with a difference in methylation ranging from 0.10-0.43 between the two groups (Table 3.7). Two of these hypermethylated fragments were located within the gene body of protein coding genes. Two fragments were located between 5 kb upstream and1 kb downstream from the transcription start site (TSS) of a protein coding gene, and therefore are classed as being located within the promoter regions. The remaining four fragments were further than 5 kb from a protein coding gene and therefore are considered intergenic.

Table 3.7: Differentially methylated fragments which show hypomethylation in ADPKD
Fragments are identified by their location within the chromosome and the nearest protein-coding gene. The difference in methylation is calculated by subtracting the methylation score of non-ADPKD from that of ADPKD. The relationship to TSS refers to the distance between the fragment location and the transcription start site of the nearest protein-coding gene.

Chromosome coordinates:	Difference in Methylation:	Nearest protein-coding gene:	Relationship to TSS:
4:793929 - 794143	-0.1786	*CPLX1*	Within intron 2
7:6511106 - 6511153	-0.1000	*DAGLB/KDELR2*	Within an intron of the gene body
9:130768316 - 130768366	-0.4308	*FAM102A*	~26 kb upstream
12:3862528 - 3862582	-0.1664	*EFCAB4B (CRACR2A)*	Within promoter
17:40683608 - 40683711	-0.2329	*NAGLU*	Within promoter flank; 4.2 kb upstream
18:10271529 - 10271568	-0.2000	*APCDD1*	~183 kb upstream
19:789109 - 789181	-0.3608	*PTBP1*	~8 kb upstream
20:2513933 - 2513980	-0.2333	*ZNF343*	~9 kb upstream

The genes associated with the hypomethylated DMFs are as follows: Complexin-1 (*CPLX1*), Diacylglycerol Lipase Beta (*DAGLB*), KDEL Endoplasmic Reticulum Protein Retention Receptor 2 (*KDELR2*), Family With Sequence Similarity 102 Member A (*FAM102A*), Calcium Release Activated Channel Regulator 2A (*EFCAB4B*), N-Acetyl-Alpha-Glucosaminidase (*NAGLU*), Adenomatosis Polyposis Coli Down-Regulated 1 (*APCDD1*), Polypyrimidine Tract Binding Protein 1 (*PTBP1*) and Zinc Finger Protein 343 (*ZNF343*).

There were five hypermethylated DMFs, with a difference of 0.19-1.00 (Table 3.8). Four of these are situated within the gene body of their respective genes, and one is intergenic.

Table 3.8: Differentially methylated fragments which show hypermethylation in ADPKD
Fragments are identified by their location within the chromosome and the nearest protein-coding gene. The difference in methylation is calculated by subtracting the methylation score of non-ADPKD from that of ADPKD. The relationship to TSS refers to the distance between the fragment location and the transcription start site of the nearest protein-coding gene.

Chromosome coordinates:	Difference in Methylation:	Nearest protein-coding gene:	Relationship to TSS:
2:771624 - 771728	0.3425	*TMEM18*	~94 kb upstream
7:922176 - 922249	0.5226	*GET4*	Within intron 9
7:6495310 - 6495378	0.2208	*KDELR2*	Within an intron of the gene body
10:14979087 - 14979136	1.0000	*DCLRE1C*	Within an intron of the gene body
22:37499386 - 37499523	0.1810	*TMPRSS6*	Spanning intron 1 and exon 2

These genes associated with the hypomethylated DMFs are: Transmembrane Protein 18 (*TMEM18*), Golgi To ER Traffic Protein 4 (*GET4*), KDEL Endoplasmic Reticulum Protein Retention Receptor 2 (*KDELR2*), DNA Cross-Link Repair 1C (*DCLRE1C*) and Transmembrane Serine Protease 6 (*TMPRSS6*).

As the function of these fragments is not limited to the nearest protein coding gene, *in silico* analysis was performed with GenHancer [150] to predict if they play a regulatory role, potentially to genes not located linearly across the genome (Table 3.9). Five of the DMFs were predicted to contain an enhancer element.

Table 3.9: Predicted regulatory roles of DMFs
DMFs were analysed to identify any associated enhancer elements (GenHancer). DMFs below contained at least one regulatory element.

Chromosome coordinates:	Closest Protein-Coding Gene:	Differential Methylation:	Predicted Enhancer:
7:922176 - 922249	*GET4*	0.5226	*PRKAR1B, DNAAF5*
9:130768316 - 130768366	*FAM102A*	-0.4308	*LOC101929314, FAM102A, PIP5KL1*
12:3862528 – 3862582	*EFCAB4B*	-0.1664	*CRACR2A, TULP3*
17:40683608 – 40683711	*NAGLU*	-0.2329	*TUBG2, NAGLU, RABC5C, HSD17B1, MLX, BECN1*
19:789109 - 789181	*PTBP1*	-0.3608	*CFD*

3.3 Expression of DMF-associated genes in ADPKD

Eight of the DMFs found in this study were investigated with qPCR for gene expression analysis in ADPKD tissue isolated from five individuals (Table 3.1). Both *NAGLU* and *GET4* demonstrated significant differential expression between ADPKD and non-ADPKD tissue samples (Fig. 3.7).

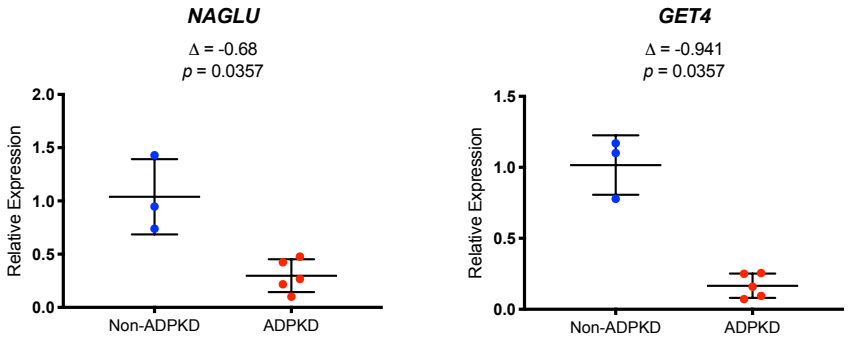


Figure 3.7: Expression of DMF-associated genes *NAGLU* and *GET4*
Change in expression (Δ) was reported from qPCR data, using reference genes *B2M* and *EEF1A1*. Both genes had a statistically significant reduction in ADPKD (p = 0.0357, Mann-Whitney U Test). Bars indicate mean expression values ± SD.

Correlation studies between gene expression and DMF methylation values did not show a statistically significant correlation in *NAGLU* and the DMF (17:40683608 – 40683711), however they did identify a strong correlation between the expression of *GET4* and the methylation of the DMF (7:922176 – 922249) (Fig. 3.8).

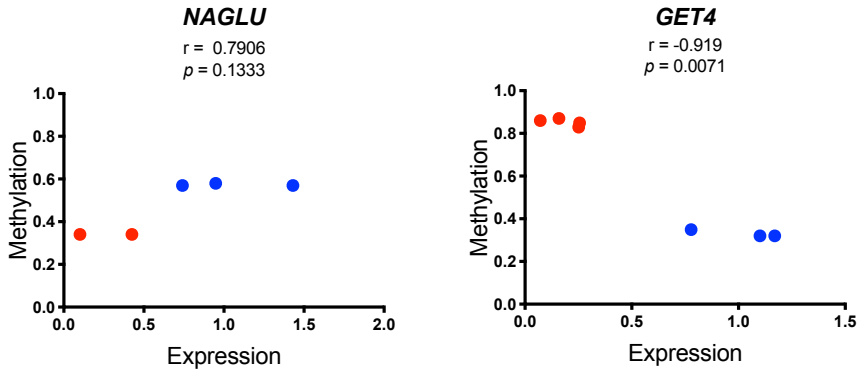


Figure 3.8: Correlation of DMF methylation and gene expression
Spearman's rank correlation was used to determine the correlation between the methylation and expression of DMF-associated genes *NAGLU* and *GET4* in ADPKD (red) and non-ADPKD (blue) kidney tissue. This association was statistically significant in *GET4.*

3.4 *PKD1* in ADPKD tissue

PKD1 is the gene most attributable to ADPKD. Previous investigations of *PKD1* have identified both increased and decreased expression of the gene to be associated with ADPKD, and published methylation analysis has identified the correlation between hypermethylation of the *PKD1* gene body and our reduction of *PKD1* expression. We decided to investigate this within our samples of kidney tissue, using fragments associated with *PKD1* in the RRBS data.

3.4.1 Methylation surrounding the ADPKD-associated gene *PKD1*

There were 52 fragments in the analysed fragment group that were associated with *PKD1*, spanning a 31,060 bp region (Fig. 3.9A). None of these fragments were classified as DMFs, as

although there were some differences in methylation between the two groups, the *p* value calculated to account for multiple comparisons was greater than 0.05.

The 11 fragments at the 3' end of *PKD1* (spanning 16:2164268-2170942) demonstrated a difference in methylation between 0.02-0.12 in the two tissue groups is visible as a different pattern of methylation on the histogram (Fig 3.9B). Additionally, the methylation of ADPKD samples varied greatly, lowering the power and therefore the ability to identify this region as being statistically differentially methylated. Another irregularity between the samples was a fragment in this middle of the analysed *PKD1* region, for which the non-ADPKD methylation value is much lower.

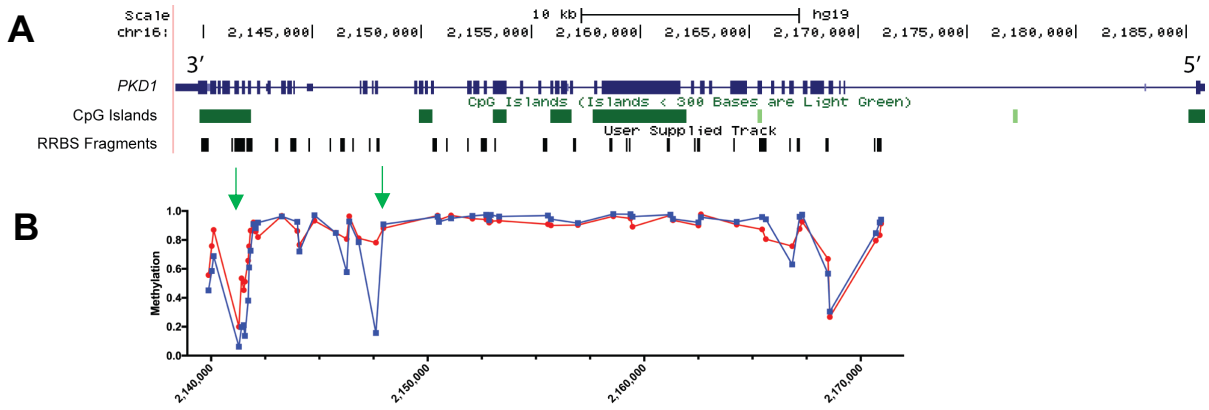


Figure 3.9: Coverage of *PKD1* with RRBS fragments
A) There were 52 fragments analysed between the two tissue groups. The fragments cover a 31 kb region from intron 1 to exon 46. **B)** The median methylation score at each fragment in ADPKD (red) and non-ADPKD (blue) groups. Methylation between the two disease states is largely consistent. There is a small amount of hypermethylation at the 3' end of the gene, and an additional point of differential methylation at 16:2147611-2147671 as indicated by green arrows.

3.4.2 Expression of *PKD1* in ADPKD

In order to assess if these modest methylation differences between ADPKD and non-ADPKD tissue translate into functional differences in the expression of the disease-related gene *PKD1*, qPCR was performed on cDNA derived from samples of fresh-frozen kidney tissue.

The three non-ADPKD tissues were used as a reference on which to normalise the data from five ADPKD samples, which included tissue samples from the four patients in the above RRBS analysis, as well as a sample from an additional ADPKD patient.

The expression of *PKD1* was tightly regulated in non-ADPKD tissue, with all samples yielding very similar Cq values. However, in ADPKD samples, the expression of *PKD1* varied widely between samples. Overall there was a 0.504 increase in ADPKD tissue (p = 0.036) when compared with non-ADPKD tissue (Fig. 3.10). Of the ten RRBS fragments located within *PKD1* with a nominal $p < 0.05$, the methylation of five showed correlation with *PKD1* gene expression (Supplementary Table H.1)

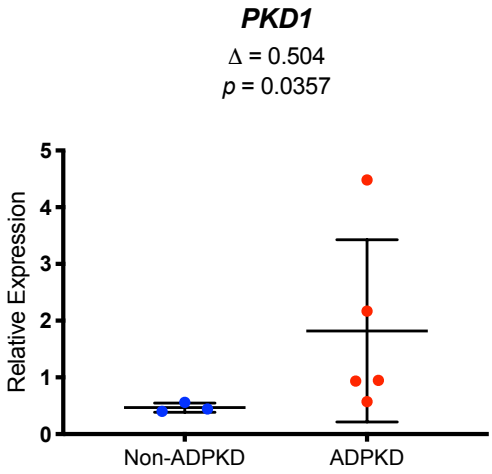


Figure 3.10: Expression of *PKD1*
The change in *PKD1* expression (Δ) is reported relative to the reference genes *B2M* and *EEF1A1*. *PKD1* had a statistically significant reduction in ADPKD patients (p = 0.0357, Mann-Whitney U Test). Bars indicate mean expression values ± SD.

3.5 miRNA in ADPKD

There are three microRNA within the promoter and gene body of *PKD1*: miR-1225, miR-3180 and miR-4516. Because miRNAs are involved in the post-translational regulation of gene expression, the miRNA within *PKD1* were selected for qPCR analysis to determine if the

expression of these genes were altered in ADPKD tissue. Five ADPKD tissue samples and three non-ADPKD samples were analysed, using the TaqMan Advanced miRNA Assay kit as per Section 2.4.5, with expression normalised to the reference gene miR-191-5p.

The analysis of these miRNA indicated that both miR-1225 and miR-3180 have very little expression in the kidney, and there is no evidence that the expression is altered in ADPKD. Preliminary analysis of miR-4516, however, demonstrates varied expression in ADPKD (Fig. 3.11).

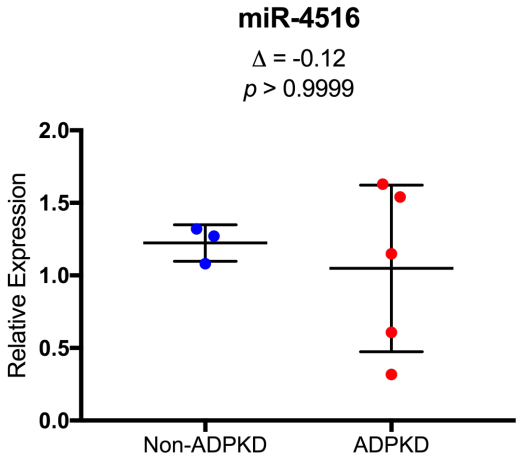


Figure 3.11: Expression of miR-4516
Change in relative expression (Δ) between ADPKD and non-ADPKD tissue was reported from qPCR data, using reference gene miR-191-5p. There was no significant change in expression in ADPKD.

Two of the ADPKD samples (ADPKD 07, D) had reduced levels of miR-4516 expression, however the other three ADPKD samples had similar (ADPKD W) or increased (ADPKD 05, 08) expression.

Chapter 4: Analysis of the methylome of individual ADPKD cysts

The defining feature of ADPKD is the number of cysts that develop within the kidneys. It is hypothesized that these all arise from individual molecular events. To investigate whether changes seen in the ADPKD kidney are universal within unique ADPKD cysts, the methylome of ADPKD cysts were analysed. Reduced representation bisulfite sequencing (RRBS) libraries were generated from the walls of unique cysts from a nephrectomized ADPKD kidney.

4.1 Source material

4.1.1 ADPKD cell lines

Cell lines WT 9-7 and WT 9-12 were grown from passage 16 and 25, respectively. They were cultured on collagen coated plates as per the protocol in section 2 for 10 days. WT 9-12 was noted to be slightly more confluent than WT 9-7, but both cell cultures were polygonal in appearance, which is typical of kidney epithelia [151] (Fig. 4.1).

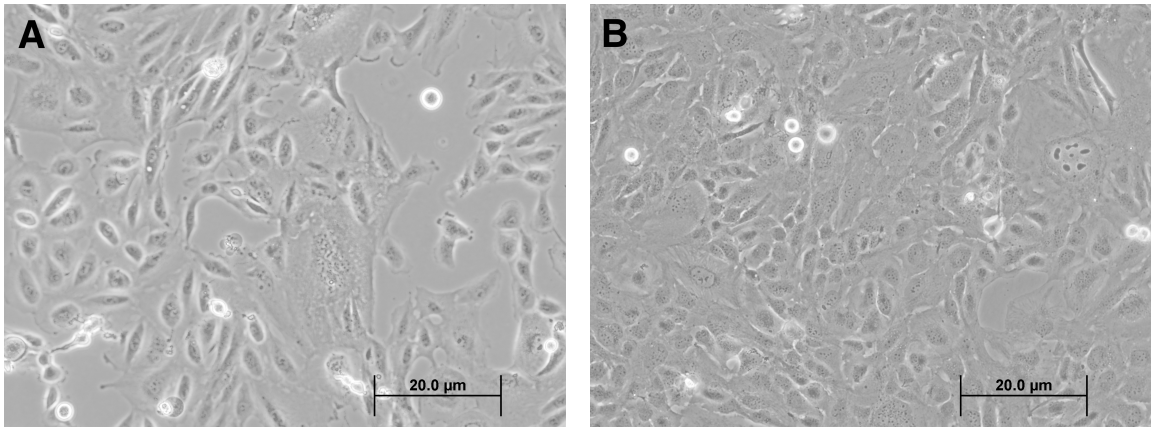


Figure 4.1: Cultured immortalised ADPKD cells
WT 9-7 **(A)** and WT 9-12 **(B)** were grown on collagen coated plates for ten days. WT 9-12 was observed to be more proliferative than WT 9-7.

4.1.2 Isolated ADPKD cyst walls

Sections of cyst walls were collected from the right kidney of patient "ADPKD D". This 1,184 g kidney had been preserved in neutral buffered formalin (NBF) for one week prior to dissection. Although the colour of tissue changes following formalin preservation, it appeared to contain adipose tissue within the centre of the kidney (much like the adipose tissue observed in the left kidney in Chapter 3). Because of the manner in which the kidney had been removed, spatial proximity of cyst samples was not recorded.

4.1.3 Generation of RRBS libraries

Because cyst wall samples were so small, very low quantities of DNA were eluted from the samples, and thus libraries were generated from 0.5 µg of DNA. During library preparation many of these samples had high amounts of adaptor contamination (Fig. 4.2) for which they required an additional round of size selection with bead purification. Ten cyst walls were prepared for analysis but only eight of these had sufficient DNA quantity for further sequencing.

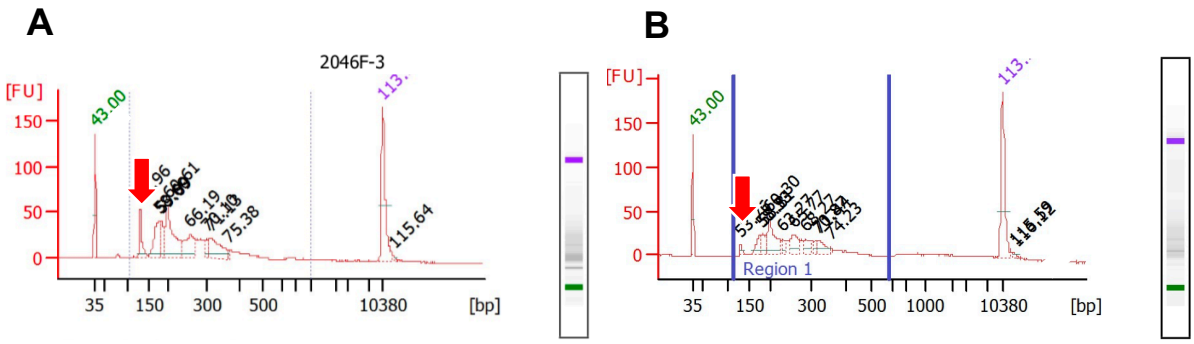


Figure 4.2: Bioanalyzer trace before and after adaptor clean up
(A) ADPKD Cyst 3 had adaptor contamination (peak at 129 bp as identified with arrow). The RRBS library is between 178- 323 bp with the highest peak at 199 bp. **(B)** The adaptor contamination of the Cyst 3 RRBS library was reduced following bead purification.

The mapping, bisulfite conversion and quality scores of the RRBS libraries produced from the cyst samples were comparable to those produced from whole tissue samples that had been produced from the 0.5 µg DNA RRBS protocol. These samples all had lower genome coverage and mapping efficiency than those prepared with 2.5 µg of DNA (Table 4.1).

Table 4.1: Genome coverage and efficiency of RRBS libraries
Seven of the libraries were prepared with previously published RRBS protocol requiring 2.5 µg of input DNA. The remaining 12 libraries were prepared with the modified 0.5 µg protocol. Libraries prepared with the 2.5 µg protocol had greater coverage of the genome and higher mapping efficiencies.

DNA input (µg):	Sample:	Sequences analysed in total:	No. of alignments with a unique best hit from the different alignments:	Mapping efficiency:	Total number of C's analysed:	
2.5	**WT 9-7**	3.75E+07	2.40E+07	63.90%	5.03E+08	Cell
	WT 9-12	2.83E+07	1.80E+07	63.60%	3.66E+08	
	Non-ADPKD E	4.25E+07	2.59E+07	61.00%	5.50E+08	Frozen
	Non-ADPKD G	3.50E+07	2.34E+07	67.00%	4.92E+08	
	Non-ADPKD H	4.15E+07	2.83E+07	68.30%	5.93E+08	
	ADPKD 07	4.69E+07	3.13E+07	66.70%	6.38E+08	
	ADPKD 08	5.72E+07	3.78E+07	66.20%	8.06E+08	
0.5	**ADPKD 05**	1.86E+07	1.01E+07	54.20%	2.01E+08	
	ADPKD D I	9.95E+06	4.75E+06	47.70%	6.88E+07	
	ADPKD D II	1.43E+07	7.26E+06	50.80%	1.31E+08	
	ADPKD D III	1.42E+07	7.70E+06	54.20%	1.37E+08	
	Cyst 1	2.85E+07	1.44E+07	50.60%	2.76E+08	Formalin fixed
	Cyst 2	3.25E+07	1.71E+07	52.50%	3.11E+08	
	Cyst 3	2.11E+07	1.14E+07	54.20%	2.31E+08	
	Cyst 4	8.93E+06	4.65E+06	52.00%	5.78E+07	
	Cyst 5	1.93E+07	8.20E+06	42.50%	1.57E+08	
	Cyst 6	3.88E+07	1.88E+07	48.40%	3.52E+08	
	Cyst 7	2.84E+07	1.11E+07	39.10%	1.97E+08	
	Cyst 8	1.43E+07	6.15E+06	43.00%	1.07E+08	
	Median	2.84E+07	1.44E+07	54.20%	2.76E+08	
	2.5 µg Median	4.15E+07	2.59E+07	66.20%	5.50E+08	
	0.5 µg Median	1.89E+07	9.12E+06	50.70%	1.77E+08	

4.1.4 Concordance of sequencing data

This analysis created RRBS libraries from three DNA sources: fresh-frozen tissue, formalin-fixed tissue and immortalised cell lines. Because these tissues undergo different chemical processes which have the potential to alter epigenetic tags and affect the quality of DNA, it is important to check that the sequencing data is concordant. To do this, the correlation coefficients between the libraries were calculated for the cell lines (Table 4.2) and cysts (Table 4.3), as visualised in Fig. 4.3.

Table 4.2: PCC of cell line RRBS libraries
Mean PCC are calculated from individual paired correlation values; full table in Appendix H.

	WT 9-7	WT 9-12
ADPKD Tissue ADPKD 05, 07, 08, D I, D II, D III	0.944	0.890
Non-ADPKD Tissue Non-ADPKD E, G, H	0.954	0.891
Cell Lines Correlation between two cell lines	0.935	

Table 4.3: PCC between various ADPKD libraries
Mean PCC are calculated from individual paired correlation values; heatmap of data in Fig. 4.3.

	Mean PCC value
ADPKD Tissue ADPKD 05, 07, 08, D I-III	0.980
Biological Replicates ADPKD D I-III	0.979
ADPKD Cysts All analyses between Cysts 1-8	0.987
Non-ADPKD Tissue Non-ADPKD E, G, H	0.987

The cell lines show a low correlation with tissue samples (both ADPKD and non-ADPKD kidney). Additionally, there was low correlation between the two libraries despite these being

biological replicates (cell lines derived from the same patient). These data indicated that immortalised cell lines would not be suitable for this research as they did not resemble kidney tissue. The formalin-fixed samples had consistent correlation with ADPKD tissue and were considered to accurately represent ADPKD tissue.

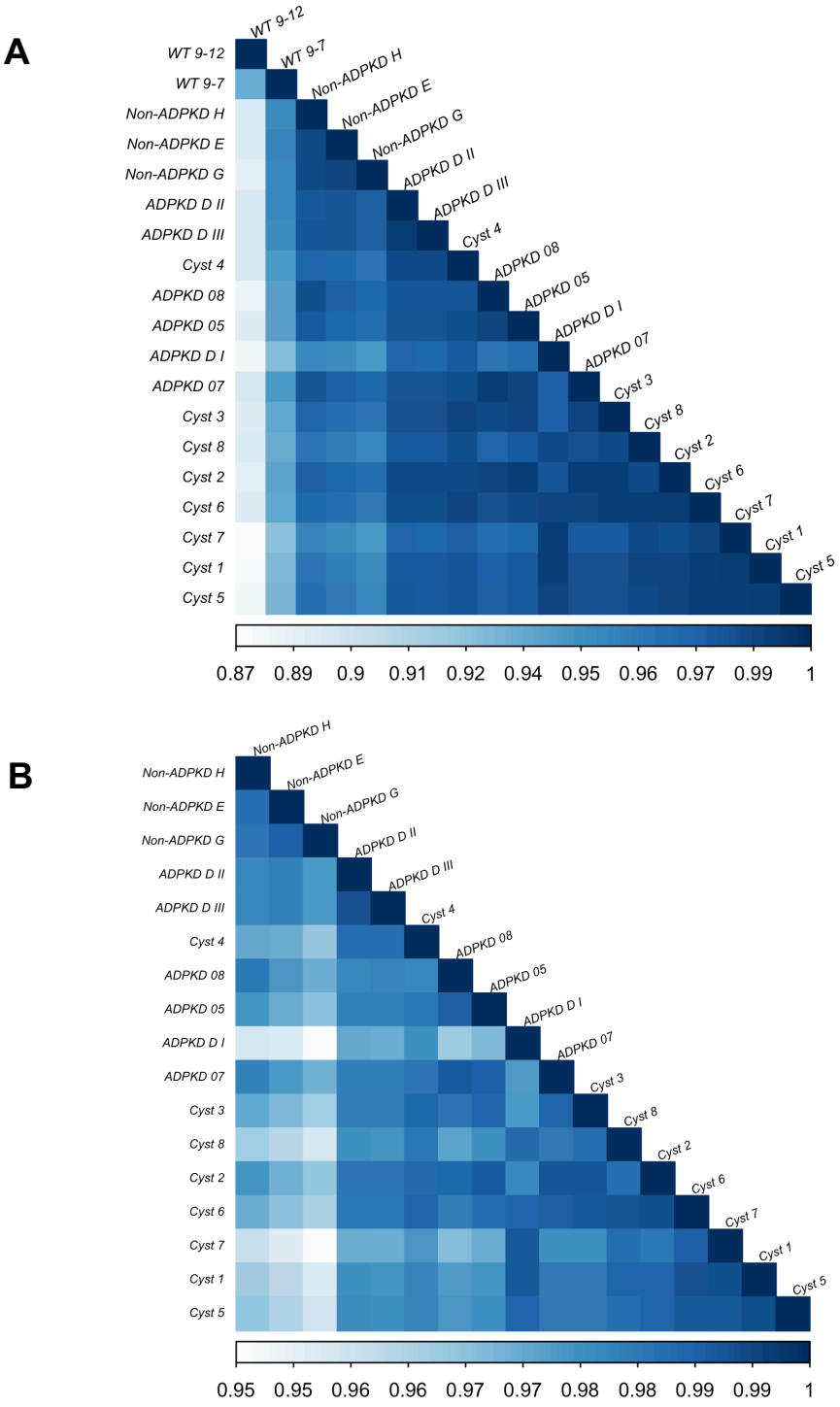


Figure 4.3: Heatmap of correlation between RRBS libraries

PCC values plotted on a heat map to illustrate relative concordance. **(A)** All 19 RRBS libraries plotted. The immortalised cell lines had significantly lower concordance than any of the tissue-sourced libraries. **(B)** The RRBS libraries, excluding the cell lines. Correlation varied between samples, but non-ADPKD samples cluster together as they have the lowest concordance in this set of data. Data extracted from Supplementary Table H.1 and the 'single' method of clustering was used.

4.2 Clustering analysis

4.2.1 Unsupervised hierarchical clustering of RRBS libraries

Using the same pool of common analysed fragments which were used to calculate PCC (n = 38,139), unsupervised hierarchical clustering was performed on all libraries prepared in this research (Fig. 4.4).

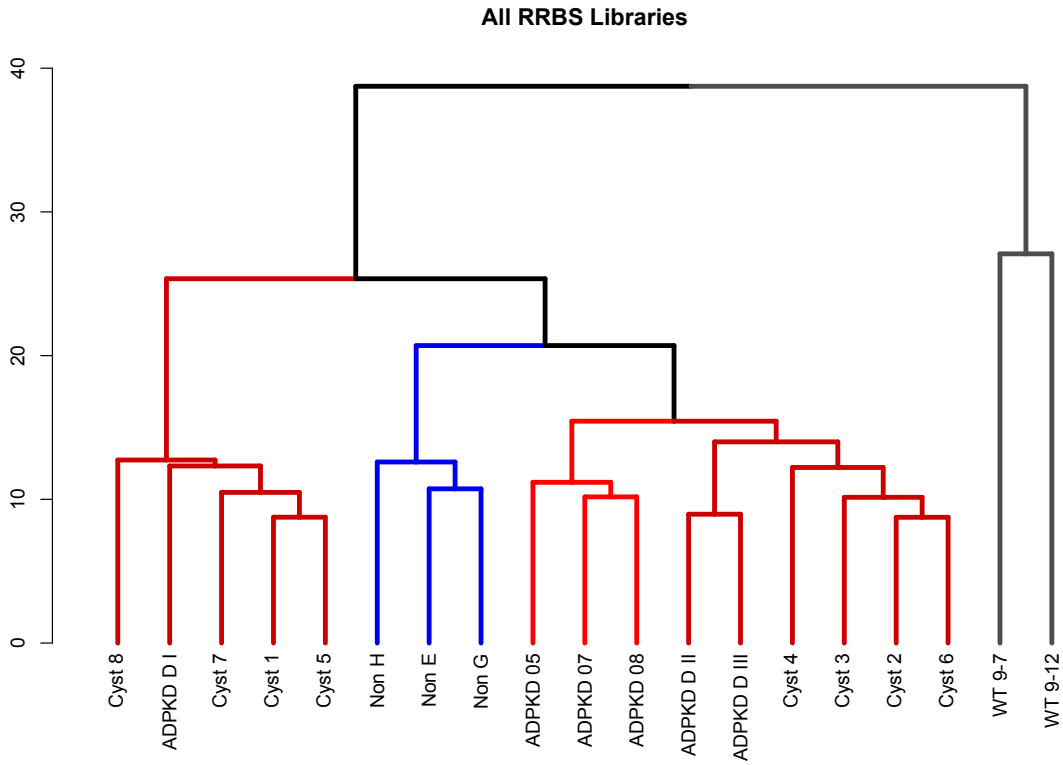


Figure 4.4: Unsupervised hierarchical clustering of all RRBS libraries
Clustering was performed on all common analysed fragments between all 19 RRBS libraries (n = 36,824). Cell lines (grey) cluster independent of all tissue samples. There are varying degrees of relatedness between non-ADPKD (blue) and ADPKD (red) tissue. The biological replicates from patient ADPKD D (whole tissue and cyst) are a darker red, and cluster into two groups. The 'complete' method of clustering was used.

As expected, the cell lines WT 9-7 and WT 9-12 clustered independently and at the highest height in the analysis, confirming the dissimilarities between these and the kidney tissue libraries.

The non-ADPKD tissue samples E, G and H were clustered together, as were the whole tissue samples 05, 07 and 08. Samples from patient ADPKD D were seen to cluster into two distinct groups. The technical replicates ADPKD D II and III had clustered very closely together, confirming the precision of the RRBS analysis. They were within a cluster that contained four of the eight cyst samples (cysts 2, 3, 4 and 6). The other ADPKD D whole tissue sample was in a distinct cluster with the remaining cysts (1, 5, 7 and 8).

Additionally, when hierarchical clustering was performed on solely the ADPKD cyst samples (based on common fragments between only these libraries) the same clustering pattern was seen (Fig. 4.5).

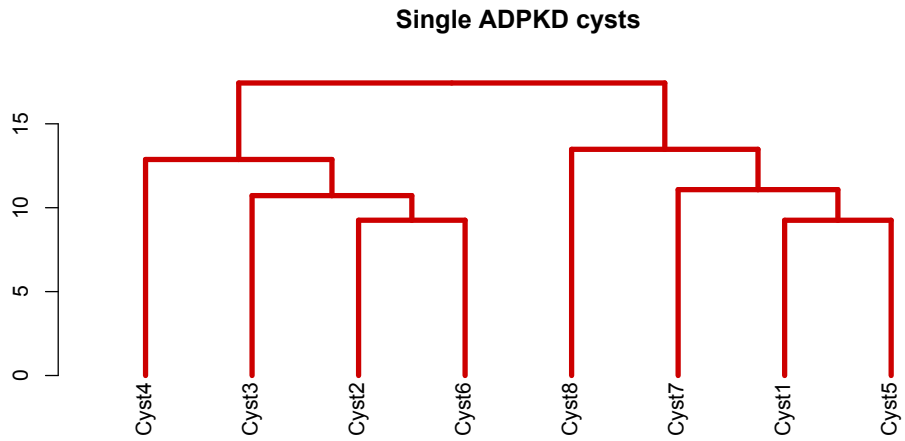


Figure 4.5: Unsupervised hierarchical clustering of single cyst libraries
Clustering was performed on common analysed fragments between all eight cysts (n = 39,805). The 'complete' method of clustering was used.

4.2.2 Consensus clustering of ADPKD cysts

The individual ADPKD cyst libraries were analysed with the program ConsensusClusterPlus (Fig. 4.6), which is an algorithm that assists in defining the suitable cluster count in an unsupervised analysis [145].

Consensus matrices were produced for each k value (where k is the predetermined number of clusters). Under each cluster, the more strongly items cluster together, the higher the consensus is (illustrated by depth of blue). Plots $k = 3$ and $k = 4$ have the strongest consensus matrix, with no overlaps between the determined clusters. However, in $k = 4$, there is not as strong consensus between items in the same clusters, therefore $k = 3$ is the stronger candidate for accurate clustering. $k = 6$ also has strong consensus between clusters, but the member count of each matrix shown in this plot includes only one sample, therefore this is not an informative clustering strategy.

The cumulative distribution function (CDF) plot displays the distribution of consensus for each k value. In this analysis, sample $k = 3$ indicates the maximum stability: $k = 2$ has a much lower CDF value, and $k = 4$, 5 and 6 demonstrate that further divisions are equivalent to random picks.

The delta area graph shows the relative change in area under the CDF curve, illustrating the relative increase in consensus between clusters. In this plot, there is no appreciable increase in consensus after $k = 4$.

The cumulation of this analysis shows that the data clusters well into three groups. While there is no appreciable increase in consensus after $k = 4$, the $k = 4$ consensus matrix shows that two of the clusters contain only one cyst, therefore this would not be an informative division of clustering. In fact, the tracking plot demonstrates that one of the clusters from $k = 3$, light green, splits into two clusters in $k = 4$ (pink and dark green).

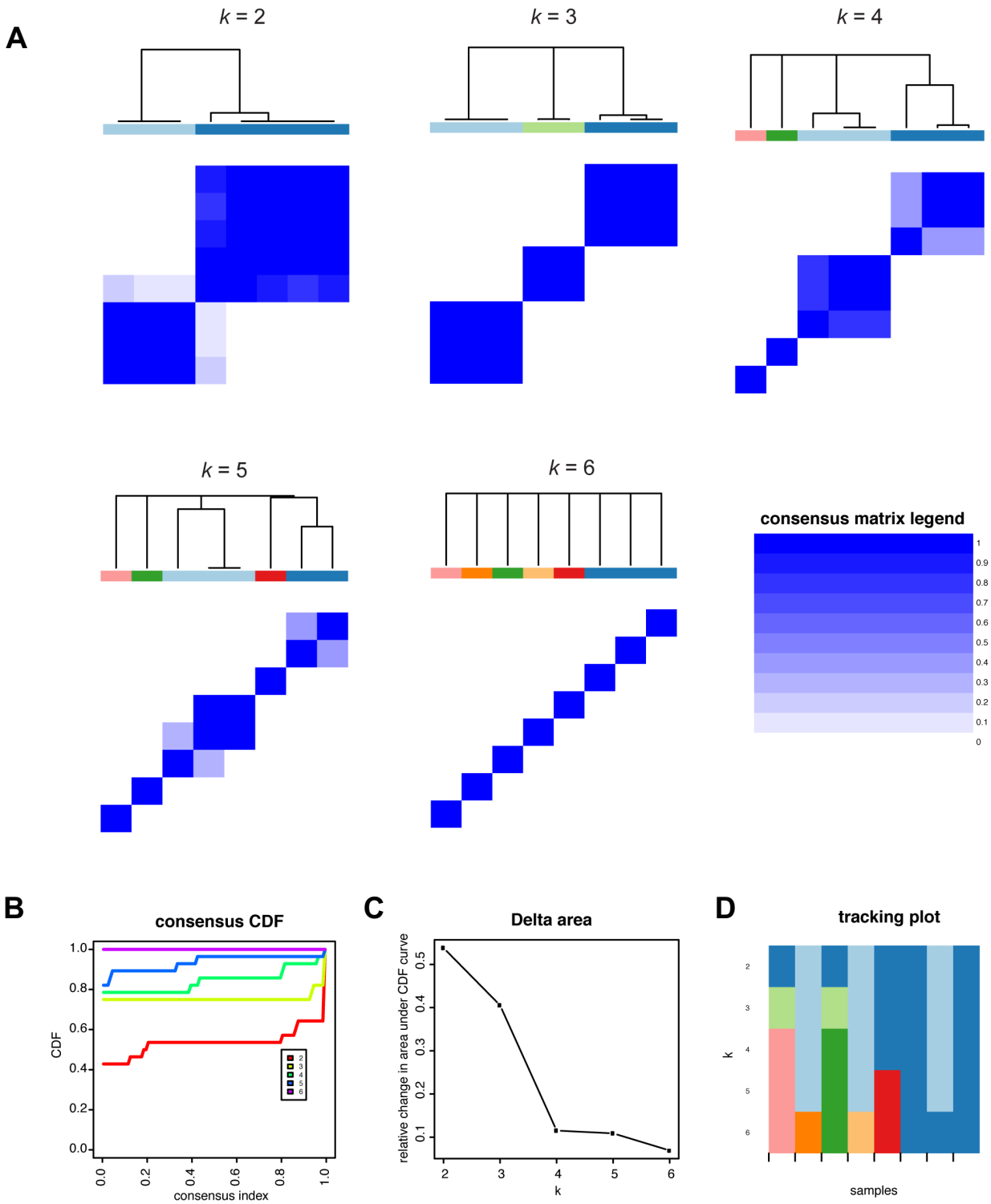


Figure 4.6: ConsensusClusterPlus matrices

The ConsensusClusterPlus algorithm visualises the consensus values between samples (between 0 and 1). Consensus Matrices with clustering dendrograms are generated for each *k* value. *k* = 3 is the 'cleanest' consensus matrix **(A)** and demonstrates the most appropriate maximum stability on the CDF plot **(B)**. *k* = 4 has the least increase in consensus **(C)** but as two of its clusters contain individual samples **(D)** it is not the most appropriate clustering value.

4.3 Global methylation in ADPKD cysts

The cyst samples were derived from tissue from patient ADPKD. Previous comparison of the median methylation of whole tissue samples had identified that this patient's genome was significantly hypomethylated compared to other samples in this analysis. The global methylation of each RRBS library used in the analysis of ADPKD tissue (excluding cell lines) was plotted (Fig. 4.7A).

This shows samples ADPKD D I, II and III had a considerably lower median methylation than other whole tissue counterparts, and that the median was also lower in cyst samples, particularly cysts 4 and 8. However, when the global methylome of each sample was plotted based on only the common analysed fragments between the 17 libraries (n = 36,888), the median value of methylation was very similar across all samples (< 0.1) (Fig. 4.7B). This suggests that the proportion of genomic fragments analysed was different in these hypomethylated libraries. To investigate this, the proportion of genomic elements in each library were calculated (Fig. 4.8) and the methylation of genomic elements were also plotted (Fig. 4.9).

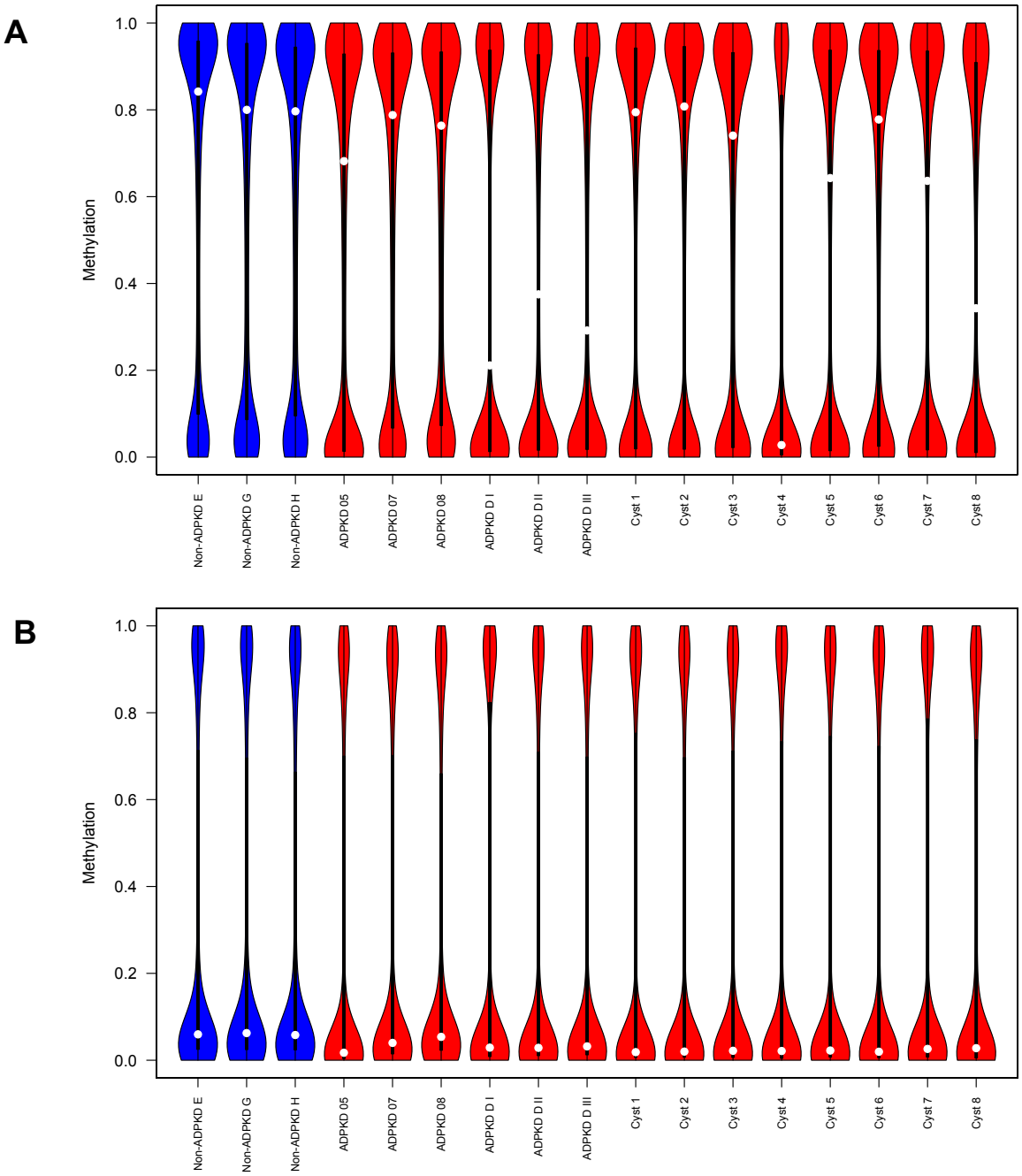


Figure 4.7: Global methylation of all RRBS libraries

The median methylation of each library is marked by the bar, interquartile values are represented by the bars, whiskers extend to minimum and maximum values. **A)** The total number of sequenced fragments in each library, n ranges from 51,805 – 397,796. **B)** Global methylome of RRBS libraries, based only on fragments common in all 17 libraries. n = 36,888.

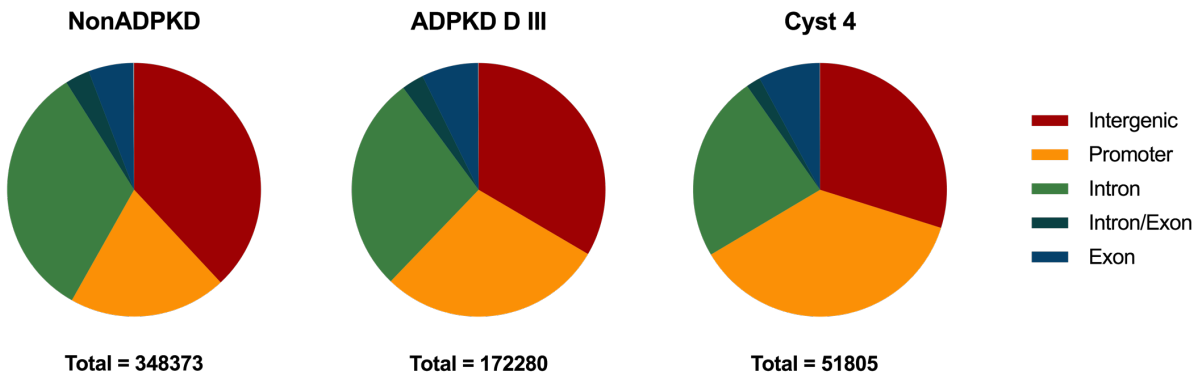


Figure 4.8: Proportion of genomic elements in RRBS libraries
The non-ADPKD reference library, ADPKD D III and Cyst 4 libraries demonstrate that the libraries that demonstrate low global methylation values have a higher proportion of promoter-associated fragments.

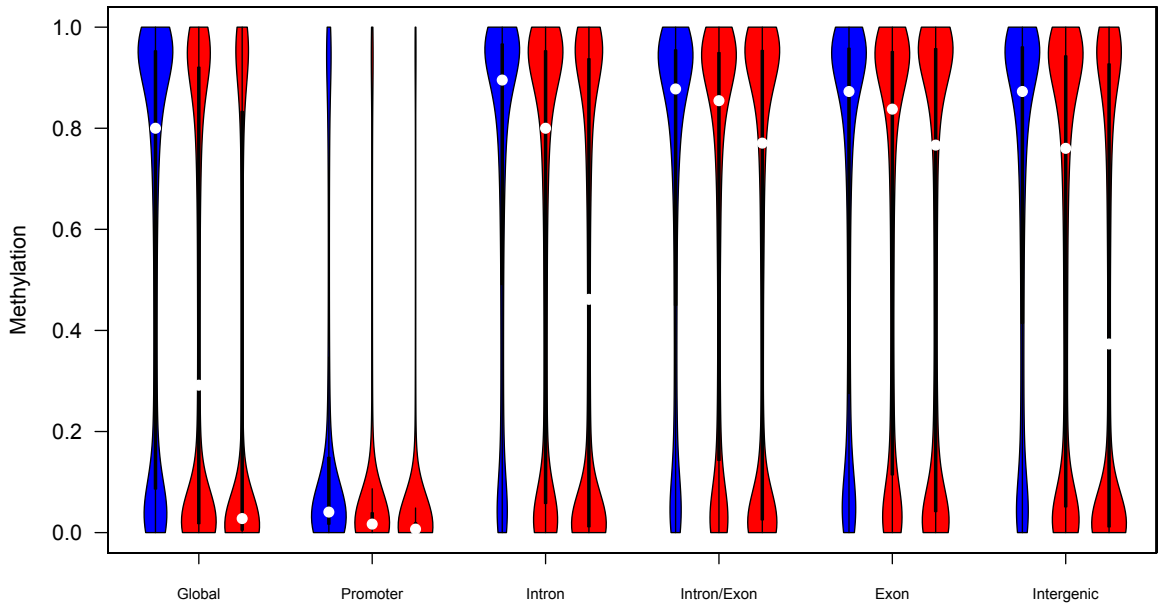


Figure 4.9: Methylation of genomic elements in RRBS libraries
The non-ADPKD reference (blue), ADPKD D III and Cyst 4 libraries (red) were analysed at each genome feature. This demonstrates that the low methylation values are not due to the overrepresentation of promoter fragments (which are typically hypomethylated).

From this it is apparent that while the hypomethylated libraries (i.e. ADPKD D, Cyst 4) have a higher proportion of promoter coverage, this is not (the sole) contributor of the hypomethylation pattern seen in the samples. Across all genomic elements, ADPKD samples had considerable

lower median methylation values (and therefore, the finding in Chapter 3 than ADPKD is globally hypomethylated is valid).

When ADPKD cysts were paired to the (grouped) non-ADPKD methylome by common analysed fragments, global methylation patterns were plotted (Fig. 4.10). In five of these cysts, the median methylation is higher than that of the non-ADPKD group. This is in contrast to the previously generated data we have produced from whole tissue, which demonstrates hypomethylation in ADPKD. Of the remaining three cyst libraries, one demonstrates hypomethylation and the other two have median methylation equal to that of non-ADPKD. In seven of the eight cysts, however, the interquartile range in ADPKD trends lower than that of non-ADPKD.

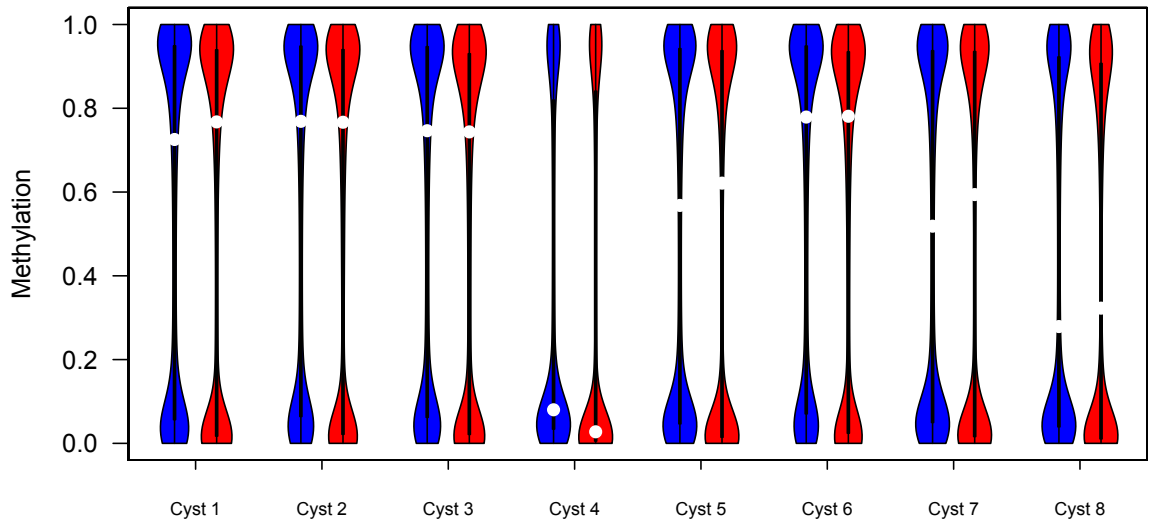


Figure 4.10: Global methylation of paired non-ADPKD and cyst RRBS libraries
Cyst libraries (red) were paired against the non-ADPKD reference genome (blue), and the methylation values of only the common analysed fragments in each pair are plotted on the violin plot. Five of the cyst libraries (1, 5, 6, 7 and 8) are statistically hypermethylated, as the median value (represented by the black bar) is higher in the cyst library.

4.3.1 Methylation status of previously identified DMFs in cysts

The specific DMFs identified within the whole ADPKD tissue were investigated in the cyst data by performing ANOVA analysis between the three non-ADPKD samples and eight cyst samples in order to provide consistent comparisons between the groups. Methylation scores were compared across three tissue types: non-ADPKD tissue, ADPKD whole tissue and ADPKD cysts. Of the 13 DMFs, 11 had coverage in at least two of the eight cysts; there was no coverage in the cysts at the *CPLX1* and *TMPRSS6*-associated fragments.

Eight of these fragments were determined to be significantly differentially methylated in the cysts (FDR-adjusted $p < 0.05$). The remaining three fragments (*KDELR2, DCLRE1C* and *APCDD1*-associated fragments) did not show the same difference in methylation as seen in whole tissue (Fig. 4.11).

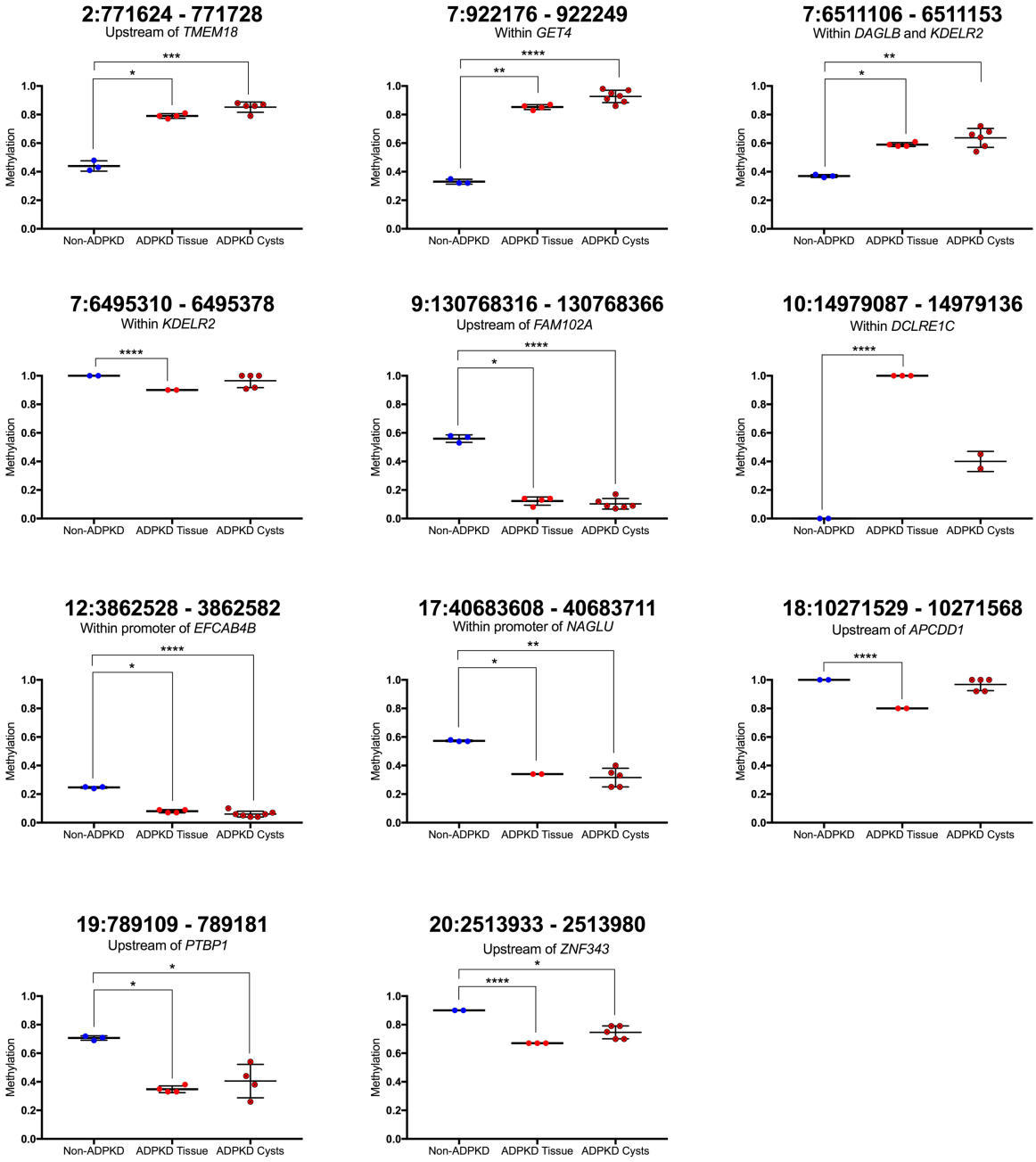


Figure 4.11: Methylation of the whole tissue-identified DMFs in cyst libraries
Statistical significance between non-ADPKD libraries and ADPKD cysts was carried out with ANOVA. Eight of the DMFs identified in ADPKD whole tissue (Section 3.2) were significantly differentially methylated in the group of individual ADPKD cysts. FDR-adjusted p values: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Chapter 5: Discussion

There are very few published data on DNA methylation in ADPKD. This study set out to identify the global methylation changes associated with ADPKD, and to identify changes within specific regions that may be involved in the disease. Additionally, it set out to determine if the analysis of whole tissue from ADKPD kidneys is able to interrogate the ADPKD methylome, or if unique changes occurred within individual ADPKD cysts that cannot be mapped by bulk analysis.

5.1 Quality of data

The RRBS libraries generated in this analysis demonstrated high Phred scores across the whole read lengths, which shows that the base calling was accurate, and that the methylation data observed was likely to accurately represent the methylome of the samples analysed.

RRBS has the capacity to cover ~4 million CpG sites across 647,902 fragments [144, 152], however, historically in experimental protocols it is typical to yield approximately 340,000 or fewer fragments [153]. This is because the established statistical analysis techniques in our group require there to be coverage at least ten reads deep on fragments containing two or more CpG sites. While these techniques reduce the quantity of fragments available for analysis, it reduces the likelihood of sequencing errors.

Bisulfite conversion was > 98% in the RRBS libraries, which is approximately the level expected based upon previous studies using this methodology [153] and indicates that the protocol was carried out successfully. The samples prepared from 0.5 μg of DNA had low

mapping efficiencies (< 60%), which is lower than expected for libraries prepared with this protocol. This may have been due to the reduction in DNA input to the modified protocol. The PCR amplification process typically results in the overrepresentation of adaptor dimers [154] which need to be removed with additional bead amplification, and two of the ten cyst samples originally collected could not be sequenced due to low DNA concentrations after the final elution. A suspected reason for this is that the protocol in this study used *PfuTurbo* DNA Polymerase for PCR amplification of the library. The methodology has since been updated in our laboratory to replace *Pfu* with KAPA Polymerase, which has shown an increased efficiency of library amplification during the PCR stage of the protocol. The reason for this increased efficiency is unclear as both polymerases are designed for high-fidelity PCR.

Another factor affecting the mapping efficiencies is the source material itself. The majority of samples with low mapping efficiencies and genome coverage were the individual cyst samples, which were very small sections of tissue dissected from neutral buffered formalin-fixed tissue. Formalin influences DNA solubility, resulting in lower yields of DNA than from fresh tissue [155]. It is also known to contribute to DNA fragmentation, resulting in a reduction in the quality of the yield [156, 157]. The kidney used for this collection had been treated in formalin for seven days prior to our access to the tissue, therefore the DNA was almost certainly degraded to some degree.

5.2 The ability of RRBS to assess the methylome of kidney tissue

RRBS is a reproducible and validated research technique for assessing the genome wide methylation of samples of DNA [152, 158-161]. The reliability of the methodology in this study was assessed using Pearson's Correlation Coefficients (PCC) between each RRBS sample. As

the RRBS library coverage largely varied between samples, the PCC was calculated from the pool of common fragments, to avoid any biases due to sampling.

Using parameters previously described [149], a PCC 0.95-0.99 suggests there is substantial concordance between the data, and PCC > 0.99 is almost perfect concordance. It is expected that technical replicates (ADPKD D II and III) should have a PCC approaching 0.99; this was seen in this study as the technical replicates had a 0.993 concordance value. This illustrates that the protocol and sequencing method was accurate in reporting the methylation of the tissue.

Sample ADPKD D I was a biological replicate of the two above samples, as it was prepared from an independent tissue section from the same patient kidney. The PCC between biological replicates was 0.971 (comparison between I and II) and 0.970 (comparison between I and III). While still indicating a substantial level of concordance between the samples, the 3% variation in these samples shows there is some heterogeneity between different sections of the ADPKD kidney. Each tissue sample was selected from different sites from the kidney therefore the exact cellular makeup of each tissue section is likely to be different.

Surprisingly, the biological replicates above have a slightly lower correlation than the correlation between unrelated ADPKD samples (mean PCC = 0.983). This may mean that ADPKD D I tissue has a different composition of cells compared to the other ADPKD tissue samples. However, this difference is still relatively small, as the methylomes still have substantial correlation, and additional biological replicates of the other ADPKD samples (if performed) may also show this variation.

The non-ADPKD samples are also closely related to each other, with a mean PCC between pairs 0f 0.987. The small amount of variance between samples is likely because each patient has their own unique methylation. An inter-individual study found a range of concordance between DNA methylation of individuals was between 0.93 and > 0.99, dependent on the tissue type (renal tissue was not reported) [162]. The non-ADPKD samples suggest that approximately 98% of the methylome is required for all kidney cells and remains invariant.

It was noted that the immortalised cell lines originating from ADPKD single cysts (WT 9-7 and WT 9-12) displayed different methylation patterns than any of the kidney samples. According to the previously mentioned parameters [149], between 0.90 – 0.95 is moderate concordance, and PCC < 0.90 is poor concordance. Cell line WT 9-7 had a mean PCC of 0.937 and WT 9-12 had a mean PCC of 0.882. While these cells are morphologically characteristic of epithelial cells (polygonal [163]) and characteristic of ADPKD [141], the immortalisation process may have altered the epigenetic makeup of the cells in order to evade senescence. There is experimental evidence to suggest that SV40 immortalisation of the cell lines means they cannot be relied upon to show methylation changes [164], and DNA methylation progressively changes with increased passages [165, 166]. Investigation of paired primary cells from primary and metastatic tumours shows no change in global methylation profiles [159], suggesting the biggest factor in the observed discrepancy in our data is immortalisation.

Additionally, there was a discrepancy between the two cell lines themselves. As they were from the same patient, they may be expected to have a high level of similarity, however, the concordance between the two cells is 0.951. The origin of the cells may contribute to global methylation differences between the samples; the cell line WT 9-7 has been characterised as originating from a proximal tubule, while WT 9-12 expresses markers for both proximal and

distal tubules [141]. Another possibility is that the immortalisation process has caused the methylomes of these replicating cell lines to be dysregulated, or the number of passages the cells have undergone prior to arriving in our laboratory may have contributed to changes in DNA methylation. While the exact cause is unclear, it does show that the cell lines no longer accurately portray the *in vivo* ADPKD methylome.

The ADPKD single cyst RRBS data were highly concordant with each other (mean PCC = 0.987). The cysts were biological replicates of each other, and biological replicates for the whole tissue samples ADPKD D I, II and III. The concordance between the cysts was higher than the concordance between the whole tissue biological replicates (mean PCC = 0.978). As alluded to above, the whole tissue samples are comprised of many cell types, and likely multiple cysts. The cell types of the kidney comprise of nephrogenic cells (renal tubules are formed of epithelial cells) [167] and the stroma, which includes fibroblasts and smooth muscle [168]. The individual cyst samples, however, were dissected with the aim of reducing the heterogeneity of the samples, both by reducing the number of cysts in each sample, and by reducing the proportion of stromal cells. Therefore, it could be concluded that the individual cyst samples better represent cystic ADPKD kidney tissue.

5.3 Hierarchical clustering

Hierarchical clustering (Fig. 4.4) of the common analysed fragments of all 19 libraries concurred with the correlation and variation values seen above. The immortalised cell lines WT 9-7 and WT 9-12 clustered the furthest away from all other RRBS libraries. As this cluster was so dissimilar to the remaining tissue samples, this supported the exclusion of immortalised cell line data from the methylation analysis.

The non-ADPKD tissue clustered separately from the ADPKD samples, demonstrating a difference in methylation based upon phenotype. The whole tissue samples 05, 07 and 08 were in one cluster, and the samples from patient ADPKD D (three whole tissue samples and eight individual cysts) were split into two clusters.

Technical replicates clustered tightly together, along with four of the cyst samples. The remaining whole tissue sample and four cysts were in an independent cluster. This suggests that there is some level of heterogeneity within patient ADPKD D to cause the splitting of clusters. One possible explanation for this is that these samples – particularly the individual cell lines – come from cysts that have originated from different cell types. It is known that cysts can arise from any region of the nephron [5], so it is possible that the characteristics of the originating tubule are preserved in the methylome.

However, there are no data on the precise origin of each sample (due to the distorted nature of the ADPKD kidney and limited amount of tissue for immunohistochemistry staining), and there are lots of variables in the collection of cystic tissue (such as cyst size) so this hypothesis cannot be followed up in the present tissue. Furthermore, tubule origin is unlikely to contribute to the difference seen between whole tissue samples, as each tissue sample likely contained a more than one cyst which may have originated from different regions of the nephron.

The deviation of methylation patterns in the clustering analyses could alternatively indicate independent pathways of DNA that might relate to the somatic mutation or change that initiated cyst formation. Previous data on prostate cancer has used DNA methylation alongside copy number data to map the unique clonal evolution pathway of tumours [169]. A similar analysis of intra-individual variation could be applied to the cyst data to investigate this. Animal models

provide a tempting method of investigating cystic clonal development and intra individual variation, however the discrepancies of gene dosage in these models (Table 1.1) and differences between *PKD1* and *Pkd1* sequences indicate that this data may not accurately reflect human ADPKD mechanisms.

5.4 DNA methylation analysis of whole ADPKD tissue samples

5.4.1 Methylome-wide sequencing in ADPKD tissue

The median methylation values of fragments in the libraries ADPKD 05 and D were significantly lower than the other ADPKD and non-ADPKD libraries, with reduced lower quartile methylation values. It was unclear if this was a biological feature of the two ADPKD libraries, or if there was an underlying methodology difference. These two whole tissue samples were generated by myself using the low-DNA input methodology, while the other five whole tissue samples were generated by Michael Bates with five times the amount of starting material. The data was exhibited across a variety of genomic factors, such as genomic elements and CpG island features (Appendix G), which revealed the difference in median methylation was consistent across investigated regions.

Additionally, when all the kidney tissue libraries were compared at only the common analysed fragments (n = 36,824; Fig. 4.7), all libraries now had a much lower and similar median methylation value, in line with the least methylated library (Cyst 4). This suggests that the lower-coverage libraries had a bias to generate sequences for unmethylated (or low methylated) fragments. As Supplementary Fig. G.1 shows, the reduced median methylation of ADPKD D tissue did not influence the overall global hypomethylation pattern seen in ADPKD, as this library simply did not have the higher methylated fragments. As such these data do not

influence the interpretation of the methylation data seen here but do provide limited power for a substantial analysis.

Figure 4.7 indicates that the libraries generated from low-DNA input had lower median methylation even when adjusted for common analysed fragments only. These data, combined with the histograms in Appendix G (Supplementary Fig. G.2), suggest the reduced DNA methodology has generated libraries with smaller fragments, concomitant to the reduction of median methylation. This is further discussed in Section 5.9.3.

5.4.2 Genome-wide methylation in ADPKD tissue

When the RRBS methylome is assessed by calculating the median methylation of all 345,711 fragments across the two tissue groups, the ADPKD genome has a methylation score 0.02 (2%) lower than the non-ADPKD genome. This difference in methylation is consistent across all genomic elements (Fig. 3.4). The Wilcoxon rank sum *p* value was < 2.2E-16 in all of these comparisons, so the value reported here is highly statistically significant and unlikely to be due to random chance. As ADPKD D has shown considerably lower methylation than the other samples in the analysis, the comparison between tissue groups was repeated excluding sample D (Supplementary Figure G.1). The difference between tissue groups was still 0.02, which confirmed that the low median methylation of sample D was due to the fragments in the ADPKD D genome originating from largely low-methylated fragments, as demonstrated in Section 5.4.1.

Although there was reduced coverage in ADPKD D compared to the other samples, this sample was left in the whole tissue analysis as it added strength to the ANOVA between ADPKD and non-ADPKD tissue and was not skewing the perceived global methylome of ADPKD tissue.

Global hypomethylation is a feature commonly associated with cancer and is believed to initiate chromosome instability and activate oncogenes in tumour cells [129]. Some oncogenes have been demonstrated to be aberrantly expressed in ADPKD tissue, and overexpression of these genes is sufficient to cause cystic disease in mice [170-172]. However, no DMFs were identified within any oncogenes in our ADPKD tissue analysis. Genomic instability has been suggested as a feature of ADPKD, including the development of somatic mutations in one of the *PKD* genes in order to initiate cystogenesis [97]. A hypomethylated genome, as is demonstrated here, could be sufficient to create an environment where somatic mutation occurs at a higher rate than in healthy tissue.

Despite this, it should be recognised that the amount of hypomethylation is only 0.02, or 2%. Global hypomethylation has been described in cancer, with studies reporting as much as 90% of the genome to be hypomethylated [173-176]. Thus, the 2% difference in methylation observed in our data is marginal, and so biological consequences, if any, arising as a result of this hypomethylation are unclear based upon the current understanding of global DNA methylation in cancer.

Due to the ambiguities of single-nucleotide genome alignment, RRBS does not have extensive coverage of repetitive elements [177]. As repetitive elements are often characterised as being areas that harbour significant amounts of hypomethylation in cancer [178], it is possible that if there was more coverage at these repetitive elements they would also be hypomethylated in the ADPKD tissue, making the magnitude of hypomethylation potentially more pronounced than that seen in this analysis. Other methods of methylation analysis, such as an EPIC-array analysis with predictive algorithms [179], could be used to address this hypothesis.

These global methylation data are also contradictory to the only other analysis of global methylation in ADPKD. Woo et al. report that 91% of CpG methylation sites with differential methylation (dCMES) were hypermethylated in their analysis [139]. A direct comparison between the two analyses cannot be made, as the estimate by Woo et al. is made by analysis of only the dCMES and not the genome at whole. Furthermore, our analysis is the summation of all fragments in the analysis, regardless of the statistical significance of each fragment between the two groups.

Woo et al. used the MIRA-Seq technique for the analysis of global methylation in their three ADPKD samples. MIRA-Seq pyrosequencing works by exploiting the affinity of two methyl-binding proteins for methylated CpGs, thus resulting in a higher proportion of methylated fragments in the final analysis than would be reported using RRBS. Additionally, this protocol is less likely to cover methylated CpG islands [180], and as it is an enrichment-based methodology, it is more prone to errors in methylation estimation especially when samples have high copy number variation, such as that seen in cancer [181, 182].

A cost-effective sequencing alternative to whole genome bisulfite sequencing (which has not yet been applied to ADPKD), RRBS uses a restriction enzyme digest that targets CpGs due to the cut site of MspI which is located within a CpG (C^CGG), irrespective of methylation. This method has an advantage over enrichment sequencing methodologies, which have an affinity for methylated cytosines [183]. Additionally, this methodology allowed us to analyse a greater number of CpG sites across both tissue groups, at single nucleotide resolution: Woo et al. generated 1 million CpGs in non-ADPKD and 0.7 million in ADPKD compared to our 1.7 million CpGs in both tissue groups (Table 3.4).

Considering the different sequencing approaches, our data indicating global hypomethylation provide a more conservative estimate of the ADPKD tissue methylome (with some data suggesting this is variable across individual cysts). Our study has a number of strengths, including similar age-matched samples and single nucleotide resolution.

5.5 Differentially methylated fragments

We identified 13 fragments with significant differential methylation in ADPKD tissue. These regions have not been identified in the previously published DNA methylation research in ADPKD [139], although this paper did not publish a full list of all differentially methylated sites and therefore we cannot rule out a similarity between experimental findings.

Thirteen is a relatively low number of loci to identify in an RRBS analysis. Woo et al. identified more regions in their 2014 paper using a different sequencing approach. It is possible that different methodologies contributed to the difference in quantity of DMFs. As mentioned above, MIRA-Seq analysis is based upon enrichment analysis, which is much more prone to errors [182] and does not have the same CpG resolution as RRBS.

Another possibility for these differences is that the ANOVA test did not have sufficient power to identify additional DMFs due to the lower coverage in ADPKD libraries 05 and D I. This appears to be the result of using the adjusted RRBS methodology with reduced input amount of DNA (Section 5.2).

Additionally, each patient may have a different methylation landscape due to other factors influencing their disease pathology. It is possible that there are several different factors

influencing or causing this pathology, and the DNA methylation seen is a consequence of, or a response to, these factors. As ADPKD is a heterogeneous disease, these factors could even differ between patients. Regardless of what causes the variation, this has led to higher FDR-adjusted *p* values which in turn renders a lot of the additional observed differences between the disease-state groups as statistically insignificant.

There were more hypomethylated DMFs (eight) than hypermethylated (five). There was a slight enrichment of DMFs within intergenic regions and intron/exon regions, although due to the low number of total DMFs this limits the significance of any observed trends.

5.5.1 Intergenic DMFs

Five of the DMFs were located in regions outside of the protein-coding region (i.e. not within the promoter or gene body). As these fragments are as far as 183 kb from the nearest protein-coding gene, they provide tenuous links to the closest genes and thus only brief summaries of these gene functions and their potential roles in ADPKD are presented below.

APCDD1 is an inhibitor in the *Wnt* signalling pathway, where increased expression has been associated with colorectal tumourigenesis [184, 185]. Hypermethylation of the *APCDD1* has also been implicated the increased expression of the gene in osteosarcoma [186]. *APCDD1* has shown expression in skin and fat tissue among others, but relatively low expression in kidney [187]. As previously described, the ADPKD tissue samples were noted to contain high levels of adipose deposits, particularly within the medullary regions and surrounding the external kidney walls. Therefore, the altered methylation value may be associated with the presence of adipocytes in the tissue samples.

There was methylation data available from five of the eight single ADPKD cysts for this *APCDD1* fragment. These data were not as hypomethylated as the whole tissue samples, and ANOVA analysis did not find this fragment differentially methylated in the ADPKD cysts.

The transcription repressor *TMEM18* has various roles reported in literature, including neural cell migration and obesity [188, 189]. One possible link to its role in ADPKD tissue is that a 2014 paper identified that the promoter methylation of *TMEM18* and subsequent reduced mRNA levels were related to the deposition of visceral adipose tissue [190]. Much like *APCDD1*, the altered methylation associated with this gene may be linked to the presence of adipose tissue in ADPKD.

FAM102A is a poorly characterised gene, believed to play a role in estrogen action [191]. There is very little evidence to connect it to ADPKD, with the exception that as it is a hormone-related gene, sex differences between the ADPKD and non-ADPKD tissue groups in this analysis (discussed further in Section 5.9.4) may cause the differences in methylation seen here, rather than disease itself.

ZNF343 is ubiquitously expressed in all tissue as it is required for the regulation of gene expression in all cells [187]. It is important for nucleic acid binding, and there are no syndromes associated with it. Therefore, any role it may be playing specifically within ADPKD is unclear.

PTBP1 is a translational regulator in the cell, involved in pre-mRNA and alternate splicing. Regulated by miRNA, expression of this gene is associated with several forms of carcinoma [192]. Like *ZNF343*, there is no clear connection to ADPKD for this gene. Additionally, *PTBP1* was the only gene with an intergenic association to a DMF to be investigated for gene

expression as PCR primers were readily available, however no difference was seen between tissue groups. While analysis of the single ADPKD cysts found the difference in methylation at this fragment to be statistically significant, there was a wide spread in the methylation of this fragment across the single cyst samples. This range of DNA methylation was not seen in the whole tissue samples.

Both of the DMFs associated with *FAM102A* and *PTBP1* were predicted to contain enhancer elements, supporting the belief that the genes closest to the DMFs themselves may not be significant to ADPKD, and instead suggesting that these genetic regions are associated with regulating other active pathways.

5.5.2 Intragenic DMFs

There were eight DMFs associated with protein-coding genes: within the promoter (two) or gene body (six). Although the function of all eight genes are discussed below, the involvement of these genes in ADPKD is unclear based upon methylome studies alone. Many of these genes also showed no change in RNA expression, and thus there is a possibility that the methylation changes are consequential of the disease state, not factors in the cystogenesis.

5.5.2.1 *CPLX1*

Essential for motor function, *CPLX1* is a cytosolic protein involved in the exocytosis of synaptic vesicles. The product of *CPLX1* binds to a SNAP receptor complex at the cellular membrane, allowing transmitter release when there is sufficient action potential (Ca^{2+}) at the synapse [193]. Given that it is an important regulator in neural synapse, *CPLX1* is expressed in the brain in higher quantities than other organs [187].

The link between ADPKD and a neural protein is unclear, however one feature of this protein does relate to ADPKD: *CPLX1* action is dependent on the concentration of Ca^{2+} - a molecule that has had a lot of attention in ADPKD as *PKD2* is a calcium channel, and the concentration of Ca^{2+} is thought to be involved in cystic disease.

A qPCR targeting this gene was performed on ADPKD tissue, however no difference was seen between non-ADPKD and ADPKD tissue.

5.5.2.2 *DAGLB*

Diacylgylerol lipase beta (*DAGLB*) is an enzyme that catalyses the hydrolysis of diacylglycerol to 2-arachinodonoyl-glycerol. This is the most abundant endocannabinoid in human tissue. This is the second of the genes associated with ADPKD that plays a role in neuronal signalling (the other being *CPLX1*). It is required for retrograde synaptic signalling at mature synapses – when post-synaptic cell bodies release a retrograde messenger to travel 'backwards' to bind at the axon terminal of a presynaptic neuron [194]. *DAGLB* is ubiquitously expressed in human tissue, with higher expression in some tissue types such as brain [187], but no noticeable renal significance. qPCR was not performed for this gene as primers with specificity to the gene sequence were not generated in the timeframe.

In addition to retrograde signalling, DAGLB is required in the developmental period for axonal growth [195]. It is unclear why an axonal gene may be triggered in ADPKD, but the involvement of a gene required for development is not necessarily surprising, as the ADPKD cystic cell has been characterised to resemble fetal tissue in many aspects [5].

5.5.2.3 *KDELR2*

This gene is a member of the KDEL (Lys-Asp-Glu-Leu) receptor family, which recognises and binds to specific resident soluble proteins in the Golgi body, returning them to the lumen of the endoplasmic reticulum. The KDEL family is responsible for the retention of the luminal ER proteins, and combined, determine the specificity of this system. This is required within cells throughout the body, and thus *KDELR2* is ubiquitously expressed in human tissue [187, 196].

The DMF 7:6511106 – 6511153 is located within the transcripts of both *KDELR2* and the aforementioned *DAGLB* as these genes run on opposing strands. Additionally, both genes have multiple splicing transcripts. Therefore, it is possible that *KDELR2* is not involved in ADPKD pathology, just close in proximity to a region of differential methylation. This is substantiated by gene expression observed by qPCR, which shows almost no difference between ADPKD and non-ADPKD tissue groups, with the exception of one outlier. Further evidence is found in the methylation of the data of the ADPKD cysts at this DMF, which do not show statistically significant methylation at this loci when compared to the non-ADPKD tissue.

5.5.2.4 *EFCAB4B*

Most commonly referred to in the literature as *CRACR2A*, *EFCAB4B* is a calcium release activated channel regulator, playing a key role in the store-operated entry of Ca^{2+} at the calcium release activated channel (CRAC), particularly in immune cells. CRAC channels require EFCAB4B and STIM1 binding to facilitate the flow of Ca^{2+} into the cell. STIM1 is located on the ER membrane and is sensitive to the stored Ca^{2+} within the ER, while EFCAB4B is a cytosolic protein, dissociating the CRAC-STIM1 complex in high cytosolic concentrations of Ca^{2+} [197].

While broadly expressed in human tissue, the expression of *EFCAB4B* in kidney tissue is recorded as being low [187]. In ADPKD tissue no statistical difference in expression was seen (Appendix I).

The *EFCAB4B*-associated DMF, which is located in the promoter has been identified to contain a binding site for the transcription factor *EGR1* (Early growth response protein 1), which is required for development, and thought to be a tumour suppressor gene . *EGR1* binds to genes irrespective of cytosine methylation [198], and plays a role in cellular growth, proliferation and death, possibly by means of regulating DNA damage [199].

Additionally, *EFCAB4B* is predicted to be an enhancer of the gene *TULP3*, a negative regulator of a signalling transduction pathway which is recruited to primary cilia [200]. *TULP3* regulates ciliary G protein-coupled receptor trafficking, and has only recently been reported as playing a regulatory role in ADPKD [201].

5.5.2.5 *DCLRE1C*

As *DCLRE1C* was found to contain a DMF with 100% hypermethylation in ADPKD tissue, the potential functional role of this gene in PKD is very interesting. As a cross-link repair gene encoding the endonuclease ARTEMIS, it is involved in V(D)J recombination and DNA repair for double-stranded breaks.

DCLRE1C may be required most within the immune system, given that it has higher levels of expression in the lymph nodes than other human tissues, and null *DCLRE1C* mutations causing ARTEMIS depletion result in severe combined immunodeficiency, and hypomorphic mutations give rise to a number of other immunodeficiency disorders [202].

The hypermethylated DMF in question is within intron 4 of *DCLRE1C*. According to the paradigm of methylation, this would be predicted to result in a sustained or increased expression of the gene. In contrast to patients with immunodeficiency above, the hypermethylation of the *DCLRE1C* gene body could increase the efficiency of ARTEMIS production, in turn increasing the efficiency of DNA repair. One possibility to explain why this may occur is that an increase in production of ARTEMIS could be an attempt to compensate for the genomic instability caused by the cystic disease.

There was only methylation data from two ADPKD cyst samples at this site, which did not agree with the whole tissue data. While the cyst samples had more methylation than the non-ADPKD tissue, they were not as methylated as the whole tissue, and were not statistically hypermethylated compared to the non-ADPKD tissue. It is unclear why these samples had such lower methylation when the whole tissue samples showed a strong binary pattern between the two disease states. There is no data at this site from the whole tissue from patient D (the same patient from which the cyst samples come from).

Interestingly, when qPCR was performed with *DCLRE1C* primers, there was no significant difference in expression between the two tissue groups. With the exception of one ADPKD outlier with markedly higher expression, the expression of *DCLRE1C* was almost the same between ADPKD and non-ADPKD groups. As alluded to above, methylation of a gene body is theorised to support the ability of transcriptional mechanisms to interact with the gene to express it, but is not always associated with an increase in gene expression.

5.5.2.6 *TMPRSS6*

TMPRSS6 encodes a protease found on the cell surface, required for matrix remodelling of the liver. *TMPRSS6* hydrolyses type I collagen, fibrinogen and fibronectin as part of its role in degradation of the extracellular matrix. Additionally, *TMPRSS6* plays a role in hepcidin production (which usually occurs within the liver). The expression of hepcidin is also tied to inflammation and infection; the reduction of hepcidin production is intended to restrict the iron available for pathogenic growth [203].

This gene shows highest expression in the liver, with very low expression in the kidney. *TMPRSS6* is hypermethylated within the gene body in ADPKD tissue (across intron 1 and exon 2), which could indicate that expression is increased, or at least stabilised, in ADPKD cells. The expression data generated in our laboratory showed no difference between the two groups.

5.5.2.7 *NAGLU*

The product of *NAGLU* is an enzyme that degrades heparan sulphate (HS) by hydrolysis. While ubiquitously expressed in human tissue, *NAGLU* is highly expressed in kidney tissue [204] [187]. Mutations in this gene result in the accumulation of HS, causing the disease mucopolysacchardosis type IIIB (MSP-IIIB), a neurodegenerative disorder [205]. Lower cellular temperature results in a higher activity of mutant enzymes in MSP-IIIB fibroblasts, although the mechanism controlling this is unknown [206].

This fragment was hypomethylated in the promoter in ADPKD tissue as evidenced by the analysis of two whole tissue samples. Although the ADPKD single cyst samples showed a slightly wider range in methylation at this fragment, the methylation was considered significant and supports hypomethylation in ADPKD.

In preliminary expression analysis with qPCR, *NAGLU* was shown to be reduced (Δ = -0.68, p = 0.0357) in ADPKD (Fig. 3.7). Hypomethylation of a promoter is typically associated with the increased expression of the gene in question, so it is not as common to observe reduced gene expression (although this phenomenon has been described before [129]). There is no evidence to explain a direct relationship between the promoter of a gene and the subsequent reduction of gene expression, although it is plausible that the DMF methylation is leading to the overexpression of a repressor of *NAGLU.* The possible consequences of this are that the concentration of *NAGLU* in the cystic tissue is reduced, in turn increasing HS. HS regulates a wide range of biological activities, including development, angiogenesis (blood vessel growth) and tumour metastasis. It is a component of the extracellular matrix and is present on the surface of fibroblasts to allow cellular adhesion [207].

Additional methylation analysis on more ADPKD tissue samples should be performed to elucidate whether there is a significant correlation between the methylation of this promoter and the reduction of *NAGLU* mRNA. Further experimental research is also required to identify whether *NAGLU* expression is altered due to the presence of inflammatory responses, as previously proposed [208], or due to molecular mechanisms regulated by HS. The addition more samples in this test would also help elucidate whether these changes are correlated with the methylation of the gene promoter.

5.5.2.8 *GET4*

GET4 is a chaperone protein, ubiquitously expressed across all tested human tissue types [187]. This protein works in a complex alongside BAG6 and UBL4A, which regulate ER-associated degradation [209]. This is a cytosolic protein quality control complex which mediates DNA

damage signalling and cell death. GET4 and UBL4A are chaperones to BAG6, and the loss of both proteins is required to elicit resistance to cellular death. However, the combination of all three genes is required for optimal DNA damage response signalling, wherein the complex recruits BRCA1 to sites of DNA damage.

Initial expression analysis of *GET4* demonstrates a significant reduction in gene expression in ADPKD tissue ($\Delta = -0.941$, $p = 0.0357$), with a strong negative correlation with the methylation of the DMF ($r = -0.919$, $p = 0.0071$). The *GET4*-associated DMF is strongly hypermethylated, with a difference of 0.52 between the non- and ADPKD tissue groups, a difference that has not been previously described in ADPKD. This fragment is located within the gene body (intron 1), which is purported to strengthen the transcription of a gene, which is conflicting with the data presented here. However the fragment, while not defined as a promoter in this study (it is further than 1kb into the gene transcript), is at the 5' end of the gene. It is typical that the methylation of the gene body enforcing gene transcription is typically seen towards the 3' end, thus it is plausible that this methylated fragment is operating as a promoter with regulatory elements.

Not only is there strong evidence to suggest a probable link between the methylation and expression of this gene in ADPKD, there has been a recent publication which identified another ADPKD-associated gene involved in ER-associated protein degradation (*DNAJB11*). *DNAJB11* is another co-chaperone molecule in which novel mono allelic mutations cause atypical ADPKD with maturation and tracking defects, causing ADPKD but not causing enlargement of the kidneys [20, 210, 211].

5.5.3 Summary of DMFs

Of the DMFs identified in this analysis, many are involved in signalling pathways or the structure of the cellular environment. These functions are expected in ADPKD, as signalling pathways including the roles of *PKD1* and *PKD2* have long been associated with the disease, as have physical characteristics such as increased fibrinogen, thick basement membranes, and increased collagen adhesion properties.

Other DMFs, such as *DCLRE1C* and *GET4*, have roles in DNA damage and protein quality control. A review from Johnson and Collis [212] has highlighted the relationship between cellular stress in ciliopathies, and the role of the DNA damage repair (DDR) pathway. The breakdown of the DDR pathway could be postulated to cause the dysregulation of cellular control (thus leading to cellular proliferation), and investigation into the potential function of these genes in ADPKD requires further research.

However, there are DMFs identified through this analysis for which the roles of the associated genes are within seemingly unrelated pathways. The genes *CPLX1* and *DAGLB* are the most obvious examples of this, as they are both neuronal signalling genes. There is limited evidence to suggest that PC1 plays a role in the epithelial lining of the cerebral cortex of the brain, and mutation within the genomic neighbour of *PKD1*, *TSC2*, causes tuberous sclerosis complex disorder in which tumours form in the brain. However, given that these genes (*CPLX1* and *DAGLB*) were identified as containing differential methylation in kidney tissue, completely isolated from brain tissue or spinal fluid, it is unlikely that this network of genes is at play here.

The most reasonable explanation for this is that these methylation patterns are arising from the distal synapses of neurons located within the kidney. An alternative explanation could be that

these regions of the genome have transcriptional roles that have not been identified in the analyses performed here (neither the *CPLX1* nor *DAGLB*-associated fragment contain transcription factor sequences nor enhancer elements).

The most promising genes identified in this analysis were *NAGLU* and *GET4*, which demonstrate differential methylation and concomitant reduction of mRNA. An appropriate follow up on this data would be to investigate whether the relationship between these observations is correlated; i.e. whether the differential methylation of these fragments is sufficient to cause reduced gene expression, possibly by the use of DNMT inhibition.

5.6 Global methylation in ADPKD cysts

The single cysts in this analysis are derived from patient ADPKD D (as single cyst tissue was not available for the other samples). The tissue sample from this patient in the whole tissue analysis (ADPKD D III) has the lowest global median methylation value of all the tissue samples. Both ADPKD D III and ADPKD 05 had lower median methylation and slightly lower interquartile methylation ranges, which may indicate that the altered RRBS protocol had a bias in which fragments with high DNA methylation may not be as easy to amplify and sequence from low amounts of DNA such as that used in the individual cyst analysis.

The ADPKD cyst libraries largely had fewer fragments with high methylation values, and subsequently had lower median methylation values (Fig. 4.7A). When each cyst sample was paired by common analysed fragments with the non-ADPKD sample (Fig. 4.10), the non-ADPKD genome reflected the methylation profile of the cyst in question, which shows that it was a technical aspect of the sequencing process which has caused the fragments with lower methylation values to be sequenced. The cysts were generated from small amounts of formalin-

fixed tissue, which may have contributed to the lower median methylation values. This is discussed further in Section 5.9.3.

Unlike the above analysis on whole tissue, the individual cysts did not all demonstrate global hypomethylation when compared to the non-ADPKD sample. Five of the cysts demonstrate statistically significant hypermethylation. One cyst demonstrated hypomethylation (as seen in whole tissue) and the remaining two samples have median methylation equal to that of the non-ADPKD sample. This pattern of DNA methylation changes is reflected in the different grouping of cyst libraries by the hierarchical clustering.

The differences between individual cysts may reflect different tubules of origin, such as the proximal or distal tubule. If this analysis were to be repeated, it would be interesting to perform immunohistochemistry on additional sections of the same cyst wall, in order to determine if these differences reflect the presence of proximal or distal tubule epithelia. Other reasons for the differences in methylation may include the disease trajectory of independent cysts.

5.7 *PKD1*

In contrast to previous data on the DNA methylation of *PKD1*, which showed hypermethylation of the gene body [139], this study found no statistically significant differential methylation surrounding *PKD1* in ADPKD, although there are some regions trending towards a difference. While the data does not fully span the 50 kb gene, there is a 31 kb region (exons 2 – 46) with RRBS coverage.

There is not sufficient coverage at the promoter in our data to define its methylation status, however it has previously been reported that there is little methylation in this region in both

non- and ADPKD tissue. Woo et al. report that within a portion of the gene body (exon 43), ADPKD is significantly hypermethylated. There was RRBS coverage flanking this exon, and the methylation profiles of these fragments, while not statistically significant, support the notion that ADPKD has increased DNA methylation at this region.

There are fewer fragments in the ADPKD single cyst samples that cover the *PKD1* region, but due to consistent methylation values between samples, several fragments were identified as being significantly methylated when comparing all cyst samples against all non-ADPKD tissue with ANOVA. These data concur with the observation in whole tissue and with Woo et al.'s published data; ADPKD is hypermethylated at the 3' end of the gene.

While the data in the current study shows the same trend in DNA methylation of *PKD1* as Woo et al., the expression of the gene does not. It was demonstrated by Woo et al. that the hypermethylation of the *PKD1* gene body was correlated with the reduction of *PKD1* expression, through treating cell lines with demethylating agents to promote hypomethylation and a reduction in cyst formation, leading the authors to conclude that increased gene body methylation (within exon 43) reduces the expression of *PKD1*, sufficient to initiate cystogenesis in ADPKD. The relationship between *PKD1* methylation and expression was demonstrated on cell lines including WT 9-12. However, these changes were not seen in ADPKD tissue in our analysis, with a trend of increased *PKD1* expression in ADPKD.

The disparity between results is puzzling, but not altogether surprising as there are conflicting data on the dosage of *PKD1* in ADPKD in published literature (Table 1.1). Adult human kidney tissue typically has little-to-no *PKD1* expression (*PKD1* is required in fetal renal cells [213]), but staining of ADPKD tissue has consistently demonstrated the presence of *PKD1*, including

in the cytoplasm. However this is not always the case, with multiple papers reporting a significant minority of tissues sampled containing no increased *PKD1* expression [6, 62].

One reason for the differences may simply be the samples used for gene expression. The initial *PKD1* expression data were generated from whole tissue samples as in this analysis, and the data from Woo et al. suggest a wide range of expression was seen between these samples, similar to what was observed in our data (Fig. 3.10). Woo et al.'s subsequent analyses on DNA methylation were performed on transformed cell lines, which, as demonstrated in Section 4.2, appear to have an altered methylome when compared with samples isolated from fresh patient tissue. Additionally, the 5-aza-dC treatment, used as a demethylating agent, acts by reducing the activity of the DNMT1 enzyme, and thus does not specifically target the gene of interest [214]. Therefore, *PKD1* will not have been the only gene influenced by demethylation and there is a huge number of genes which would have been influenced by DNMT inhibition to reduce cyst growth that were not investigated.

While we agree with the previous data that shows there is hypermethylation in the gene body of *PKD1,* it is not yet clear whether this is having an effect on gene expression. As the current expression data in this study has only been generated from mRNA, further investigations of the protein using western blots may demonstrate complexities of *PKD1* transcription and translation not seen by qPCR, such as whether the transcribed gene is functional.

5.8 microRNA

Analysis of the microRNA located within the *PKD1* gene was performed as these genes are associated with CpG islands, and have the ability to work as post-translational regulators. Given that *PKD1* is such a significant gene in ADPKD, and there is evidence to suggest that DNA

methylation occurs within the gene, it was decided to investigate the three miRNA located within the gene promoter and body to determine if DNA methylation had any influence on their expression.

It was observed that two of the genes (miR-1225-5p and miR-3180-5p) have little-to-no expression in the kidney, and therefore they were not altered by any epigenetic influences occurring in ADPKD.

However, miR-4516 was found to be expressed in the kidney tissue of both non- and ADPKD samples. miR-4516 is found within intron 1 of *PKD1*, 2.6 kb from the TSS. The preliminary analysis of miR-4516 expression in ADPKD tissue showed that the expression varied widely between samples. While this did not provide statistical evidence to support the loss or gain of miR-4516 in ADPKD, it does appear that the gene is dysregulated.

Currently there is little information on the function of this miRNA, with literature suggesting that the downregulation of this gene leads to the accumulation of fibronectin and thickening of extracellular matrix (by negatively regulating fibronectin 1 – *FN1*) in psoriasis [215, 216]. Although psoriasis is an epidermal disorder, increased fibronectin and thickening ECM is typical of ADPKD cysts [217, 218]. While there is insufficient evidence currently to confirm that miR-4516 is influenced by *PKD1* expression, or that this is consequential to the cystogenesis of ADPKD kidneys, this is a mechanism which may require further investigation.

5.9 Experimental limitations

5.9.1 Immortalised cell lines

Human immortalised cell lines are the most readily available and renewable source of nucleic acids that can be used in the molecular analysis of ADPKD. Both of the cell lines used in this project, WT 9-7 and WT 9-12, have been studied in a multitude of ADPKD studies, including the only other analysis of DNA methylation in ADPKD [139, 219-222]. Therefore, they are an established and well characterised sample for which direct comparisons can be made across studies.

However, these cell lines are immortalised with wild type simian adenovirus 40, which enables them to proliferate indefinitely. The integration of the virus into the cell has had a consequence on the methylation of the cell line, as demonstrated by the correlation coefficients between the cell lines and whole tissue (Section 4.2). Loghmann-Adham et al. demonstrated that WT 9-12 cells grow twice as fast as their counterparts immortalised with a temperature sensitive virus [141]. Therefore, in this study we have identified that the immortalised cell lines have limitations on their application in the analysis of the methylome, and possibly other epigenetic studies in ADPKD.

5.9.2 Limitations of whole tissue sections

The first part of this study utilised tissue taken from whole cystic human kidney. Whole tissue ADPKD sections likely contain several cell types from cystic and interstitial tissue [223]. Additionally, it is difficult to determine from which region of the nephron a dissected tissue sample has originated from, because the kidney morphology becomes distorted due to the cystic overgrowth. Assumptions about the tubule of origin can be made from staining a tissue sample, however this was not performed in this study.

Isolating tissue from individual cyst walls reduces the number of cell types and the amount of interstitial tissue in each sample, and the hope of generating RRBS libraries from this was that these would better illustrate similarities and differences between cysts. However, these walls were dissected with surgical scissors, and the selection was 'blunt' (as compared to laser dissection or similar techniques [224]) meaning that there is likely still neighbouring connective tissue in each wall sample, rather than pure epithelial cells. Additionally, it was not possible to determine exactly where in the kidney they originated from (i.e. cortex or medulla), due to the distorted nature of the cystic ADPKD kidney morphology. However, these samples can be said to be enriched for epithelial cyst wall cells.

5.9.3 Formalin-fixed tissue provides experimental complications

The individual cyst samples were generated from tissue which had been fixed in formalin for seven days prior to our access to the tissue. Although care was taken during DNA extraction to optimise the yield of DNA (Section 2.1.1), NBF treatment can significantly degrade DNA. This occurs due to cross-links formed by formaldehyde [225].

It has been shown that leaving tissue in NBF for longer than 24 hours reduces the yield of DNA from cDNA [155]. It is possible that this study is assessing the ability to amplify from formalin fixed DNA, as the means of assessment was qPCR, and formalin is known to increase the proportion of enzyme inhibitors [226]. Given that the RRBS protocol requires PCR amplification, it may be that the amplification of DNA is inhibited, rather than a lack of sufficient DNA from the formalin fixed tissue.

Additionally, NBF-treated tissue yields largely fragmented DNA [227]. As shown in Appendix G, the cyst libraries typically had a skewed distribution of fragment lengths, predominantly featuring small fragments. This likely had an effect on the coverage of the aligned fragments across the methylome, which could in part explain why these libraries had lower coverage than other libraries in the analysis.

Furthermore, NBF may cause the deamination of DNA, including 5-methylcytosine residues which deaminate to thymine in the presence of formalin-fixation at a higher rate than to uracil [228, 229]. Perhaps the deamination of these 5-methylcytosine residues reduces the ability of MspI to cleave the recognition site CCGG, leading in turn to reporting a higher proportion of unmethylated fragments. This hypothesis is supported by data produced in Chapter 4 (Fig. 4.7A), which demonstrated that the amount of coverage in the RRBS libraries contributed to the median methylation value.

It could be hypothesised that the fixation of the tissues had an impact on the methylation, as the cyst libraries largely had lower median methylation than other RRBS libraries. Yet, some of the samples from whole tissue also displayed lower median methylation (Fig. 4.7A). There are two factors which possibly contribute to this observation. The first is that these low-median samples are almost exclusively from patient ADPKD D. There is a small chance that this patient is truly, globally hypomethylated at a higher percentage than other individuals. The second and more significant factor is that these hypomethylated-median libraries were generated from a protocol using five times less input DNA. While there are other published protocols for generating RRBS libraries from low input DNA that generate ~4 million CpG reads [160, 161], this analysis suggests caution should be used when generating libraries from formalin-fixed tissue, particularly when working on small amounts of input DNA.

5.9.4 Sample population differences

The composition of environmental factors in each sample population will likely have an influence on the DNA methylation. To limit this, it is best to draw data from populations which have common backgrounds (such as age-matched samples). Unfortunately, the complication of nephrectomy is such that kidneys are rarely removed from patients, and thus there is a limited pool of available tissue for analysis.

The most significant contributors to variation between populations are likely to be age and sex (outlined in Appendix A). All the samples in this analysis were within one decade, minimising the variation due to age. Unfortunately, the only non-ADPKD tissue available through the tissue bank was female cortical renal tissue, and our ADPKD samples are predominantly male. This creates a potential environment for bias in the analysis, as females have one X chromosome silenced through DNA methylation. While none of the DMFs identified in this study were sex-linked, the global methylation patterns between the two tissue groups were reassessed by eliminating sex-linked chromosomes from the overall comparison (Appendix G). No change in global methylation was seen in any of the tissue samples, and ADPKD tissue remained globally hypomethylated by 2%. Therefore we can be assured that any effects of X chromosome silencing did not influence this trend.

5.9.5 Considering the medical history of patients

A consideration when performing studies on the epigenetic landscape in human tissue is that these patients were exposed to environmental factors which may be influencing the epigenetic profile of individuals. A multitude of unique aspects such as diet, physical activity and other lifestyle factors contribute to the unique profile of an individual. As the ADPKD patients have a chronic medical condition, which has likely spanned over at least four decades, this is likely

to have been a factor modifying their epigenetic landscape [230]. While this would not alter the bioinformatic assessment of DNA methylation, understanding the clinical history of patient samples provides greater context for elucidating whether changes are likely to be involved in initiating or maintaining disease, or arise in response to cystogenesis. Unfortunately due to the anonymisation of tissue samples, the patients' medical history is unavailable, and we can only speculate on what medical interventions they underwent prior to nephrectomy.

Based on common practices in treating ADPKD [231, 232], it is plausible that the patients were on medications, likely for the long-term management of symptoms. Like all compounds, pharmaceuticals have the potential to cause epigenetic changes. Some of these epigenetic changes are the direct mechanism by which a medication works (such as HDAC inhibition negatively regulating Wnt signalling pathways at therapeutic levels [233]), however many more changes are the indirect result of pharmaceuticals, believed to often manifest as side effects [234].

According to clinical reviews, the most commonly used medications in ADPKD include statins, beta-blockers and ACE inhibitors, which are all used to combat cardiovascular-related symptoms such as hypertension [235]. There is limited observational evidence to speculate that statins and beta-blockers could initiate epigenetic changes in other conditions: beta-blockers can increase the risk of diabetes [236] which is known to have an epigenetic components [237], and statins cause gene expression alterations thought to arise from epigenetic changes [234, 238]. However, whether this could result in epigenetic changes to the kidney tissue, or play a broader role in the epigenetics of all tissue, is uncertain. There are not sufficient data to determine whether any medications could contribute to the DMFs identified in this research.

5.9.6 Limitations of advanced-stage ADPKD tissue

All of the ADPKD tissue and cell lines used in this research were derived from the kidneys of patients with end-stage renal disease. While this ensures that the tissue that was used in this analysis was (truly) cystic, the analysis might be less likely to be able to elucidate any changes required to initiate cystogenesis, as these changes may have occurred four or more decades earlier. Therefore, the differences between ADPKD and non-ADPKD tissue may be highlighting the differences between 'healthy' and dysregulated tissue from chronically ill patients. These could also include changes in molecular pathways that occur in response to the disease; a consequence of disease instead of a causative factor.

An example of epigenetic changes occurring in response to disease has been recently investigated in a metastudy performed on patients with chronic kidney disease (CKD) by way of epigenome-wide association studies [239]. Of 19 CpG sites identified in the study as being significantly differentially methylated in CKD, five of these regions were associated with renal fibrosis in patients. This suggests that pathological changes to the kidney, such as the fibrosis observed in chronic kidney disease, is linked to epigenetic changes. However, a comparison of these data with the data generated in our ADPKD study has shown no overlap between the differentially methylated regions in this CKD paper and the DMFs identified in our analysis. This suggests that the data identified in our study is likely demonstrating changes which are specific to ADPKD and not as a result of CKD.

It is possible that the differences in sequencing methodologies has also contributed to this lack of overlap. However, there is an overlap of the CKD paper data (generated from Illumina 450k) and Woo et al. data (MIRA-Seq). This means that even with different platforms, similarities

can be seen in genome-wide analyses, and the lack of overlap is more likely to be biological differences.

Despite our data not fitting with the chronic kidney disease data, it is unlikely that the changes to DNA methylation seen in this study are directly contributing to the initiating of cyst growth, considering they are end-stage kidneys. However the results seen here may provide information about the molecular pathways modified in APDKD, providing potential new targets for therapeutic intervention.

A way to identify cystogenesis-associated epigenetic changes could be to sample early-stage ADPKD epithelial cells. These can be isolated and characterised from urine, as humans shed renal epithelial cells via their urine [240, 241]. This methodology is not commonly employed in ADPKD and would need further validation. Additionally, current published protocols predominantly use SV40 viruses to immortalise cells [242]. In order to accurately assess DNA methylation changes in ADPKD this protocol would need to be altered to avoid viral effects on epigenetic changes as seen in the immortalised ADPKD cell lines used in this study. However, unlike nephrectomy, this technique is non-invasive and would provide access to cellular material from a wider range of ADPKD patients, including various ages and patients with milder disease (such as *PKD2*-related disease) who never reach ESRD and therefore never require nephrectomy.

Other PKD models commonly used to study molecular changes associated with disease include pluripotent stem cell lines, knockdown and knockout cell lines [243]. However, these do not accurately paint a picture of the environment in which cells are originating or growing within

a patient, and therefore would have a limited application for studying the epigenetic landscape of ADPKD.

5.10 Future Directions

In the future it would be important to extend the research covered in this thesis, especially to include more tissue samples. Human ADPKD tissue is limited both because there is a small population with the disease, and secondly because tissue is only available when kidneys are removed – this only happens when the kidneys grow to such an extent that they cause unmanageable pain or impede the function of other organs.

In addition to the expansion of the data set presented here, further analysis should be performed on the individual ADPKD cyst libraries, including the identification of differentially methylated fragments in each cyst library. This research was outside the bounds of the project presented here, but further work will be done to analyse whether fragments are consistently methylated in ADPKD cysts across the kidney, or whether heterogeneity plays a role in the DNA methylation of the ADPKD kidney.

Further assessment into the expression of genes in the individual cyst samples would be useful. It is interesting that within the eight individual cyst samples, the global trend in methylation compared to non-ADPKD tissue was variable, demonstrating both hyper- and hypomethylation. This indicates that there is heterogeneity within the cysts of the ADPKD kidney, and has implications for both the process of cystogenesis and the response to potential epigenetic therapies. While qPCR of the DMF-associated genes could be performed, it may be more advantageous to perform RNA-Seq on these samples, given that there is very limited tissue

available for analysis, and that analysis to identify DMFs may bring forward a number of additional genes to analyse.

Despite variation at the global level, there were also similarities between the whole tissue data and the single cyst data within DMFs, which suggests there are fundamental differences in the methylome of ADPKD. The expression of these DMF-associated genes should be followed up in cyst tissue, to further identify whether the methylation of these genes is linked to their expression.

GET4 expression was found to be significantly reduced in ADPKD, which had a strong negative correlation with the methylation of the DMF in the gene body. Further research needs to be performed to validate whether this is a causal relationship, and additionally to identify if the gene plays a role in the cystogenesis of the kidney in ADPKD. Suggested experimental approaches include the treatment of primary cell lines with demethylating agents to measure the effects of methylation on the gene expression.

5.11 Conclusion

We have shown that the methylome of kidney tissue is hypomethylated in ADPKD patients, and we have identified 13 fragments with statistically significant methylation in these patients. Of these, at least one gene has a change in expression correlating with this methylation. We have found no significant differential methylation in the ADPKD-causing gene *PKD1*, although there was a trend towards hypermethylation at the 3' end of the gene, which correlated with an increase of *PKD1* expression by qPCR. We have found no significant changes in the expression of *PKD1*-associated miRNA.

We have also demonstrated that the ADPKD kidney is heterogenous with regards to DNA methylation, as illustrated by varying global methylation patterns across eight individual ADPKD cysts from a single patient. Furthermore, the cysts fall into two distinct, isolated clusters when performing unsupervised hierarchical clustering. We have confirmed that of the 13 DMFs in whole tissue, eight of these are significantly differentially methylated in ADPKD single cyst tissue.

DNA hypomethylation has not been previously reported in ADPKD, and there are currently no published data on the methylome of individual ADPKD cysts. This research provides new data on the methylation landscape of ADPKD tissue, and identifies two genes (*NAGLU* and *GET4*) as dysregulated in ADPKD and subsequently will require further investigation for their role in cystogenesis.

Appendix A: Clinical data

Supplementary Table A.1: Clinical data from samples used in this analysis

Due to anonymisation of donated tissue, there is limited information on the patient history of historical ADPKD tissue. Blanks left where information unavailable. PKD1 mutations are not typically investigated, thus there is only disease-causing mutant information on patient ADPKD D – as this patient had private genetic screening, and the commercially available cell lines which have been previously documented.

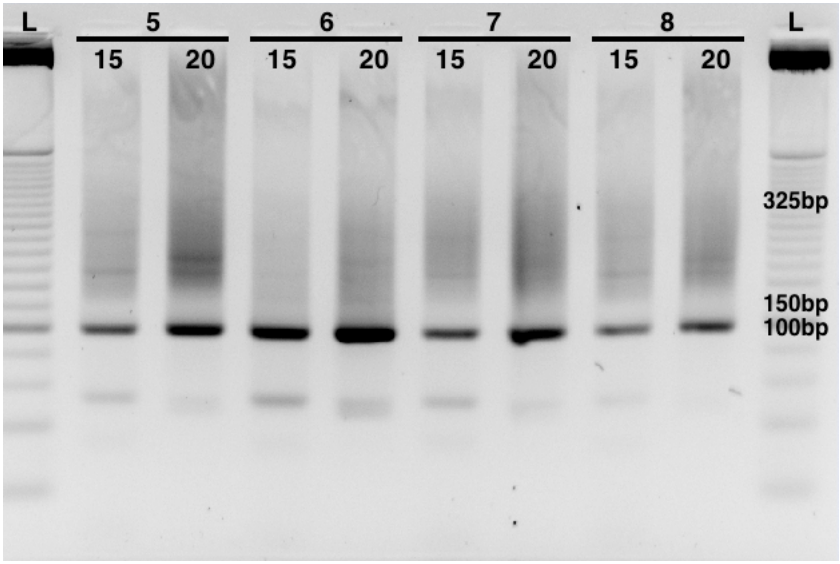
Sample ID	Non-ADPKD E	Non-ADPKD G	Non-ADPKD H	ADPKD 05	ADPKD 07	ADPKD 08	ADPKD D	Cysts	ADPKD W	WT 9-7	WT 9-12
Disease state	Non-ADPKD	Non-ADPKD	Non-ADPKD	ADPKD	ADPKD	ADPKD	ADPKD	ADPKD	ADPKD	ADPKD	ADPKD
PKD1 mutation	n/a	n/a	n/a				D1249fs	D1249fs		Q2556X	Q2556X
Tissue Type	Whole tissue	Whole tissue	Whole tissue	Whole tissue	Whole tissue	Whole tissue	Whole tissue	Cyst wall^	Whole tissue^	Cell Line	Cell Line
RRBS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Patient age§	53	51	55	52			60	60	56		
Patient sex	F	F	F	M			M	M	M	F	F
Year collected	2010	2009	2015	2005	2007	2008	2016	2016	2016	1999*	1999*
qPCR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No

§ Patient age at time of nephrectomy

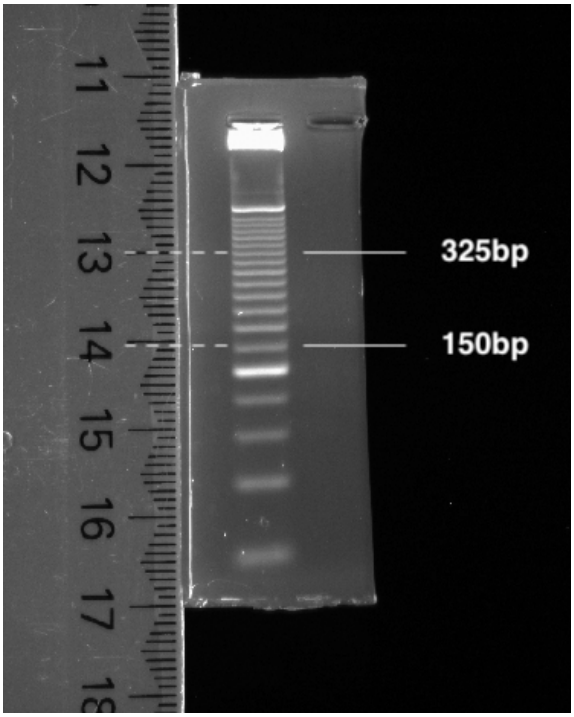
^ NBF-treated prior to our access to tissue

* Immortalised cell lines generated in 1999. WT 9-7 was cultured from P16, WT 9-12 was cultured from P25.

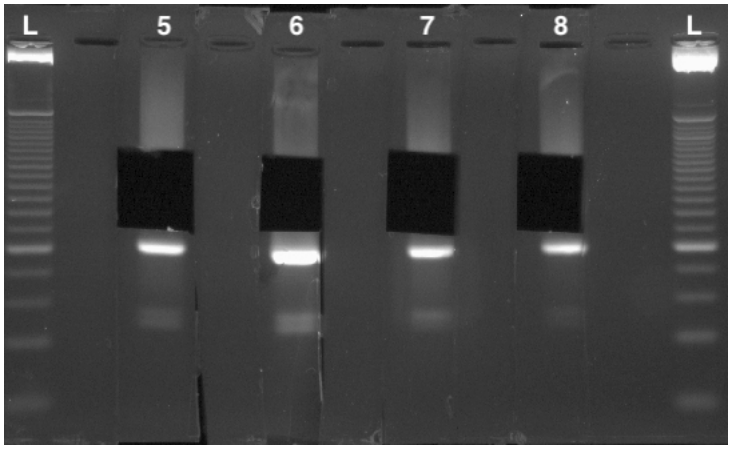
Appendix B: Preparation of RRBS libraries



Supplementary Figure B.1: Semi-quantitative PCR gel
Semi-quantitative PCR amplifications of the RRBS libraries prepared using n = 15 and n = 20 PCR cycles were loaded onto NuSieve agarose gel (Section 2.2.4) in pairs. Following gel electrophoresis, the gel was imaged under UV light. This gel is of RRBS libraries from cysts 5, 6, 7 and 8 (L = 1 kb ladder). The number of PCR cycles to amplify each sample is assessed by determining how much product is between 150 – 325 bp without non-specific amplification. All of these samples required 20 cycles of PCR. The dark bands at 100 kb are the adaptor dimers.



Supplementary Figure B.2: Measurement of 1 Kb ladder
The ladder from one side of the gel was excised and imaged next to a ruler to measure the exact location of the 150 bp and 325 bp bands. The lower edge of the gel was lined up with the 17 cm mark on the ruler. The 150 bp band was at 14.05 cm, the 325 cm band was at 13.00 cm.



Supplementary Figure B.3: Large-scale amplification gel
Once the RRBS libraries were excised from the gel for DNA extraction (Section 2.2.6), the gel was pieced together and imaged under UV. This shows that the RRBS library was cut out from the gel, leaving the adaptor dimer band (~100 bp) intact. The smear of DNA product above and below the excised section of the gel is non-specific amplification as it is outside the targeted region (150 – 325 bp).

Appendix C: Sequencing submission

Supplementary Table C.1: Sequencing submission from low-input libraries
Libraries sent to Otago Genomics Facility (OGF) in 2016, sequenced across three lanes.

Library	Adaptor	Average fragment size (bp)	Conc. (nM)	Volume (uL)	Adaptor trimming
ADPKD D 1	AD005	292	10	10	Yes
ADPKD D 2	AD006	245	2.65	5	Yes
ADPKD D 3	AD002	276	6.86	5	Yes
ADPKD D 4	AD004	217	2.68	7	
ADPKD D 5	AD012	313	10	10	Yes
ADPKD D 6	AD007	325	10	10	
ADPKD D 7	AD019	335	10	10	Yes
ADPKD D 8	AD010	249	5.7	7	Yes
ADPKD D I	AD001	203	10	10	Yes
ADPKD D II	AD011	203	10	10	Yes
ADPKD D III	AD003	221	10	10	Yes
ADPKD 05	AD016	239	10	10	

Supplementary Table C.2: Sequencing submission from cell lines
Libraries sent to OGF in 2016, sequenced across two lanes alongside three other RRBS libraries. These libraries were prepared with Jackie Ludgate.

Library	Adaptor	Average fragment size (bp)	Conc. (nM)	Volume (uL)
WT 9-7	AD011	306	10	30
WT 9-12	AD003	298	10	30

Supplementary Table C.3: Sequencing submission from RRBS libraries generated in previous research
Libraries sent to OGF in 2015, sequenced across three lanes alongside 10 other RRBS libraries. These libraries were prepared by Michael Bates [140].

Library	Adaptor	Average fragment size (bp)	Conc. (nM)	Volume (uL)
ADPKD 07	AD004	324	10	25
ADPKD 08	AD002	340	10	25
Non-ADPKD E	AD006	305	10	25
Non-ADPKD G	AD007	269	10	25
Non-ADPKD H	AD005	305	10	25

Appendix D: RRBS data processing

RRBS sequences were returned as fastq files which were processed for analysis in command line using DMAP programs and in RStudio. Below are examples of the code used to generate this data in command line.

```
# Trim fragments and concatenate into single file
cleanadaptors –I contam.fa –t 3 –T 3 –x 4 HNYL2BCXY–2046F–01–14–
1_S1_L001_R1_001.fastq > ADPKD05.fastq
cleanadaptors –I contam.fa –t 3 –T 3 –x 4 HNYL2BCXY–2046F–01–14–
1_S1_L002_R1_001.fastq >> ADPKD05.fastq
cleanadaptors –I contam.fa –t 3 –T 3 –x 4 HVJWYBCXY–2046F–01–14–
1_S1_L001_R1_001.fastq >> ADPKD05.fastq


# Quality check
fastqc –o qc ADPKD05.fastq

# Bismark alignment
bismark –n 1 GRCh37 ADPKD05.fastq

# Merge non–ADPKD libraries to create a reference file
samtools merge nonADPKD.bam
/home/sbowden/HuKidneyE1_R1_adtr.fastq_bismark_sorted.bam
/home/sbowden/HuKidneyG2_R1_adtr.fastq_bismark_sorted.bam
/home/sbowden/HuKidneyH3_R1_adtr.fastq_bismark_sorted.bam

# Generation of list of fragments
diffmeth –F 2 –t 10 –g Homo_sapiens.GRCh37.dna.chromosome. –L 40,220 –z ADPKD05.fastq >
ADPKD05_list.txt

# ANOVA
diffmeth –F 2 –t 10 –g Homo_sapiens.GRCh37.dna.chromosome. –L 40,220 –z
–R NonE.fastq –R NonG.fastq –R NonH.fastq –S ADPKD05.fastq –S ADPKD07.fastq –S
ADPKD08.fastq –S ADPKDDIII.fastq > ANOVA.txt

# Gene annotation
identgeneloc –i –Q –U –R –B “protein_coding” –p GRCh37dat –s “.dat” –r ANOVA.txt >
ANOVA_gene.txt
```

The following are examples of the programs and commands used in RStudio to perform data analysis in Chapters 3, 4 and Appendix G.

```
## RStudio script for analysis of RRBS data
# Correlation
```{r}
# create data matrix
corr_data = data.matrix(common18[,2:ncol(common18)])
corrrownames = c("ADPKD 05", "ADPKD 07", "ADPKD 08", "ADPKD D I", "ADPKD D II", "ADPKD D III", "Cyst 1", "Cyst 2", "Cyst 3", "Cyst 4", "Cyst 5", "Cyst 6", "Cyst 7", "Cyst 8",
"Non-ADPKD E", "Non-ADPKD G", "Non-ADPKD H", "WT 9-7", "WT 9-12")

# calculate Pearson's Correlation Coefficient
PCC = cor(corr_data)
rownames(PCC) = corrrownames
colnames(PCC) = corrrownames

```
Correlation heatmap
```{r}
library(corrplot)
corrplot(PCC, order = "hclust",method = "color", hclust.method="single", font=3, tl.col = "black", type = "lower", tl.srt = 30, tl.cex=0.6, is.corr = FALSE)
```
Global methylation (box plots)
```{r}
# Boxplot generates total methylation from each data frame
boxplot(NonE$Non_E, NonG$Non_G, NonH$Non_H,ADPKD05$ADPKD_05, ADPKD07$ADPKD_07, ADPKD08$ADPKD_08, ADPKDDIII$ADPKD_D_III,
        col = c("blue","blue","blue","red","red","red","red"), las = 1, ylab = "Methylation")
axis(side = 1, at = c(1,2,3,4,5,6,7),
     labels = c("Non-ADPKD E", "Non-ADPKD G", "Non-ADPKD H", "ADPKD 05", "ADPKD 07", "ADPKD 08", "ADPKD D"),
     las = 2, cex.axis = 0.7)
```
Unsupervised hierarchical clustering
```{r}
# create data matrix
dendro_matrix = data.matrix(common18[,2:ncol(common18)])
colnames(dendro_matrix) = corrrownames
dendro_data_2 = t(dendro_matrix)

#hierarchical clustering
dendro_hc = hclust(dist(dendro_data_2)) # uses 'common' clustering method
```
ConsensusClusterPlus
```{r}
# create data matrix
allcyst = read.csv("CystMerged.csv")
allcyst2 = na.omit(allcyst)
rnames = allcyst2[ ,1]
cystmatrix = data.matrix(allcyst2[ ,2:ncol(allcyst2)])
rownames(cystmatrix) = rnames

cystmads = apply(cystmatrix,1,mad)
cystdata = cystmatrix[rev(order(cystmads))[1:5000],]
cystmedian = sweep(cystdata,1, apply(cystdata,1,median, na.rm=T))

library(ConsensusClusterPlus)
results = ConsensusClusterPlus(cystmedian,maxK = 6,reps = 1000, pItem = 0.8, pFeature = 1, title = "ADPKD Cysts", clusterAlg="hc", distance = "pearson")
```
```

Appendix E: Primer sequences

Supplementary Table E.1: Primer sequences for qPCR

Primer sequences for qPCR analysis. RefSeq accession numbers for each gene version are reported [244].

PKD1 NM_000296.4	F	5-GCCGCGCCATCCCTTTCTGT-3
	R	5-GCTCCGGCTGTCCACCCCATAC-3
B2M NM_004048.3	F	5-GAGTGCTGTCTCCATGTTTGATGT-3
	R	5-AAGTTGCCAGCCCTCCTAGAG-3
EEF1A1 NM_001402.6	F	5-CTGCCACCCCACTCTTAATCA-3
	R	5-GGCCAATTGAAACAAACAGTTCT-3
GET4 NM_015949.3	F	5-GAGCCCTGAAGTGGTCCAG-3
	R	5-GTGCAGAAAATGATACCTCGACT-3
NAGLU NM_000263.4	F	5-TGGCACATCAAGCAGCTTTA-3
	R	5-GTGACAGCCTCGGGAACAT-3
CPLX1 NM_006651.4	F	5-AGTTTGTGATGAAGCAGGCTCT-3
	R	5-GTCTGGGTCCTTCTCCTCGT-3
DCLRE1C NM_022487.4	F	5-ACAGGAGACTTCAGATTG-3
	R	5-CACTCCTCCCGACTTGGAATT-3
EFCAB4B NM_032680.4	F	5-CTTCAGGAATAACCCAAGTCAGG-3
	R	5-TTTGGGCTCCAAGTCTGTCCATCA-3
KDELR2 NM_001100603.2	F	5-GATTTCTCTCCTCTTGAGATCCTCT-3
	R	5-ACAGTATAGGATGGTCTGGACTACG-3
PTBP1 NM_002819.5	F	5-ATCAGGCCTTCATCGAGATGCACA-3
	R	5-TGTCTTGAGCTCCTTGTGGTTGGA-3
TMPRSS6 NM_153609.3	F	5-AACGGCAGCGACGAAGAGCA-3
	R	5-TCACACTGCGGGTTGGGCTT-3

Appendix F: Publication of data derived from this thesis

The following paper manuscript describes global DNA methylation as observed in ADPKD from whole tissue samples (derived from work in Chapter Three):

Bowden, S. A., Rodger, E. J., Bates, M., Chatterjee, A., Eccles, M. R., & Stayner, C. (2018). Genome-Scale Single Nucleotide Resolution Analysis of DNA Methylation in Human Autosomal Dominant Polycystic Kidney Disease. *Am J Nephrol, 48*(6), 415-424. doi:10.1159/000494739

Abstract

Background: Autosomal dominant polycystic kidney disease (ADPKD) is characterised by the formation of fluid-filled cysts in the kidney and end stage renal disease by the fourth or fifth decade of life. Mutations in the *PKD1* gene account for 85% of all cases of ADPKD. No curative therapy exists for patients affected by this disease and an underexplored avenue for the treatment of ADPKD is the targeting of epigenetic changes that occur during cystogenesis. Limited data exists on alterations in DNA methylation that are associated with ADPKD. Given the similarities between cyst growth and neoplasia, and the fact that two DNA methylation inhibitors are already FDA-approved for the treatment of myelodysplastic syndrome, we hypothesized that global DNA methylation patterns in ADPKD would parallel that observed in neoplasia, and which may provide an opportunity to treat ADPKD with epigenetic inhibitors. To address this hypothesis we undertook a global DNA methylation analysis of human ADPKD kidney.

Methods: We generated single nucleotide resolution methylomes of cortical kidney tissue from individuals with ADPKD, and from non-ADPKD kidney tissue, using reduced representation bisulfite sequencing (RRBS). Using quantitative reverse transcription polymerase chain reaction (qRT-PCR), we investigated expression of the *PKD1* gene in both ADPKD and non-ADPKD kidney.

Results: We have shown that ADPKD-derived genomic DNA exhibits global hypomethylation when compared with non-ADPKD kidney, a pattern commonly observed in DNA methylation studies of tumour derived tissue. We have also identified thirteen discrete regions that are significantly differentially methylated in ADPKD compared to non-ADPKD, and eight of these are gene specific. The *PKD1* gene shows an increase in methylation at the 3' end of the gene body, but in contrast to previously published data, this is not associated with a decrease in *PKD1* mRNA expression.

Conclusion: This genome-scale single nucleotide resolution analysis of DNA methylation in human polycystic kidneys suggests that DNA methylation differences at specific loci are associated with ADPKD. Further exploration into the significance of these preliminary results may shed light on the disease process, and potentially reveal new therapeutic possibilities.

Background

Autosomal dominant polycystic kidney disease (ADPKD) is the most common form of inherited kidney disease, affecting 1-5 per 10,000 individuals[1]. ADPKD is characterised by the formation of fluid-filled cysts within the kidney, with renal function declining in the fourth or fifth decade of life, leading to end stage kidney disease. Extra-renal clinical features include cysts in the liver, hypertension, and an increased risk of intracranial aneurysms.

A germline mutation in one copy of either the *PKD1* or *PKD2* gene causes ADPKD, with 85% of cases attributable to a mutation in *PKD1*. Polycystin-1 and -2, are the protein products of the *PKD1* and *PKD2* genes, respectively, and have been shown to interact at their C-terminal domains in renal tubule epithelia[2]. Both are transmembrane proteins, with Polycystin-2 able to act as a nonselective cation channel on the plasma and/or the endoplasmic reticulum membrane[3]. Polycystin-1 has a large extra-cellular N-terminal domain and on the primary cilia is postulated to act as a mechanosensor that responds to renal flow, subsequently regulating Ca^{2+} ion flow via Polycystin-2[4].

Recent data investigating epigenetic modification during cystogenesis has suggested a potential role for histone modification[5,6] and/or DNA methylation[7,8] in modulating cyst formation. DNA methylation occurs at cytosine residues within CpG dinucleotides, and is catalyzed by DNA methyltransferases (DNMTs). Methylation at the promoter of a gene is frequently associated with a loss of gene expression, while methylation in the gene body, particularly at the 3' end of the gene, correlates with active expression of that gene[9]. There are, however, many exceptions to this generalized observation[10-13].

Aberrant DNA methylation patterns are associated with many types of cancer, and accordingly, several DNMT inhibitors are currently FDA-approved for the treatment of myelodysplastic syndrome. Given that cysts are effectively benign neoplasms[14], and ADPKD itself has been described as 'neoplasia in disguise'[15], there may also exist parallels in global DNA methylation patterns that could be exploited therapeutically. A previous study by Woo et al.,[7] profiled DNA methylation patterns in ADPKD tissues using MIRA-seq (methylated-CpG island recovery assay). They identified a global hypermethylation in the genome of ADPKD-

derived DNA, including hypermethylation of the *PKD1* gene body (but not the *PKD1* gene promoter), correlating with a decrease in *PKD1* mRNA levels.

To address the hypothesis that global DNA methylation patterns in ADPKD parallels that observed in neoplasia, which may provide an opportunity to treat ADPKD with epigenetic inhibitors, we have undertaken RRBS (reduced representation bisulfite sequencing) analysis of tissues from both ADPKD and non-ADPKD kidneys in the first genome-scale single nucleotide resolution analysis of DNA methylation in human polycystic kidneys.

Methods

Collection and processing of tissues

The three non-ADPKD cortical kidney tissues were collected from adults aged 50-60yrs by the Christchurch Cancer Society Tissue Bank following nephrectomy to remove a renal tumour. Wildtype tissue was sampled as far from the tumour site as possible, and appeared normal histologically. The ADPKD tissue samples were cortical samples taken from cystic regions of the kidney of male patients with ADPKD (aged 50-60yrs). Clinical data for these patients was not collected as these data were anonymised. However patients were on dialysis prior to nephrectomy, which was performed around the time of kidney transplantation. All procedures for collection of renal tissue was approved by the Otago Human Ethics Committee H15/110.

Preparation and sequencing of RRBS libraries

RRBS libraries of kidney tissue were prepared according to previously published methods [16-19]. Briefly, genomic DNA was extracted from frozen tissue using the QIAamp DNA Mini Kit (Qiagen) with an extended 16 hour Proteinase K treatment at 56 °C. Genomic DNA was digested overnight with MspI enzyme (New England Biolabs, Ipswich, MA) and the resulting

fragments were end-repaired, A-tailed and ligated to methylated adaptors (Illumina, San Diego, CA). Fragments were size selected (150 to 330 bp post-adaptor-ligation size) on 3% Nusieve agarose gels (Lonza, Basel, Switzerland) and bisulfite converted using the EZ DNA methylation kit (Zymo Research, Irvine, CA). Bisulfite converted libraries were amplified by PCR (16-18 cycles) and size-selected again to eliminate primer contamination. Libraries were quantified using the Qubit fluorometer (Life Technologies, Grand Island, NY) and quality assessed on the 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) using the high sensitivity DNA chip according to the manufacturer's instructions. RRBS libraries were sequenced on an Illumina HiSeq2500 with single-ended 100 bp reads. Base-calling was performed with Illumina Real Time Analyzer (RTA) software [17].

Quality assessment, alignment and DNA methylation analysis

Quality of the sequenced reads for each sample was assessed using the FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) program. Adaptor sequences and unmethylated CpG bases at the 3′ end of the reads that were added during end-repair step were removed using our in-house cleanadaptors program [20]. The sequenced reads were aligned to the complete human reference genome GRCh37 with the Bismark v0.6.4 alignment tool [21], allowing for only one mismatch in the seed. The DMAP package [22] was used to generate MspI fragment-based methylomes and perform differential methylation analysis of the two groups. Only fragments that had at least 2 CpG sites and covered by 10 or more sequenced reads in all libraries were included in an F test (ANOVA) to test for significant methylation differences between non-ADPKD and ADPKD kidney tissue. We used a false discovery rate of <0.05 and fragments with an absolute methylation difference of ≥ 0.10 between tissue groups were identified as differentially methylated fragments (DMFs).

Generating single nucleotide resolution methylomes of ADPKD and non-ADPKD kidney

Reduced representation bisulfite sequencing (RRBS) was used to generate DNA methylomes of renal cortex tissue from four ADPKD patients and three non-ADPKD kidneys. For these seven samples, a total of 243 million 100 bp sequencing reads were generated. The sequenced reads were mapped to the GRCh37 human reference genome using Bismark [21], with unique alignment efficiency ranging from 54.2% to 68% (median=66.2%, Supplementary Table S1). The median bisulfite conversion rate was calculated to be >98%, assuming that the total quantified non-CpG methylation was due to inefficient bisulfite conversion (Supplementary Table S1). We determined the CpG DNA methylation (on a scale of 0 to 1) using MspI fragments (40-220 bp) as the unit of analysis. Size-selected MspI fragments were the input for RRBS sequencing libraries and we have previously described the utility of this approach for analysing genome-scale DNA methylation patterns in human tissues [19,22-24].

Gene enhancer and transcription factor binding prediction

To identify DMFs overlapping predicted enhancer regions, the GeneHancer database of genome-wide enhancer-to-gene associations [25] was downloaded from the GeneCards website (https://genecards.weizmann.ac.il/geneloc/genehancer.xlsx) and converted to GRCh37 coordinates using the UCSC genome browser Liftover tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). BEDTools (v2.26.0) intersect was used to identify enhancer-to-gene associations that overlapped with the DMFs we identified in ADPKD. The UCSC table browser was used to retrieve details of predicted transcription factor binding sites from the ORegAnno database [26] that overlap with DMFs we identified in ADPKD.

Gene expression analysis

Gene expression analysis was performed on the same 3 non-ADPKD and 4 ADPKD tissue

samples used for RRBS, as well as an additional ADPKD sample. RNA was extracted from frozen tissue using the miRNeasy mini kit (Qiagen), and cDNA was synthesized using the high-capacity cDNA reverse transcription kit (Applied Biosystems) without RNase Inhibitor. The qPCR analysis was performed on cDNA using the SYBR Premix Ex Taq (Tli RNaseH Plus) kit (Takara) in a LightCycler 480 system. All qPCRs were performed in triplicate with *PKD1* primers designed to target the 3′ end of the cDNA, with primers targeting *B2M* and *EEF1A1* as housekeeping genes. Data analysis was carried out using the software qbase+ to determine the relative expression of *PKD1* in ADPKD compared to non-ADPKD tissue.

Results

ADPKD shows global hypomethylation compared to non-ADPKD kidneys

We compared global methylation patterns between the ADPKD and non-ADPKD kidneys. For this comparison, we required that each fragment fulfilled the coverage criteria of 10 or more reads and ≥ 2 CpG sites in each fragment) in both analysed groups. Subsequently referred to as "*common analysed fragments*", there were 345,711 fragments (336,121 from autosomes and 9590 from sex chromosomes) that contained a total of 1,785,585 CpG sites. The distribution of global methylation patterns revealed a bimodal distribution in both non-ADPKD and ADPKD samples (Supplementary Fig. S1). This observation is consistent with RRBS methylomes observed in somatic cells [23,24].

The global methylation profiles of the common analysed fragments showed slight, yet significant, hypomethylation in ADPKD compared to non-ADPKD kidney (Wilcoxon rank sum test, *P*-value < 2.2e-16, median methylation = 0.76 and 0.78 in ADPKD and non-ADPKD kidney respectively). We wanted to determine whether the hypomethylated fragments were enriched in particular genomic regions and therefore we performed a similar global analysis on

individual genomic elements: gene promoters (-5kb to +1 kb from TSS), introns, exons, intron-exon boundaries and intergenic regions (Figure 1A and Supplementary Table S2). We also performed global analysis on CpG island (CGI) features: CGI cores (contains the core CpG island), CGI shores (flanking 2 kb either side of the CGI core), CGI shelfs (flanking 2 kb either side of the CGI shore) and open sea (isolated CpGs in the genome) (Figure 1B and Supplementary Table S3). Hypomethylation of ADPKD was observed in all these genomic elements and CGI features (~2%), but more so in exons (3%) and to a lesser extent in fragments overlapping promoters (1%), intron/exon boundaries (0.3%) and CGI cores (0.8%).

Region-specific differential methylation in ADPKD

Differential methylation analysis was used to determine region-specific methylation differences between ADPKD and non-ADPKD kidney tissue. RRBS MspI fragments were used as the unit of analysis as previously described [19,23,24]. An ANOVA test followed by corrrection for false discovery rate (FDR <0.05) and the requirement for an absolute difference between kidney and ADPKD methylation of ≥0.10 identified 13 significantly differentially methylated fragments (DMFs) among the 345,711 common analysed fragments (Figure 2A, Table 1). Of the DMFs, 62% were hypomethylated in ADPKD compared to non-ADPKD kidney (eight hypomethylated; five hypermethylated). In relation to genomic features, two DMFs overlapped with gene promoters (-5kb to +1 kb from the transcription start site, TSS), six overlapped with gene bodies (five in introns and one in an intron/exon boundary) and five were intergenic (>5kb upstream of any gene). In comparison to the common analysed fragments (Figure 2B), there was an enrichment of DMFs overlapping with intergenic regions and intron/exon boundaries (Figure 2C). The two genes containing promoter DMFs were: *NAGLU* (which encodes an alpha-N-acetylglucosaminidase enzyme) and *EFCAB4B* (a calcium channel regulator; Figure 2D, Table 1). The six genes with gene body DMFs were: *CPLX1* (a regulator of cytoplasmic

vesicle exocytosis) *GET4* (a part of the BAG6/BAT3 cytosolic protein quality control complex), *DAGLB* (a diacylglycerol lipase enzyme), *KDELR2* (a receptor required for retention of proteins in the endoplasmic reticulum), *DCLRE1C* (a DNA cross-link repair enzyme), and *TMPRSS6* (a type II transmembrane serine proteinase; Figure 2D, Table 1). Some of the DMFs, including two that were intergenic, overlapped predicted gene enhancer regions (Supplementary Table S4) and transcription factor binding sites (Supplementary Table S5).

The methylation landscape of *PKD1* in ADPKD

PKD1 is a large gene (47.2 kb in length with 46 exons and 47 introns) that plays a central role in the pathogenesis of AKPKD. To our knowledge there is currently no single nucleotide resolution DNA methylation data available for the whole *PKD1* gene in ADPKD. Therefore, here we have particularly focused on characterizing the methylation status of this gene. For the *PKD1* gene, we obtained high quality methylation information for 68 RRBS fragments containing 374 CpG sites (Supplementary Table S6). Our RRBS libraries did not have sufficient sequencing read coverage upstream of the TSS, with the exception of one 67 bp fragment containing 9 CpGs (3.6 kb from TSS) that was fully methylated in both ADPKD and non-ADPKD but in contrast, exon 1 and part of intron 1 was almost fully unmethylated. The gene body of *PKD1* was largely fully methylated in both non-ADPKD and ADPKD kidney tissue (Figure 3A and Supplementary Table S6). These findings are consistent with the pattern observed in RRBS libraries of kidney tissue and kidney epithelial cells from the ENCODE project (ENCSR938TEC) [27]. There were no statistically significant DMFs in *PKD1* using a FDR adjusted P-value <0.05, but with a nominal raw P-value threshold of <0.05 from the ANOVA, the ADPKD tissue had ten fragments in *PKD1* with methylation differences greater than 0.10 compared to non-ADPKD kidney (Figure 3B and Supplementary Table S6). Three of the fragments in intron 1 of *PKD1* and another on the boundary of intron 9 and exon 10 were

less methylated in ADPKD (14-44% difference in methylation) compared to non-ADPKD kidney. The other six fragments in intron 32, overlapping introns 40-41 and exons 41-42, and the intron 45/exon 46 junction at the 3' end of the *PKD1* gene body were more methylated in ADPKD (Figure 3B and Supplementary Table S6). One of the intron one fragments overlapped a predicted binding site for the transcriptional repressor CTCF (Supplementary Figure S2). Five of the 3′ fragments overlapped a predicted enhancer (Genehancer ID: GH16H002086, gene-enhancer score = 12.2; Supplementary Table S6) and four of these fragments overlapped a predicted binding site for POLR2A, which potentially regulates transcription of *PKD1* (Supplementary Figure S2).

A previous study had correlated an increase in *PKD1* gene body methylation with a decrease in *PKD1* mRNA expression in human ADPKD kidney[7]. We investigated the mRNA expression level of *PKD1* in five ADPKD kidney samples, (including the four samples for which we had RRBS data) and compared this with the three non-ADPKD kidney samples. None of the ADPKD samples showed a decrease in *PKD1* expression when compared with non-ADPKD kidney, with most showing a higher level of *PKD1* mRNA. Overall the *PKD1* gene had a significant 2-fold higher expression (*P*-value = 0.036) in ADPKD compared to non-ADPKD kidney (Figure 3C). Notably, the level of methylation in three of the 3′ fragments was significantly correlated (*P*-value < 0.05) with *PKD1* expression (Supplementary Table S7).

Discussion

By generating and analysing single nucleotide resolution methylomes of cortical kidney tissue from individuals with autosomal dominant polycystic kidney disease (ADPKD) and from non-ADPKD kidney tissue, we have shown that ADPKD exhibits global hypomethylation when compared with normal kidney. We have identified a gain of methylation at the 3' end of the

PKD1 gene body, which is associated with a significant increase in *PKD1* mRNA expression. In addition, thirteen loci show statistically significant differential gain or loss of methylation in ADPKD renal tissue.

In cancer, global hypomethylation and site-specific gain of DNA methylation is observed [28]. Our results indicate that the methylation landscape of ADPKD compared to normal kidney is more likely to resemble the pattern observed in human cancers. However, the specific method used for global methylation profiling has implications in interpreting these results. We have used RRBS in this study, which enriches for functionally important CpG-rich regions of the genome but does not include repetitive elements at sufficient coverage. Repetitive elements are generally silenced by high amounts of DNA methylation. In cancers, several types of these repeat elements tend to lose methylation marks, leading to global hypomethylation [9]. Based on our data, it is plausible that if whole genome methylation were analysed, the degree of hypomethylation is likely to be even more pronounced in ADPKD samples compared to normal kidney tissue.

In contrast to these RRBS data, genome–wide methylation profiling of ADPKD tissues by Woo et al.,[7] using MIRA seq, indicated a global hypermethylation, with only 9% of the ADPKD genome being hypomethylated. In agreement with the RRBS data, they also observed hypermethylation of the *PKD1* gene body. However, Woo et al have correlated this *PKD1* hypermethylation with a decrease in *PKD1* gene expression levels, whereas our *PKD1* gene expression analysis did not indicate a loss of *PKD1* expression in the ADPKD tissues, with all ADPKD samples showing the same or higher mRNA levels when compared with non-ADPKD renal tissue. Overall, there was a 2-fold higher expression of *PKD1* in the ADPKD samples (Figure 3C). Our expression data are consistent with a number of earlier studies that have noted

an increase in *PKD1* mRNA and protein expression in ADPKD kidney.[29-33] In addition, the level of *PKD1* expression we observed in each ADPKD sample significantly correlated with the extent of DNA methylation in several fragments at the 3' end of the *PKD1* gene body (Supplementary Table S7).

The difference in the findings of our study and Woo et al. could be largely explained by the platform used for profiling methylation patterns. There are at least twenty different techniques that have been developed to interrogate genome-wide DNA methylation patterns[34], which provide snapshots of the genome based on their design. The most commonly-used of these include HumanMethylation450 BeadChip (450K) DNA methylation profiling, enrichment-based methods include Methylation-dependent Immunoprecipitation followed by sequencing (MeDIP-Seq) and MIRA-seq[35], reduced representation bisulfite sequencing (RRBS) and whole genome bisulfite sequencing (WGBS). While WGBS potentially allows investigation of every CpG site in the genome, cost and data management considerations have advanced the use of alternative methods. RRBS is an alternative to WGBS, that through the use of a methylation insensitive restriction digest and size selection process, enriches for CpG sites, CpG islands (~ 30-fold for human) and gene promoters and gene bodies[36]. The 450K BeadChip approach is the most popular method for genome-wide DNA methylation due to its low cost and ease of data analysis. However 450K arrays allow investigation of only 1.7% of the CpG sites in the genome, one tenth the sites examined using RRBS. Enrichment-based methods have a number of limitations and do not allow investigation of single CpG sites, with a lower accuracy rate when compared to RRBS and WGBS[37].

Woo et al. used MIRA (methylated-CpG island recovery assay), which exploits the affinity of two methyl binding domain proteins (MBD2b and its homologue, MBD3L1). These MBD

proteins form a complex and have higher affinity to bind to methylated DNA (in the CpG context). The binding of this complex is dependent on the CpG density of the genomic locus [38]. These bound DNA fragments (enriched for CpG-methylated DNA) are sequenced followed by analysis of methylation. On the other hand, RRBS is based on MspI digestion followed by massively parallel sequencing to provide a genome-wide methylation map [20]. A recent study of B-cell subsets using MIRA-Seq demonstrated that the majority of methylation peaks are present in intergenic and intronic regions, and are less frequent in CpG islands [39], while RRBS highly enriches for CpG island fragments. We hypothesise that the difference in these two approaches results in analysis of different CpGs and therefore dissimilar differential methylation patterns. The two platforms also substantially differ in their CpG coverage. For example, Woo et al obtained 1 million distinct CpG sites (per million sequence reads) for the pooled non-ADPKD tissue, and 0.7 million unique CpG sites for the three ADPKD tissues. In contrast, we analysed 1,525,770 unique CpG sites in both groups using RRBS. MIRA is based on the principle of affinity and enrichment, and a limitation of this approach is that they do not allow investigation of single CpG sites, and the efficiency of the methylation calls is affected by the CpG density of the genomic locus [40]. Moreover, they are more prone to error if the samples have copy number variation. Large-scale studies have demonstrated that enrichment-based methods have lower accuracy compared to RRBS and Whole Genome Bisulfite Sequencing (WGBS) [37,41].

Another factor that may explain differences found between these two studies is that of tissue heterogeneity. While both Woo et al.,[7] and the current study have sampled cystic tissues from the renal cortex of ADPKD patients, there is likely to be tissue variation between samples. In addition, the potential for different cell types to be present within or between samples could lead to an alternative set of differentially methylated fragments being identified. The possibility

also exists that the DNA methylation differences noted between ADPKD and non-ADPKD kidney are related to changes associated with end-stage kidney disease, rather than cystic growth. A recent study by Chu et al., 2017[42], identified DNA methylation changes associated with chronic kidney disease (CKD) in two large cohorts of CKD patients. However, none of the 243 eGFR-associated CpGs identified overlap with our DMFs, leading us to conclude that the DNA methylation changes uncovered in the current study are not related to CKD. A further limitation in this study is the small number of samples analysed, which limits the statistical power for detecting significant differences between the two groups. Future studies using a higher number of ADPKD samples will provide further insights about the differential methylation landscape in this disease. However, this first genome-scale single nucleotide resolution analysis of DNA methylation in human polycystic kidney using RRBS, has revealed a number of loci that are epigenetically altered in cystic tissue, and further elucidation of the potential role these play in cystogenesis will be the subject of future studies.

Conclusions

Methylation profiles of ADPKD kidney using RRBS showed global hypomethylation in ADPKD compared to non-ADPKD kidney. This is in contrast to previously published work but is consistent with the observed similarities between ADPKD and neoplasia. Thirteen significantly differentially methylated fragments were identified, eight of which are associated with a specific gene. In addition, ten fragments in the *PKD1* gene showed differences in methylation between ADPKD and non-ADPKD kidney. However, none of the ADPKD tissue samples showed a decrease in *PKD1* expression when compared with non-ADPKD kidney, with most showing a higher level of *PKD1* mRNA. These data suggest that DNA methylation differences are associated with ADPKD and further exploration into the significance of these changes may shed new light on the disease process.

List of abbreviations

ADPKD: autosomal dominant polycystic kidney disease; ANOVA: analysis of variance; CGI: CpG island; DMFs: differentially methylated fragments; DNMTs: DNA methyltransferases; FDA: Food and Drug Administration; MBD: methyl binding domain; MeDIP-Seq: methylation-dependent immunoprecipitation followed by sequencing; MIRA: methylated-CpG island recovery assay; RRBS: reduced representation bisulfite sequencing; RTA: Real Time Analyzer; TSS: transcription start site; WGBS: whole genome bisulfite sequencing.

Declarations

Ethics approval and consent to participate

This project and the collection of tissue was approved by the University of Otago Human Ethics Committee, reference code H15/110. Written informed consent was provided to ME or to curators of the Cancer Society Tissue Bank, Christchurch, New Zealand, from which the non-ADPKD samples were obtained (application 15106SEC).

Availability of data and materials

The DNA methylation data generated by RRBS that supports the conclusions of this article have been submitted to the NCBI Gene Expression Omnibus under accession number GSE110057.

Funding

Funding for this research was supplied by the HS and JC Anderson Charitable Trust, and the PKD Foundation of Australia. AC and MRE would like to thank the New Zealand Institute for Cancer Research Trust for their salary support.

Authors' contributions

SB and MB carried out the experimental work. ER, SB and AC carried out the data analysis and, along with CS, substantially contributed in writing the manuscript. ME provided critical revision for the publication. All authors read and approved the final manuscript.

Acknowledgements

We thank all patients who provided material for the current study, along with the nephrologists and hospital staff who facilitated the tissue collection. We also acknowledge Peter Stockwell for his assistance with analysis of the RRBS data.

References

1 Solazzo A, Testa F, Giovanella S, Busutti M, Furci L, Carrera P, Ferrari M, Ligabue G, Mori G, Leonelli M, Cappelli G, Magistroni R: The prevalence of autosomal dominant polycystic kidney disease (adpkd): A meta-analysis of european literature and prevalence evaluation in the italian province of modena suggest that adpkd is a rare and underdiagnosed condition. PLoS One 2018;13:e0190430.

2 Qian F, Germino FJ, Cai Y, Zhang X, Somlo S, Germino GG: Pkd1 interacts with pkd2 through a probable coiled-coil domain. Nat Genet 1997;16:179-183.

3 Stayner C, Zhou J: Polycystin channels and kidney disease. Trends Pharmacol Sci 2001;22:543-546.

4 Nauli SM, Zhou J: Polycystins and mechanosensation in renal and nodal cilia. Bioessays 2004;26:844-856.

5 Cao Y, Semanchik N, Lee SH, Somlo S, Barbano PE, Coifman R, Sun Z: Chemical modifier screen identifies hdac inhibitors as suppressors of pkd models. Proc Natl Acad Sci U S A 2009;106:21819-21824.

6 Li X: Epigenetics and autosomal dominant polycystic kidney disease. Biochim Biophys Acta 2011;1812:1213-1218.

7 Woo YM, Bae JB, Oh YH, Lee YG, Lee MJ, Park EY, Choi JK, Lee S, Shin Y, Lyu J, Jung HY, Lee YS, Hwang YH, Kim YJ, Park JH: Genome-wide methylation profiling of adpkd identified epigenetically regulated genes associated with renal cyst development. Hum Genet 2014;133:281-297.

8 Woo YM, Shin Y, Hwang JA, Hwang YH, Lee S, Park EY, Kong HK, Park HC, Lee YS, Park JH: Epigenetic silencing of the mupcdh gene as a possible prognostic biomarker for cyst growth in adpkd. Sci Rep 2015;5:15238.

9 Jones PA: Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat Rev Genet 2012;13:484-492.

10 Chatterjee A, Stockwell PA, Ahn A, Rodger EJ, Leichter AL, Eccles MR: Genome-wide methylation sequencing of paired primary and metastatic cell lines identifies common DNA methylation changes and a role for ebf3 as a candidate epigenetic driver of melanoma metastasis. Oncotarget 2017;8:6085-6101.

11 Jia N, Wang J, Li Q, Tao X, Chang K, Hua K, Yu Y, Wong KK, Feng W: DNA methylation promotes paired box 2 expression via myeloid zinc finger 1 in endometrial cancer. Oncotarget 2016;7:84785-84797.

12 Guillaumet-Adkins A, Richter J, Odero MD, Sandoval J, Agirre X, Catala A, Esteller M, Prosper F, Calasanz MJ, Buno I, Kwon M, Court F, Siebert R, Monk D: Hypermethylation of the alternative awt1 promoter in hematological malignancies is a highly specific marker for acute myeloid leukemias despite high expression levels. J Hematol Oncol 2014;7:4.

13 Seynnaeve B, Lee S, Borah S, Park Y, Pappo A, Kirkwood JM, Bahrami A: Genetic and epigenetic alterations of tert are associated with inferior outcome in adolescent and young adult patients with melanoma. Sci Rep 2017;7:45704.

14 Grantham JJ: Clinical practice. Autosomal dominant polycystic kidney disease. N Engl J Med 2008;359:1477-1485.

15 Grantham JJ: Polycystic kidney disease: Neoplasia in disguise. Am J Kidney Dis 1990;15:110-116.
16 Chatterjee A, Ozaki Y, Stockwell PA, Horsfield JA, Morison IM, Nakagawa S: Mapping the zebrafish brain methylome using reduced representation bisulfite sequencing. Epigenetics 2013;8:979-989.
17 Chatterjee A, Rodger EJ, Stockwell PA, Weeks RJ, Morison IM: Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. Journal Of Biomedicine & Biotechnology 2012;2012:741542.
18 Chatterjee A, Stockwell, PA, Horsfield, JA, Morison, IM, Nakagawa, S: Base-resolution DNA methylation landscape of zebrafish brain and liver. Genomics Data 2014;2:342–344.
19 Chatterjee A, Macaulay EC, Rodger EJ, Stockwell PA, Parry MF, Roberts HE, Slatter TL, Hung NA, Devenish CJ, Morison IM: Placental hypomethylation is more pronounced in genomic loci devoid of retroelements. G3 (Bethesda) 2016
20 Chatterjee A, Stockwell PA, Rodger EJ, Morison IM: Comparison of alignment software for genome-wide bisulphite sequence data. Nucleic Acids Research 2012;40:e79.
21 Krueger F, Andrews SR: Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics 2011;27:1571-1572.
22 Stockwell PA, Chatterjee A, Rodger EJ, Morison IM: Dmap: Differential methylation analysis package for rrbs and wgbs data. Bioinformatics 2014;30:1814-1822.
23 Chatterjee A, Stockwell PA, Rodger EJ, Morison IM: Genome-scale DNA methylome and transcriptome profiling of human neutrophils. Sci Data 2016;3:160019.
24 Chatterjee A, Stockwell PA, Rodger EJ, Duncan EJ, Parry MF, Weeks RJ, Morison IM: Genome-wide DNA methylation map of human neutrophils reveals widespread inter-individual epigenetic variation. Sci Rep 2015;5:17328.
25 Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, Lancet D, Cohen D: Genehancer: Genome-wide integration of enhancers and target genes in genecards. Database (Oxford) 2017;2017
26 Lesurf R, Cotto KC, Wang G, Griffith M, Kasaian K, Jones SJ, Montgomery SB, Griffith OL, Open Regulatory Annotation C: Oreganno 3.0: A community-driven resource for curated regulatory annotation. Nucleic Acids Res 2016;44:D126-132.
27 Consortium EP: An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.
28 Sharma S, Kelly TK, Jones PA: Epigenetics in cancer. Carcinogenesis 2010;31:27-36.
29 Ward CJ, Turley H, Ong AC, Comley M, Biddolph S, Chetty R, Ratcliffe PJ, Gattner K, Harris PC: Polycystin, the polycystic kidney disease 1 protein, is expressed by epithelial cells in fetal, adult, and polycystic kidney. Proc Natl Acad Sci U S A 1996;93:1524-1528.
30 Peters DJ, Spruit L, Klingel R, Prins F, Baelde HJ, Giordano PC, Bernini LF, de Heer E, Breuning MH, Bruijn JA: Adult, fetal, and polycystic kidney expression of polycystin, the polycystic kidney disease-1 gene product. Lab Invest 1996;75:221-230.

31 Palsson R, Sharma CP, Kim K, McLaughlin M, Brown D, Arnaout MA: Characterization and cell distribution of polycystin, the product of autosomal dominant polycystic kidney disease gene 1. Mol Med 1996;2:702-711.
32 Griffin MD, Torres VE, Grande JP, Kumar R: Immunolocalization of polycystin in human tissues and cultured cells. Proc Assoc Am Physicians 1996;108:185-197.
33 Geng L, Segal Y, Peissel B, Deng N, Pei Y, Carone F, Rennke HG, Glucksmann-Kuis AM, Schneider MC, Ericsson M, Reeders ST, Zhou J: Identification and localization of polycystin, the pkd1 gene product. J Clin Invest 1996;98:2674-2682.
34 Chatterjee A, Rodger EJ, Morison IM, Eccles MR, Stockwell PA: Tools and strategies for analysis of genome-wide and gene-specific DNA methylation patterns. Methods Mol Biol 2017;1537:249-277.
35 Rauch TA, Pfeifer GP: The mira method for DNA methylation analysis. Methods Mol Biol 2009;507:65-75.
36 Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 2008;454:766-770.
37 Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A: Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 2010;28:1106-1114.
38 Rauch T, Li H, Wu X, Pfeifer GP: Mira-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. Cancer Res 2006;66:7939-7947.
39 Almamun M, Levinson BT, Gater ST, Schnabel RD, Arthur GL, Davis JW, Taylor KH: Genome-wide DNA methylation analysis in precursor b-cells. Epigenetics 2014;9:1588-1595.
40 Laird PW: Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet 2010;11:191-203.
41 Matarese F, Carrillo-de Santa Pau E, Stunnenberg HG: 5-hydroxymethylcytosine: A new kid on the epigenetic block? Mol Syst Biol 2011;7:562.
42 Chu AY, Tin A, Schlosser P, Ko YA, Qiu C, Yao C, Joehanes R, Grams ME, Liang L, Gluck CA, Liu C, Coresh J, Hwang SJ, Levy D, Boerwinkle E, Pankow JS, Yang Q, Fornage M, Fox CS, Susztak K, Kottgen A: Epigenome-wide association studies identify DNA methylation associated with kidney function. Nat Commun 2017;8:1286.

Figures

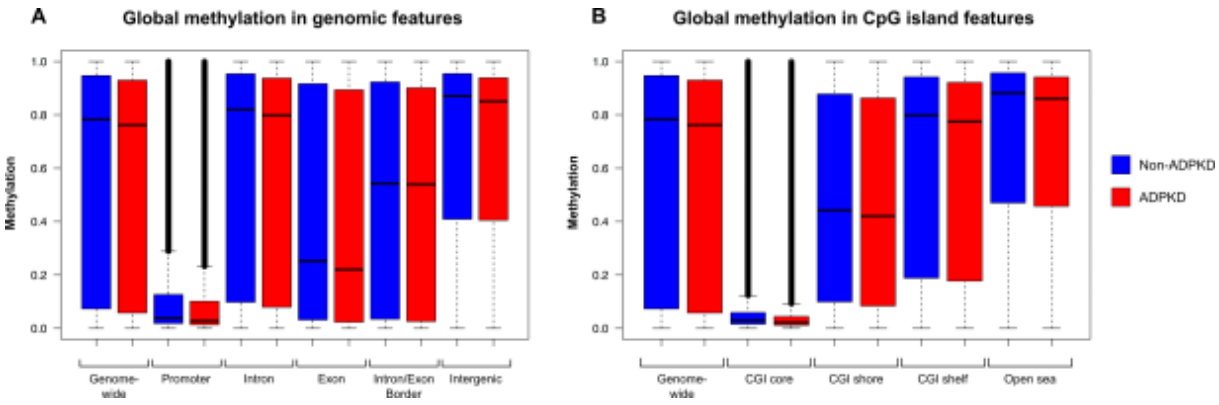


Figure 1: Global DNA methylation patterns of different genomic and CpG island features in ADPKD. DMAP software was used to generate MspI fragment-based methylomes from the renal cortex of non-ADPKD (n=3, shown in blue) and ADPKD kidney (n=4, shown in red). A. In all common analysed fragments (n=345,711), the median methylation in ADPKD (0.76) was significantly lower (Wilcoxon rank sum test, *P*-value < 2.2e-16) than non-ADPKD kidney (0.78) and this trend was also observed in the different genomic features: gene promoters, introns, exons and intergenic regions. B. The median methylation in ADPKD was also lower in all the different CpG island (CGI) features: CGI core (contains the core CpG island), CGI shore (flanking 2 kb either side of the CGI core), CGI shelf (flanking 2 kb either side of the CGI shore) and open sea (isolated CpGs in the genome). Boxplots show the distribution of methylation: minimum, maximum, interquartile range (box), and median (black bar).

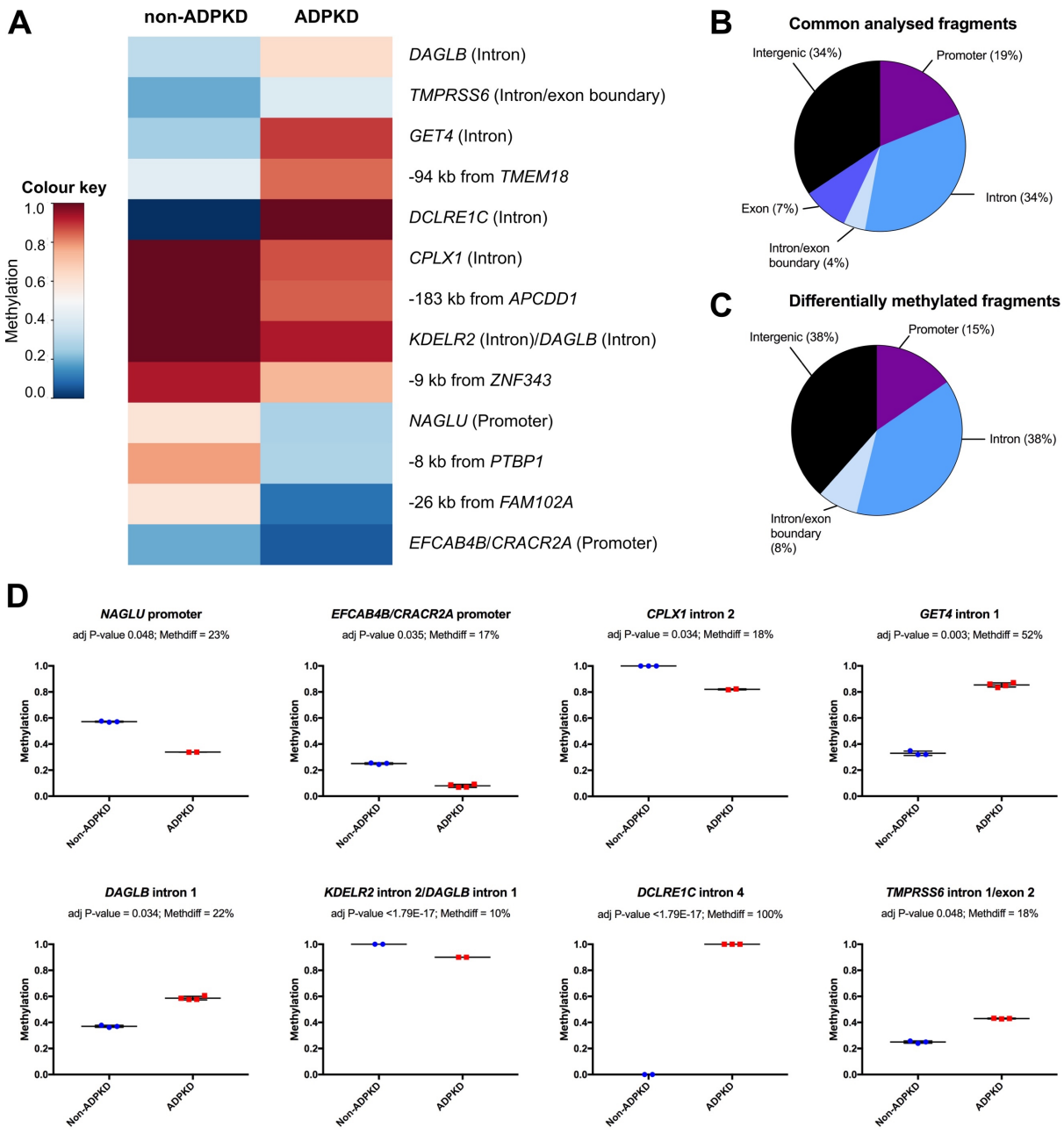


Figure 2: Differentially methylated fragments identified in ADPKD. DMAP software was used to determine region-specific methylation differences between non-ADPKD (n=3) and ADPKD (n=4) kidney using the MspI fragments as the unit of analysis. An ANOVA test followed by correction for false discovery rate (FDR<0.05) and inclusion of fragments with a methylation difference ≥10%. was used to identify significantly differentially methylated fragments (DMFs). A. Methylation of each of the 13 DMFs is plotted on a heatmap from 0 (dark blue) to 1 (dark red). B. The genomic distribution of all 345,711 common analysed fragments. C. The

genomic distribution of the 13 DMFs. D. Scatter dot plots of the 8 DMFs overlapping promoter regions or gene bodies. Bars show mean ± SD.

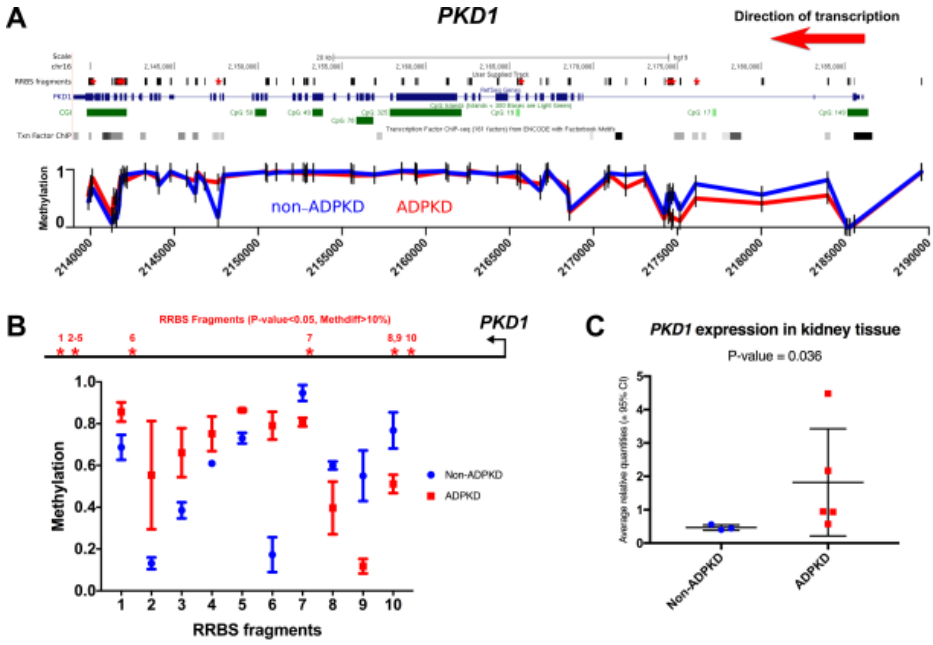


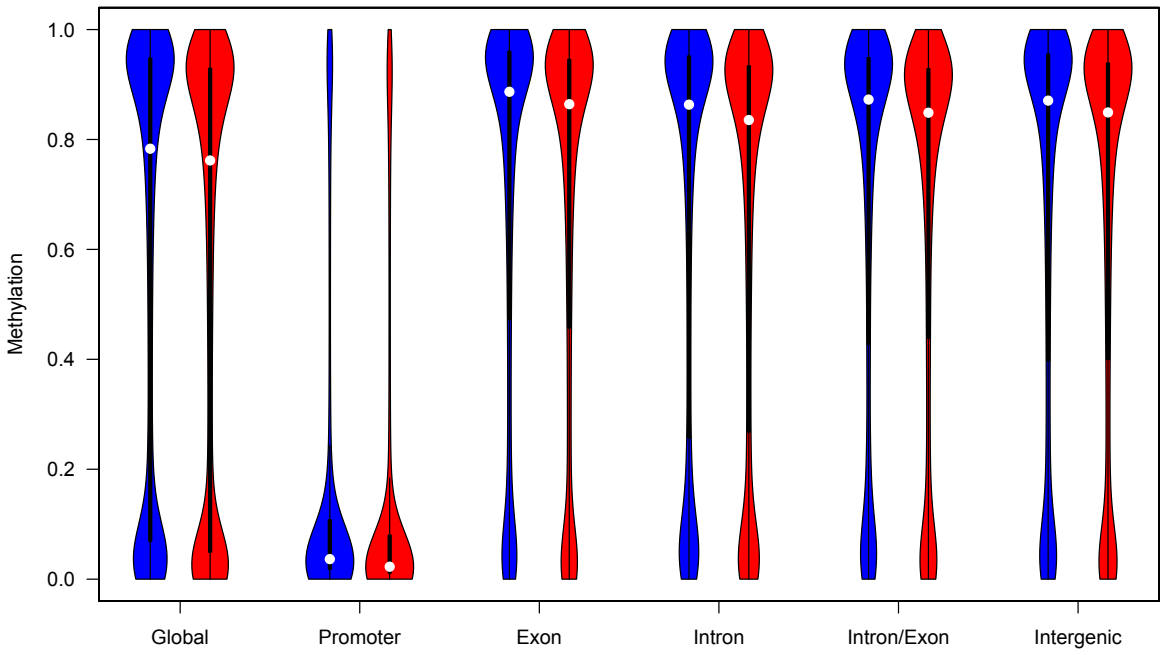
Figure 3: The methylation profile of the *PKD1* gene in ADPKD compared to non-ADPKD kidney. A. UCSC genome browser view of RRBS fragments (in black track; DMFs shown with red asterisk) overlapping the *PKD1* gene (blue track), CpG islands (green track) and ENCODE transcription factor binding sites (grey/black track). The MspI fragment-based methylomes from renal cortex of non-ADPKD (n=3, blue line) and ADPKD kidney (n=4, red line) had 68 fragments (black bars, median length = 86 bp) across the *PKD1* gene. B. Using a nominal raw P-value threshold of <0.05 from the ANOVA, the ADPKD tissue had ten fragments in *PKD1* with methylation differences greater than 10% compared to non-ADPKD kidney (loci indicated in panel A with red asterisks). C. *PKD1* gene expression in ADPKD was significantly higher (Mann Whitney U test, *P*-value < 0.05) than non-ADPKD kidney. An additional ADPKD sample to those analysed by RRBS was included in the expression analysis (n=5). Bars show mean ± SD.

Table 1. Differentially methylated fragments identified in ADPKD

Chr	Fragment Start and End	FDR adjusted P-value (FDR<0.05)	Non-ADPKD Methylation	ADPKD Methylation	Methdiff[1]	Closest protein-coding gene and function
2	771624 – 771728	0.049	0.44	0.79	0.34	-94kb from *TMEM18*
4	793929 – 794143	0.034	1.00	0.82	-0.18	*CPLX1* (intron 2) Cytoplasmic vesicle exocytosis
7	922176 – 922249	0.003	0.33	0.85	0.52	*GET4* (intron 1)[3] Cytosolic protein quality control
7	6495310 – 6495378	0.034	0.37	0.59	0.22	*DAGLB* (intron 1) Diacylgylcerol lipase enzyme
7	6511106 – 6511153	<1.79E-17	1.00	0.90	-0.10	*KDELR2* (intron 2) Protein retention in the ER *DAGLB* (intron 1) Diacylgylcerol lipase enzyme
9	130768316 – 130768366	0.038	0.56	0.13	-0.43	-26 kb from *FAM102A*[2, 3]
10	14979087 – 14979136	<1.79E-17	0.00	1.00	1.00	*DCLRE1C* (intron 4) DNA crosslink repair
12	3862528 – 3862582	0.035	0.25	0.08	-0.17	*EFCAB4B* (promoter)[3] Calcium channel regulator
17	40683608 – 40683711	0.048	0.57	0.34	-0.23	*NAGLU* (promoter)[3] Alpha-N-acetylglucosaminidase
18	10271529 – 10271568	<1.79E-17	1.00	0.80	-0.2	-183kb from *APCDD1*
19	789109 – 789181	0.034	0.71	0.35	-0.36	-8kb from *PTBP1*[2, 3]
20	2513933 – 2513980	1.79E-17	0.90	0.67	-0.23	-9kb from *ZNF343*
22	37499386 – 37499523	0.048	0.25	0.43	0.18	*TMPRSS6* (intron 1/exon 2) Transmembrane serine proteinase

[1] Methdiff = average methylation difference between non-ADPKD and ADPKD kidney tissue (a negative value indicates hypomethylation in ADPKD tissue compared to non-ADPKD).
[2] Fragment overlaps a transcription factor binding site for this gene, as predicted by OregAnno analysis.
[3] Fragment overlaps predicted enhancer regions reported by Genehancer.

Appendix G: Additional data



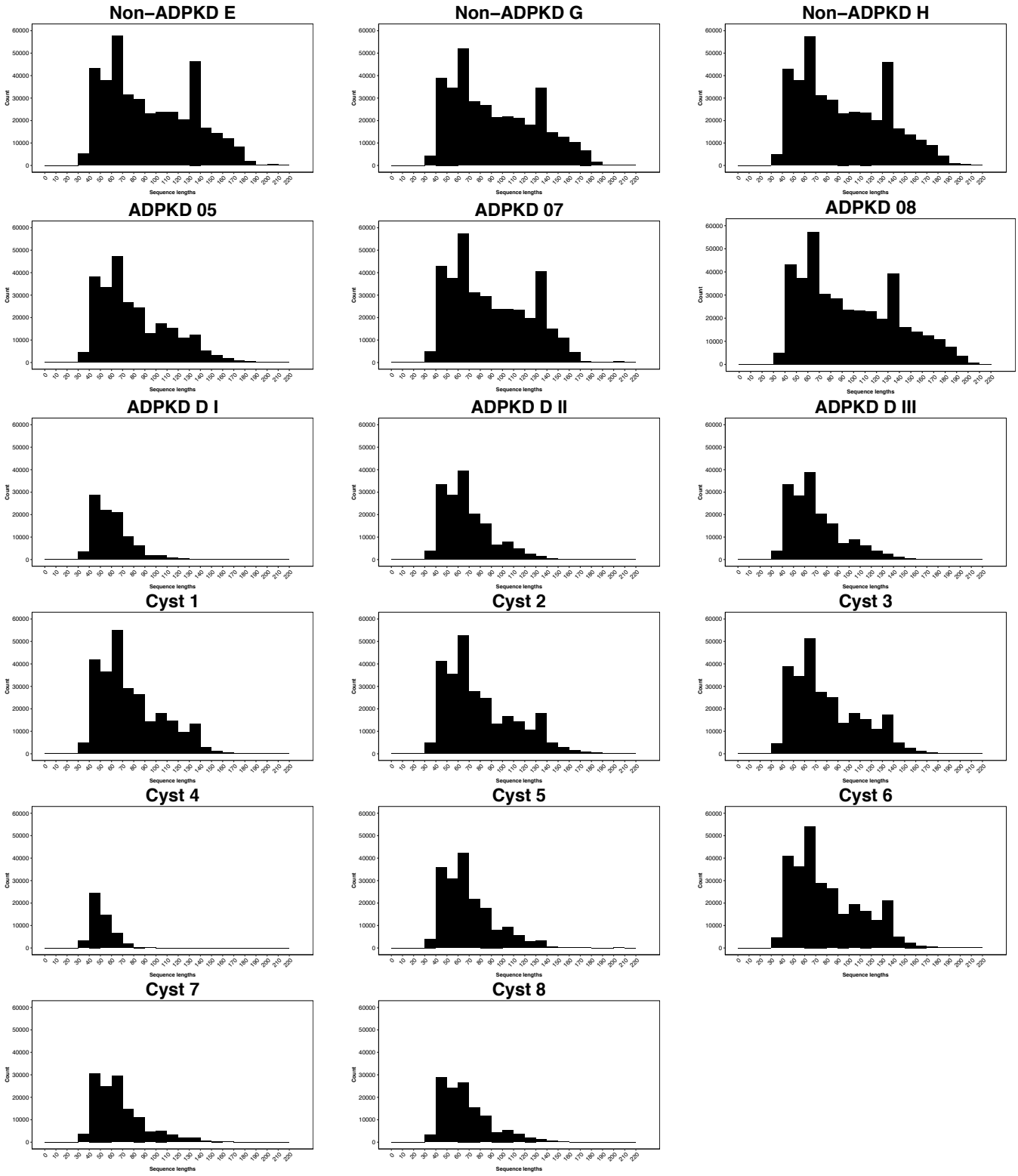
Supplementary Figure G.1: Median global methylation of ADPKD and non-ADPKD tissue excluding sample ADPKD D III

Sample ADPKD D III was excluded from the global analysis to confirm that the lower median methylation of this fragment was not contributing to the pattern of hypomethylation seen in ADPKD. Non-ADPKD $n = 3$ (blue), ADPKD $n = 3$ (red). $p < 2.2E\text{-}16$.

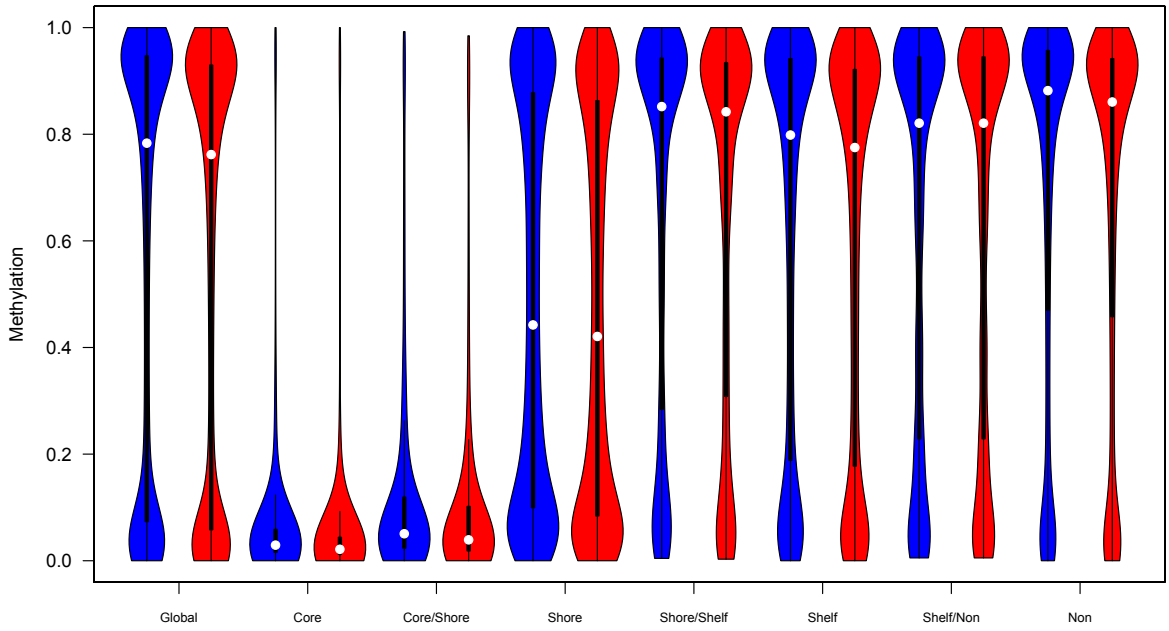
Supplementary Table G.1: Correlation between DMF methylation and *PKD1* expression in ADPKD and non-ADPKD tissue

Correlation calculated with Spearman's correlation: * = $p < 0.05$, ** = $p < 0.01$. XY pairs refers to the number of comparisons between DMF methylation and *PKD1* expression in each Spearman's correlation calculation at each coordinate.

Chromosome coordinates:	r (*p*):	XY pairs:
16:2140121-2140224	0.9286 (**)	7
16:2141559-2141709	1 (**)	6
16:2141710-2141753	0.8214 (*)	7
16:2141754-2141820	0.5714	7
16:2141821-2141911	0.9	5
16:2147611-2147671	0.8857 (*)	6
16:2165619-2165775	-0.6	4
16:2174437-2174483	-0.9643 (**)	7
16:2174683-2174753	-0.7714	6
16:2176100-2176196	-0.8286	6



Supplementary Figure G.2: Distribution of fragment sizes in RRBS libraries
Depiction of the distribution of RRBS fragment lengths. Fragments are no shorter than 40 bp and no greater than 220 bp. Libraries generated from 0.5 µg of DNA do not have a distinct peak at 140 bp unlike their 2.5 µg counterparts.



Supplementary Figure G.3: Median methylome analysis at CpG islands

The median methylation of the non-ADPKD reference sample (blue) and ADPKD D (red) in relation to CpG islands. ADPKD was significantly hypomethylated at all regions, therefore the reason for global hypomethylation could not be due to the over-representation of fragments within all in a CpG core, for example.

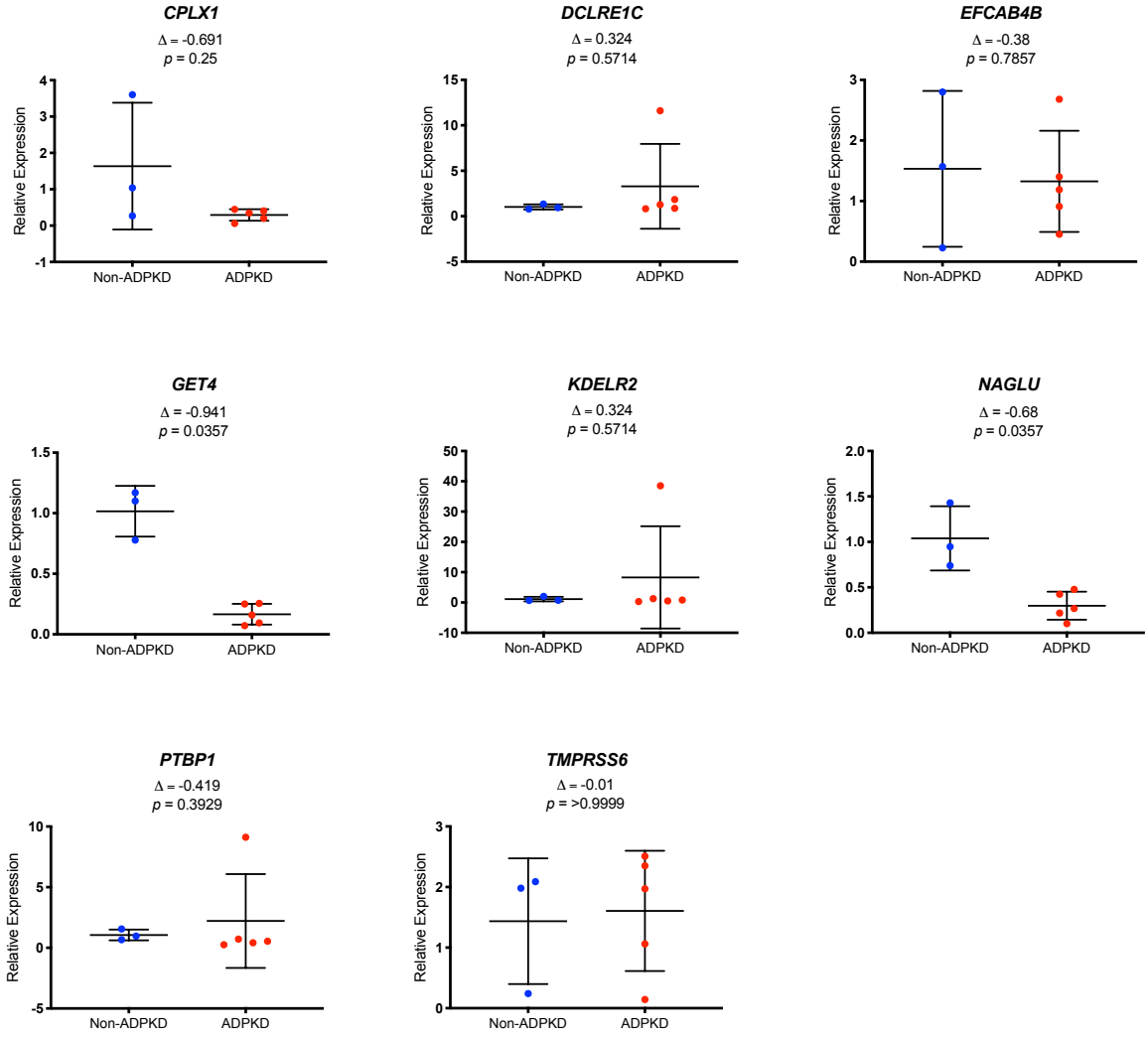
Appendix H: Correlation of RRBS libraries

Supplementary Table H.1: Pearson's correlation coefficient between all 19 RRBS libraries

n = *36*,824 fragments, $p < 2.2E\text{-}16$ for all observations.

	ADPKD 05	ADPKD 07	ADPKD 08	ADPKD D I	ADPKD D II	ADPKD D III	Cyst 1	Cyst 2	Cyst 3	Cyst 4	Cyst 5	Cyst 6	Cyst 7	Cyst 8	WT 9-7	WT 9-12	Non-ADPKD E	Non-ADPKD G	Non-ADPKD H
ADPKD 05	1.000																		
ADPKD 07	0.989	1.000																	
ADPKD 08	0.990	0.990	1.000																
ADPKD D I	0.969	0.976	0.965	1.000															
ADPKD D II	0.982	0.982	0.981	0.973	1.000														
ADPKD D III	0.982	0.982	0.981	0.972	0.993	1.000													
Cyst 1	0.978	0.983	0.975	0.991	0.978	0.978	1.000												
Cyst 2	0.990	0.992	0.988	0.980	0.985	0.985	0.989	1.000											
Cyst 3	0.989	0.989	0.985	0.976	0.983	0.983	0.983	0.992	1.000										
Cyst 4	0.984	0.984	0.980	0.978	0.986	0.986	0.981	0.987	0.988	1.000									
Cyst 5	0.979	0.984	0.977	0.988	0.979	0.978	0.994	0.989	0.984	0.981	1.000								
Cyst 6	0.986	0.990	0.983	0.988	0.984	0.984	0.993	0.993	0.991	0.988	0.991	1.000							
Cyst 7	0.971	0.979	0.969	0.991	0.973	0.972	0.993	0.984	0.978	0.977	0.991	0.989	1.000						
Cyst 8	0.979	0.983	0.974	0.987	0.978	0.978	0.988	0.986	0.986	0.984	0.987	0.992	0.986	1.000					
WT 9-7	0.941	0.945	0.941	0.927	0.953	0.953	0.929	0.940	0.939	0.945	0.931	0.939	0.923	0.936	1.000				
WT 9-12	0.890	0.894	0.884	0.881	0.895	0.894	0.877	0.888	0.891	0.894	0.880	0.890	0.874	0.891	0.935	1.000			
Non-ADPKD E	0.972	0.976	0.977	0.953	0.982	0.982	0.960	0.971	0.969	0.972	0.961	0.967	0.953	0.960	0.956	0.892	1.000		
Non-ADPKD G	0.967	0.971	0.972	0.945	0.976	0.977	0.953	0.966	0.964	0.966	0.955	0.962	0.946	0.954	0.954	0.889	0.989	1.000	
Non-ADPKD H	0.977	0.982	0.984	0.955	0.980	0.980	0.964	0.977	0.974	0.973	0.966	0.972	0.957	0.964	0.953	0.892	0.987	0.985	1.000

Appendix I: DMF-associated expression



Supplementary Figure I.1: qPCR of whole tissue DMF identified genes
Eight genes associated with DMFs were investigated with qPCR in whole tissue from renal samples. *n*(non-ADPKD) = 3, *n* (ADPKD) = 5. Only *GET4* and *NAGLU* had significant differential expression ($p < 0.05$).

References

1. Solazzo, A., et al., *The prevalence of autosomal dominant polycystic kidney disease (ADPKD): A meta-analysis of European literature and prevalence evaluation in the Italian province of Modena suggest that ADPKD is a rare and underdiagnosed condition*. PLoS One, 2018. **13**(1): p. e0190430.
2. Gabow, P.A., et al., *Factors affecting the progression of renal disease in autosomal-dominant polycystic kidney disease*. Kidney Int, 1992. **41**(5): p. 1311-9.
3. Fick, G.M., et al., *Characteristics of very early onset autosomal dominant polycystic kidney disease*. J Am Soc Nephrol, 1993. **3**(12): p. 1863-70.
4. Reed-Gitomer, B., *Autosomal dominant polycystic kidney disease: genetics, epidemiology, and treatment*. Advances in Genomics and Genetics, 2014. **4**: p. 173-183.
5. Wilson, P.D., *Polycystic kidney disease*. N Engl J Med, 2004. **350**(2): p. 151-64.
6. Geng, L., et al., *Identification and localization of polycystin, the PKD1 gene product*. J Clin Invest, 1996. **98**(12): p. 2674-82.
7. de Lemos Barbosa, C.M., et al., *Regulation of CFTR Expression and Arginine Vasopressin Activity Are Dependent on Polycystin-1 in Kidney-Derived Cells*. Cell Physiol Biochem, 2016. **38**(1): p. 28-39.
8. Rangan, G.K., et al., *Recent advances in autosomal-dominant polycystic kidney disease*. Intern Med J, 2016. **46**(8): p. 883-92.
9. Gabow, P.A., *Autosomal dominant polycystic kidney disease*. N Engl J Med, 1993. **329**(5): p. 332-42.
10. Rinkel, G.J., et al., *Prevalence and risk of rupture of intracranial aneurysms: a systematic review*. Stroke, 1998. **29**(1): p. 251-6.
11. Shin Y.B., P.J.H., *Recent Trends in ADPKD Research*, in *Cystogenesis. Advances in Experimental Medicine and Biology*, A.C. Park J., Editor. 2016, Springer: Singapore.
12. Grantham, J.J., *Clinical practice. Autosomal dominant polycystic kidney disease*. N Engl J Med, 2008. **359**(14): p. 1477-85.
13. Braun, W.E., *Autosomal dominant polycystic kidney disease: emerging concepts of pathogenesis and new treatments*. Cleve Clin J Med, 2009. **76**(2): p. 97-104.
14. ANZDATARegistry, *41st Report, Chapter 4: Haemodialysis*., in *Australia and New Zealand Dialysis and Transplant Registry*. 2018: Adelaide, Australia.
15. Torres, V.E., et al., *Tolvaptan in patients with autosomal dominant polycystic kidney disease*. N Engl J Med, 2012. **367**(25): p. 2407-18.
16. Stayner, C., et al., *Targeted Therapies for Autosomal Dominant Polycystic Kidney Disease*. Curr Med Chem, 2018.
17. Chang, M.Y. and A.C. Ong, *New treatments for autosomal dominant polycystic kidney disease*. Br J Clin Pharmacol, 2013. **76**(4): p. 524-35.
18. Hateboer, N., et al., *Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group*. Lancet, 1999. **353**(9147): p. 103-7.
19. Porath, B., et al., *Mutations in GANAB, Encoding the Glucosidase IIalpha Subunit, Cause Autosomal-Dominant Polycystic Kidney and Liver Disease*. Am J Hum Genet, 2016. **98**(6): p. 1193-1207.

20. Cornec-Le Gall, E., et al., *Monoallelic Mutations to DNAJB11 Cause Atypical Autosomal-Dominant Polycystic Kidney Disease*. Am J Hum Genet, 2018. **102**(5): p. 832-844.
21. Reed, B., et al., *Presence of de novo mutations in autosomal dominant polycystic kidney disease patients without family history*. Am J Kidney Dis, 2008. **52**(6): p. 1042-50.
22. Gout, A.M., et al., *PKDB: Polycystic Kidney Disease Mutation Database--a gene variant database for autosomal dominant polycystic kidney disease*. Hum Mutat, 2007. **28**(7): p. 654-9.
23. Rossetti, S., et al., *Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications*. Am J Hum Genet, 2001. **68**(1): p. 46-63.
24. Kiktev, D.A., et al., *GC content elevates mutation and recombination rates in the yeast Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A, 2018. **115**(30): p. E7109-e7118.
25. Blaszak, R.T., et al., *DNA structural transitions within the PKD1 gene*. Nucleic Acids Res, 1999. **27**(13): p. 2610-7.
26. The European Polycystic Kidney Disease Consortium, C.J.W., et al., *The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16*. Cell, 1994. **77**(6): p. 881-894.
27. Kent, W.J., et al., *The human genome browser at UCSC*. Genome Res, 2002. **12**(6): p. 996-1006.
28. Casper, J., et al., *The UCSC Genome Browser database: 2018 update*. Nucleic Acids Res, 2018. **46**(D1): p. D762-d769.
29. Lohning, C., U. Nowicka, and A.M. Frischauf, *The mouse homolog of PKD1: sequence analysis and alternative splicing*. Mamm Genome, 1997. **8**(5): p. 307-11.
30. Xu, H., et al., *Tissue-specific expression and splicing of the rat polycystic kidney disease 1 gene*. DNA Seq, 2001. **12**(5-6): p. 361-6.
31. *Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. The International Polycystic Kidney Disease Consortium*. Cell, 1995. **81**(2): p. 289-98.
32. Bogdanova, N., et al., *Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes*. Genomics, 2001. **74**(3): p. 333-41.
33. Cooke, S.L., et al., *Processed pseudogenes acquired somatically during cancer development*. Nat Commun, 2014. **5**: p. 3644.
34. Mochizuki, T., et al., *PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein*. Science, 1996. **272**(5266): p. 1339-42.
35. Hackmann, K., et al., *A splice form of polycystin-2, lacking exon 7, does not interact with polycystin-1*. Hum Mol Genet, 2005. **14**(21): p. 3249-62.
36. Gonzalez-Perrett, S., et al., *Polycystin-2, the protein mutated in autosomal dominant polycystic kidney disease (ADPKD), is a Ca2+-permeable nonselective cation channel*. Proc Natl Acad Sci U S A, 2001. **98**(3): p. 1182-7.
37. Nauli, S.M., et al., *Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells*. Nat Genet, 2003. **33**(2): p. 129-37.
38. Carone, F.A., et al., *Cyst-derived cells do not exhibit accelerated growth or features of transformed cells in vitro*. Kidney Int, 1989. **35**(6): p. 1351-7.
39. Wilson, P.D., et al., *The PKD1 gene product, "polycystin-1," is a tyrosine-phosphorylated protein that colocalizes with alpha2beta1-integrin in focal clusters in adherent renal epithelia*. Lab Invest, 1999. **79**(10): p. 1311-23.
40. Drummond, I.A., *Polycystins, focal adhesions and extracellular matrix interactions*. Biochim Biophys Acta, 2011. **1812**(10): p. 1322-6.

41. Kreidberg, J.A., et al., *Alpha 3 beta 1 integrin has a crucial role in kidney and lung organogenesis*. Development, 1996. **122**(11): p. 3537-47.
42. Lo, S.H., et al., *Progressive kidney degeneration in mice lacking tensin*. J Cell Biol, 1997. **136**(6): p. 1349-61.
43. Wright, F.S. and G. Giebisch, *Renal potassium transport: contributions of individual nephron segments and populations*. Am J Physiol, 1978. **235**(6): p. F515-27.
44. Wilson, P.D., et al., *Reversed polarity of Na(+) -K(+) -ATPase: mislocation to apical plasma membranes in polycystic kidney disease epithelia*. Am J Physiol, 1991. **260**(3 Pt 2): p. F420-30.
45. Wilson, P.D., *Epithelial cell polarity and disease*. Am J Physiol, 1997. **272**(4 Pt 2): p. F434-42.
46. Stocklin, E., F. Botteri, and B. Groner, *An activated allele of the c-erbB-2 oncogene impairs kidney and lung function and causes early death of transgenic mice*. J Cell Biol, 1993. **122**(1): p. 199-208.
47. Lu, W., et al., *Perinatal lethality with kidney and pancreas defects in mice with a targetted Pkd1 mutation*. Nat Genet, 1997. **17**(2): p. 179-81.
48. Boulter, C., et al., *Cardiovascular, skeletal, and renal defects in mice with a targeted disruption of the Pkd1 gene*. Proc Natl Acad Sci U S A, 2001. **98**(21): p. 12174-9.
49. Lantinga-van Leeuwen, I.S., et al., *Kidney-specific inactivation of the Pkd1 gene induces rapid cyst formation in developing kidneys and a slow onset of disease in adult mice*. Hum Mol Genet, 2007. **16**(24): p. 3188-96.
50. Natoli, T.A., et al., *Pkd1 and Nek8 mutations affect cell-cell adhesion and cilia in cysts formed in kidney organ cultures*. Am J Physiol Renal Physiol, 2008. **294**(1): p. F73-83.
51. Bhunia, A.K., et al., *PKD1 induces p21(waf1) and regulation of the cell cycle via direct activation of the JAK-STAT signaling pathway in a process requiring PKD2*. Cell, 2002. **109**(2): p. 157-68.
52. Shibazaki, S., et al., *Cyst formation and activation of the extracellular regulated kinase pathway after kidney specific inactivation of Pkd1*. Hum Mol Genet, 2008. **17**(11): p. 1505-16.
53. Starremans, P.G., et al., *A mouse model for polycystic kidney disease through a somatic in-frame deletion in the 5' end of Pkd1*. Kidney Int, 2008. **73**(12): p. 1394-405.
54. Piontek, K.B., et al., *A functional floxed allele of Pkd1 that can be conditionally inactivated in vivo*. J Am Soc Nephrol, 2004. **15**(12): p. 3035-43.
55. Piontek, K., et al., *A critical developmental switch defines the kinetics of kidney cyst formation after loss of Pkd1*. Nat Med, 2007. **13**(12): p. 1490-5.
56. Lantinga-van Leeuwen, I.S., et al., *Lowering of Pkd1 expression is sufficient to cause polycystic kidney disease*. Hum Mol Genet, 2004. **13**(24): p. 3069-77.
57. Happe, H., et al., *Cyst expansion and regression in a mouse model of polycystic kidney disease*. Kidney Int, 2013. **83**(6): p. 1099-108.
58. Jiang, S.T., et al., *Defining a link with autosomal-dominant polycystic kidney disease in mice with congenitally low expression of Pkd1*. Am J Pathol, 2006. **168**(1): p. 205-20.
59. Wang, E., et al., *Progressive renal distortion by multiple cysts in transgenic mice expressing artificial microRNAs against Pkd1*. J Pathol, 2010. **222**(3): p. 238-48.
60. Kurbegovic, A. and M. Trudel, *Progressive development of polycystic kidney disease in the mouse model expressing Pkd1 extracellular domain*. Hum Mol Genet, 2013. **22**(12): p. 2361-75.

61. Kurbegovic, A., et al., *Pkd1 transgenic mice: adult model of polycystic kidney disease with extrarenal and renal phenotypes*. Hum Mol Genet, 2010. **19**(7): p. 1174-89.
62. Ibraghimov-Beskrovnaya, O., et al., *Polycystin: in vitro synthesis, in vivo tissue expression, and subcellular localization identifies a large membrane-associated protein*. Proc Natl Acad Sci U S A, 1997. **94**(12): p. 6397-402.
63. Ong, A.C., et al., *Polycystin-1 expression in PKD1, early-onset PKD1, and TSC2/PKD1 cystic tissue*. Kidney Int, 1999. **56**(4): p. 1324-33.
64. Ward, C.J., et al., *Polycystin, the polycystic kidney disease 1 protein, is expressed by epithelial cells in fetal, adult, and polycystic kidney*. Proc Natl Acad Sci U S A, 1996. **93**(4): p. 1524-8.
65. Palsson, R., et al., *Characterization and cell distribution of polycystin, the product of autosomal dominant polycystic kidney disease gene 1*. Mol Med, 1996. **2**(6): p. 702-11.
66. Van Adelsberg, J., S. Chamberlain, and V. D'Agati, *Polycystin expression is temporally and spatially regulated during renal development*. Am J Physiol, 1997. **272**(5 Pt 2): p. F602-9.
67. Happe, H. and D.J. Peters, *Translational research in ADPKD: lessons from animal models*. Nat Rev Nephrol, 2014. **10**(10): p. 587-601.
68. Qian, F., et al., *The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I*. Cell, 1996. **87**(6): p. 979-87.
69. Grantham, J.J., et al., *Volume progression in polycystic kidney disease*. N Engl J Med, 2006. **354**(20): p. 2122-30.
70. Irazabal, M.V., et al., *Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials*. J Am Soc Nephrol, 2015. **26**(1): p. 160-72.
71. Kriz, W., Elger, M, *Renal Anatomy*, in *Comprehensive Clinical Nephrology*, R.J.J. J. Floege, and J. Feehally, Editor. 2010, Mosby: Philadelphia. p. 3-14.
72. Badenas, C., et al., *Loss of heterozygosity in renal and hepatic epithelial cystic cells from ADPKD1 patients*. Eur J Hum Genet, 2000. **8**(7): p. 487-92.
73. Reeders, S.T., *Multilocus polycystic disease*. Nat Genet, 1992. **1**(4): p. 235-7.
74. Brasier, J.L. and E.P. Henske, *Loss of the polycystic kidney disease (PKD1) region of chromosome 16p13 in renal cyst cells supports a loss-of-function model for cyst pathogenesis*. J Clin Invest, 1997. **99**(2): p. 194-9.
75. Koptides, M., et al., *Loss of heterozygosity in polycystic kidney disease with a missense mutation in the repeated region of PKD1*. Hum Genet, 1998. **103**(6): p. 709-17.
76. Rossetti, S., et al., *Incompletely penetrant PKD1 alleles suggest a role for gene dosage in cyst initiation in polycystic kidney disease*. Kidney Int, 2009. **75**(8): p. 848-55.
77. Vujic, M., et al., *Incompletely penetrant PKD1 alleles mimic the renal manifestations of ARPKD*. J Am Soc Nephrol, 2010. **21**(7): p. 1097-102.
78. Torres, M., et al., *Pax-2 controls multiple steps of urogenital development*. Development, 1995. **121**(12): p. 4057-65.
79. Stayner, C., et al., *Pax2 gene dosage influences cystogenesis in autosomal dominant polycystic kidney disease*. Hum Mol Genet, 2006. **15**(24): p. 3520-8.
80. Shillingford, J.M., et al., *The mTOR pathway is regulated by polycystin-1, and its inhibition reverses renal cystogenesis in polycystic kidney disease*. Proc Natl Acad Sci U S A, 2006. **103**(14): p. 5466-71.
81. Parrot, C., et al., *C-MYC is a regulator ofthe PKD1 gene and PC1-induced pathogenesis*. Hum Mol Genet, 2018.

82. Basten, S.G. and R.H. Giles, *Functional aspects of primary cilia in signaling, cell cycle and tumorigenesis*. Cilia, 2013. **2**(1): p. 6.
83. Lin, F., et al., *Kidney-specific inactivation of the KIF3A subunit of kinesin-II inhibits renal ciliogenesis and produces polycystic kidney disease*. Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5286-91.
84. Pazour, G.J., et al., *Polycystin-2 localizes to kidney cilia and the ciliary level is elevated in orpk mice with polycystic kidney disease*. Curr Biol, 2002. **12**(11): p. R378-80.
85. Nauli, S.M., et al., *Loss of polycystin-1 in human cyst-lining epithelia leads to ciliary dysfunction*. J Am Soc Nephrol, 2006. **17**(4): p. 1015-25.
86. Bergmann, C., *ARPKD and early manifestations of ADPKD: the original polycystic kidney disease and phenocopies*. Pediatr Nephrol, 2015. **30**(1): p. 15-30.
87. Torres, V.E. and P.C. Harris, *Mechanisms of Disease: autosomal dominant and recessive polycystic kidney diseases*. Nat Clin Pract Nephrol, 2006. **2**(1): p. 40-55; quiz 55.
88. Luo, F. and Y.H. Tao, *Nephronophthisis: A review of genotype-phenotype correlation*. Nephrology (Carlton), 2018. **23**(10): p. 904-911.
89. Hateboer, N., et al., *Confirmation of a gene locus for medullary cystic kidney disease (MCKD2) on chromosome 16p12*. Kidney Int, 2001. **60**(4): p. 1233-9.
90. Fuchshuber, A., et al., *Refinement of the gene locus for autosomal dominant medullary cystic kidney disease type 1 (MCKD1) and construction of a physical and partial transcriptional map of the region*. Genomics, 2001. **72**(3): p. 278-84.
91. Li, A., et al., *Mutations in PRKCSH cause isolated autosomal dominant polycystic liver disease*. Am J Hum Genet, 2003. **72**(3): p. 691-703.
92. Davila, S., et al., *Mutations in SEC63 cause autosomal dominant polycystic liver disease*. Nat Genet, 2004. **36**(6): p. 575-7.
93. Grantham, J.J., *Polycystic kidney disease: neoplasia in disguise*. Am J Kidney Dis, 1990. **15**(2): p. 110-6.
94. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell, 2011. **144**(5): p. 646-74.
95. Seeger-Nukpezah, T., et al., *The hallmarks of cancer: relevance to the pathogenesis of polycystic kidney disease*. Nat Rev Nephrol, 2015. **11**(9): p. 515-34.
96. Li, X., *Epigenetics and autosomal dominant polycystic kidney disease*. Biochim Biophys Acta, 2011. **1812**(10): p. 1213-8.
97. Li, M., et al., *Genomic instability in patients with autosomal-dominant polycystic kidney disease*. J Int Med Res, 2013. **41**(1): p. 169-75.
98. Nishio, S., et al., *Pkd1 regulates immortalized proliferation of renal tubular epithelial cells through p53 induction and JNK activation*. J Clin Invest, 2005. **115**(4): p. 910-8.
99. Fan, L.X., et al., *Inhibition of histone deacetylases targets the transcription regulator Id2 to attenuate cystic epithelial cell proliferation*. Kidney Int, 2012. **81**(1): p. 76-85.
100. Veis, D.J., et al., *Bcl-2-deficient mice demonstrate fulminant lymphoid apoptosis, polycystic kidneys, and hypopigmented hair*. Cell, 1993. **75**(2): p. 229-40.
101. Zeier, M., et al., *Renal histology in polycystic kidney disease with incipient and advanced renal failure*. Kidney Int, 1992. **42**(5): p. 1259-65.
102. Bello-Reuss, E., K. Holubec, and S. Rajaraman, *Angiogenesis in autosomal-dominant polycystic kidney disease*. Kidney Int, 2001. **60**(1): p. 37-45.
103. Wei, W., et al., *Evidence of angiogenesis and microvascular regression in autosomal-dominant polycystic kidney disease kidneys: a corrosion cast study*. Kidney Int, 2006. **70**(7): p. 1261-8.

104. Mao, Z., G. Xie, and A.C. Ong, *Metabolic abnormalities in autosomal dominant polycystic kidney disease*. Nephrol Dial Transplant, 2015. **30**(2): p. 197-203.
105. Harris, P.C. and V.E. Torres, *Genetic mechanisms and signaling pathways in autosomal dominant polycystic kidney disease*. J Clin Invest, 2014. **124**(6): p. 2315-24.
106. Negrini, S., V.G. Gorgoulis, and T.D. Halazonetis, *Genomic instability--an evolving hallmark of cancer*. Nat Rev Mol Cell Biol, 2010. **11**(3): p. 220-8.
107. Lewis, B.P., et al., *Prediction of mammalian microRNA targets*. Cell, 2003. **115**(7): p. 787-98.
108. Esquela-Kerscher, A. and F.J. Slack, *Oncomirs - microRNAs with a role in cancer*. Nat Rev Cancer, 2006. **6**(4): p. 259-69.
109. Kehl, T., et al., *About miRNAs, miRNA seeds, target genes and target pathways*. Oncotarget, 2017. **8**(63): p. 107167-107175.
110. Hammond, S.M., *An overview of microRNAs*. Adv Drug Deliv Rev, 2015. **87**: p. 3-14.
111. Trionfini, P. and A. Benigni, *MicroRNAs as Master Regulators of Glomerular Function in Health and Disease*. J Am Soc Nephrol, 2017. **28**(6): p. 1686-1696.
112. Woo, Y.M., et al., *Profiling of miRNAs and target genes related to cystogenesis in ADPKD mouse models*. Sci Rep, 2017. **7**(1): p. 14151.
113. Sun L, Z.J., Wu M, Sun H, Zhou C, Fu L, Xu C, Mei C., *Inhibition of MiR-199a-5p reduced cell proliferation in autosomal dominant polycystic kidney disease through targeting CDKN1C*. Med Sci Monit, 2015. **15**(21): p. 195-200.
114. Kim, D.Y., et al., *Impact of miR-192 and miR-194 on cyst enlargement through EMT in autosomal dominant polycystic kidney disease*. Faseb j, 2019. **33**(2): p. 2870-2884.
115. de Stephanis, L., et al., *MicroRNA501-5p induces p53 proteasome degradation through the activation of the mTOR/MDM2 pathway in ADPKD cells*. J Cell Physiol, 2018. **233**(9): p. 6911-6924.
116. Patel, V., et al., *miR-17~92 miRNA cluster promotes kidney cyst growth in polycystic kidney disease*. Proc Natl Acad Sci U S A, 2013. **110**(26): p. 10765-70.
117. Yheskel, M., et al., *Anti-microRNA screen uncovers miR-17 family within miR-17~92 cluster as the primary driver of kidney cyst growth*. Sci Rep, 2019. **9**(1): p. 1920.
118. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. Nature, 2000. **403**(6765): p. 41-5.
119. Chen, H.P., Y.T. Zhao, and T.C. Zhao, *Histone deacetylases and mechanisms of regulation of gene expression*. Crit Rev Oncog, 2015. **20**(1-2): p. 35-47.
120. Van Bodegom, D., et al., *The polycystic kidney disease-1 gene is a target for p53-mediated transcriptional repression*. J Biol Chem, 2006. **281**(42): p. 31234-44.
121. Xia, S., et al., *Polycystin-dependent fluid flow sensing targets histone deacetylase 5 to prevent the development of renal cysts*. Development, 2010. **137**(7): p. 1075-84.
122. Yanda, M.K., et al., *Histone deacetylase 6 inhibition reduces cysts by decreasing cAMP and Ca(2+) in knock-out mouse models of polycystic kidney disease*. J Biol Chem, 2017. **292**(43): p. 17897-17908.
123. West, A.C. and R.W. Johnstone, *New and emerging HDAC inhibitors for cancer treatment*. J Clin Invest, 2014. **124**(1): p. 30-9.
124. Ozdag, H., et al., *Differential expression of selected histone modifier genes in human solid cancers*. BMC Genomics, 2006. **7**: p. 90.
125. Qi, Y., et al., *HEDD: the human epigenetic drug database*. Database (Oxford), 2016.
126. Cao, Y., et al., *Chemical modifier screen identifies HDAC inhibitors as suppressors of PKD models*. Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21819-24.
127. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences*. Nature, 2009. **462**(7271): p. 315-22.

128. Ehrlich, M., et al., *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*. Nucleic Acids Res, 1982. **10**(8): p. 2709-21.
129. Herman, J.G. and S.B. Baylin, *Gene silencing in cancer in association with promoter hypermethylation*. N Engl J Med, 2003. **349**(21): p. 2042-54.
130. Watt, F. and P.L. Molloy, *Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter*. Genes Dev, 1988. **2**(9): p. 1136-43.
131. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex*. Nature, 1998. **393**(6683): p. 386-9.
132. Flanagan, J.M. and L. Wild, *An epigenetic role for noncoding RNAs and intragenic DNA methylation*. Genome Biol, 2007. **8**(6): p. 307.
133. Rauch, T.A., et al., *A human B cell methylome at 100-base pair resolution*. Proc Natl Acad Sci U S A, 2009. **106**(3): p. 671-8.
134. Zhou, J., et al., *Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation*. BMC Genomics, 2017. **18**(1): p. 724.
135. Migeon, B.R., *Choosing the Active X: The Human Version of X Inactivation*. Trends Genet, 2017. **33**(12): p. 899-909.
136. Tang, W.W., et al., *Specification and epigenetic programming of the human germ line*. Nat Rev Genet, 2016. **17**(10): p. 585-600.
137. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer*. Cell, 2007. **128**(4): p. 683-92.
138. Jones, P.A., J.P. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy*. Nat Rev Genet, 2016. **17**(10): p. 630-41.
139. Woo, Y.M., et al., *Genome-wide methylation profiling of ADPKD identified epigenetically regulated genes associated with renal cyst development*. Hum Genet, 2014. **133**(3): p. 281-97.
140. Bates, M., *DNA Methylation in Polycystic Kidney Disease and Bio-distribition of the C3-G12 peptide*. 2016, University of Otago: Dunedin, New Zealand.
141. Loghman-Adham, M., et al., *Immortalized epithelial cells from human autosomal dominant polycystic kidney cysts*. Am J Physiol Renal Physiol, 2003. **285**(3): p. F397-412.
142. Chatterjee, A., et al., *Genome-scale DNA methylome and transcriptome profiling of human neutrophils*. Sci Data, 2016. **3**: p. 160019.
143. Illumina, I., *TruSeq® DNA Sample Preparation Guide*. 2012: San Diego, United States of America. p. 45.
144. Stockwell, P.A., et al., *DMAP: differential methylation analysis package for RRBS and WGBS data*. Bioinformatics, 2014. **30**(13): p. 1814-22.
145. Wilkerson, M.D. and D.N. Hayes, *ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking*. Bioinformatics, 2010. **26**(12): p. 1572-3.
146. Vandesompele, J., et al., *Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes*. Genome Biol, 2002. **3**(7): p. Research0034.
147. Hellemans, J., et al., *qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data*. Genome Biol, 2007. **8**(2): p. R19.
148. Bowden, S.A., et al., *Genome-Scale Single Nucleotide Resolution Analysis of DNA Methylation in Human Autosomal Dominant Polycystic Kidney Disease*. Am J Nephrol, 2018. **48**(6): p. 415-424.

149. McBride, G.B., *A Proposal for Strength-of-Agreement Criteria for Lins Concordance Correlation Coefficient*. 2005, National Institute of Water & Atmospheric Research Ltd: www.medcalc.org.
150. Fishilevich, S., et al., *GeneHancer: genome-wide integration of enhancers and target genes in GeneCards*. Database (Oxford), 2017. **2017**.
151. Wilson, P.D., et al., *A new method for studying human polycystic kidney disease epithelia in culture*. Kidney Int, 1986. **30**(3): p. 371-8.
152. Smith, Z.D., et al., *High-throughput bisulfite sequencing in mammalian genomes*. Methods, 2009. **48**(3): p. 226-32.
153. Chatterjee, A., et al., *Genome-wide DNA methylation map of human neutrophils reveals widespread inter-individual epigenetic variation*. Sci Rep, 2015. **5**: p. 17328.
154. Ludgate, J.L., et al., *A streamlined method for analysing genome-wide DNA methylation patterns from low amounts of FFPE DNA*. BMC Med Genomics, 2017. **10**(1): p. 54.
155. Bresters, D., et al., *The duration of fixation influences the yield of HCV cDNA-PCR products from formalin-fixed, paraffin-embedded liver tissue*. J Virol Methods, 1994. **48**(2-3): p. 267-72.
156. Patel, P.G., et al., *Reliability and performance of commercial RNA and DNA extraction kits for FFPE tissue cores*. PLoS One, 2017. **12**(6): p. e0179732.
157. Lu, X.J.D., et al., *Using ddPCR to assess the DNA yield of FFPE samples*. Biomol Detect Quantif, 2018. **16**: p. 5-11.
158. Carmona, J.J., et al., *Empirical comparison of reduced representation bisulfite sequencing and Infinium BeadChip reproducibility and coverage of DNA methylation in humans*. NPJ Genom Med, 2017. **2**: p. 13.
159. Chatterjee, A., et al., *Genome-wide methylation sequencing of paired primary and metastatic cell lines identifies common DNA methylation changes and a role for EBF3 as a candidate epigenetic driver of melanoma metastasis*. Oncotarget, 2017. **8**(4): p. 6085-6101.
160. Garrett-Bakelman, F.E., et al., *Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution*. J Vis Exp, 2015(96): p. e52246.
161. Gu, H., et al., *Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling*. Nat Protoc, 2011. **6**(4): p. 468-81.
162. Lokk, K., et al., *DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns*. Genome Biol, 2014. **15**(4): p. r54.
163. Yaoita, E. and Y. Yoshida, *Polygonal epithelial cells in glomerular cell culture: Podocyte or parietal epithelial origin?* Microsc Res Tech, 2002. **57**(4): p. 212-6.
164. Milavetz, B., et al., *Virion-mediated transfer of SV40 epigenetic information*. Epigenetics, 2012. **7**(6): p. 528-34.
165. Bork, S., et al., *DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells*. Aging Cell, 2010. **9**(1): p. 54-63.
166. Saferali, A., et al., *Cell culture-induced aberrant methylation of the imprinted IG DMR in human lymphoblastoid cell lines*. Epigenetics, 2010. **5**(1): p. 50-60.
167. Li, Y. and R.A. Wingert, *Regenerative medicine for the kidney: stem cell prospects & challenges*. Clin Transl Med, 2013. **2**(1): p. 11.
168. Rowan, C.J., S. Sheybani-Deloui, and N.D. Rosenblum, *Origin and Function of the Renal Stroma in Health and Disease*. Results Probl Cell Differ, 2017. **60**: p. 205-229.
169. Brocks, D., et al., *Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer*. Cell Rep, 2014. **8**(3): p. 798-806.

170. Trudel, M., *c-Myc Signalling in the Genetic Mechanism of Polycystic Kidney Disease*, in *Polycystic Kidney Disease*, X. Li, Editor. 2015, Codon Publications Copyright: The Author.: Brisbane (AU).
171. Hughes, P., et al., *Loss of PKD1 and loss of Bcl-2 elicit polycystic kidney disease through distinct mechanisms*. Cell Death Differ, 2006. **13**(7): p. 1123-7.
172. Trudel, M., et al., *C-myc-induced apoptosis in polycystic kidney disease is Bcl-2 and p53 independent*. J Exp Med, 1997. **186**(11): p. 1873-84.
173. Pei, L., et al., *Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia*. Epigenetics, 2012. **7**(6): p. 567-78.
174. Chan, K.C., et al., *Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing*. Proc Natl Acad Sci U S A, 2013. **110**(47): p. 18761-8.
175. Legendre, C., et al., *Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer*. Clin Epigenetics, 2015. **7**: p. 100.
176. Carpenter, B.L., et al., *Integrin alpha6beta4 Upregulates Amphiregulin and Epiregulin through Base Excision Repair-Mediated DNA Demethylation and Promotes Genome-wide DNA Hypomethylation*. Sci Rep, 2017. **7**(1): p. 6174.
177. Treangen, T.J. and S.L. Salzberg, *Repetitive DNA and next-generation sequencing: computational challenges and solutions*. Nat Rev Genet, 2011. **13**(1): p. 36-46.
178. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond*. Nat Rev Genet, 2012. **13**(7): p. 484-92.
179. Zheng, Y., et al., *Prediction of genome-wide DNA methylation in repetitive elements*. Nucleic Acids Res, 2017. **45**(15): p. 8697-8711.
180. Almamun, M., et al., *Genome-wide DNA methylation analysis in precursor B-cells*. Epigenetics, 2014. **9**(12): p. 1588-95.
181. Bock, C., et al., *Quantitative comparison of genome-wide DNA methylation mapping technologies*. Nat Biotechnol, 2010. **28**(10): p. 1106-14.
182. Laird, P.W., *Principles and challenges of genomewide DNA methylation analysis*. Nat Rev Genet, 2010. **11**(3): p. 191-203.
183. Jung, M., et al., *MIRA-seq for DNA methylation analysis of CpG islands*. Epigenomics, 2015. **7**(5): p. 695-706.
184. Takahashi, M., et al., *Isolation of a novel human gene, APCDD1, as a direct target of the beta-Catenin/T-cell factor 4 complex with probable involvement in colorectal carcinogenesis*. Cancer Res, 2002. **62**(20): p. 5651-6.
185. Schatoff, E.M., B.I. Leach, and L.E. Dow, *Wnt Signaling and Colorectal Cancer*. Curr Colorectal Cancer Rep, 2017. **13**(2): p. 101-110.
186. Han, W. and J. Liu, *Epigenetic silencing of the Wnt antagonist APCDD1 by promoter DNA hyper-methylation contributes to osteosarcoma cell invasion and metastasis*. Biochem Biophys Res Commun, 2017. **491**(1): p. 91-97.
187. Fagerberg, L., et al., *Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics*. Mol Cell Proteomics, 2014. **13**(2): p. 397-406.
188. Jurvansuu, J., et al., *Transmembrane protein 18 enhances the tropism of neural stem cells for glioma cells*. Cancer Res, 2008. **68**(12): p. 4614-22.
189. Almen, M.S., et al., *The obesity gene, TMEM18, is of ancient origin, found in majority of neuronal cells in all major brain regions and associated with obesity in severely obese children*. BMC Med Genet, 2010. **11**: p. 58.
190. Rohde, K., et al., *Adipose tissue depot specific promoter methylation of TMEM18*. J Mol Med (Berl), 2014. **92**(8): p. 881-8.

191. Wang, D.Y., et al., *Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1*. Mol Endocrinol, 2004. **18**(2): p. 402-11.
192. Taniguchi, K., et al., *PTBP1-associated microRNA-1 and -133b suppress the Warburg effect in colorectal tumors*. Oncotarget, 2016. **7**(14): p. 18940-52.
193. Lai, Y., et al., *Complexin inhibits spontaneous release and synchronizes Ca2+-triggered synaptic vesicle fusion by distinct mechanisms*. Elife, 2014. **3**: p. e03756.
194. Mackie, K., *Cannabinoid receptors as therapeutic targets*. Annu Rev Pharmacol Toxicol, 2006. **46**: p. 101-22.
195. Brittis, P.A., et al., *Fibroblast growth factor receptor function is required for the orderly projection of ganglion cell axons in the developing mammalian retina*. Mol Cell Neurosci, 1996. **8**(2-3): p. 120-8.
196. Ruggiero, C., et al., *A Golgi-based KDELR-dependent signalling pathway controls extracellular matrix degradation*. Oncotarget, 2015. **6**(5): p. 3375-93.
197. Srikanth, S., et al., *A novel EF-hand protein, CRACR2A, is a cytosolic Ca2+ sensor that stabilizes CRAC channels in T cells*. Nat Cell Biol, 2010. **12**(5): p. 436-46.
198. Zandarashvili, L., et al., *Structural impact of complete CpG methylation within target DNA on specific complex formation of the inducible transcription factor Egr-1*. FEBS Lett, 2015. **589**(15): p. 1748-53.
199. Sukhatme, V.P., et al., *A novel early growth response gene rapidly induced by fibroblast, epithelial cell and lymphocyte mitogens*. Oncogene Res, 1987. **1**(4): p. 343-55.
200. Hong, C.J. and B.A. Hamilton, *Zfp423 Regulates Sonic Hedgehog Signaling via Primary Cilium Function*. PLoS Genet, 2016. **12**(10): p. e1006357.
201. Legue, E. and K.F. Liem, Jr., *Tulp3 Is a Ciliary Trafficking Gene that Regulates Polycystic Kidney Disease*. Curr Biol, 2019. **29**(5): p. 803-812.e5.
202. Volk, T., et al., *DCLRE1C (ARTEMIS) mutations causing phenotypes ranging from atypical severe combined immunodeficiency to mere antibody deficiency*. Hum Mol Genet, 2015. **24**(25): p. 7361-72.
203. Dion, S.P., et al., *Transcriptome analysis reveals TMPRSS6 isoforms with distinct functionalities*. J Cell Mol Med, 2018. **22**(4): p. 2498-2509.
204. Zhou, L.T., et al., *Are Urinary Tubular Injury Markers Useful in Chronic Kidney Disease? A Systematic Review and Meta Analysis*. PLoS One, 2016. **11**(12): p. e0167334.
205. Li, J., H. Xie, and Y. Jiang, *Mucopolysaccharidosis IIIB and mild skeletal anomalies: coexistence of NAGLU and CYP26B1 missense variations in the same patient in a Chinese family*. BMC Med Genet, 2018. **19**(1): p. 51.
206. Meijer, O.L.M., et al., *Processing of mutant N-acetyl-alpha-glucosaminidase in mucopolysaccharidosis type IIIB fibroblasts cultured at low temperature*. Mol Genet Metab, 2017. **122**(1-2): p. 100-106.
207. Zulueta, M.M.L., C.L. Chyan, and S.C. Hung, *Structural analysis of synthetic heparan sulfate oligosaccharides with fibroblast growth factors and heparin-binding hemagglutinin*. Curr Opin Struct Biol, 2018. **50**: p. 126-133.
208. Schiattarella, G.G., et al., *The Murine Model of Mucopolysaccharidosis IIIB Develops Cardiopathies over Time Leading to Heart Failure*. PLoS One, 2015. **10**(7): p. e0131662.
209. Krenciute, G., et al., *Nuclear BAG6-UBL4A-GET4 complex mediates DNA damage signaling and cell death*. J Biol Chem, 2013. **288**(28): p. 20547-57.

210. Harris, P.C. and V.E. Torres, *Polycystic Kidney Disease, Autosomal Dominant*, in *GeneReviews((R))*, M.P. Adam, et al., Editors. 1993, University of Washington, Seattle
University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.: Seattle (WA).
211. Allison, S.J., *DNAJB11: another player in ADPKD*. Nat Rev Nephrol, 2018. **14**(8): p. 476.
212. Johnson, C.A. and S.J. Collis, *Ciliogenesis and the DNA damage response: a stressful relationship*. Cilia, 2016. **5**: p. 19.
213. Chauvet, V., et al., *Expression of PKD1 and PKD2 transcripts and proteins in human embryo and during normal kidney development*. Am J Pathol, 2002. **160**(3): p. 973-83.
214. Gnyszka, A., Z. Jastrzebski, and S. Flis, *DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer*. Anticancer Res, 2013. **33**(8): p. 2989-96.
215. Chowdhari, S. and N. Saini, *hsa-miR-4516 mediated downregulation of STAT3/CDK6/UBE2N plays a role in PUVA induced apoptosis in keratinocytes*. J Cell Physiol, 2014. **229**(11): p. 1630-8.
216. Chowdhari, S., K. Sardana, and N. Saini, *miR-4516, a microRNA downregulated in psoriasis inhibits keratinocyte motility by targeting fibronectin/integrin alpha9 signaling*. Biochim Biophys Acta Mol Basis Dis, 2017. **1863**(12): p. 3142-3152.
217. Kovacs, J., et al., *Differential growth factor-induced modulation of proteoglycans synthesized by normal human renal versus cyst-derived cells*. J Am Soc Nephrol, 1994. **5**(1): p. 47-54.
218. Song, C.J., et al., *Inflammation and Fibrosis in Polycystic Kidney Disease*. Results Probl Cell Differ, 2017. **60**: p. 323-344.
219. Lee, E.J., et al., *Blockade of interleukin-8 receptor signalling inhibits cyst development in vitro, via suppression of cell proliferation in autosomal polycystic kidney disease*. Nephrology (Carlton), 2014. **19**(8): p. 471-8.
220. Tan, M., et al., *Novel inhibitors of nuclear transport cause cell cycle arrest and decrease cyst growth in ADPKD associated with decreased CDK4 levels*. Am J Physiol Renal Physiol, 2014. **307**(11): p. F1179-86.
221. Ta, M.H., D. Liuwantara, and G.K. Rangan, *Effects of pyrrolidine dithiocarbamate on proliferation and nuclear factor-kappaB activity in autosomal dominant polycystic kidney disease cells*. BMC Nephrol, 2015. **16**: p. 212.
222. de Stephanis, L., et al., *Double inhibition of cAMP and mTOR signalling may potentiate the reduction of cell growth in ADPKD cells*. Clin Exp Nephrol, 2017. **21**(2): p. 203-211.
223. Norman, J., *Fibrosis and progression of autosomal dominant polycystic kidney disease (ADPKD)*. Biochim Biophys Acta, 2011. **1812**(10): p. 1327-36.
224. Patil, A., et al., *Unique interstitial miRNA signature drives fibrosis in a murine model of autosomal dominant polycystic kidney disease*. World J Nephrol, 2018. **7**(5): p. 108-116.
225. Howat, W.J. and B.A. Wilson, *Tissue fixation and the effect of molecular fixatives on downstream staining procedures*. Methods, 2014. **70**(1): p. 12-9.
226. Dietrich, D., et al., *Improved PCR performance using template DNA from formalin-fixed and paraffin-embedded tissues by overcoming PCR inhibition*. PLoS One, 2013. **8**(10): p. e77771.
227. Cronin, M., et al., *Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay*. Am J Pathol, 2004. **164**(1): p. 35-42.

228. Do, H., et al., *Reducing sequence artifacts in amplicon-based massively parallel sequencing of formalin-fixed paraffin-embedded DNA by enzymatic depletion of uracil-containing templates*. Clin Chem, 2013. **59**(9): p. 1376-83.
229. Do, H. and A. Dobrovic, *Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization*. Clin Chem, 2015. **61**(1): p. 64-71.
230. Ladd-Acosta, C. and M.D. Fallin, *The role of epigenetics in genetic and environmental epidemiology*. Epigenomics, 2016. **8**(2): p. 271-83.
231. Lanktree, M.B. and A.B. Chapman, *New treatment paradigms for ADPKD: moving towards precision medicine*. Nat Rev Nephrol, 2017. **13**(12): p. 750-768.
232. Lantinga, M.A., et al., *Management of renal cyst infection in patients with autosomal dominant polycystic kidney disease: a systematic review*. Nephrol Dial Transplant, 2017. **32**(1): p. 144-150.
233. Phiel, C.J., et al., *Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen*. J Biol Chem, 2001. **276**(39): p. 36734-41.
234. Csoka, A.B. and M. Szyf, *Epigenetic side-effects of common pharmaceuticals: a potential new field in medicine and pharmacology*. Med Hypotheses, 2009. **73**(5): p. 770-80.
235. Clark, L.A., et al., *Cost-effectiveness of angiotensin-converting enzyme inhibitors versus angiotensin II receptor blockers as first-line treatment in autosomal dominant polycystic kidney disease*. J Med Econ, 2017. **20**(7): p. 715-722.
236. Alderman, M.H., *New onset diabetes during antihypertensive therapy*. Am J Hypertens, 2008. **21**(5): p. 493-9.
237. Liu, L., Y. Li, and T.O. Tollefsbol, *Gene-environment interactions and epigenetic basis of human diseases*. Curr Issues Mol Biol, 2008. **10**(1-2): p. 25-36.
238. Hanai, J., et al., *The muscle-specific ubiquitin ligase atrogin-1/MAFbx mediates statin-induced muscle toxicity*. J Clin Invest, 2007. **117**(12): p. 3940-51.
239. Chu, A.Y., et al., *Epigenome-wide association studies identify DNA methylation associated with kidney function*. Nat Commun, 2017. **8**(1): p. 1286.
240. Lazzeri, E., et al., *Human Urine-Derived Renal Progenitors for Personalized Modeling of Genetic Kidney Disorders*. J Am Soc Nephrol, 2015. **26**(8): p. 1961-74.
241. Ben-Dov, I.Z., et al., *Urine microRNA as potential biomarkers of autosomal dominant polycystic kidney disease progression: description of miRNA profiles at baseline*. PLoS One, 2014. **9**(1): p. e86856.
242. Sakairi, T., et al., *Conditionally immortalized human podocyte cell lines established from urine*. Am J Physiol Renal Physiol, 2010. **298**(3): p. F557-67.
243. Ho, M.C., et al., *Generation of an induced pluripotent stem cell line, IBMS-iPSC-014-05, from a female autosomal dominant polycystic kidney disease patient carrying a common mutation of R803X in PKD2*. Stem Cell Res, 2017. **25**: p. 38-41.
244. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation*. Nucleic Acids Res, 2016. **44**(D1): p. D733-45.