npj | genomic medicine

Article

Published in partnership with CEGMR, King Abdulaziz University

https://doi.org/10.1038/s41525-025-00477-5

Targeted long-read sequencing enables higher diagnostic yield of ADPKD by accurate *PKD1* genetic analysis

Qian Sun[1,2,3,4,5,6,7,12], Peiwen Xu[1,2,3,4,5,6,7,12], Aiping Mao[8], Sexin Huang[1,2,3,4,5,6,7], Jie Li[1,2,3,4,5,6,7], Libao Chen[8], Jing Li[1,2,3,4,5,6,7], Haopeng Kan[1,2,3,4,5,6,7], Ju Huang[1,2,3,4,5,6,7], Wenkai Ji[1,2,3,4,5,6,7], Dayong Si[9], Junhao Yan[1,2,3,4,5,6,7], Zi-Jiang Chen[1,2,3,4,5,6,7,10,11], Xuan Gao[1,2,3,4,5,6,7] ✉ & Yuan Gao[1,2,3,4,5,6,7] ✉

Genetic diagnosis of ADPKD has been challenging due to the variant heterogeneity, presence of duplicated segments, and high GC content of exon 1 in PKD1. In our reproductive center, 40 patients were still genetically undiagnosed or diagnosed without single-nucleotide resolution after testing with a short-read sequencing panel in 312 patients with ADPKD phenotype. A combination of long-range PCR and long-read sequencing approach for PKD1 was performed on these 40 patients. LRS additionally identified 10 pathogenic or likely pathogenic PKD1 variants, including four patients with microgene conversion (c.160_166dup, c.2180T>C, and c.8161+1G>A) between PKD1 and its pseudogenes, three with indels (c.-49_43del, c.2985+2_2985+4del, and c.10709_10760dup), one with likely pathogenic deep intronic variant (c.2908-107G>A) and two with large deletions. LRS also identified nine PKD1 CNVs and precisely determined the breakpoints, while SRS failed to identify two of these CNVs. Therefore, LRS enables higher diagnostic yield of ADPKD and provides significant benefits for genetic counseling.

Autosomal dominant polycystic kidney disease (ADPKD), with an incidence of approximately 0.1–0.25% worldwide, is one of the most common inherited kidney disorders and responsible for 5–10% of end-stage renal disease[1,2]. Disease-causing variants in *PKD1* and *PKD2* account for approximately 85% and 15% of ADPKD, respectively[3]. In general, ADPKD is clinically diagnosed by ultrasonography based on age-related criteria[4]. However, the diagnosis made by imaging may be uncertain, particularly in young individuals less than 30 years old[4]. Therefore, molecular genetic testing is required to achieve a definite diagnosis[5]. The genetic diagnosis of ADPKD can provide significant benefits for patients who seek genetic counseling during assisted reproductive treatments and reduce the possibility of at-risk pregnancies[6].

However, the genetic diagnosis of ADPKD has been particularly challenging due to several factors. First, *PKD1* is a large gene that contains 46 exons spanning 47.2 kb of genomic DNA[5]. Second, genetic analysis for *PKD1* is complicated by the presence of 39.9-kb duplicated segment that encompasses exons 1–33 and shares 98% sequence similarity to six pseudogenes (*PKD1P1* to *PKD1P6*)[7,8]. Microgene conversions between *PKD1* and its pseudogenes can be caused by the transfer of sequences from high homologous pseudogenes to functional genes, which have been reported to be associated with ADPKD[9–11]. Third, the high GC content of exon 1 in *PKD1* makes it difficult to amplify and sequence using standard sequencing methods[2]. Lastly, a high level of allelic heterogeneity has been observed in disease-causing variants of *PKD1* and *PKD2*, including more than one thousand pathogenic or likely pathogenic single-nucleotide variants (SNVs) and insertions/deletions (indels), and the majority of these variants are exclusive to a single pedigree[12]. In addition, approximately 2–6% of ADPKD are caused by

[1]State Key Laboratory of Reproductive Medicine and Offspring Health, Center for Reproductive Medicine, Institute of Women, Children and Reproductive Health, Cheeloo College of Medicine, Shandong University, Jinan, China. [2]National Research Center for Assisted Reproductive Technology and Reproductive Genetics, Shandong University, Jinan, China. [3]Key Laboratory of Reproductive Endocrinology (Shandong University), Ministry of Education, Jinan, China. [4]Shandong Technology Innovation Center for Reproductive Health, Jinan, China. [5]Shandong Provincial Clinical Research Center for Reproductive Health, Jinan, China. [6]Shandong Key Laboratory of Reproductive Medicine, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China. [7]Research Unit of Gametogenesis and Health of ART-Offspring, Chinese Academy of Medical Sciences (No.2021RU001), Jinan, China. [8]Department of Research and Development, Berry Genomics Corporation, Beijing, China. [9]School of Life Science, Jilin University, Changchun, China. [10]Shanghai Key Laboratory for Assisted Reproduction and Reproductive Genetics, Shanghai, China. [11]Department of Reproductive Medicine, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [12]These authors contributed equally: Qian Sun, Peiwen Xu. ✉e-mail: gaoxuan@sduivf.com; gaoyuan@sduivf.com

copy number variants (CNVs), which consist of single to multiple exon deletions or duplications[5,13–15].

Currently, a few genetic testing methods, such as long-range PCR, short-read sequencing (SRS) and multiplex ligation-dependent probe amplification (MLPA), have been used in the diagnosis of ADPKD[16–20]. Due to high sequence homology between *PKD1* and its pseudogenes, SRS approaches could lead to false positive or false negative genotyping due to the incorrect calling of variants from pseudogenes[21–23]. Although targeted SRS combined with long-range PCR for locus-specific *PKD1* provides a robust approach for ADPKD genetic testing, extra input for exon 1 in *PKD1* is required to achieve balanced sequencing depth due to GC bias during SRS[24]. MLPA can only be employed to detect large deletions and duplications in *PKD1* and *PKD2*, which could increase the turnaround time and diagnostic cost[15,24]. These traditional methods have limitations in accurately detecting all types of variants in *PKD1* and *PKD2*, particularly in the duplicated regions of *PDK1*. This has led to a significant diagnostic gap in ADPKD genetic testing, with some patients remaining undiagnosed or incompletely diagnosed despite a clear clinical phenotype. At present, the diagnostic rate in ADPKD is approximately 80 to 90%[1,20,25–28]. Therefore, it is necessary to improve the diagnosis for ADPKD patients, especially when they are seeking assisted reproductive treatments.

In recent years, single-molecule GC-unbiased long-read sequencing (LRS) has emerged as a promising approach to overcome these challenges, and improved genetic diagnosis for diseases with complicated molecular genetics such as thalassemia[29,30], congenital adrenal hyperplasia[31,32], spinal muscular atrophy[33,34], fragile-X syndrome[35], hemophilia A[36], and ADPKD[27]. LRS-based approach generates reads spanning large genomic regions of *PKD1* and *PKD2*, improves detection of structural variants and the resolution of repetitive sequences, and makes incremental diagnostic rate compared to LR-PCR, SRS and MLPA combined approach[27].

In this study, we aimed to evaluate the effectiveness of LRS-based approach in improving the genetic diagnosis of ADPKD, particularly for patients who had remained undiagnosed or incompletely diagnosed after conventional SRS testing.

Results

Improvement of diagnostic yield in *PKD1* and *PKD2* by LRS

A total of 40 clinically diagnosed ADPKD patients without full genetic characterization by SRS were retrospectively enrolled in this study (Fig. 1). The cohort included 28 male and 12 female patients, with a mean age of 25.8 ± 5.5 years. Targeted LRS for *PKD1* and *PKD2* genetic analysis were performed for all 40 patients. LRS identified 9 P/LP and 17 VUS SNVs/indels in *PKD1*, as well as 9 P/LP CNVs in *PKD1* (Fig. 1, Table 1). Compared to SRS panel-based diagnostic workflow, LRS approach identified pathogenic or likely pathogenic *PKD1* variants in ten more patients, which increased the genetic diagnosis rate from 20.0% (8/40) to 45.0% (18/40) and the variant detection rate from 62.5% (25/40) to 87.5% (35/40) among previously undiagnosed cases. This represents a significant improvement in diagnostic yield for this challenging subset of patients. Among all the cases, LRS could presumably increase the genetic diagnosis rate from 89.7% (280/312) to 92.8% (290/312) and the variant detection rate from 95.2% (297/312) to 98.4% (307/312).

LRS detections for *PKD1* and *PKD2* SNVs/indels

Eight patients were found to have pathogenic or likely pathogenic SNVs/indels in PKD1 using targeted LRS approach. These eight variants included four microgene conversions between *PKD1* and its pseudogenes, three indels, and one deep intronic variant (Table 1). All eight variants were confirmed by Sanger sequencing (Supplementary Figs. 1 and 2).

Among the four patients with microgene conversions, two patients harbored the same frameshift variant *PKD1*:c.160_166dup, which was a duplication of 7 nucleotides in exon 1 and existed in four pseudogenes (*PKD1P1/3/5/6*) (Fig. 2A). There was no detected SRS reads to support the variant in these samples (Fig. 2A). Patient KPK003 had a missense variant *PKD1*:c.2180T>C in exon 11, and this variant also located in five pseudogenes (*PKD1P1/2/3/4/5*) (Fig. 2B). The splice site variant *PKD1*:c.8161+1G>A, identified in patient KPK004, was affecting intron 22 and located in three pseudogenes (*PKD1P1/3/4*) (Fig. 2C). The SRS reads of these two variants were below the 20% detection threshold, which could be due to the alignment of these large segments to their pseudogenes (Fig. 2B, C). Visual inspection of the SRS bam files supported the existence of these two variants.

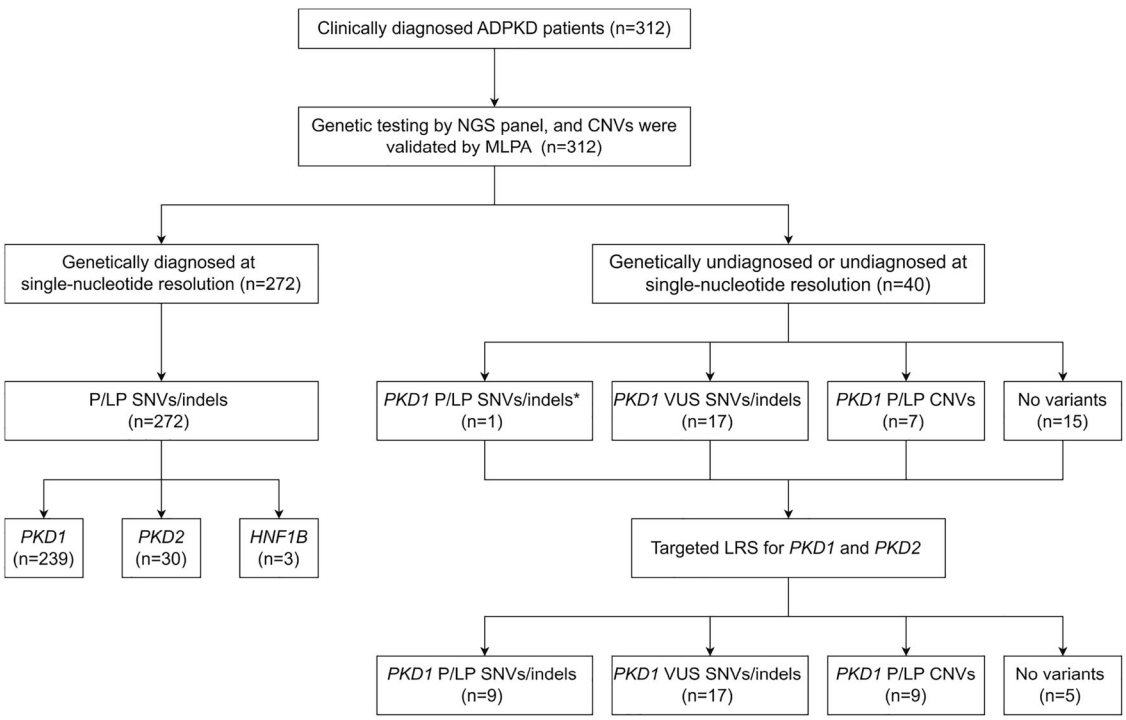

Fig. 1 | Flowchart of study design. *The variant *PKD1*:c.12048C>T in KPK035 was classified as VUS at the time of enrollment and was upgraded to pathogenic with added evidence in this study.

Table 1 | *PKD1* variants identified by LRS and SRS in the 40 ADPKD probands

Sample	With family history	SRS	LRS						Concordance analysis
		Nucleotide	Nucleotide	Amino acid	Region	Type	Pathogenicity	Novel	
KPK001	Yes	ND	c.160_166dup	p.L56Pfs*60	Exon1	Frameshift	P	No	Discordant
KPK002	Yes	ND	c.160_166dup	p.L56Pfs*60	Exon1	Frameshift	P	No	Discordant
KPK003	Yes	ND	c.2180T>C	p.L727P	Exon11	Missense	LP	No	Discordant
KPK004	Yes	ND	c.8161+1G>A	NA	Intron22	Splicing	P	No	Discordant
KPK005	Yes	ND	c.-49_43del	NA	Exon1	Start codon loss	P	Yes	Discordant
KPK006	No	ND	c.2985+2_2985+4del	NA	Intron12	Splicing	P	Yes	Discordant
KPK007	Yes	ND	c.10709_10760dup	p.W3588Gfs*56	Exon36	Frameshift	P	Yes	Discordant
KPK008	Yes	ND	c.2098-107G>A	NA	Intron10	Splicing	VUS	Yes	Discordant
KPK009	Yes	c.10984C>T, c.12408G>T	c.11712+81_11713-100del, c.10984C>T/ c.12408G>T	NA p.R3662C, p.R4136S	Intron42 Exon37, Exon45	Splicing Missense, Missense	VUS VUS, VUS	Yes Yes, Yes	Concordant*
KPK010	Yes	c.8949-14C>G, c.9193G>A, c.9789G>T	c.8949-14C>G, c.9193G>A, c.9789G>T	NA p.V3065M, p.W3263C	Intron24, Exon25, Exon29	Splicing, Missense, Missense	VUS, VUS, VUS	No	Concordant*
KPK011	Yes	ND	Del40.9kb (chr16:2,132,723-2,173,682)	NA	Upstream-Intron1	Deletion	P	Yes	Discordant
KPK012	No	ND	Del2.1kb (chr16:2,088,749-2,090,850)	NA	Exon44-Intron46	Deletion	P	Yes	Discordant
KPK013	Yes	Del(Upstream, E3,5-7,9-14)	Del39.9kb (chr16:2,112,048-2,151,914)	NA	Upstream-Intron14	Deletion	P	Yes	Concordant*
KPK014	Yes	Del(E1-14)	Del39.9kb (chr16:2,112,048-2,151,914)	NA	Upstream-Intron14	Deletion	P	Yes	Concordant*
KPK015	Yes	Del(Upstream,E3,5-7,9-15)	Del90.6kb (chr16:2,109,835-2,200,473)	NA	Upstream-Exon15	Deletion	P	Yes	Concordant*
KPK016	Yes	Del(E10-11)	Del2.4kb (chr16:2,113,448-2,115,805)	NA	Intron9-Intron11	Deletion	P	Yes	Concordant*
KPK017	Yes	Del(E15)	Del3.2kb (chr16:2,108,095-2,111,259)	NA	Exon15-Intron15	Deletion	P	Yes	Concordant*
KPK018	Yes	Del(E31-34)	Del1.7kb (chr16:2,096,810-2,098,505)	NA	Intron30-Intron34	Deletion	P	Yes	Concordant*
KPK019	Yes	Dup(E43-44)	Dup612bp (chr16:2,090,688-2,091,265, chr16:2,091,204-2,091,237)	NA	Intron42-Exon44	Duplication	P	Yes	Concordant*
KPK020	No	c.191G>C	c.191G>C	p.R64P	Exon1	Missense	VUS	No	Concordant
KPK021	Yes	c.386G>A	c.386G>A	p.C129Y	Exon4	Missense	VUS	Yes	Concordant
KPK022	No	c.1386-11C>G	c.1386-11C>G	NA	Intron6	Splicing	VUS	Yes	Concordant
KPK023	Yes	c.1676C>T	c.1676C>T	p.P559L	Exon8	Missense	VUS	No	Concordant
KPK024	Yes	c.1781T>C	c.1781T>C	p.F594S	Exon9	Missense	VUS	No	Concordant
KPK025	Yes	c.2153A>C	c.2153A>C	p.Q718P	Exon11	Missense	VUS	Yes	Concordant
KPK026	Yes	c.4573G>T	c.4573G>T	p.V1525F	Exon15	Missense	VUS	Yes	Concordant
KPK027	Yes	c.4999A>C	c.4999A>C	p.T1667P	Exon15	Missense	VUS	Yes	Concordant
KPK028	Yes	c.6170T>A	c.6170T>A	p.V2057D	Exon15	Missense	VUS	Yes	Concordant
KPK029	Yes	c.7587G>C	c.7587G>C	p.K2529N	Exon19	Missense	VUS	Yes	Concordant
KPK030	Yes	c.8501T>C	c.8501T>C	p.F2834S	Exon23	Missense	VUS	Yes	Concordant
KPK031	No	c.9559G>A	c.9559G>A	p.D3187N	Exon27	Missense	VUS	Yes	Concordant
KPK032	Yes	c.9758T>C	c.9758T>C	p.L3253P	Exon29	Missense	VUS	Yes	Concordant

Table 1 (continued) | *PKD1* variants identified by LRS and SRS in the 40 ADPKD probands

Sample	With family history	SRS	LRS						Concordance analysis
		Nucleotide	Nucleotide	Amino acid	Region	Type	Pathogenicity	Novel	
KPK033	Yes	c.10785C>G	c.10785C>G	p.S3595R	Exon36	Missense	VUS	Yes	Concordant
KPK034	Yes	c.11588T>C	c.11588T>C	p.L3863P	Exon42	Missense	VUS	Yes	Concordant
KPK035	Yes	c.12048C>T	c.12048C>T	p.G4016=	Exon44	Splicing	VUS→P	No	Concordant
KPK036	No	ND	c.1607-76C>T	NA	Intron7	Splicing	LB	Yes	Concordant
KPK037	No	ND	ND	NA	NA	NA	NA	NA	Concordant
KPK038	No	ND	ND	NA	NA	NA	NA	NA	Concordant
KPK039	No	ND	ND	NA	NA	NA	NA	NA	Concordant
KPK040	No	ND	ND	NA	NA	NA	NA	NA	Concordant

ND not detected, *NA* not applicable, *P* pathogenic, *LP* likely pathogenic, *VUS* variants of uncertain significance, *LB* likely benign
*LRS determined the phasing or had single-nucleotide resolution; reference build, GRCh38/hg38; PKD1 transcript, NM_001009944.3.

LRS also detected two small deletions of *PKD1*:c.-49_+43del and *PKD1*:c.2985+2_2985+4del, and one small duplication of *PKD1*:c.10709_10760dup in three individual samples (Fig. 3A–C). These three variants were missed by traditional SRS due to limited read length or low coverage in specific regions. Of note, LRS also identified a benign 34-bp duplication of *PKD1*:c.11713-30_11713-63dup in sample KPK007 (Fig. 3C). Further analysis showed that this region had a variable number of tandem repeat and occurred in 92.5% (37/40) of the samples (Supplementary Fig. 2D).

In addition, LRS identified two deep intronic variants, *PKD1*:c.1607-76C>T in patient KPK036 and *PKD1*:c.2908-107G>A in patient KPK008 (Fig. 3D). The SRS coverage of the *PKD1*:c.2908-107G>A region in patient KPK008 was only 15x (Fig. 3D), which hampered the calling of the variant. RT-PCR sequencing proved that *PKD1*:c.1607-76C>T did not affect the splicing and was a likely benign variant (Supplementary Fig. 2G). However, RT-PCR analysis revealed that *PKD1*:c.2908-107G>A created a new splice acceptor site, leading to the inclusion of a 111-nucleotide pseudoexon. This aberrant splicing event results in an inframe insertion of 37 amino acids and premature termination codon (Fig. 3E, F). Moreover, pedigree analysis of 15 family members from three generations showed *PKD1*:c.2908-107G>A was likely pathogenic because the variant was co-segregated with ADPKD phenotype within the family (Fig. 3G).

Both LRS and SRS identified the variant *PKD1*:c.12048C>T in patient KPK035 (Table 1). Although this was synonymous variant, RT-PCR analysis showed the variant affected splicing by creating a new exonic splice enhancer site. This led to a 92-bp deletion transcript and resulted in a frameshift (Supplementary Fig. 3A, B). In addition, this variant was also co-segregated with ADPKD phenotype through pedigree analysis (Supplementary Fig. 3C). Thus, *PKD1*:c.12048C>T was reclassified as a pathogenic variant by considering the newly added evidence.

LRS can perform the phasing analysis for SNVs/indels. In patient KPK009, the variants of *PKD1*:c.10984C>T and c.11712+81_11713-100del were found to be located on one allele, while variant c.12408G>T was located on the other allele (Fig. 3H). LRS also determined variants *PKD1*:c.8949-14C>G, c.9193G>A and c.9789G>T were in cis-configuration in patient KPK010 (Fig. 3I), which has been reported in another ADPKD patient[27]. Considering the effect of hypomorphic variants in *PKD1* gene, the compound heterozygous variants generally cause more severe manifestations[37,38]. Therefore, phasing analysis can enrich the mutational spectrum of ADPKD genes and serve as a reference for future patients.

LRS detections for large deletions and duplications in *PKD1*

Six large deletions in *PKD1* were identified by SRS and confirmed by MLPA (Table 1, Supplementary Fig. 4A–F). Besides these six deletions (Fig. 4A–C), LRS identified two more deletions in two samples (Fig. 4D, E), which were also confirmed by MLPA (Supplementary Fig. 4G, H). Moreover, LRS was able to precisely determine the breakpoints of these eight large deletions in CNVs.

SRS and MLPA can detect the duplication of exons 43 and 44 in *PKD1* (Fig. 5A). However, the exact sequence and inserted position of these duplications cannot be determined by traditional sequencing methods. LRS demonstrated that the duplication was a 578-bp sequence (hg38 chr16:2,090,688-2,091,265) encompassing partial intron 42, exon 43, intron 43, and partial exon 44 (Fig. 5B, C). The breakpoint of duplication was validated by Sanger sequencing (Fig. 5D). RT-PCR analysis showed that the duplication led to an alternative transcript with duplication of exon 43 and thus an inframe insertion of 97 amino acids with skipping the duplicated partial exon 44 (Fig. 5C, E, F). Therefore, LRS can directly detect large duplications in *PKD1* at single-nucleotide resolution. In addition, this duplication was found to co-segregate with ADPKD phenotype in a three-generation pedigree (Fig. 5G).

Discussion

For the past eight years, our assisted reproductive center has utilized SRS panel for genetic testing of 312 ADPKD patients. Compared to SRS panel, targeted LRS additionally identified eight P/LP SNVs/indels and two P/LP CNVs in *PKD1*, as well as determined single-nucleotide information for all the CNVs. Our study demonstrated the significant advantages of using long-read sequencing for the genetic diagnosis of ADPKD, particularly in cases where conventional SRS approaches have failed to provide a definitive diagnosis. The ability of LRS to span large genomic regions and resolve complex structural variations has allowed us to identify a spectrum of variants that were previously undetectable or difficult to characterize. The improved diagnostic yield of LRS for ADPKD patients is particularly important for assisted reproductive treatments since genetic diagnosis is required for those who seek preimplantation genetic testing for monogenetic diseases (PGT-M) to lower the pregnancy risk.

The identification of microgene conversions between *PKD1* and its pseudogenes is particularly noteworthy to utilize LRS approach. These events, which involve the transfer of genetic information from pseudogenes to the functional *PKD1* gene, can be easily missed by SRS technologies due to the high sequence similarity between *PKD1* and its pseudogenes[7]. In the past, microgene conversions were generally detected by short PCR followed by Sanger sequencing, which could lead to drop-out of variants with microgene conversions. In this study, LRS was used to identify four microgene conversions ranging from 200 to 1000 bp that were missed by SRS. Our results highlighted the importance of considering such mechanisms in the pathogenesis of ADPKD and underscored the requirement for sequencing technologies that can accurately distinguish between the functional gene and its pseudogenes.

Although various bioinformatics pipelines have been developed to call SNVs/indels from SRS data, accurate detection of indels over 50 bp is still very challenging[27,39,40]. Many partially mapped reads of these large indels can

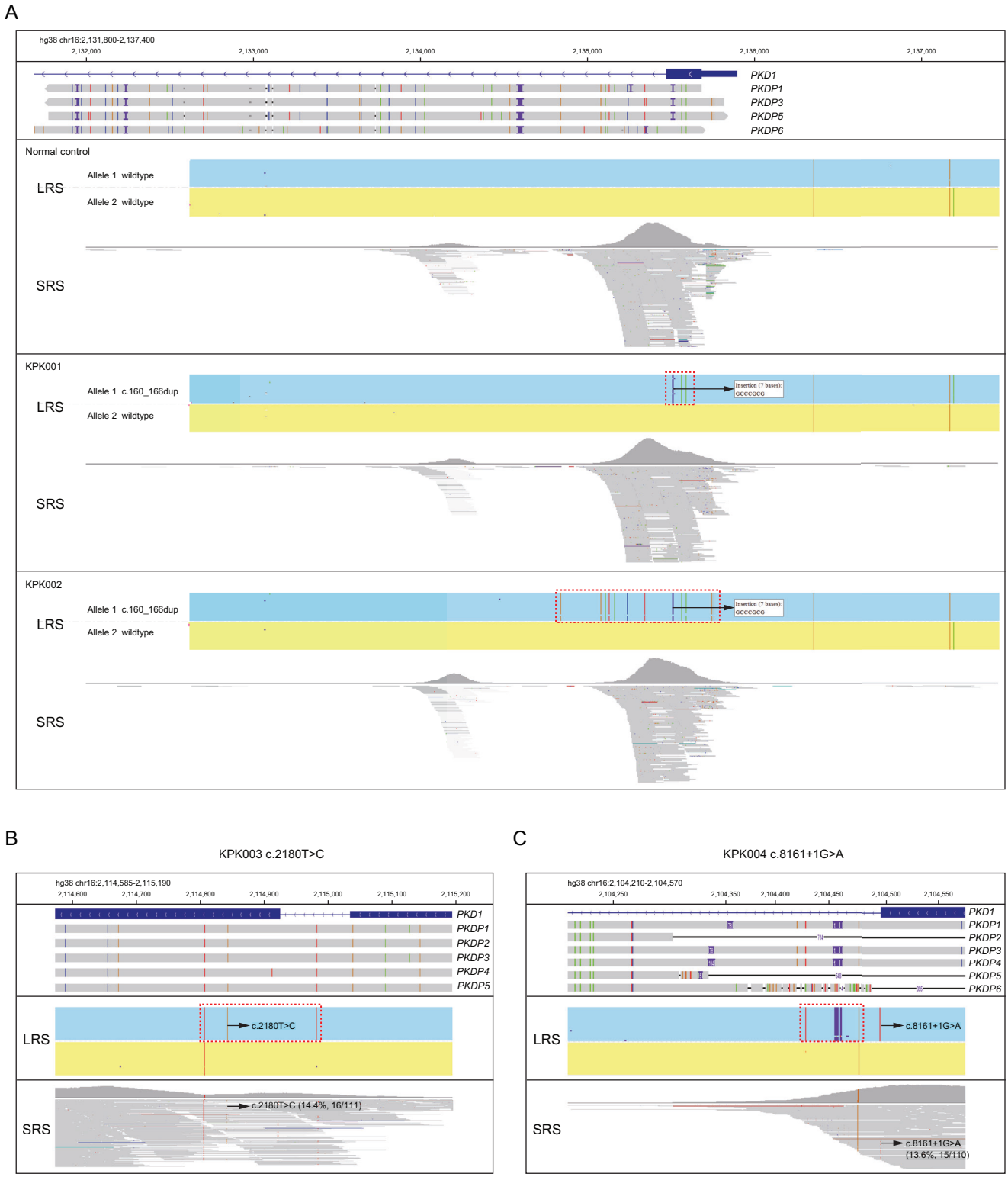

Fig. 2 | LRS additionally identified four *PKD1* SNVs/indels caused by microgene conversion between *PKD1* and pseudogenes. A–**C** IGV plots of LRS CCS reads and SRS reads showing *PKD1*:c.160_166dup (**A**), *PKD1*:c.2180T>C (**B**), and *PKD1*:c.8161+1G>A (**C**). Pseudogenes were aligned to *PKD1* and those homologous to the displayed region were shown in the IGV plots. The dotted boxes highlighted regions with microgene conversion.

be lost in unmatched fragments[27,39,40]. In this cohort study, SRS panel failed to detect two intermediate sizes of indels including *PKD1*:c.-49_43del and *PKD1*:c.10709_10760dup. Surprisingly, SRS also failed to detect a 3-bp deletion *PKD1*:c.2985+2_c.2985+4del. Manual inspection on the original SRS data suggested the presence of 3-bp deletion variant. However, the sequencing depth was only 19x in this specific area, and 20x sequencing depth is generally required to make a call in the bioinformatics pipeline, which is a common issue for large SRS panels. In SRS, the quality control for average sequencing depth is usually acceptable, but it does not guarantee the sequencing depth in each specific locus. The targeted LRS assay utilized full-length reads of each amplicon to do variant calling, therefore all the regions within the coverage can be successfully analyzed with ≥30x coverage.

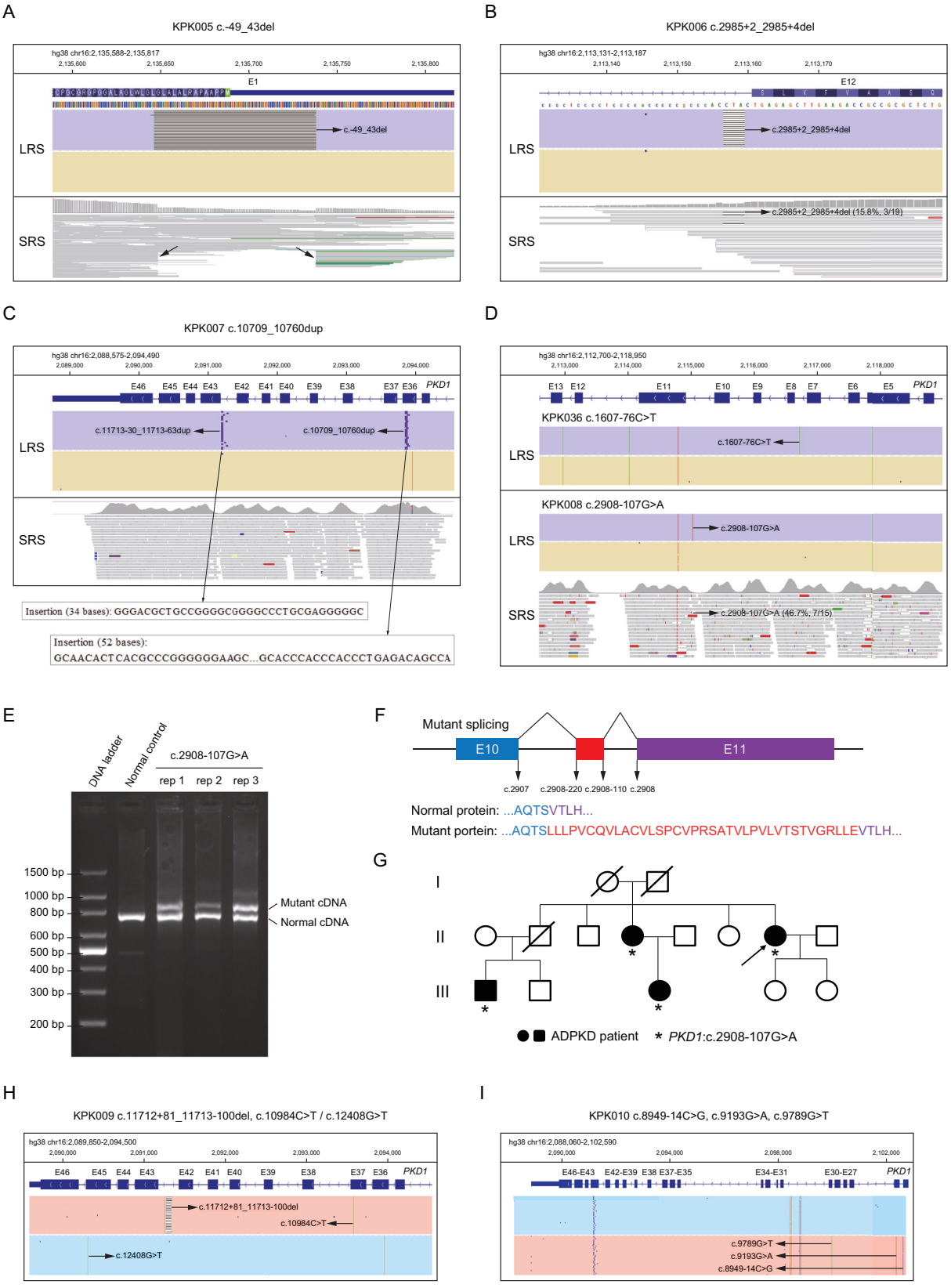

Fig. 3 | LRS additionally identified three indels and two deep intronic variants in *PKD1*. **A**–**C** IGV plots of LRS CCS reads and SRS reads showing *PKD1*:c.-49_43del (**A**), *PKD1*:c.2985+2_2985+4del (**B**), and *PKD1*:c.10709_10760dup (**C**). **D** IGV plots of LRS CCS reads and SRS reads showing two deep intronic variants, *PKD1*:c.1607-76C>T and *PKD1*:c.2908-107G>A. **E** Amplification of cDNA showing the existence of an alternative transcript caused by the variant *PKD1*:c.2908-107G>A. **F** Diagram showing the transcription of 111-bp pseudoexon led to inframe insertion of 37 amino acids caused by the variant *PKD1*:c.2908-107G>A. **G** The pedigree, phenotypes and genotypes of the family of KPK008 with the variant *PKD1*:c.2908-107G>A. **H**, **I** IGV plots of CCS reads showing the phasing the three variants in KPK009 (**H**) and KPK010 (**I**).

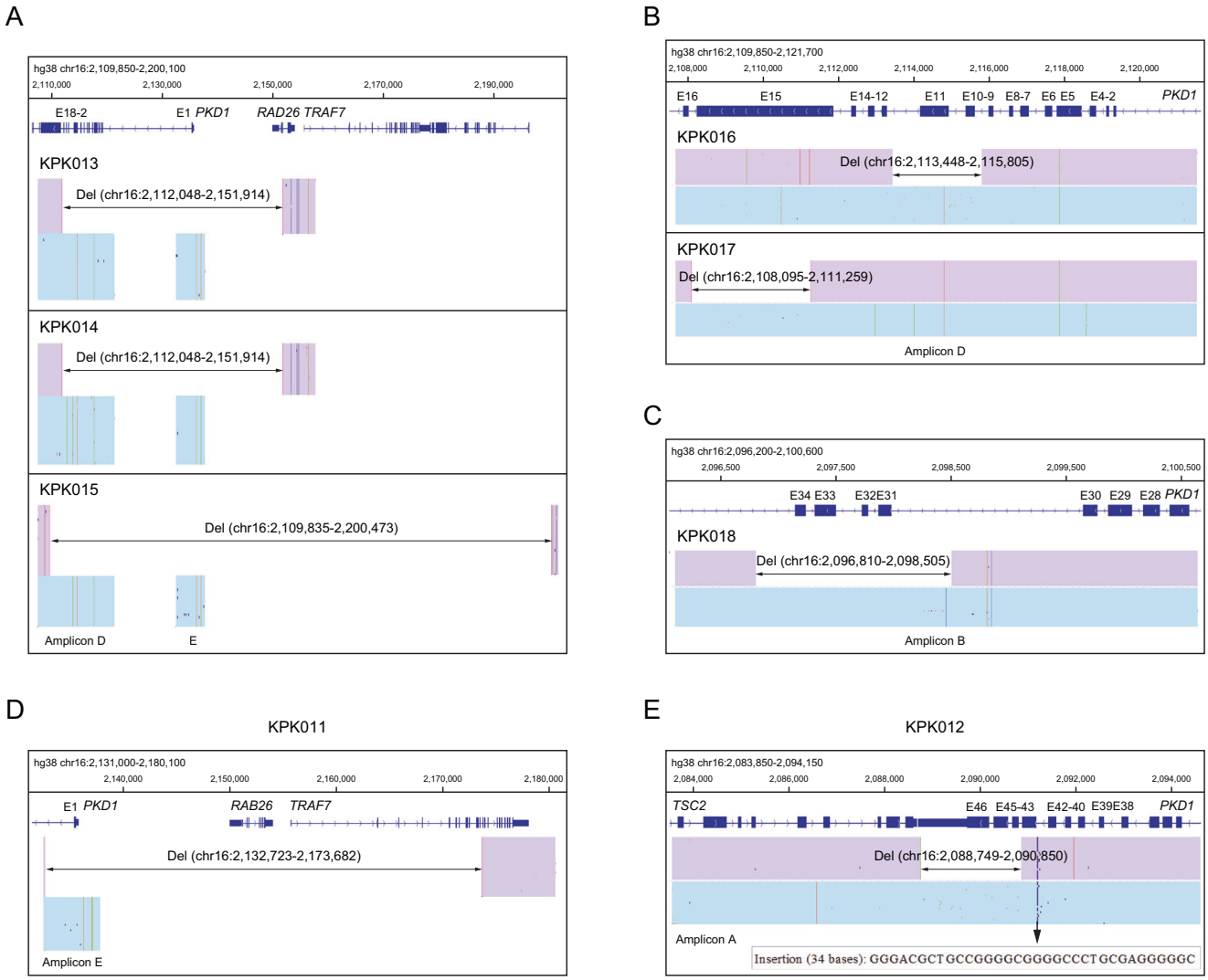

Fig. 4 | LRS identified eight large deletions in *PKD1*. **A**–**E** IGV plots of CCS reads showing the deletions in KPK013, KPK014, and KPK015 (**A**), KPK016 and KPK017 (**B**), KPK018 (**C**), KPK011 (**D**), and KPK012 (**E**). The purple and blue boxes represented two alleles, and the purple alleles contained the deletions.

In recent years, more and more disease-causing deep intronic variants have been discovered due to the rapid development of sequencing technologies[41,42]. Such variants can be easily overlooked by exon-focused sequencing approaches but may have significant implications for disease pathogenesis[43]. In this study, LRS identified two novel deep intronic variants *PKD1*:c.1607-76C>T and *PKD1*:c.2908-107G>A that were undetectable in SRS panel. RT-PCR analysis showed that *PKD1*:c.1607-76C>T did not affect the splicing, but *PKD1*:c.2908-107G>A led to the inclusion of a 111-nucleotide pseudoexon with an inframe insertion of 37 amino acids. Pedigree analysis proved *PKD1*:c.2908-107G>A was co-segregated with ADPKD phenotype and was a likely pathogenic variant. The identification of the c.2908-107G>A variant and its effect on splicing illustrated the potential of LRS to expand our understanding of the mutational spectrum in ADPKD.

Since the majority of *PKD1* and *PKD2* variants are exclusive to single families, novel and VUS variants can be identified in a cohort of ADPKD patients[25–27]. Precise genetic diagnosis is required for PGT-M to select embryos without disease-causing variants, therefore functional and pedigree analysis are critical to accumulate evidence to determine the pathogenicity of each variant according to the AMCG guidelines. Here, we found that the synonymous variant *PKD1*:c.12048C>T caused a 62-bp deletion transcript and thus frameshift of translation by RT-PCR. In combination of co-segregation analysis, the pathogenicity of this variant was upgraded from VUS to pathogenic. Moreover, LRS can determine the cis and trans-configuration of variants within the same PCR amplicon without the need for pedigree analysis.

Besides SNVs/indels, SRS has been gradually and widely used for CNV analysis[44,45]. In our study, LRS identified two more *PKD1* CNVs compared to SRS, including one deletion encompassing exon 44 to 46 and another deletion encompassing exon 1 to upstream regions. This highlighted the superior sensitivity of LRS for CNV detection in complex genomic regions. The improved detection of CNVs by LRS addressed a significant limitation of SRS-based approaches. SRS typically relies on read depth analysis to infer CNVs, which can be challenging in regions with pseudogenes or repetitive sequences[46]. In addition, the precise determination of breakpoints by LRS offered valuable insights into the molecular mechanisms underlying these variants, which is crucial for understanding genotype-phenotype correlations and may have implications for predicting disease severity[47].

The improved diagnostic yield achieved by LRS in our study (25% of previously undiagnosed cases and 3.2% of all cases) has significant clinical implications. Precise genetic counseling to patients and their families can allow for informed decision-making regarding family planning and genetic testing of at-risk relatives[48]. Furthermore, preimplantation genetic testing for couples undergoing assisted reproduction can potentially reduce the transmission of ADPKD to future generations[49]. The major limitation of SMRT LRS technology is the high cost for instrument and sequencing[50].

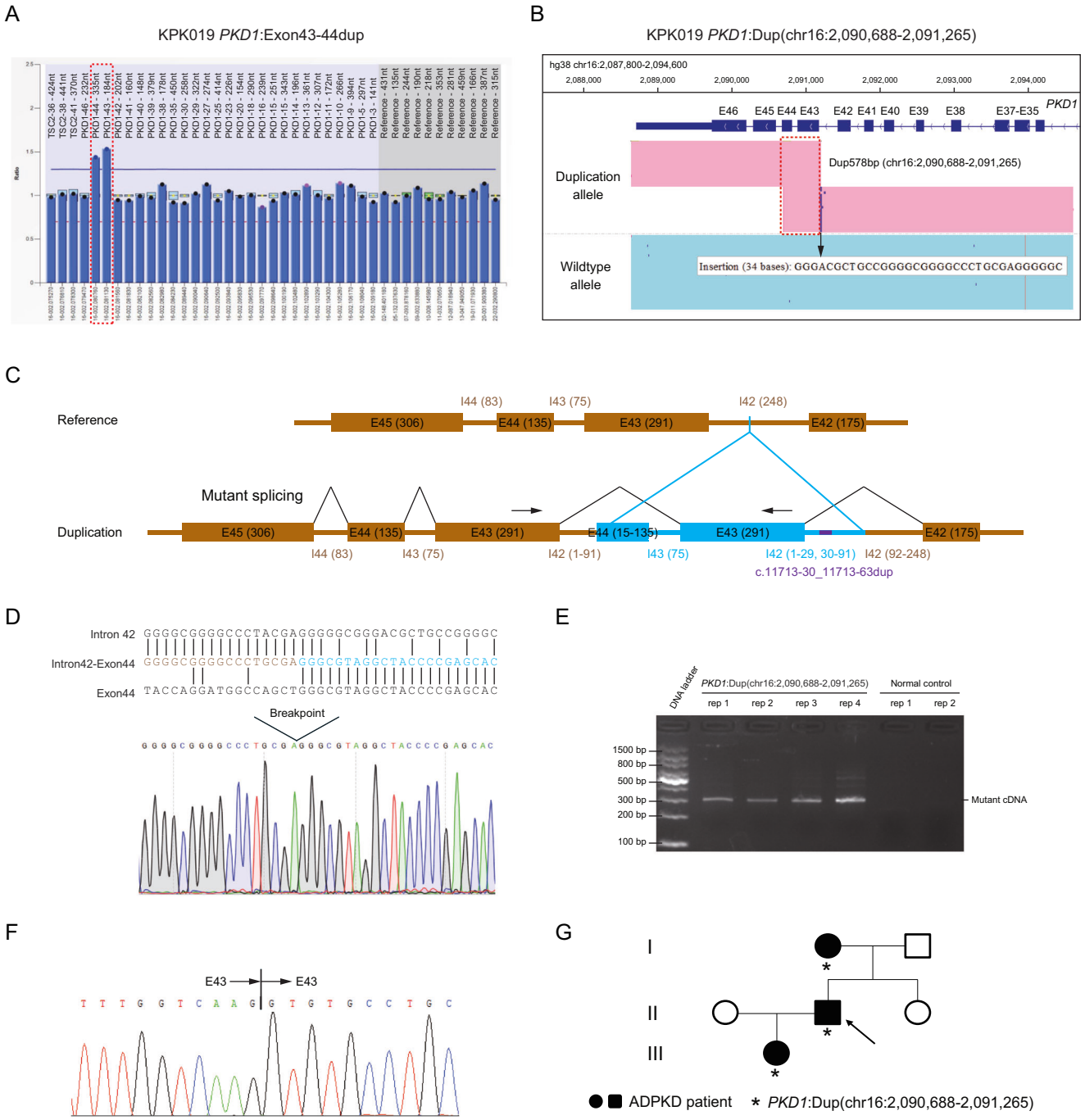

Fig. 5 | LRS identified a large duplication in *PKD1*. **A** MLPA showing the duplication of probes targeting exons 43 and 44 of *PKD1* in KPK019. **B** IGV plots of CCS reads showing the 578-bp duplication (chr16:2,090,688-2,091,265) encompassing partial intron 42, exon 43, intron 43 and partial exon 44 in KPK019. **C** Diagram showing the duplication arrangement. The arrows indicated the design of primers of cDNA analysis. **D** Sanger sequencing validated the breakpoints of duplication identified by LRS. **E** Amplification of cDNA showing the existence of an alternative transcript caused by the 578-bp duplication. **F** Sanger sequencing of the cDNA showing the duplicated transcription of exon 43, and the partially duplicated exon 44 was skipped. **G** The pedigree, phenotypes and genotypes of the family of KPK019 with the 578-bp duplication.

However, with the development of more cost-effective benchtop instruments to decrease the overall sequencing costs, LRS technology would have better clinical feasibility in the near future.

Future applications of LRS technology should focus on several directions. First, integrating LRS into routine clinical genetic testing for ADPKD, including the development of standardized protocols and bioinformatics pipelines optimized for *PKD1* and *PKD2* analysis. Second, investigating the utility of LRS in other genetically complex disorders, particularly those involving large genes, pseudogenes, or repetitive regions. Third, developing improved methods for functional characterization of novel variants identified by LRS, including high-throughput assays for assessing the impact of deep intronic or regulatory variants.

Our study demonstrated that targeted long-read sequencing enabled a higher diagnostic yield for ADPKD through accurate and precise genetic analysis of SNVs, indels, and CNVs in *PKD1* and *PKD2*. This approach has significant implications for improving genetic diagnosis, particularly in cases where conventional sequencing methods have been inconclusive. As LRS technologies continue to evolve and become more accessible, their integration

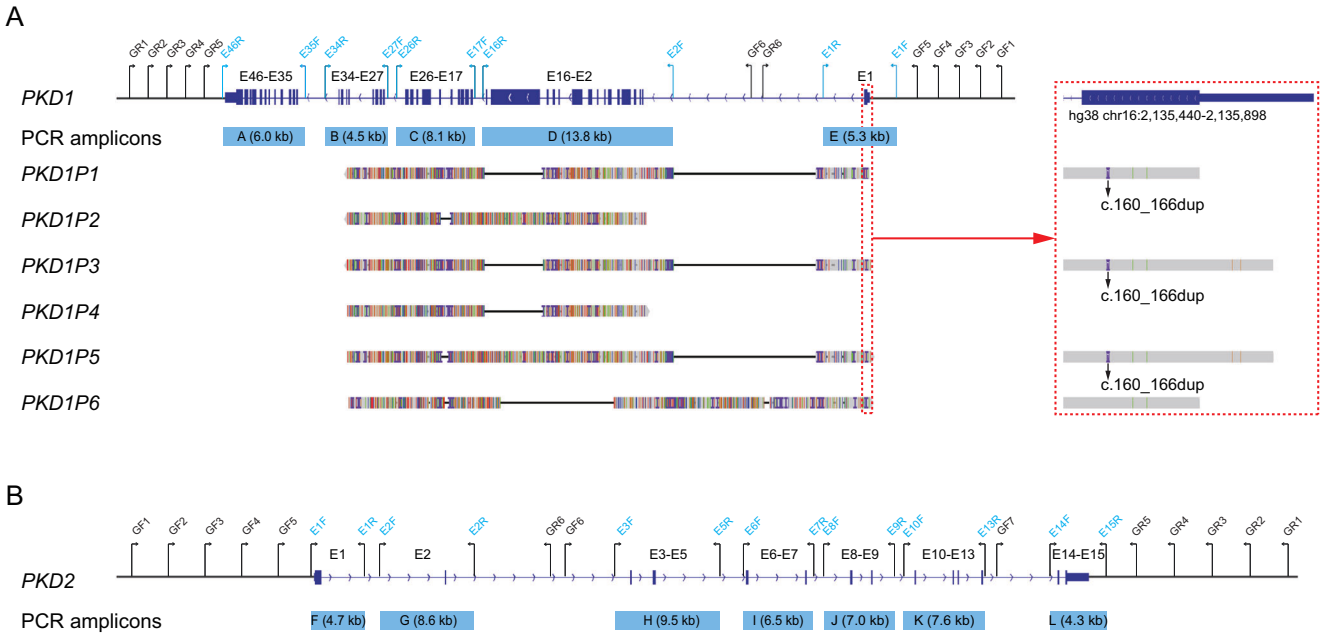

Fig. 6 | Design of multiplex long-range PCR primers for *PKD1* and *PKD2*. **A** PCR design for *PKD1* and alignment of six pseudogenes (*PKD1P1*-P6) to *PKD1*. The *PKD1P1-P6* were individually aligned to *PKD1* to show the sequence difference, as indicated by vertical horizontal lines. The green, red, blue, orange, and purple vertical lines represented A, T, C, G, and insertions, respectively. The horizontal lines represented deletions. **B** PCR design for *PKD2*. GF forward gap primer, GR reverse gap primer.

into routine clinical practice has the potential to improve genetic counseling and offer preimplantation genetic testing for more ADPKD patients.

Methods

Study participants

From January 2016 to December 2023, 312 patients with clinical diagnosis of ADPKD have undergone genetic testing at our assisted reproductive center using a targeted SRS panel, which included thousands of kidney disease-related genes such as *PKD1*, *PKD2*, *PKHD1* and *HNF1B*. The pathogenicity of SNV/indels was classified according to the American College of Medical Genetics and Genomics guidelines[51]. From this cohort, a total of 272 patients were genetically diagnosed with pathogenic (P) or likely pathogenic (LP) SNVs/indels, including 239 cases with variants in *PKD1*, 30 cases with variants in *PKD2*, and 3 cases with variants in *HNF1B* (Fig. 1). The remaining 40 patients were either genetically undiagnosed or diagnosed without single-nucleotide resolution at the time of enrollment in our study, which included 18 cases with variants of uncertain significance (VUS) in *PKD1*, 7 cases with *PKD1* CNVs, and 15 cases with no identified variants (Fig. 1). Genomic DNA from these 40 patients were sent to Berry Genomics Corporation for genetic analysis of *PKD1* and *PKD2* with a targeted LRS approach (Fig. 6). This study adhered to the Declaration of Helsinki and was approved by the institutional review board of Center for Reproductive Medicine at Shandong University. Informed written consent was obtained from all the participants.

ADPKD genetic diagnosis by SRS

Genomic DNA from peripheral blood was extracted using the Blood DNA extraction kits (ZEESAN, China). Targeted SRS, with IDT xGen Exome Research Panel (Integrated DNA Technologies, USA), was performed using xGen SRS Hybridization Capture (Integrated DNA Technologies, USA) and sequencing was performed on the NovaSeq 6000 platform (Illumina, USA). High-quality sequencing reads were selected and aligned to the reference genome hg19 using the BWA algorithm with default settings[52]. SNVs and indels were called by GATK[53]. Copy-number variations (CNVs) were determined as previously described[54]. For samples with SNVs/indels identified by SRS, Sanger sequencing was performed to confirm the variants. For samples with CNVs in *PKD1* and *PKD2* identified by SRS, MLPA analysis was performed to confirm the CNVs (MRC-Holland, the Netherlands) using probemix P351 and P352 *PKD1-PKD2*.

ADPKD genetic diagnosis by LRS

Targeted LRS for *PKD1* and *PKD2* genetic analysis was performed similarly as previously described[27]. Genomic DNA was subjected to multiplex LR-PCR in 50-μL reactions containing 10–100 ng of genomic DNA, 1 x PCR buffer for KOD FX Neo, 0.4 mM of each dNTP, 1 μM of primer mixture and 1 μL of KOD FX Neo (Toyobo, Japan). PCR cycling conditions were 94 °C for 5 min (1 cycle), 98 °C for 15 s and 68 °C for 12 min (32 cycles) and 68 °C for 10 min (1 cycle). The PacBio single-molecule real-time dumbbell (SMRTbell) libraries were prepared by a one-step end-repair and ligation reaction to add barcoded adapters, followed by digestion with exonucleases to remove failed ligated DNA. The uniquely barcoded libraries were purified, quantified and then pooled with equal mass. SMRTbell sequencing library was prepared using the Sequel II Binding Kit 3.2 (Pacific Biosciences, USA) and sequenced under circular consensus sequencing (CCS) mode with Sequel IIe platform (Pacific Biosciences, USA) for 30 h. After sequencing, the raw subreads were converted to high-fidelity CCS reads, debarcoded and aligned to reference genome build hg38 in the SMRT Link analysis software suite (version 10.1.0.119588, Pacific Biosciences, USA). The average yield of CCS reads per run was approximately two million. Up to 96 LRS libraries for ADPKD were pooled for sequencing in one run, and the average CCS read number was 20,000. The aligned CCS reads were then subjected to an in-house developed bioinformatics pipeline to identify SNVs/indels, deletions and duplications (Fig. 6A, B)[27]. The hg38-aligned CCS reads were then filtered to obtain reads with primer sequences used in the panel by blasting to remove misaligned reads[55]. For the study participants, the average coverage of *PKD1* was 1221x (range: 138–3920x), and the average coverage of *PKD2* was 922x (range: 83–5378x). For CCS reads with full-length fragments A-L, SNVs and indels of *PKD1/2* were called by FreeBayes1.3.4 with read depth ≥30 (Biomatters, Inc., San Diego, CA). For other CCS reads, kernel density estimation method was applied to identify peaks of reads with deletions or duplications. The exact breakpoints of deletions or duplications were determined by pysam and displayed in the Integrative Genomics Viewer (IGV)[56]. The CCS reads of representative variants were displayed in the Integrative Genomics Viewer (IGV). The

pathogenicity of variants identified by LRS was classified according to the American College of Medical Genetics and Genomics guidelines[51].

Validation of variants additionally detected by LRS

Sanger sequencing was performed to validate discordant SNVs/indels between LRS and SRS, with primers listed in Supplementary Table 1. Moreover, MLPA was utilized to confirm the CNVs additionally identified by LRS.

Identification of splicing junctions by RT-PCR sequencing

For participants with variants that could potentially disrupt mRNA splicing, total RNA was extracted from peripheral blood using RNAiso Plus (Takara, Japan). cDNA was synthesized using PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara, Japan) according to the manufacturer's instructions. PCR amplification was performed using primers flanking the putative splicing junctions, and products were analyzed by Sanger sequencing to identify the exact junction sites (Supplementary Table 1).

Data availability

All other data and materials are available from the corresponding authors upon reasonable request.

Code availability

PacBio ccs is available at https://github.com/PacificBiosciences/ccs. PacBio lima is available at https://github.com/PacificBiosciences/barcoding. PacBio pbmn2 is available at https://github.com/PacificBiosciences/pbmm2. FreeBayes1.3.4 is available at https://github.com/freebayes/freebayes. IGV v.2.17.4 is available at https://software.broadinstitute.org/software/igv/download.

Received: 9 October 2024; Accepted: 3 February 2025;
Published online: 11 March 2025

References

1. Harris, P. C. & Rossetti, S. Molecular diagnostics for autosomal dominant polycystic kidney disease. *Nat. Rev. Nephrol.* **6**, 197–206 (2010).
2. Lanktree, M. B., Haghighi, A., di Bari, I., Song, X. & Pei, Y. Insights into autosomal dominant polycystic kidney disease from genetic studies. *Clin. J. Am. Soc. Nephrol.* **16**, 790–799 (2021).
3. Barua, M. et al. Family history of renal disease severity predicts the mutated gene in ADPKD. *J. Am. Soc. Nephrol.* **20**, 1833–1838 (2009).
4. Pei, Y. et al. Unified criteria for ultrasonographic diagnosis of ADPKD. *J. Am. Soc. Nephrol.* **20**, 205–212 (2009).
5. Rossetti, S. et al. Comprehensive molecular diagnostics in autosomal dominant polycystic kidney disease. *J. Am. Soc. Nephrol.* **18**, 2143–2160 (2007).
6. Chaperon, J. L. et al. Preimplantation genetic testing for kidney disease-related genes: a laboratory's experience. *Am. J. Nephrol.* **52**, 684–690 (2021).
7. Bogdanova, N. et al. Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes. *Genomics* **74**, 333–341 (2001).
8. Loftus, B. J. et al. Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics* **60**, 295–308 (1999).
9. Bischof, J. M. et al. Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum. Mutat.* **27**, 545–552 (2006).
10. Watnick, T. J. et al. An unusual pattern of mutation in the duplicated portion of PKD1 is revealed by use of a novel strategy for mutation detection. *Hum. Mol. Genet.* **6**, 1473–1481 (1997).
11. Watnick, T. J., Gandolph, M. A., Weber, H., Neumann, H. P. & Germino, G. G. Gene conversion is a likely cause of mutation in PKD1. *Hum. Mol. Genet.* **7**, 1239–1243 (1998).
12. Gout, A. M., Martin, N. C., Brown, A. F. & Ravine, D. PKDB: Polycystic Kidney Disease Mutation Database—a gene variant database for autosomal dominant polycystic kidney disease. *Hum. Mutat.* **28**, 654–659 (2007).
13. Suzuki, Y. et al. Mutation analysis of autosomal-dominant polycystic kidney disease patients. *Genes* **14**, 443 (2023).
14. Ariyurek, Y. et al. Large deletions in the polycystic kidney disease 1 (PKD1) gene. *Hum. Mutat.* **23**, 99 (2004).
15. Consugar, M. B. et al. Characterization of large rearrangements in autosomal dominant polycystic kidney disease and the PKD1/TSC2 contiguous gene syndrome. *Kidney Int.* **74**, 1468–1479 (2008).
16. Wojcik, M. H. et al. Genome sequencing for diagnosing rare diseases. *N. Engl. J. Med.* **390**, 1985–1997 (2024).
17. Adams, D. R. & Eng, C. M. Next-generation sequencing to diagnose suspected genetic disorders. *N. Engl. J. Med.* **379**, 1353–1362 (2018).
18. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
19. Renkema, K. Y., Stokman, M. F., Giles, R. H. & Knoers, N. V. Next-generation sequencing for research and diagnostics in kidney disease. *Nat. Rev. Nephrol.* **10**, 433–444 (2014).
20. Song, X., Haghighi, A., Iliuta, I. A. & Pei, Y. Molecular diagnosis of autosomal dominant polycystic kidney disease. *Expert Rev. Mol. Diagn.* **17**, 885–895 (2017).
21. Hu, H. Y. et al. Comprehensive strategy improves the genetic diagnosis of different polycystic kidney diseases. *J. Cell. Mol. Med.* **25**, 6318–6332 (2021).
22. Ali, H. et al. PKD1 duplicated regions limit clinical utility of whole exome sequencing for genetic diagnosis of autosomal dominant polycystic kidney disease. *Sci. Rep.* **9**, 4141 (2019).
23. Claes, K. B. M., Rosseel, T. & De Leeneer, K. Dealing with pseudogenes in molecular diagnostics in the next generation sequencing era. *Methods Mol. Biol.* **2324**, 363–381 (2021).
24. Mochizuki, T. et al. Mutation analyses by next-generation sequencing and multiplex ligation-dependent probe amplification in Japanese autosomal dominant polycystic kidney disease patients. *Clin. Exp. Nephrol.* **23**, 1022–1030 (2019).
25. Tan, A. Y. et al. Molecular diagnosis of autosomal dominant polycystic kidney disease using next-generation sequencing. *J. Mol. Diagn.* **16**, 216–228 (2014).
26. Rossetti, S. et al. Identification of gene mutations in autosomal dominant polycystic kidney disease through targeted resequencing. *J. Am. Soc. Nephrol.* **23**, 915–933 (2012).
27. Xu, D. et al. Comprehensive analysis of PKD1 and PKD2 by long-read sequencing in autosomal dominant polycystic kidney disease. *Clin. Chem.* **70**, 841–854 (2024).
28. Rossetti, S. et al. A complete mutation screen of the ADPKD genes by DHPLC. *Kidney Int.* **61**, 1588–1599 (2002).
29. Huang, W. et al. Nanopore third-generation sequencing for comprehensive analysis of hemoglobinopathy variants. *Clin. Chem.* **69**, 1062–1071 (2023).
30. Liang, Q. et al. A more universal approach to comprehensive analysis of thalassemia alleles (CATSA). *J. Mol. Diagn.* **23**, 1195–1204 (2021).
31. Tantirukdham, N. et al. Long-read amplicon sequencing of the CYP21A2 in 48 Thai patients with steroid 21-hydroxylase deficiency. *J. Clin. Endocrinol. Metab.* **107**, 1939–1947 (2022).
32. Liu, Y. et al. Comprehensive analysis of congenital adrenal hyperplasia using long-read sequencing. *Clin. Chem.* **68**, 927–939 (2022).
33. Li, S. et al. Comprehensive analysis of spinal muscular atrophy: SMN1 copy number, intragenic mutation, and 2 + 0 carrier analysis by third-generation sequencing. *J. Mol. Diagn.* **24**, 1009–1020 (2022).
34. Li, S. et al. An effective and universal long-read sequencing-based approach for SMN1 2 + 0 carrier screening through family trio analysis. *Clin. Chem.* **69**, 1295–1306 (2023).

35. Liang, Q. et al. Comprehensive analysis of Fragile X Syndrome: full characterization of the FMR1 locus by long-read sequencing. *Clin. Chem.* **68**, 1529–1540 (2022).
36. Liu, Y. et al. Comprehensive analysis of hemophilia A (CAHEA): towards full characterization of the F8 gene variants by long-read sequencing. *Thromb. Haemost.* **123**, 1151–1164 (2023).
37. Durkie, M., Chong, J., Valluru, M. K., Harris, P. C. & Ong, A. C. M. Biallelic inheritance of hypomorphic PKD1 variants is highly prevalent in very early onset polycystic kidney disease. *Genet. Med.* **23**, 689–697 (2021).
38. Gilbert, R. D., Sukhtankar, P., Lachlan, K. & Fowler, D. J. Bilineal inheritance of PKD1 abnormalities mimicking autosomal recessive polycystic disease. *Pediatr. Nephrol.* **28**, 2217–2220 (2013).
39. Craven, K. E. et al. Optimizing insertion and deletion detection using next-generation sequencing in the clinical laboratory. *J. Mol. Diagn.* **24**, 1217–1231 (2022).
40. Shigemizu, D. et al. IMSindel: an accurate intermediate-size indel detection tool incorporating de novo assembly and gapped global-local alignment with split read analysis. *Sci. Rep.* **8**, 5608 (2018).
41. Zhang, C. et al. Identification of deep intronic variants of PAH in phenylketonuria using full-length gene sequencing. *Orphanet J. Rare Dis.* **18**, 128 (2023).
42. Luo, X. et al. Deep intronic PAH variants explain missing heritability in hyperphenylalaninemia. *J. Mol. Diagn.* **25**, 284–294 (2023).
43. Vaz-Drago, R., Custódio, N. & Carmo-Fonseca, M. Deep intronic mutations and human disease. *Hum. Genet.* **136**, 1093–1111 (2017).
44. Fromer, M. et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am. J. Hum. Genet.* **91**, 597–607 (2012).
45. Gambin, T. et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. *Nucleic Acids Res.* **45**, 1633–1648 (2017).
46. Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **16**, 172–183 (2015).
47. Cornec-Le Gall, E. et al. Type of PKD1 mutation influences renal outcome in ADPKD. *J. Am. Soc. Nephrol.* **24**, 1006–1013 (2013).
48. Chebib, F. T. & Torres, V. E. Autosomal dominant polycystic kidney disease: core curriculum 2016. *Am. J. Kidney Dis.* **67**, 792–810 (2016).
49. De Rycke, M. et al. ESHRE PGD Consortium data collection XIV-XV: cycles from January 2011 to December 2012 with pregnancy follow-up to October 2013. *Hum. Reprod.* **32**, 1974–1994 (2017).
50. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **21**, 597–614 (2020).
51. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
52. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
53. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
54. Shang, X. et al. Rapid targeted next-generation sequencing platform for molecular screening and clinical genotyping in subjects with hemoglobinopathies. *EBioMedicine* **23**, 150–159 (2017).
55. Ye, J., McGinnis, S. & Madden, T. L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res.* **34**, W6–W9 (2006).
56. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

Acknowledgements

This study was supported by the National Key Research and Development Program of China (2022YFC2703200), Guangxi Key Research and Development Plan Project (No. AB22035080), Guangxi Science and Technology Base and Talent Project (No. AC22080002), National Key Research and Development Program of China (2021YFC2700600, 2021YFC2700500, 2023YFC2705500), Key Construction Project of Shandong (2023ZLGX02), and Medical and Health Development Program of Shandong Province (202201030242). We thank all the patients and their families for their participation in this study.

Author contributions

Y.G., and X.G. involved in the study concept and design. Z-J. Chen directed the study progress. A.M. and L.C. acquired the data. Q.S., P.X., S.H. and J.L. analyzed and interpreted the data. H.K. and J.H. performed the experiments for variant confirmation. W.J. performed the bioinformatics analysis. J.Y., Jing.L. and D.S. recruited the participants. Q.S. and P.X. drafted the original manuscript. All authors reviewed and approved the final article.

Competing interests

A.M. and L.C. are employees of Berry Genomics Corporation. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-025-00477-5.

Correspondence and requests for materials should be addressed to Xuan Gao or Yuan Gao.