Short and long-read whole genome sequencing explains most undiagnosed Autosomal Dominant Polycystic Kidney Disease

Amali Mallawaarachchi (✉ a.mallawaarachchi@garvan.org.au)  
Garvan Institute of Medical Research  
https://orcid.org/0000-0002-1229-1701

Yvonne Hort  
Garvan Institute of Medical Research

Patricia Sullivan  
Children's Cancer Institute

Laura Wedd  
Garvan Institute of Medical Research

Lindsay Fowles  
Royal Brisbane and Women's Hospital  
https://orcid.org/0000-0001-6626-3004

Igor Stevanonvski  
Garvan Institute of Medical Research

Ira Deveson  
Garvan Institute of Medical Research

Cas Simons  
Garvan Institute of Medical Research

Andrew Mallett  
Townsville University Hospital  
https://orcid.org/0000-0002-8752-2551

Chirag Patel  
Genetic Health Queensland

Timothy Furlong  
Garvan Institute of Medical Research

Mark Cowley  
Childrens' Cancer Institute Sidney  
https://orcid.org/0000-0002-9519-5714

John Shine  
Garvan Institute of Medical Research

Article

Keywords: PKD1, PKD2, Polycystic Kidney Disease, ADPKD, whole genome sequencing, long-read sequencing, splicing
Short and long-read whole genome sequencing explains most undiagnosed Autosomal Dominant Polycystic Kidney Disease

Yvonne Hort¹, Patricia Sullivan²,³, Laura Wedd¹,⁴, Lindsay Fowles⁵, Igor Stevanovski⁶,⁷, Ira Deveson⁶,⁷, Cas Simons⁴,⁸, Andrew Mallett⁹,¹⁰,¹¹, Chirag Patel⁵, Timothy Furlong¹, Mark Cowley²,³, John Shine¹ & Amali Mallawaarachchi¹,¹².

¹Molecular Genetics of Inherited Kidney Disorders Laboratory, Garvan Institute of Medical Research, Sydney, Australia
²Children’s Cancer Institute, Lowy Cancer Centre, UNSW Sydney, Kensington, New South Wales, Australia
³School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney, Kensington, New South Wales, Australia
⁴Centre for Population Genomics, Garvan Institute of Medical Research and UNSW Sydney, Sydney, NSW, Australia
⁵Genetic Health Queensland, Royal Brisbane and Women’s Hospital, Herston QLD, Australia
⁶Genomic Technologies, Garvan Institute of Medical Research, Sydney, Australia
⁷Centre for Population Genomics, Garvan Institute of Medical Research and Murdoch Children’s Research Institute, Australia
⁸Centre for Population Genomics, Murdoch Children’s Research Institute, Melbourne, VIC, Australia
⁹Department of Renal Medicine, Townsville University Hospital, Townsville, Queensland, Australia
¹⁰Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia
¹¹College of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia
¹²Clinical Genetics Service, Institute of Precision Medicine and Bioinformatics, Royal Prince Alfred Hospital, Sydney, Australia
Corresponding Author:
Dr Amali Mallawaarachchi
Garvan Institute of Medical Research
384 Victoria Street, Darlington, NSW 2010, AUSTRALIA
Email: a.mallawaarachchi@garvan.org.au
Abstract

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is the most common monogenic cause of kidney failure and primarily associated with PKD1 or PKD2. Approximately 10% of patients remain undiagnosed after standard genetic testing. We aimed to utilise short and long read genome sequencing and RNA studies to investigate undiagnosed families. Patients with typical ADPKD phenotype and undiagnosed after genetic diagnostics were recruited. Probands underwent short-read genome sequencing, PKD1 and PKD2 coding and non-coding analyses and then genome-wide analysis. Targeted RNA studies investigated variants suspected to impact splicing. Those undiagnosed then underwent Oxford Nanopore Technologies long-read genome sequencing. From over 172 probands, 9 met inclusion criteria and consented. A genetic diagnosis was made in 8 of 9 (89%) families undiagnosed on prior genetic testing. Six had variants impacting splicing, five in non-coding regions of PKD1. Short-read genome sequencing identified novel branchpoint, AG-exclusion zone and missense variants generating cryptic splice sites and a deletion causing critical intron shortening. Long-read sequencing confirmed the diagnosis in one family. Most undiagnosed families with typical ADPKD have splice-impacting variants in PKD1. We describe a pragmatic method for diagnostic laboratories to assess PKD1 and PKD2 non-coding regions and validate suspected splicing variants through targeted RNA studies.

Keywords: PKD1, PKD2, Polycystic Kidney Disease, ADPKD, whole genome sequencing, long-read sequencing, splicing
Introduction

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is the most common monogenic cause of kidney failure, affecting approximately 1 in 1000 people\(^1\). The condition is primarily caused by disease-causing variants in \textit{PKD1} and \textit{PKD2}. Genetic diagnosis of ADPKD is technically challenging due to six pseudogenes that are $>97\%$ homologous in sequence to the genuine \textit{PKD1} gene\(^1\). This sequence homology has driven development of specific genetic diagnostic techniques to robustly sequence \textit{PKD1} that have focussed mainly on the analysis of the protein-coding regions of \textit{PKD1}, \textit{PKD2} and then the wider exome. These techniques include long-range PCR (LR-PCR) and Sanger sequencing, targeted next-generation sequencing using probes specific to coding regions of cystic-related genes (tNGS), exome sequencing and genome sequencing with coding-based analysis\(^1-6\). The diagnostic yield from these studies differs based on the clinical breadth of the cohort, ranging from approximately 60\% in phenotypically broad cohorts to $>90\%$ in cohorts tightly selected for features typical of \textit{PKD1} and \textit{PKD2}-mediated disease\(^1,3-5,7\). Even with the most stringent coding-based analysis, at least 7\% of ADPKD families are left without a genetic diagnosis\(^7\).

In recent years, there have been substantial advances in understanding the breadth of ADPKD gained through investigating genetically undiagnosed patients. New genes have been identified that contribute to the ADPKD spectrum, including \textit{GANAB}, \textit{DNAJB11}, \textit{IFT140}, \textit{ALG5}, \textit{ALG8} and \textit{ALG9}\(^5,7-11\). These gene-disease associations have been made through studies performing exome sequencing in patients with undiagnosed PKD and focussing on genes within pathways that impact polycystin-1 (PC1) folding and trafficking in the endoplasmic reticulum, ciliary localisation of PC1 and glycosylation pathway genes that are
thought to impact PC1 maturation\textsuperscript{5,7}. These ADPKD-spectrum genes are all associated with an atypical ADPKD phenotype, which includes a) milder disease, b) asymmetric kidney disease, c) a fibrotic phenotype that overlaps Autosomal Dominant Tubulointerstitial Kidney Disease (ADTKD) or d) a hepatic phenotype that overlaps Autosomal Dominant Polycystic Liver Disease (ADPLD). Despite these advances in disease knowledge and extensive coding-region focussed analysis, there remains a cohort of patients with a typical ADPKD phenotype who are without a genetic diagnosis. It is thus an open question as to whether this is due to technical limitations in identifying causative variants within \textit{PKD1} or \textit{PKD2}, or the existence of an unknown ‘PKD3’ gene that is associated with a typical ADPKD phenotype.

We aimed to address this question by investigating a cohort selected to have a typical ADPKD phenotype, a positive family history and be undiagnosed on standard diagnostic genetic testing. We aimed to investigate whether these patients had variants in previously unidentified genes or, as other diseases suggest, novel variants in the most likely genes of interest – \textit{PKD1} and \textit{PKD2}. To approach this challenge, we applied sequencing methods not previously extensively used in ADPKD, including short and long-read genome sequencing, combined with targeted RNA sequencing.

We have previously shown that short-read genome sequencing is a robust diagnostic method in ADPKD, allowing the detection of single nucleotide, short indel and structural variants\textsuperscript{1,4}. However, though the whole genome is sequenced, current diagnostic laboratory protocols essentially limit analysis to protein-coding regions of the genome. This particularly biases against detection of non-coding variants that may impact splicing. Another challenge for diagnostic laboratories is in clarifying the pathogenicity of identified variants of
uncertain significance (VUS). Even if non-coding variants are identified, pathogenicity confirmation of these variants typically requires functional analysis that is not routinely performed in diagnostic laboratories\textsuperscript{12}. Pathogenicity can also be clarified in some instances by phasing of variants (confirming which allele the variant is present on), which is not usually possible with short-read sequencing. More recently available long-read technologies, such as Oxford Nanopore Technologies (ONT), have been shown to inform phasing in other disease groups but this has not been previously applied in ADPKD\textsuperscript{13,14}. Our previous studies in genome sequencing diagnostics in ADPKD have informed this study and patients left undiagnosed from these previous cohorts were assessed for suitability for this study\textsuperscript{1,4}. In this study we report for the first time the combination of short and long-read genome sequencing with whole genome analysis and RNA studies to investigate ADPKD families without a diagnosis after standard diagnostic genetic testing.

**Materials and Methods:**

We enrolled patients with typical ADPKD clinical features and a family history of ADPKD, who were without a genetic diagnosis after diagnostic sequencing of \textit{PKD1} and \textit{PKD2} or a larger cystic gene panel that included the \textit{PKD1} and \textit{PKD2} genes. Patients undiagnosed from our previous studies were assessed for suitability for this study (Figure 1). In addition, patients were recruited from clinical sites across Australia. Family members were recruited as required and available. All participants provided written informed consent. Ethics approval for the study was obtained from the RPAH Human Research Ethics Committee (HREC/18/RPAH/726).
Clinical, family and imaging data was obtained during clinical review or review of medical records. Kidney lengths were based on ultrasound measurements as kidney ultrasound is the Medicare-funded imaging modality available for assessment of ADPKD patients in Australia. Kidney function was calculated using the CKD-EPI equation.

All probands underwent short-read genome sequencing using DNA extracted from peripheral blood samples. Genome sequencing was performed on the HiSeqX sequencing system (Illumina Inc., California, CA, USA) after either PCR-based library preparation (Illumina HiSeq X TruSeq Nano DNA HT Sample Prep Kit) or PCR-free library preparation (KAPA Hyper PCR-free kit, Roche). All samples were processed via a custom bioinformatics pipeline based on GATK best practice, which was optimised for identification of germline variants\(^1,4\). Reads were aligned to the hg37 reference sequence. Sequence variants were filtered using Seave\(^15\). CNV and structural variant analysis was performed using ClinSV\(^16\). Introme was used to assess for variants predicted to impact splicing (Sullivan et al., Manuscript Under Review, https://github.com/CCICB/introme). Control PKD1 splice junction usage was obtained using GTEx V8, filtered to include only kidney samples\(^17\). Initial variant analysis was targeted to coding and intronic and promoter regions of PKD1 (NM_001009944.3) and PKD2 (NM_000297.4), with all variants (ranging from predicted high to low impact) manually reviewed. Analysis was then expanded to phenotype-driven whole genome analysis. Variants were classified according to American College of Medical Genetics (ACMG) Guidelines\(^18\). Sanger sequencing (with prior LR-PCR amplification if within the PKD1-pseudogene homologous region) was performed to confirm all single nucleotide and short indel variants identified on genome sequencing and for family studies.
RNA functional studies were performed to assess variants predicted to impact splicing in *PKD1*. Total RNA was extracted from venous blood (Macherey-Nagel Nucleospin RNA Blood Kit) for RT-PCR studies. If the variant of interest was within the PKD1-pseudogene homologous region, amplification was performed with at least one of the primer pairs being unique to the *PKD1* sequence in order to avoid amplifying *PKD1*-pseudogene transcripts (see Supplementary Table 1 for primer sequences). Sanger sequencing was performed on this PCR-product. See Supplementary Methods for further details.

In families who remained negative after short-read genome sequencing or for whom phasing could inform variant interpretation and classification, long-read sequencing was performed. High molecular weight DNA was sheared to ~20kb fragment size using Covaris G-tubes. Sequencing libraries were prepared from ~1.5-5ug of sheared DNA using native library prep kits (SQK-LSK110) and sequenced for 72 hours on a PromethION (FLO-PRO002, R9.4.1) flow cell. Raw ONT sequencing data was converted to BLOW5 format with slow5tools (v0.3.0)\textsuperscript{19} then base-called using Guppy (4.0.11 or later). Resulting FASTQ files were aligned to the hg38 reference genome using minimap2 (v2.14-r883)\textsuperscript{20} and Longshot (v0.4.1)\textsuperscript{21} was used to identify and phase variants within the *PKD1* locus.

Variants identified in the study have been submitted to ClinVar (ClinVar Accessions: SCV002756451 - SCV002756459).

**Results:**

Over 172 patients were assessed for suitability for recruitment. This included 28 patients from a cohort of typical ADPKD who had undergone genome sequencing, 144 patients from
a cohort of suspected ADPKD who had undergone diagnostic genome sequencing and a
cohort of patients with typical and atypical PKD reviewed at multidisciplinary kidney
genetics clinics from across Australia who had undergone diagnostic genetic testing\textsuperscript{1,4,22}.

From this initial pool of over 172 probands, 9 families were recruited who met the study
inclusion criteria (Figure 1). Recruitment was restricted to those with a family history of
ADPKD in order to target analysis towards inherited germline, rather than mosaic, variants.
All patients had previously undergone standard diagnostic genetic testing via LR-PCR of
\textit{PKD1} and \textit{PKD2} coding regions and massively parallel sequencing of this PCR-product (2
probands) or diagnostic genome sequencing with analysis targeted to coding regions of a
cystic kidney disease gene panel (6 probands) or both (1 participant) (Supplementary Table
2). An additional seven patients met the inclusion criteria and had a VUS identified in \textit{PKD1}
on initial diagnostic genetic testing, but were not consented for further research analysis
and therefore did not proceed to this study (Figure 1).

Four of nine probands had ESKD and all had enlarged kidney lengths with numerous kidney
cysts on imaging (Table 1 & Supplementary Table 2). Five of nine probands had extra-renal
features of ADPKD reported.

After genome sequencing and whole genome analysis, a genetic diagnosis (identification of
a Pathogenic or Likely Pathogenic variant) was made in eight out of nine families, with all
having disease-causing variants in \textit{PKD1} (Table 1 and Figure 2). An additional family
(FRPA007) had a VUS identified in \textit{PKD1}. Six of the disease-causing variants were shown
through RNA-studies to impact splicing. Four of these splicing variants had been identified
on the initial diagnostic testing (including one coding variant) but classified as of uncertain
significance, with segregation studies not able to clarify pathogenicity (Supplementary Table 2).

**FRBW403, FRPA028 and FRPA014**

Patient RBW403 had a clinical diagnosis of ADPKD made at 12 years of age in the context of a known diagnosis in his father, who reached ESKD at 44yo (Table 1 and Figure 3A). He had previously undergone diagnostic genetic testing via next-generation sequencing of LR-PCR amplicons targeted to *PKD1* and *PKD2* and then genome sequencing (analysis targeted to coding regions of a virtual cystic gene panel). No pathogenic variants were reported with either test. Through this study, a novel variant was identified in intron 37 of *PKD1* (c.11017-25A>G) that was predicted by *in silico* splice prediction tool, introme, to interrupt the splicing branchpoint (Figure 3C & 3D). This variant was absent in control databases and not previously reported in ADPKD cohorts. This variant had been detected on both previously performed diagnostic tests but predicted benign based on available *in silico* tools. Introme predicted multiple potential splicing impacts. The predominant interpretation was that this branchpoint variant would largely result in skipping of exon 38, introducing a premature stop codon. An alternate interpretation was that the presence of a wildtype cryptic splice site 156 base pairs upstream of the c.11017-25A>G branchpoint variant would result in retention of 180bp of intron 37, at a reduced frequency (Figure 3D). RNA studies revealed evidence for both splicing outcomes, with the skipping of exon 38 being far more prevalent than intron retention. Review of control GTEx RNA data suggests low-level (5%) natural alternative splicing of exon 38 in kidney samples (Figure 3B & Supplementary Table 3).
Two additional patients in this cohort (RPA028 & RPA014, Table 1 and Figure 3A) were identified to have different variants in intron 37 that were also predicted to interrupt usual function of the exon 38 acceptor splice site (Figure 3C & 3D). RPA028 had a different nucleotide substitution (*PKD1* c.1107-25A>C) interrupting the same branchpoint as in RBW403. RPA014 had a variant 10 base pairs from the start of exon 38 (*PKD1* c.11017-10C>A) that impacted the acceptor splice site through the inclusion of the ‘AG’ dinucleotide in the AG Exclusion Zone\(^23\). RNA studies in both patients demonstrated similar impact to that seen in RBW403, with a combination of skipping of exon 38 and partial retention of intron 37 (Figure 3C).

**FRPA019**

Patient RPA019 had a clinical diagnosis of ADPKD made at 29yo during screening as a potential kidney donor for her affected brother (Table 1 and Supplementary Figure 1A). She was motivated for a genetic diagnosis to inform her family planning. Previous diagnostic testing had identified missense variant classified as a VUS in exon 10 of *PKD1* (c.1991C>T). RPA019 underwent genome sequencing and whole genome analysis and no additional phenotype-relevant variants were identified. Introme predicted that the c.1991C>T variant would generate a new cryptic donor site and result in an in-frame deletion of 36 amino acids (Supplementary Figure 1C). RNA studies supported this prediction (Supplementary Figure 1B). The variant was segregated to the proband’s affected mother and was absent in population datasets, though alternate amino acid substitutions at the same residue and substitutions at the same nucleotide are reported in population datasets. To our knowledge, this variant has not previously been reported in ADPKD cohorts.
**F19F00138**

19F00138 and her sister (RBW401) both had a clinical diagnosis of ADPKD, with bilateral kidney enlargement, multiple kidney cysts and multi-generational family history of ADPKD (Supplementary Figure 1D). Both were motivated for genetic diagnosis to allow access to *in vitro* fertilization (IVF) and Pre-implantation Genetic Diagnosis (PGD). Clinical short-read genome sequencing in 19F00138 had been non-diagnostic. Re-analysis of the genome sequencing data identified a 19bp deletion within intron 31 of *PKD1* that was predicted to result in shortening of the intron beneath its critical length and, therefore, intron retention\(^{24}\) (Supplementary Figure 1E). This variant was segregated to RBW401 and RNA studies in her demonstrated retention of intron 31, creating a frameshifting insertion (Supplementary Figure 1F). This variant has been reported previously in a patient with a *de novo* ADPKD phenotype\(^{25}\).

**FRPA021**

RPA021 and her brother both had a clinical diagnosis of ADPKD, with both undergoing kidney transplantation in their 50s (Supplementary Figure 2A&B). In our previous study\(^1\), a variant of uncertain significance had been identified in intron 18 of *PKD1* (c.7489+5G>A). Though suspicious for being disease-causing via disruption of the native splice site, there was insufficient evidence to confirm pathogenicity without support from functional studies. This variant has also been reported previously by our research group in an unrelated patient who was part of a cohort of patients who underwent clinical PKD testing via short-read genome sequencing\(^4\). The patient in this previous study (Pt D158) was not known to be related to FRPA021 and did not share ethnicity\(^4\). RNA studies in RPA021 demonstrated that
the c.7489+5G>A variant in \textit{PKD1} resulted in retention of 93 base pairs of intron 18 introducing a premature stop codon (Supplementary Figure 2C & 2D).

\textit{RG\_0044}

The RG\_0044 family had a multi-generational history of ADPKD (Supplementary Figure 3). Participant RG\_0044.0048 had previously undergone diagnostic genetic testing via next-generation sequencing of LR-PCR amplicons targeted to \textit{PKD1} and \textit{PKD2} and no clinically significant variants had been identified. Genome sequencing and analysis identified a previously reported, likely pathogenic missense variant in \textit{PKD1} p.(Gly960Ser) that had not been identified on the previous diagnostic testing\textsuperscript{26}. This variant appropriately segregated in six affected and unaffected family members (Supplementary Figure 3).

\textit{FRPA017}

RPA017 had a clinical diagnosis of ADPKD made at 34yo, and at 38yo had significantly enlarged kidneys and Stage 3b CKD (Figure 4 and Supplementary Table 2). She was motivated for a genetic diagnosis to inform IVF and PGD. Her mother (RPA015) was diagnosed with ADPKD in her 50’s in the context of a diagnosis in her mother and had CKD Stage 3a at diagnosis. RPA017’s father (RPA016) had normal kidney ultrasound at 72yo. Genome sequencing in RPA017 identified a nonsense variant in exon 31 of \textit{PKD1}, however segregation by LR-PCR and Sanger sequencing demonstrated that this variant was absent in her affected mother and unaffected father (paternity confirmed). Subsequent genome sequencing in RPA015 identified a nonsense variant in exon 5 of \textit{PKD2} that was absent in her affected daughter. To clarify the inheritance in this family, ONT long-read sequencing was performed to facilitate variant phasing, which demonstrated that the \textit{PKD1} variant
identified in RPA017 was present on the allele she inherited from her father. Sanger sequencing showed that the variant was absent in her father’s peripheral blood DNA, strongly suggesting this was a de novo variant in RPA017(Figure 4). Phasing also confirmed that RPA017 had not inherited the affected PKD2 allele from her mother. This information was used to inform linkage studies for PGD.

RPA007

RPA007 has a clinical diagnosis of ADPKD. She had significant hepatic cystic disease, causing local mass effect, and underwent dual kidney and liver transplantation. Her mother also had a clinical diagnosis of ADPKD and underwent kidney transplantation in her sixth decade. RPA007 underwent diagnostic short-read genome sequencing with no clear diagnosis made. A VUS, p.(Gln2824Arg), was identified in PKD1 that segregated to her affected mother. Whole genome analysis in RPA007 with both short and long read genome sequencing did not identify any additional variants of interest. Copy number variant analysis was uninformative. No splicing impact was predicted by introme. The PKD1 p.(Gln2824Arg) variant is absent in population databases, predicted pathogenic by in silico tools and has not been previously reported in PKD cohorts. This information alone is insufficient to clarify the pathogenicity of this variant.

Discussion

Pathogenic variants in PKD1 and PKD2 have been shown to be responsible for disease in most patients with a typical phenotype of ADPKD, however genetic sequencing in ADPKD cohorts consistently results in approximately ten percent of patients being left without a genetic diagnosis4,5,7. Next-generation sequencing has facilitated broad analysis of the
undiagnosed patients in these cohorts and patients with atypical phenotypes have had variants identified in new polycystic kidney disease-related genes, such as *DNAJB11*, *GANAB* and *IFT140*. However, patients with a typical ADPKD phenotype often remain undiagnosed after expanded exome-wide sequencing. This study shows that most of these undiagnosed families have variants in *PKD1* that were uncertain or undetected on standard diagnostic genetic testing. We demonstrate the value of sequencing and, importantly, analysis of the protein-coding and non-coding regions of *PKD1* and *PKD2*, combined with targeted RNA studies to confirm a genetic diagnosis. For the first time, we show the value of long-read sequencing in ADPKD to inform phasing and inheritance of variants.

A confirmed genetic diagnosis is increasingly becoming standard of care for families with genetic disorders. In nephrology, a genetic diagnosis is being utilised across all inherited kidney disease, including ADPKD. A genetic diagnosis allows early definitive diagnosis in ADPKD which can provide prognostic information and allow for early institution of treatments, including vigorous hypertension management and tolvaptan for those predicted to have more rapidly progressive disease. Genetic diagnosis also allows for informing the selection of kidney donors and for family planning. In this cohort alone, several families utilised a confirmed genetic diagnosis to inform PGD of embryos. In many jurisdictions, such as Australia, New Zealand and the United Kingdom, national health service subsidies are available for ADPKD families for IVF and PGD, highlighting that providing genetic counselling to ADPKD families is an essential aspect of their care. ADPKD is one of the most common monogenic conditions screened for in IVF and PGD. There is also increasing evidence of nocturnal hypertension in children with ADPKD, suggesting that guidelines regarding diagnosis of ADPKD in childhood may be modified in the future, to recommend early
intervention for these children\textsuperscript{32}. Imaging results can be variable in paediatric populations, whereas genetic diagnostics allows reliable, definitive diagnosis\textsuperscript{32}.

Utilising a genetic diagnosis to inform clinical care requires a definitive genetic result. Cascade testing can only be offered in families with a definitive genetic diagnosis and this is the same for using genetic results to inform family planning. This highlights the value of improving diagnostic yield for families with ADPKD and the value of the results of this study, which demonstrates that a significant proportion of undiagnosed families have variants that affect gene splicing. There is increasing evidence for similar variants across other disease groups, where, for example, RNA-sequencing in a cohort of patients with undiagnosed muscle disease identified a diagnosis through aberrant splicing in 35\% of patients\textsuperscript{33}. RNA-sequencing often requires analysis of a tissue of interest, rather than using blood RNA\textsuperscript{12,33,34}.

In kidney disease, RNA extracted from kidney tissue or urothelial cells is technically more challenging to access\textsuperscript{12}. We demonstrate a practical approach for evaluating suspected splicing variants in ADPKD using RT-PCR of total RNA extracted from peripheral blood. This pragmatic, targeted approach provides functional evidence to classify VUS that result in substantial splicing defects without requiring access to kidney tissue and is achievable for a diagnostic laboratory to replicate\textsuperscript{35}. Developing protocols for diagnostic genetic laboratories to identify and then confirm coding or non-coding aberrant splicing variants is key in improving current genomic diagnostic rates\textsuperscript{12}. The high diagnostic yield in our study highlights the importance of diagnostic laboratories analysing beyond the coding region in patients with a typical ADPKD phenotype by using robust \textit{in silico} tools, such as introme. Importantly, we also show that RT-PCR can then be used to evaluate potential splicing variants identified through this broader analysis. The homologous \textit{PKD1}-pseudogenes
produce mRNA transcripts that are approximately 97% homologous to the 5’ region of \textit{PKD1} mRNA; therefore this RT-PCR method requires use of unique primers to avoid inadvertently amplifying transcripts from the PKD1-pseudogenes\textsuperscript{36}.

Classifying VUS is currently a challenge for diagnostic laboratories, particularly in ADPKD, where variants are often private to families and multiple samples from large pedigrees are typically not available to perform segregation studies to clarify pathogenicity. We also demonstrate the additional value of long-read sequencing in understanding pathogenicity of variants in ADPKD. Long-read technologies have the additional advantage of allowing phasing of variants, which has obvious applications in confirming bi-allelic inheritance in autosomal recessive disease. In autosomal dominant disorders, long-read sequencing allows the opportunity to identify the parental allele on which a \textit{de novo} variant occurs. This is valuable in understanding inheritance in complex families, such as we demonstrate in family FRPA017. Another unique application is in informing linkage studies for couples undergoing PGD for \textit{de novo} ADPKD. For couples undergoing PGD, detailed phasing studies are performed to identify accurate markers that are then used to ascertain affected vs unaffected embryos. For patients with \textit{PKD1}-mediated disease, this requires samples from multiple generations of affected family, as direct sequencing of \textit{PKD1} is hampered by the presence of the homologous pseudogenes. This makes PGD challenging to access for patients with \textit{de novo} \textit{PKD1}-disease, who are the only affected person in their family. In this situation, which impacts 10% of patients with ADPKD, long-read sequencing can provide important phasing information that can allow these patients to access PGD of embryos\textsuperscript{37}. Long-read data may additionally be used to detect structural variants that are missed by short-read sequencing, although we did not detect any relevant events here\textsuperscript{13}. 
This cohort was collected from a larger pool of over 172 patients from across Australia who had undergone standard diagnostic genetic testing for PKD. Given the high yield of diagnostic testing in ADPKD and the strict inclusion criteria of this study, only 16 patients from this larger pool met the inclusion criteria, of which 9 consented for participation. Inclusion was deliberately limited to patients with a family history of ADPKD to focus on germline rather than mosaic variants, restricting the eligible pool of patients\textsuperscript{38}. Our results are comparable to other disease groups, and highlight that previously unrecognised or undetected splice variants may be causative in these families\textsuperscript{39}. Our smaller cohort size means there is value in applying this method to a larger cohort.

Our results show that for families with a typical ADPKD phenotype, variants are most likely to be found in \textit{PKD1} and \textit{PKD2}, rather than other PKD-associated genes. We provide evidence of the value of diagnostic laboratories expanding analysis to non-protein-coding regions to improve diagnostic yield in ADPKD. We also describe an achievable method for assessing uncertain variants that are predicted to impact splicing in this common disorder. ADPKD is the most common inherited kidney disorder and contributes to approximately ten percent of kidney failure cohorts\textsuperscript{4}. Improving diagnostic rates allows for improved management through earlier institution of treatment and access to holistic care that includes genetic counselling. Importantly, improving understanding of the underlying genetic basis for all families with ADPKD is a critical step in developing personalised therapies for this common genetic disease.
Acknowledgments

This study has been funded through support from PKD Australia, PKD Foundation, RACP Jacquot Foundation, the NSW Office of Health and Medical Research funded Luminesce Alliance and The Sylvia and Charles Viertel Charitable Foundation.

Ethics Declaration

Ethics approval for the study was obtained from the RPAH Human Research Ethics Committee (HREC/18/RPAH/726). All participants provided written informed consent as required by the Human Research Ethics Committee. The authors have received and archived written patient consent. All data included is de-identified.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability

Variants identified in this study have been submitted to ClinVar (ClinVar Accessions: SCV002756451 - SCV002756459). Other data is available upon request.

Figures:

Figure 1: Study Design

Figure 2: Overview of variants types identified in \textit{PKD1}

Figure 3: Disease-Causing Splicing Variants in \textit{PKD1} intron 37

Figure 4: Long Read Nanopore sequencing confirms de novo \textit{PKD1} variant occurred on paternal allele
References


<table>
<thead>
<tr>
<th>Family ID</th>
<th>Patient ID</th>
<th>Kidney function / Age (yrs)*</th>
<th>Gene</th>
<th>c.^</th>
<th>p.#</th>
<th>Variant type</th>
<th>Variant Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRBW403</td>
<td>RBW403</td>
<td>&gt;90/31</td>
<td>PKD1</td>
<td>c.11017-25A&gt;G</td>
<td>p.?</td>
<td>Splice variant - branchpoint variant</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>FRPA028</td>
<td>RPA028</td>
<td>&gt;90/23</td>
<td>PKD1</td>
<td>c.11017-25A&gt;C</td>
<td>p.?</td>
<td>Splice variant - branchpoint variant</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>FRPA014</td>
<td>RPA014</td>
<td>&gt;90/35</td>
<td>PKD1</td>
<td>c.11017-10C&gt;A</td>
<td>p.?</td>
<td>Splice variant – AG-Exclusion Zone variant</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>FRPA019</td>
<td>RPA019</td>
<td>&gt;90/35</td>
<td>PKD1</td>
<td>c.1991C&gt;T</td>
<td>p.?</td>
<td>Splice variant - generation of new cryptic donor site</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td></td>
<td>RPA020</td>
<td>44/69</td>
<td>PKD1</td>
<td>c.1991C&gt;T</td>
<td>p.Al664_Ser699del</td>
<td>Splice variant - generation of new cryptic donor site</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>F19F00138</td>
<td>RBW401</td>
<td>&gt;90/30</td>
<td>PKD1</td>
<td>c.10167+25_10167+43del</td>
<td>p.?</td>
<td>Splice variant - intronic deletion causing critical intron shortening</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td></td>
<td>19F00138</td>
<td>&gt;90/25</td>
<td>PKD1</td>
<td>c.10167+25_10167+43del</td>
<td>p.?</td>
<td>Splice variant - intronic deletion causing critical intron shortening</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>FRPA021</td>
<td>RPA021</td>
<td>ESKD/52</td>
<td>PKD1</td>
<td>c.7489+5G&gt;A</td>
<td>p.?</td>
<td>Splice variant – donor splice site variant</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>RG_0044</td>
<td>RG_0044</td>
<td>ESKD/58</td>
<td>PKD1</td>
<td>c.2878G&gt;A</td>
<td>p.(Gly960Ser)</td>
<td>Missense variant</td>
<td>Likely Pathogenic</td>
</tr>
<tr>
<td>FRPA017</td>
<td>RPA017</td>
<td>41/38</td>
<td>PKD1</td>
<td>c.10118C&gt;A</td>
<td>p.(Ser3373Ter)</td>
<td>Nonsense variant</td>
<td>Pathogenic</td>
</tr>
<tr>
<td></td>
<td>RPA015</td>
<td>50/71</td>
<td>PKD2</td>
<td>c.1249C&gt;T</td>
<td>p.(Arg417Ter)</td>
<td>Nonsense variant</td>
<td>Pathogenic</td>
</tr>
<tr>
<td>FRPA007</td>
<td>RPA007</td>
<td>ESKD/45</td>
<td>PKD1</td>
<td>c.8471A&gt;G</td>
<td>p.(Gln2824Arg)</td>
<td>Missense variant</td>
<td>VUS</td>
</tr>
</tbody>
</table>

*age kidney function recorded; ^PKD1 NM_001009944.3, PKD2 NM_000297.4; ^PKD1 NP_001009944.3 PKD2 NP_000288.1; ESKD = End Stage Kidney Disease; VUS = Variant of Uncertain significance.
Figure 1: Study Design

Undiagnosed Cohort (n=9)
Inclusion Criteria: Typical ADPKD phenotype AND positive family history

GS and analysis of coding and non-coding regions of PKD1 and PKD2, then genome-wide analysis

2 families remained undiagnosed
7 potential PKD1 splicing variants identified

Long read sequencing
RNA studies

PKD1 disease-causing variant in 8 out of 9 undiagnosed families with typical ADPKD and positive family history

n = 28
Genome Sequencing
Typical ADPKD phenotype 86% diagnosed

3 ineligible: no FmHx

n = 144
Clinical Genome Sequencing Cohort 47% diagnosed

51 ineligible: atypical phenotype

n = unknown
(population prevalence ~1:1000)
Clinical PKD testing across Australia
Typical and Atypical ADPKD

47% diagnosed

n = 9
undiagnosed

3 ineligible: no FmHx

15 ineligible: typical phenotype, no FmHx

7 with VUS not consented for further research

6 undiagnosed

3 ineligible: atypical phenotype

51 ineligible: typical phenotype

15 ineligible: atypical phenotype,
no FmHx

7 potential PKD1 splicing variants identified

Figure 1: Overview of patients assessed for study suitability and study method.
VUS = Variant of Uncertain Significance; GS = Genome Sequencing; FmHx = Family history
Figure 2: Overview of variants types identified in PKD1

Variant types:
- **c.1991C>T**: Generation of novel cryptic donor splice site
  - ccaggccctgc (exon 10)
- **c.7489+5G>A**: Donor splice site variant
  - AGGgtgagtg (exon 18)
- **c.10167+25_10167+43del**: Deletion causing critical shortening of intron length
  - exon 31 → exon 32
- **c.11017-10C>A**: Wildtype intron length 87bp
- **c.11017-25A>G**: Wildtype intron length 87bp
- **c.111017-10C>A**: AG-Exclusion Zone variant
- **c.11017-25A>C**: Branchpoint variants

Gene illustration developed using Protein Paint40. PKD1 NM_001009944.3
Figure 3: Disease-Causing Splicing Variants in PKD1 intron 37

A  Clinical Data:

B  Exon 37-39 Natural Alternative Splicing:

Predominant transcript: 37 38 39

Alternative transcript: 37 39

C  RNA Studies PKD1 Disease-causing variants:

RBW403

c.11017-25A>G:

Sanger sequencing of RNA product:

Control

RBW403

RPA028

c.11017-25A>C:

Sanger sequencing of RNA product:

Control

RBW403

RPA028

c.11017-10C>A:

Sanger sequencing of RNA product:

Control

RBW403

RPA014

D  Pathogenic Splicing Impact:

Branchpoint:

WT: ggctgacca
RBW403: ggctgcca
RPA028: ggctgcca

WT: cccccagtggtg
Cryptic Splice Site

RPA014: gccccccccggt
Polypyrimidine tract

Pathogenic Splice Alteration:

Exon Skipping:

AND

Intron Retention with use of Cryptic Splice Site:
Figure 4: Middle panel shows long read genome sequencing data from RPA017 over PKD1 exons 31-35, separated into maternal and paternal alleles. The top panel ‘zooms in’ over the region that includes the pathogenic *PKD1* exon 31 variant identified in RPA017. The variant is absent in RPA015 (affected mother). Bottom panel ‘zooms in’ over an intronic single nucleotide variant identified in RPA017 that short-read GS and Sanger sequencing shows is present in RPA016 (unaffected father) and absent in RPA015 (affected mother). In RPA017, this variant is seen on the same long-read sequencing as the pathogenic PKD1 variant, demonstrating that the de novo disease-causing variant in RPA017 has occurred on her paternally inherited PKD1 allele. GS = Genome sequencing.
Figure Legends

Figure 1: Study Design
Overview of patients assessed for study suitability and study method. VUS = Variant of Uncertain Significance; GS = Genome Sequencing; FmHx= Family history

Figure 2: Overview of variants types identified in \textit{PKD1}
Variants identified in the \textit{PKD1} gene in the study, including a range of different splicing variants. Gene illustration developed using Protein Paint\textsuperscript{40}. \textit{PKD1} NM_001009944.3

Figure 3: Disease-Causing Splicing Variants in \textit{PKD1} intron 37
\textbf{A} – Pedigrees and renal ultrasound images from RBW403 demonstrating bilateral kidney cysts; \textbf{B} – Natural splicing of exons 37, 38 and 39 of \textit{PKD1}, depicting that skipping of exon 38 naturally occurs at a low level. \textbf{C} – RT-PCR studies and Sanger sequencing of RT-PCR product in RBW403, RPA028 and RPA014. Additional bands are demonstrated in the affected individuals compared with controls, consistent in size with skipping of exon 38 and partial retention of exon 37. This is also reflected in Sanger sequencing of the RT-PCR product. Low-level skipping of exon 38 is evident in the controls. \textbf{D} – Illustration of the splicing impact of the different variants identified in exon 37 across the cohort. All three variants result in skipping of exon 38 and, less frequently, partial retention of exon 37 due to use of an upstream cryptic splice site.

Figure 4: Long Read Nanopore sequencing confirms de novo \textit{PKD1} variant occurred on paternal allele
Middle panel shows long read genome sequencing data from RPA017 over \textit{PKD1} exons 31-35, separated into maternal and paternal alleles. The top panel ‘zooms in’ over the region that includes the pathogenic \textit{PKD1} exon 31 variant identified in RPA017. The variant is absent in RPA015 (affected mother). Bottom panel ‘zooms in’ over an intronic single nucleotide variant identified in RPA017 that short-read GS and Sanger sequencing shows is present in RPA016 (unaffected father) and absent in RPA015 (affected mother). In RPA017, this variant is seen on the same long-read sequencing as the pathogenic \textit{PKD1} variant,
demonstrating that the de novo disease-causing variant in RPA017 has occurred on her paternally inherited PKD1 allele. GS = Genome sequencing
Supplementary Material

Supplementary Methods

Supplementary Figure 1: Disease-causing variants impacting PKD1 splicing

Supplementary Figure 2: Splicing studies in FRPA021

Supplementary Figure 3: RG0044 Pedigree

Supplementary Table 1: Primer Sequences and PCR Conditions

Supplementary Table 2: Patient Characteristics

Supplementary Table 3: GTex Data
Supplementary Methods

Methods for the Targeted RNA studies:

RNA extraction
Total RNA was extracted from peripheral blood using the Machery Nagel Nucleospin RNA Blood Kit following the manufacturer’s recommendations.

Reverse Transcription PCR and Sanger Sequencing
Superscript III (Invitrogen) first-strand synthesis system was used to make complementary DNA (cDNA) from 400ng total RNA according to the kit instructions and with the random hexamers included. All samples were also run with a Reverse Transcriptase negative control (RT-).

Amplitaq 360 DNA polymerase (applied biosystems) or LongAmp Taq DNA Polymerase (New England Biolabs) was used for PCR using the primers listed in Supplementary Table 1. Conditions for the PCR varied according to the primer sets, and are included in Supplementary Table 1.

Control cDNA was from healthy individuals, or healthy family members where available. PCR products were analyzed on 1.5 – 1.8% agarose gels.

PCR products were cleaned up using ExoSAP-IT (affymetrix USB) before Sanger Sequencing. 10-30ng PCR product was sequenced with 3.2pmol sequencing primer. Sanger sequence chromatograms were analysed using MacVector DNA sequence analysis software (MacVector Inc.).
<table>
<thead>
<tr>
<th>Region Covered</th>
<th>Primer Label</th>
<th>Direction</th>
<th>Primer Sequence 5' - 3'</th>
<th>PCR Enzymes and Conditions</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKD1 exon 37 - 39</td>
<td>PKD1_RTPCR_37FOR</td>
<td>Forward</td>
<td>gtcttgctggaagccctgtac</td>
<td>Amplitaq 360 DNA Pol + 0.5M Betaine and 5% DMSO 94°C 3 min; then 35 cycles of 95°C 30 sec, 62°C 30 sec, 72°C 30 sec; final extension of 72°C for 5 min. 1.8% Agarose gel.</td>
<td>RBW403, RPA028, RPA014</td>
</tr>
<tr>
<td></td>
<td>PKD1_RTPCR_39REV</td>
<td>Reverse</td>
<td>agctctgggctggactggtt</td>
<td>Long Amp Taq Pol + 0.5M Betaine and 5% DMSO 94°C 3 min; then 35 cycles of 95°C 30 sec, 62°C 30 sec, 72°C 30 sec; final extension of 72°C for 5 min. 1.8% Agarose gel.</td>
<td></td>
</tr>
<tr>
<td>PKD1 exon 5 - 12</td>
<td>PKD1_RTPCR_5F</td>
<td>Forward</td>
<td>ggggccccacggacctctgg</td>
<td>Touch-down protocol of initial step 95°C 2 min; followed by 14 cycles 98°C 20 sec, 69°C 15 sec with decreasing 0.5°C per cycle, 72°C 4 min; then followed by 25 cycles 98°C 20 sec, 62°C 15 sec, 72°C 4 min; final extension 72°C 5 min. Modified from Tan et al 2014. 1.5% Agarose gel.</td>
<td>RPA019</td>
</tr>
<tr>
<td></td>
<td>TANLR2(2_12)R^</td>
<td>Reverse</td>
<td>ccacggttagttgtagttcagtggtgac</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKD1 exon 31 - 34</td>
<td>Mut4F</td>
<td>Forward</td>
<td>ggactctgctcgtgctggac</td>
<td>Amplitaq 360 DNA Pol + 5% DMSO 94°C 3 min; then 35 cycles of 95°C 30 sec, 58°C 30 sec, 72°C 30 sec; final extension of 72°C for 5 min. 1.5% Agarose gel.</td>
<td>RBW401</td>
</tr>
<tr>
<td></td>
<td>AH3B2</td>
<td>Reverse</td>
<td>tccatgtaggtctctggtaggg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKD1 exon 15 - 21</td>
<td>ROSLR4(15_21)F#</td>
<td>Forward</td>
<td>agcgcaactactttggcc</td>
<td>Long Amp Taq Pol + 0.5M Betaine and 5% DMSO 94°C 1 min; then 35 cycles of 94°C 30 sec, 63°C 30 sec, 67°C 3 min; final extension 65°C 7 min. 1.8% Agarose gel.</td>
<td>RPA021</td>
</tr>
<tr>
<td></td>
<td>PKD1_21RTPCR_REV</td>
<td>Reverse</td>
<td>ctcggcgtctgctggggctt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKD1 exon 18 - 19</td>
<td>PKD1_18RTPCR_FOR</td>
<td>Forward</td>
<td>cacaagacgctggtctgg</td>
<td></td>
<td>RPA021*</td>
</tr>
<tr>
<td></td>
<td>PKD1_19RTPCR_REV</td>
<td>Reverse</td>
<td>tgtcctgttagacacacaagac</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*used for Sanger sequencing


## Supplementary Table 2: Patient Characteristics

<table>
<thead>
<tr>
<th>Family ID</th>
<th>Patient ID</th>
<th>Sex</th>
<th>CKD stage</th>
<th>eGFR/Age (years)</th>
<th>Kidney Imaging</th>
<th>Extra-renal features</th>
<th>Family History of ADPKD</th>
<th>Previous Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRBW403</td>
<td>RBW403</td>
<td>M</td>
<td>1</td>
<td>&gt;90/31</td>
<td>R kidney length 14.4cm, L kidney length 16.2cm; multiple bilateral simple renal cysts</td>
<td>multiple hepatic cysts</td>
<td>Affected father (ESKD 44yo)</td>
<td>LR-PCR PKD1 and PKD2 and MPS of PCR product (variant reported likely benign); Diagnostic GS</td>
</tr>
<tr>
<td>FRPA028</td>
<td>RPA028</td>
<td>F</td>
<td>1</td>
<td>&gt;90/23</td>
<td>R kidney length 11.9cm; L kidney length 11.9cm; multiple kidney cysts bilaterally</td>
<td>no hepatic cysts</td>
<td>Father affected; paternal grandmother affected (dialysis at 47yo)</td>
<td>Diagnostic GS (variant reported as VUS)</td>
</tr>
<tr>
<td>FRPA014</td>
<td>RPA014</td>
<td>F</td>
<td>5T</td>
<td>ESKD/56</td>
<td>bilateral, enlarged, cystic kidneys; L kidney length 15cm; R kidney length 14.8cm</td>
<td>diffuse liver cysts</td>
<td>Affected brothers, mother, maternal aunt and maternal grandmother</td>
<td>Diagnostic GS</td>
</tr>
<tr>
<td>FRPA019</td>
<td>RPA019</td>
<td>F</td>
<td>1</td>
<td>&gt;90/35</td>
<td>R kidney 12.6cm: L kidney 12.7cm; multiple cysts bilaterally</td>
<td>no hepatic cysts</td>
<td>Affected mother and brother</td>
<td>LR-PCR PKD1 and PKD2 and MPS of PCR product (variant reported as VUS)</td>
</tr>
<tr>
<td>RPA020</td>
<td>F</td>
<td>3b</td>
<td>44/69</td>
<td></td>
<td>numerous cysts both kidneys</td>
<td>not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F19F00138</td>
<td>RBW401</td>
<td>F</td>
<td>1</td>
<td>&gt;90/30</td>
<td>R kidney length 12.1cm, L kidney length 11.9cm; multiple bilateral simple renal cysts</td>
<td>not reported</td>
<td>Affected maternal grandmother. Affected mother and maternal uncle with ESKD</td>
<td>Diagnostic GS</td>
</tr>
<tr>
<td>19F00138</td>
<td>F</td>
<td>1</td>
<td>&gt;90/25</td>
<td>R kidney length 13.2cm, L kidney length 11.7cm; multiple bilateral simple renal cysts</td>
<td>not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRPA021</td>
<td>RPA021</td>
<td>F</td>
<td>5T</td>
<td>ESKD/52</td>
<td>bilateral, enlarged, cystic kidneys</td>
<td>massive liver cysts requiring liver transplant at 48yo</td>
<td>Affected Father and brother (ESKD 47yo)</td>
<td>Diagnostic GS (variant reported as VUS)</td>
</tr>
<tr>
<td>RG_0044</td>
<td>RG_0044.0048</td>
<td>F</td>
<td>5</td>
<td>ESKD/58</td>
<td>bilateral, enlarged, cystic kidneys</td>
<td>not reported</td>
<td>see Supp. Fig 2</td>
<td>LR-PCR PKD1 and PKD2 and MPS of PCR product</td>
</tr>
<tr>
<td>FRPA017</td>
<td>RPA017</td>
<td>F</td>
<td>3b</td>
<td>41/38</td>
<td>R kidney length 18cm; L kidney length 20cm; numerous bilateral simple renal cysts</td>
<td>not reported</td>
<td>Two affected maternal uncles with stage 3 CKD; maternal grandmother ESKD 81yo</td>
<td>Diagnostic GS</td>
</tr>
<tr>
<td>RPA015</td>
<td>F</td>
<td>3a</td>
<td>50/71</td>
<td>R kidney length 10.1cm; L kidney length 14.9cm; multiple bilateral simple renal cysts</td>
<td>not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| FRPA007   | RPA007     | F   | 5T        | ESKD/45         | massively enlarged kidneys bilaterally with numerous cysts | massively enlarged liver, required liver/kidney transplant | Affected mother | Diagnostic GS (variant reported as VUS) | CKD = Chronic Kidney Disease; ESKD = End Stage Kidney Disease; GS = Genome Sequencing; VUS = Variant of Uncertain Significance; MPS = Massively Parallel Sequencing; LR-PCR = Long Range PCR
Supplementary Figure 1: Disease-causing variants impacting PKD1 splicing

A - Pedigree of FRPA019; B - RNA studies from FRPA019 with RT-PCR product demonstrating wild type and truncated allele in RPA019 and RPA020. Sanger sequencing of RT-PCR product demonstrates deletion of 36 amino acids of exon 10 in RPA019 and RPA020. C - PKD1 variant generates new cryptic donor site and results in in-frame deletion of part of exon 10. D - Pedigree 19F00138; E - Visualisation of BAM files from short read genome sequencing data of 19F00138 demonstrating 19bp intronic deletion; F - RNA studies demonstrating retention of intron 31 in RBW401.
Supplementary Figure 2: Splicing studies in FRPA021:
A - Coronal section from Abdominal CT imaging showing massively enlarged native kidneys. B - Pedigree; C - RT-PCR studies; D - Sanger sequencing of RT-PCR product showing retention of *PKD1* intron 18 in one allele in RPA021.
Supplementary Figure 3: RG_044 Pedigree. Sequencing results highlighted in green. Shows PKD1 c.2878G>A variant segregates with phenotype in multiple family members.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable3.xlsx