Human Genomics

BRIEF REPORT Open Access

Minigene-based splice assays provide new insights on intronic variants of the *PKHD1* gene

Yiyin Zhang[1], Ran Zhang[1], Xiaomeng Shi[1], Xuyan Liu[1], Changying Li[1], Yan Zhang[2], Zhi Wang[2], Dan Qiao[3], Fengjiao Pan[1], Bingying Zhang[2], Ning Xu[1], Bingzi Dong[4*] and Leping Shao[5*]

Abstract

Background Autosomal Recessive Polycystic Kidney Disease (ARPKD) is a rare hereditary disorder caused by variants in *PKHD1*. Currently, aberrant splicing has been reported to play important roles in genetic disease. Our goal is to analyze intronic variants in *PKHD1* at the mRNA level.

Results The 12 candidate variants were introduced into the corresponding minigene and functionally assayed in HEK 293T and Hela cells. We identified 11 variants that induce splicing alterations, resulting in various consequences such as skipping of exons, intron retention and protein truncation.

Conclusions This underlined the importance of mRNA-level assessment for genetic diagnostics in related genetic disorders.

Keywords *PKHD1* gene, Intronic variant, Minigene, pre-mRNA splicing

Introduction

Autosomal Recessive Polycystic Kidney Disease (ARPKD, MIM#263200) is a rare genetic disorder characterized by bilateral enlargement of cystic kidneys and liver fibrosis. Cysts form predominantly in the collecting tubules, eventually leading to dilatation of the collecting tubules. Compared with Autosomal Dominant Polycystic Kidney Disease, ARPKD is much rarer, but bilateral enlarged kidneys, impaired lung formation and pulmonary hypoplasia frequently result in fetal and neonatal deaths [1]. ARPKD is caused by variants in the *PKHD1* gene, the longest open reading frame of which contains 67 exons and encodes a 4074 amino acid protein called fibrocystin/polyductin (FPC/PD1, Figure S1) [1]. To date, a total of 4,429 single nucleotide variants (SNVs) have been annotated in the ClinVar database for the *PKHD1* gene (accessed August 22, 2024).

*Correspondence:
Bingzi Dong
dongbingzi@hotmail.com
Leping Shao
lepingshao@163.com
[1]Department of Nephrology, The Affiliated Municipal Hospital of Qingdao University, No.5 Donghai Middle Road, Qingdao 266071, China
[2]School of Clinical Medicine, Shandong Second Medical University, Weifang 261000, China
[3]Department of Nephrology, Dalian Medical University, Dalian 116000, China
[4]Department of Endocrinology, The Affiliated Hospital of Qingdao University, No.16 Jangsu Road, Qingdao 266003, China
[5]Department of Nephrology, The First Affiliated Hospital of Xiamen University, Xiamen, China

Splicing is the removal of introns from a pre-mRNA transcript and the joining of exons, which involves the identification of splice sites and two successive transesterification reactions [2]. Multiple motifs are crucial during splicing process, including consensus donor and acceptor splice sites, the branchpoint, polypyrimidine tract, splicing enhancers, and silencers [3]. Sequence variations in genomic DNA that cause incorrect splicing and consequently pathogenicity are becoming more common in genetic disorders [4]. The majority of variant splicing analyses of the *PKHD1* gene concentrated on missense variants, nonsense variants, and variants located in classical splice sites [5, 6]. Consequently, intronic variants are focus in this study. Due to the lack of patient blood RNA samples, minigene splicing assay was helpful to argue the pathogenicity of intronic variants. Our goal was to explore the effect of intronic variation on splicing with minigene assays that have previously proved persuasive [7].

Materials and methods

DNA variant numbering was based on the cDNA sequence for *PKHD1* (RefSeq NM_138694.4). *PKHD1* variants were available from the ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). Bioinformatics analysis with the use of Splice AI, BDGP, Human Splicing Finder, MaxEntScan and CADD were performed to predict the gain and loss of donor or acceptor sites. Variants that met all of the following criteria were included in this study: (1) located at the intronic positions −3 to -40 and +3 to +6; (2) SpliceAI Δscore>0.2; (3) The BDGP results indicate a significant decrease in canonical splice site scores(Δscore>0.5) or a generation of novel alternative splice sites. Please refer to the Figure S2 for details on location selection.

Detailed descriptions of minigene construction, cell culture, plasmid transfection, and analysis of minigenes are provided in Data S1.

Results

A total of 661 intronic point variants in the *PKHD1* gene could be retrieved at Clinvar database (accessed August 2023). Twelve candidate variants were finally included in this study after bioinformatic filtering. See Table S1 for details of the screening and selection process. Ultimately, eleven out of the 12 selected variants induced anomalous splicing in minigene assays, where exon skipping was the most common event. Furthermore, we conducted a predictive analysis of the functional consequences of abnormal splicing caused by SNVs, and details are listed in Table 1.

Variants c.53-5T>A, c.5751+3 A>G and c.7109+3 A>C induced partial skipping of exons

These three variants exhibited normal splicing and exon skipping. Variant c.53-5T>A induced weak skipping of exon 3(11.64% in HEK293T cells and 3.41% in Hela cells) (Table 1; Fig. 1A). The predicted protein would lose 26 amino acids, which involved a part of the TIG1 domain.

Variants c.5751+3 A>G and c.7109+3 A>C resulted in significant skipping of respectively exon 35 (Fig. 1F) and exon 44 (Fig. 1H), which would result in truncated proteins (p.Ser1867Argfs*57 and p.Gly2333Valfs*8) with TIG domains and no transmembrane and cytoplasmic domains.

Variants c.2716–3 C>G and c.8302+5G>A result in complete skipping of exons

Variants c.2716–3 C>G and c.8302+5G>A resulted in complete skipping of respectively exon 26 (Fig. 1E) and exon 52 (Fig. 1I). The skipping of exon26 would lead to protein truncation (p.Val906Metfs*6). The skipping of exon52 would lead to in-frame deletions (p.Ala2725_Asn2768delinsAsp). As a result, mutated protein would lack part of G8 2 domain.

Variants c.1694–10 C>G and c.8441–5 A>G resulted in retention of introns

The c.1694–10 C>G construct produced a larger transcript of 415 bp including SD6, exon19, SA and inserted 9 bp sequences (Fig. 1C). As a result, 3 amino acids are inserted into the protein sequence after amino acid 564 (p.Arg564_Gly565insValHisPro).

We observed a 4 bp insertion in variant c.8441–5 A>G (c.8440_8441ins ACAG, Fig. 1J). The resulting protein sequence is therefore expected to be truncated (p.Gly2814Aspfs*19).

Variants c.281+3 A>G generated three transcripts

Figure 1B demonstrates that mutant minigene(c.281+3 A>G) generated two different bands. Direct sequencing showed that the larger amplicons had an extra 2 bp downstream of exon 4(c.281_282insGT), while the smaller amplicons included only 263 bp. Notably, there were overlapping peaks in the peak diagrams of the larger amplicons. We performed TA cloning and isolated another product that exhibited a splicing mode consistent with the wild type. As the routine agarose electrophoresis gel does not separate DNA fragments with small differences, the capillary electrophoresis was used for quantitative analysis (Figure S5). The predicted mutant proteins (p.Ser95Tyrfs*23 and p.Gly44Aspfs*23) only retained the part of the TIG1 domain.

Table 1 Splicing outcomes and predicted protein alterations of variants

Variants	Splicing Outcomes	the proportion of each transcript		Protein(Predicted)	Protein effect(Predicted)	Clinvar
		HEK293T	Hela			
c.53-5T > A	CT	88.36%	96.59%	No changes	No changes	LB (1)
	Ex3 skipping(c.53_130del)	11.64%	3.41%	p.Val18_Asp43del	Absence of 26 amino acids	
c.281 + 3 A > G	IVS4-insGT(c.281_282insGT)	36.85%	35.93%	p.Ser95Tyrfs*23	Truncated protein	VUS (2); LB (3)
	CT	28.37%	30.15%	No changes	No changes	
	Ex4 skipping(c.131_281del)	34.78%	33.92%	p.Gly44Aspfs*23	Truncated protein	
c.1694–10 C > G	IVS18-ins9(c.1693_1694ins TGCA TCCAG)	100.00%	100.00%	p.Arg564_Gly565insValHisPro	Insertion of 3 amino acids	VUS (1); LB (1)
c.2716–3 C > G	Ex26 skipping(c.2716_2821del)	100.00%	100.00%	p.Val906Metfs*6	Truncated protein	VUS (1)
c.5751 + 3 A > G	CT	6.88%	9.79%	No changes	No changes	P (4)
	Ex35 skipping(c.5601_5751del)	93.12%	90.21%	p.Ser1867Argfs*57	Truncated protein	
c.6332 + 3 A > C	Ex38-del76(c.6257_6332del)	22.08%	33.84%	p.Gly2086Aspfs*18	Truncated protein	VUS (1)
	Ex38 skipping(c.6122_6332del)	77.92%	66.16%	p.Gly2041Aspfs*18	Truncated protein	
c.7109 + 3 A > C	CT	15.56%	12.00%	No changes	No changes	LP (2)
	Ex44 skipping(c.6997_7109del)	84.44%	88.00%	p.Gly2333Valfs*8	Truncated protein	
c.8302 + 5G > A	Ex52 skipping(c.8174_8302del)	100%	100%	p.Ala2725_Asn2768delinsAsp	Absence of 44 amino acids and aspartate insertion	VUS (1)
c.8441–5 A > G	IVS53-ins4(c.8440_8441ins ACAG)	100%	100%	p. Gly2814Aspfs*19	Truncated protein	VUS (1)
c.11175-10T > G	IVS61-ins9(c.11174_1175ins ATC TGTCAG)	63.74%	63.51%	p.Gly3725_Asn3726insSerValArg	Insertion of 3 amino acids	VUS (1); LB (1)
	IVS61-ins9 + IVS62-ins14 (c.11174_11175ins ATCTGTCA G + c.11310_11311ins GTAACTTT TCACAG)	36.26%	36.49%	p.Gly3725_Asn3726insSerValArg and p.Asn3771Valfs*9	Truncated protein	
c.11310 + 3 A > G	IVS62-ins14(c.11310_11311ins GT GACTTTTCACAG)	100%	100%	p.Asn3771Valfs*9	Truncated protein	VUS (1)

Note: CT: Canonical transcript

ClinVar last accessed March 2024. LB (Likely Benign), VUS (variant of uncertain significance), LP (Likely Pathogenic), P (Pathogenic). In brackets, N submitters supporting each classification

Variant c.6332 + 3 A > C generated two transcripts

Variant c.6332 + 3 A > C (Fig. 1G) induced a dual out-of-frame effect; we observed both a deletion of the entire exon 38 (−211 nt) as well as a band representing the deletion of part of exon 38 (−76 nt). The predicted protein (p.Gly2086Aspfs*18, Gly2041Aspfs*18) retains only the TIG domains.

Variants c.11175-10T > G and c.11310 + 3 A > G exhibit complex splicing patterns

BDGP demonstrated that there is an alternative splice site (agGTaaca, score:0.83) downstream of the donor splice site of exon 62(score:0.89). The recombinant WT plasmid produces two types of transcription products: one is the normal spliced transcript including exon 62 with a size of 399 bp, another was 413 bp containing exon 62 and 14 bp intron sequence. As predicted, the variant c.11175-10T > G created a new acceptor splice site upstream of exon 62, resulting in 9 bp intron inclusion on the base of two WT products. Besides, c.11310 + 3 A > G was demonstrated to reduce the score of the donor site of exon 62 from 0.89 to 0.35 by BDGP, and minigene analysis of the variant c.11310 + 3 A > G resulted in only one product that matched in size with 413 bp transcript generated by the WT minigene (Fig. 1K, Figure S5).

Variant c.2593-39G > T did not alter pre-mRNA splicing

We did not find any differences in transcript processing between wild-type and mutant constructions (Fig. 1D) for variant c.2593-39G > T.

Discussion

The next-generation sequencing technology is revealing an increasing number of variations, and bioinformatics predictions and functional studies are crucial for the accurate classification of these variants. Here, we focused on potentially spliceogenic *PKHD1* intronic variants reported in the ClinVar database, and checked 12 variants to dissect splicing effects using pSPL3 vector in HEK293T and Hela cells. Remarkably, a considerable percentage of tested variants (11/12, 91.67%) impaired splicing, underlining the specificity of our selection criteria. Variants presented similar splicing patterns and slightly different proportions of transcripts in two cell

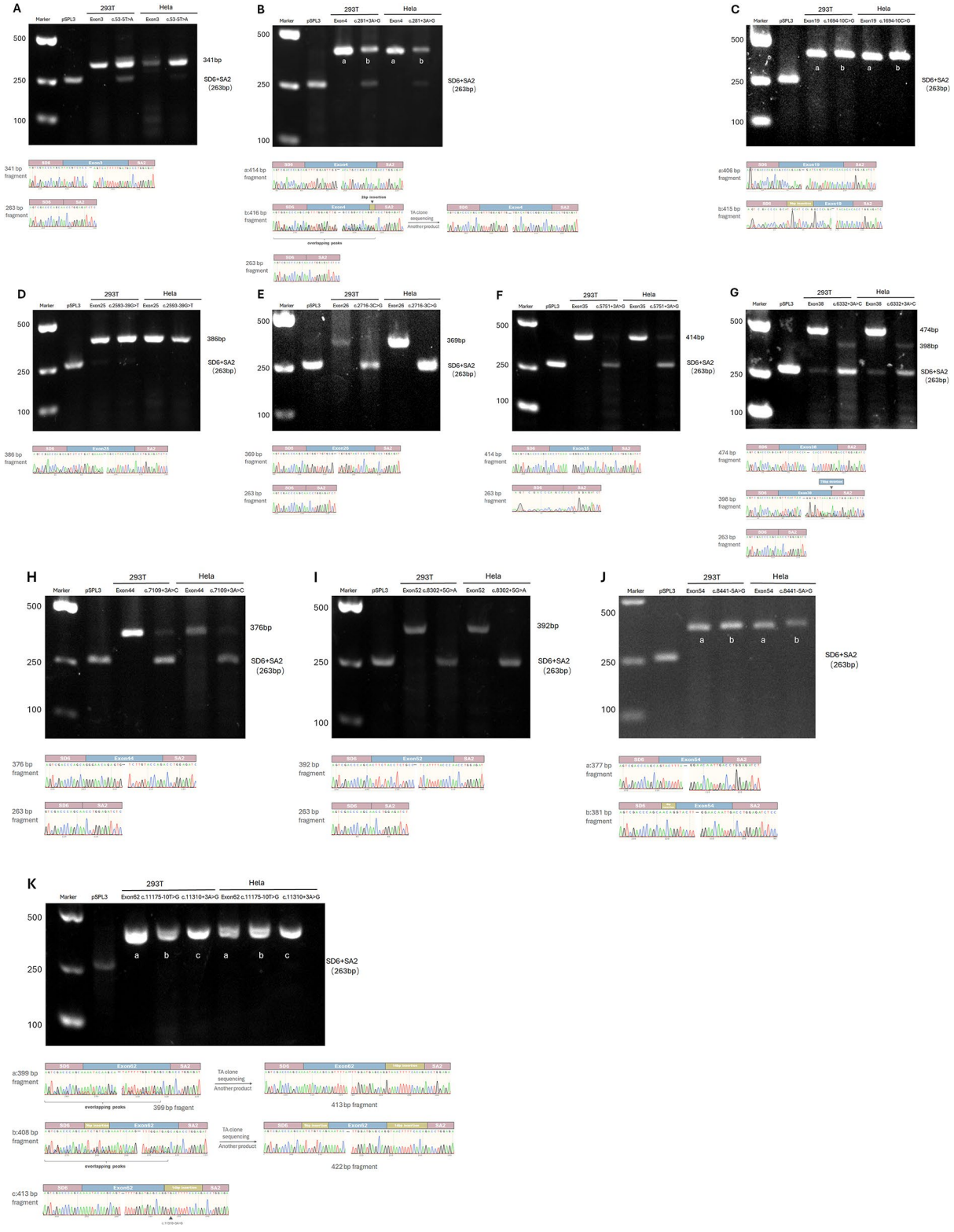

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Agarose gel electrophoresis and sanger sequences of RT-PCR products in HEK 293T and Hela cells. All gel images were representative results of three independent experiments. (**A**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex3 (341 bp); Lane 4 and 6: c.53-5T > A (341 bp and 263 bp). (**B**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex4 (414 bp); Lane 4 and 6: c.281 + 3 A > G (414 bp, 416 bp and 263 bp). (**C**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex19 (406 bp); Lane 4 and 6: c.1694–10 C > G (415 bp). (**D**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex25 (386 bp); Lane 4 and 6: c.2593-39G > T (386 bp). (**E**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex26 (369 bp); Lane 4 and 6: c.2716–3 C > G (263 bp). (**F**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex35 (414 bp); Lane 4 and 6: c.5751 + 3 A > G (414 bp and 263 bp). (**G**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex38 (474 bp and 263 bp); Lane 4 and 6: c.6332 + 3 A > C (398 bp and 263 bp). (**H**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex44 (376 bp); Lane 4 and 6: c.7109 + 3 A > C (376 bp and 263 bp). (**I**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex52 (376 bp); Lane 4 and 6: c.8302 + 5G > A (263 bp).(**J**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 5: PSPL3 Ex54 (377 bp); Lane 4 and 6: c.8441–5 A > G (381 bp). (**K**) Lane 1: marker; lane 2: PSPL3 (263 bp); Lane 3 and 6: PSPL3 Ex62 (399 bp and 413 bp); Lane 4 and 7: c.11175-10T > G (408 bp and 422 bp); Lane 5 and 8: c.11310 + 3 A > G (413 bp). The sequencing plots of the mRNA products that resulted from transfection is visualized below the electropherogram

lines, which may be attributed to differences in splicing factor expression.

According to the prediction software and their location, eight variants disrupted the canonical 3' and 5' splice sites (c.53-5T > A, c.281 + 3 A > G, c.2716–3 C > G, c.5751 + 3 A > G, c.6332 + 3 A > C, c.7109 + 3 A > C, c.8302 + 5G > A, c.11310 + 3 A > G), five created novel active splice sites(c.281 + 3 A > G, c.1694–10 C > G, c.6332 + 3 A > C, c.8441–5 A > G, c.11175-10T > G). Notably, the wild-type (WT) minigene of Exon 62 generated a transcript of 413 bp (IVS62-ins14) coexist with the canonical transcript which might reflect a complex alternative splicing mechanism of *PKHD1*.

Functional studies of variants provide key information for their clinical interpretation. FPC undergoes a complicated pattern of Notch-like proteolytic processing. The latest research revealed that C-terminal cleavage product ICD15 translocate to mitochondria, thereby enhancing mitochondrial respiration in renal epithelial cells [8]. In this research, the predicted FPC encoded by aberrant transcripts resulting from eight variants of c.281 + 3A > G, c.2716-3C > G, c.5751 + 3A > G, c.6332 + 3A > C, c.7109 + 3A > C, c.8441-5A > G, c.11175-10T > G and c.11310 + 3A > G would be truncated at different position of extracellular regions, resulting in the loss of transmembrane domains and cytoplasmic domains. These predicted proteins have lost their basic structures and are unable to recognize signals and undergo cleavage and processing, ultimately impairing mitochondrial function and leading to a severe phenotype. Qiu LR had reported that the patients with compound heterozygous mutations of c.5751 + 3A > G and EX6_11DEL can present with renal cysts and reduced kidney function [9], which is consistent with our results. Although the remaining seven variants were listed as 'LP' or 'VUS' in the Clinvar database, our findings indicate that they are 'detrimental' variants. Moreover, the surveillance systems may recognize premature termination codons and recruit enzymes to degrade the aberrant mRNAs, which requires further study.

Variant c.53-5T > A causes a partial loss of IPT1 domain which may bind to undetermined ligands, modulating oligomerization of the protein and mediating adhesive interaction [10]. Consistent with "LB" presented in Clinvar database, the abnormal transcripts accounted for only a small proportion, which may be associated with a milder phenotype. Unlike the "VUS" presented in Clinvar database, the FPC caused by c.8302 + 5G > A would loss the part of G8 domains. It is speculated that the G8 domains may be related to scaffolding, cell signaling and adhesion. Thus, the above function of the mutant protein may be weakened. While it is true that variant c.1694–10 C > G will impact splicing, the resulting alteration is not necessarily pathogenic. Insertions (c.1693_1694insTGCATCCAG) are multiples of three base pairs (9 bp) and thus do not cause frameshifting.

The results predicted by SpliceAI and BDGP are basically consistent with minigene results, except for the variant c.2593-39G > T. BDGP demonstrated that there is an alternative splice site (tgacAGaatt, score:0.13) upstream of the acceptor splice site of exon 25(score:0.24), and the variants c.2593-39G > T increased the score of the alternative site from 0.13 to 0.51. However, the results of minigene assays were negative, which may be attributed to limitations of bioinformatics tools and the differential sequence context in regulating splicing.

It is worth noting that the true splicing effect of these variants needs to be confirmed by analyzing the mRNA from the patient. In addition, the functional activities and the cell surface expression of these mutant FPCs need further research.

In summary, our studies were very helpful in reclassification according to ACMG criteria and developing novel targeted gene therapy. Antisense oligonucleotide (ASO) therapy brings new hope for the treatment of ARPKD, which has made pleasing progress in neurological disorders. A research group has successfully designed three ASOs targeting the *PKHD1* gene variant c.2141-3T > C, and they confirmed that the ASOs can block the abnormal pseudo exon insertion [6]. For this study, it is expected that ASO will be designed to block pseudo-exon insertion associated with variant c.11310 + 3 A > G. Though beyond the scope of this study, future efforts are needed in this direction.

Conclusions

We have carried out a comprehensive analysis of intronic variants in *PKHD1* using bioinformatics tools and minigene assays. Eleven variants (c.53-5T>A, c.281+3 A>G, c.1694–10 C>G, c.2716–3 C>G, c.5751+3 A>G, c.6332+3 A>C, c.7109+3 A>C, c.8302+5G>A, c.8441–5 A>G, c.11175-10T>G, c.11310+3 A>G) were confirmed to affect pre-mRNA splicing. These findings highlight the importance to assess the effects of single nucleotide variants at the mRNA level. Moreover, the pSPL3-based minigenes are effective tools for molecular diagnostics and genetic counseling of ARPKD or other genetic disorders as well as for the basic research on the splicing process.

Abbreviations

ARPKD	Autosomal recessive polycystic kidney disease
BDGP	Berkeley Drosophila Genome Project
HEK	293T Human epithelial kidney 293T
mRNA	Messenger RNA

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40246-024-00675-9.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We thank all the subjects for their participation.

Author contributions

Leping Shao and Yiyin Zhang conceived and designed and the experiments. Ran Zhang supervised this work. Yiyin Zhang, Yan Zhang, Dan Qiao, Fengjiao Pan, Bingying Zhang and Ning Xu conducted experiments. Xuyan Liu, Zhi Wang conducted in silico prediction of variants. Ran Zhang, Xiaomeng Shi, Changying Li contributed to the data analysis. Yiyin Zhang wrote the manuscript. Bingzi Dong and Leping Shao revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No. 821707171).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 3 May 2024 / Accepted: 18 September 2024
Published online: 09 November 2024

References

1. Wilson PD. Polycystic kidney disease. N Engl J Med. 2004;350:151–64.
2. Rogalska ME, Vivori C, Valcárcel J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. Nat Rev Genet. 2023;24:251–69.
3. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. J Appl Genet. 2018;59:253–68.
4. Ramanagoudr-Bhojappa R, Carrington B, Ramaswami M, Bishop K, Robbins GM, Jones M, et al. Multiplexed CRISPR/Cas9-mediated knockout of 19 fanconi anemia pathway genes in zebrafish revealed their roles in growth, sexual development and fertility. PLoS Genet. 2018;14:e1007821.
5. Ishiko S, Morisada N, Kondo A, Nagai S, Aoto Y, Okada E, et al. Clinical features of autosomal recessive polycystic kidney disease in the Japanese population and analysis of splicing in PKHD1 gene for determination of phenotypes. Clin Exp Nephrol. 2022;26:140–53.
6. Li H, Wang C, Che R, Zheng B, Zhou W, Huang S, et al. A potential therapy using antisense oligonucleotides to treat autosomal recessive polycystic kidney disease. JCM. 2023;12:1428.
7. Liu X, Shi X, Xin Q, Liu Z, Pan F, Qiao D et al. Identified eleven exon variants in PKD1 and PKD2 genes that altered RNA splicing by minigene assay. BMC Genomics. 2023;24:407.
8. Walker RV, Yao Q, Xu H, Maranto A, Swaney KF, Ramachandran S, et al. Fibrocystin/Polyductin releases a C-terminal fragment that translocates into mitochondria and suppresses cystogenesis. Nat Commun. 2023;14:6513.
9. Qiu L, Xu R, Tang J, Zhou J. Possible PKHD1 hot-spot mutations related to early kidney function failure or hepatofibrosis in Chinese children with ARPKD: a retrospective single center cohort study and literature review. CURR MED SCI. 2020;40:835–44.
10. Bannell TAK, Cockburn JJB. The molecular structure and function of fibrocystin, the key gene product implicated in autosomal recessive polycystic kidney disease (ARPKD). Ann Hum Genet. 2024;88:58–75.

Publisher's note