

# mRNA Analysis Identifies deep Intronic Splicing Variants Leading to Alport Syndrome and Overcomes the Problem of Negative Results of Exome Sequencing

Xiaoyuan Wang (✉ [prettywxy@163.com](mailto:prettywxy@163.com))

Peking University First Hospital

Yanqin Zhang

Peking University First Hospital

Jie Ding

Peking University First Hospital

Fang Wang

Peking University First Hospital

---

## Research Article

**Keywords:** Alport syndrome, next-generation sequencing, diagnosis, urine, deep intronic variant

**Posted Date:** January 21st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-150525/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

Mutations in COL4A3, COL4A4 and COL4A5 genes lead to Alport syndrome (AS). However, pathogenic variants in some AS patients are not detected by exome sequencing. The aim of this study was to identify the underlying genetic causes of five unrelated AS probands with negative NGS test results. Urine COL4A3–5 mRNAs were analyzed in the probands with an uncertain inherited mode of AS, and COL4A5 mRNA of skin fibroblasts was analyzed in the probands with X-linked AS. RT-PCR and direct sequencing were performed to detect mRNA abnormalities. PCR and direct sequencing were used to analyze the exons with flanking intronic sequences corresponding to mRNA abnormalities. Nine novel deep intronic splicing variants in COL4A4 and COL4A5 genes that cannot be captured by exome sequencing were identified in the five AS probands. Skipping of an exon was caused by four intronic variants, and retention of an intron fragment led to the remaining variant. Our results reveal that mRNA analysis for AS genes from either urine or skin fibroblasts can resolve genetic diagnosis in AS patients with negative NGS results. We recommend analyzing COL4A3–5 mRNA from urine as the first choice for these patients because it is feasible and non-invasive.

## Introduction

Alport syndrome (AS) is a hereditary nephritis characterized by hematuria, proteinuria, and progressive renal failure and is sometimes accompanied by sensorineural deafness and ocular abnormal<sup>1</sup>. The three genetic forms of Alport syndrome depend on the mode of inheritance: X-linked AS (XLAS), autosomal recessive AS (ARAS), and autosomal dominant AS (ADAS)<sup>2,3</sup>. XLAS is caused by pathogenic variants in the *COL4A5* gene, while ARAS and ADAS are caused by pathogenic variants in the *COL4A3* or *COL4A4* gene<sup>4</sup>. Pathogenic variants in *COL4A3*, *COL4A4* or *COL4A5* genes lead to abnormal  $\alpha 3$ ,  $\alpha 4$  or  $\alpha 5$  chains of type IV collagen in the glomerular basement membrane (GBM). The gold standard for clinical diagnosis of AS is the characteristic changes in GBM, including irregular thickening, splitting and “basket-weave” changes, seen under an electron microscope<sup>5</sup>. Genetic testing for pathogenic variants in *COL4A3*, *COL4A4*, or *COL4A5* genes is currently used frequently in the diagnosis of AS because of the increasing availability of next-generation sequencing (NGS), including targeted NGS and whole exome sequencing, in the clinic<sup>6</sup>. NGS detects approximately 82–86% of pathogenic *COL4A3–5* gene variants<sup>6,7</sup>. However, some genetic changes that cause AS, such as deep intronic splicing variants, somatic mosaicism, and copy number variants, are not detectable by NGS<sup>8–10</sup>.

mRNA sequencing is an effective method to identify intronic splicing variants. Several *COL4A5* gene deep intronic splicing variants have been reported in studies that analyzed mRNA from skin fibroblasts<sup>11</sup>, peripheral blood lymphocytes<sup>12</sup>, hair root<sup>13</sup>, or renal tissue<sup>14</sup>. In our clinical practice, since 2000, simultaneous examinations of  $\alpha 5(\text{IV})$  staining in skin and *COL4A5* mutation screening using mRNA extracted from cultured skin fibroblasts have been routinely performed in patients with suspected XLAS. However, this approach cannot be applied for Alport syndrome patients with autosomal inherited patterns, since  $\alpha 3(\text{IV})$  and  $\alpha 4(\text{IV})$  are not expressed in skin. Two deep intronic variants in the *COL4A3* gene were identified by analysis of mRNA from blood or urine<sup>15</sup>. However, to our knowledge, the value of detecting mutations in *COL4A5* and *COL4A4* genes in mRNA isolated from urine has not been adequately studied.

The aim of this study was to identify the genetic etiologies of five unrelated AS patients with negative NGS results. We used our developed approach for analysis of the entire coding regions of *COL4A3*, *COL4A4*, and *COL4A5* mRNAs isolated from urine and *COL4A5* mRNA extracted from cultured skin fibroblasts and identified deep intronic splicing variants in the enrolled patients. These findings indicate that our developed approach may help guide medical practitioners and genetic counselors to provide personalized management of AS.

## Results

### Clinical features of AS patients

Five unrelated Alport syndrome probands were enrolled according to the inclusion and exclusion criteria listed in the Material and Methods. Patient clinical information and pedigrees are shown in Table 1 and Fig. 1, respectively. Patient 1 was diagnosed with AS based on characteristic AS features in GBM; the inheritance pattern was uncertain because of a negative family history and normal staining of  $\alpha 5(\text{IV})$  chain in skin tissue. Patient 2 was highly suspected of having XLAS based on a positive family history of end stage renal disease; however, normal staining of  $\alpha 5(\text{IV})$  chain in skin tissue did not support the diagnosis. XLAS was diagnosed in patients 3–5 with abnormal staining of  $\alpha 5(\text{IV})$  chain in skin specimens.

Table 1  
Clinical features and analyzed samples of the 5 probands in this study.

Proband	Gender	Age of onset (y)	Hematuria	Proteinuria	serum creatinine* (umol/L)	family history**	Renal EM	$\alpha 5(\text{IV})$ in skin EBM	Mode of inheritance	Skin mRNA	Urine mRNA	Blood DNA
1	F	25	+	+	91	-	AS	positive	uncertain	-	+	+
2	F	30	+	-	NA	+	ND	positive	uncertain	-	+	+
3	M	3	+	-	30.9	+	ND	negative	X-Link	+	-	+
4	M	6.8	+	+	78	-	AS	negative	X-Link	+	-	+
5	F	33	+	-	NA	+	ND	segmental positive	X-Link	+	-	+

\* Age at which serum creatinine tests were carried out in the three probands was 25, 10 and 14 years, respectively; \*\* Positive family history of hematuria and/or end stage renal disease. EM: electron microscopy; EBM: epithelial basement membrane; NA: not available; ND: not done. Patients 1 and 4 did not undergo staining of type IV collagen  $\alpha 5$  chain in their renal specimen; hearing loss and ocular changes were not detected in five patients.

## Gene mutations in the five probands

In proband 1, cDNA analysis showed that no abnormal transcripts were detected in *COL4A3* and *COL4A5* mRNAs isolated from proband 1's urine (Suppl.1 and 2). However, agarose gel electrophoresis revealed a smaller *COL4A4* mRNA transcript (fragment 1; Fig. 2A) in proband 1's urine. Sequencing of 10 RT-PCR products revealed a heterozygous skipping of exons 3 and 26 (r.72\_r113del, p.Trp24\*; r.1804\_r1987del, p.Gly602Valfs\*8; Fig. 2B and D) and an insertion of a 109 bp sequence of intron 23 between exons 23 and exon 24 (r.1623\_r.1624 ins [1623 + 590\_1623 + 698], p.Gly542Alafs\*29; Fig. 2C). To further confirm the anomalies detected at the cDNA level, *COL4A4* exons 2–4, 23, and 25–27 with the sequences of flanking introns 2–3, 23, and 25–26 were amplified by PCR from genomic DNA of proband 1 and her parents. Sequence analysis demonstrated that proband 1 and her mother were heterozygous for the variants in intron 2 c.72 – 26\_72-23delTAAT, intron 23 c.1623 + 570A > G, and intron 25 c.1804–158A > G (Fig. 2E, F and H). Proband 1 and her father were heterozygous for the variant in intron 23 c.1623 + 702T > A (Fig. 2G). Therefore, these results confirmed that the variants located in introns 2 and 25 resulted in aberrant splicing, causing skipping of exons 3 and 26, respectively. To determine the impact of the two variants in intron 23 on splicing, *COL4A4* mRNA from urine of proband 1's father was analyzed (urine from proband 1's mother was not available). Sequencing of the fragment 4 RT-PCR product showed that the father also had a heterozygous insertion of a 109 bp sequence of intron 23 between exons 23 and exon 24 (r.1623\_r.1624 ins [1623 + 590\_1623 + 698], p.Gly542Alafs\*29) (data not shown), which indicated that the variant c.1623 + 702T > A located in intron 23 was pathogenic.

In proband 2, cDNA analysis showed that no abnormal transcripts were detected in *COL4A3* and *COL4A4* mRNAs isolated from proband 2's urine (Suppl.3 and 4). Agarose gel electrophoresis revealed a smaller *COL4A5* mRNA transcript (fragment 5, Fig. 3A) in proband 2's urine. Sequencing of 10 RT-PCR products revealed that exon 32 of *COL4A5* gene was skipped heterozygously (r.2678\_r2767del), which led to an in-frame deletion (p.Thr894\_Gly923del) (Fig. 3B). *COL4A5* exons 31–33 with the sequences of the flanking introns 31–32 were amplified by PCR from genomic DNA for proband 2, her husband, and her daughter and then sequenced. Proband 2 and her daughter were heterozygous for the variant intron 31 c.2677 + 487C > A and c.2677 + 646C > T (Fig. 3C and D). Neither of the two variants were identified in her husband.

In proband 3, agarose gel electrophoresis of RT-PCR fragment 2 products of *COL4A5* mRNA from skin fibroblasts showed an abnormal transcript in addition to the wild-type transcript (Fig. 4A). Sequencing of 10 RT-PCR products revealed that a 128 bp sequence from intron 10 was inserted between exon 10 and exon 11 (r.609\_r.610 ins[609 + 751\_609 + 878]) (Fig. 4B), which led to premature termination of  $\alpha 5(\text{IV})$  chain (p.Gly204Valfs\*7). *COL4A5* intron 10 was amplified by PCR from genomic DNA in proband 3 and his mother. Sequencing revealed an A to G change in intron 10 at 879 bp downstream from exon 10 (IVS10 c.609 + 879 A > G) (Fig. 4C) in proband 3 and his mother.

In proband 4, agarose gel electrophoresis revealed a bigger *COL4A5* mRNA transcript (fragment 5, Fig. 4D) in proband 4's skin fibroblasts. Sequencing of 10 RT-PCR products revealed a 53 bp sequence from intron 29 of *COL4A5* gene in *COL4A5* mRNA from skin fibroblasts inserted between exons 29 and exon 30 (r.2395\_r.2396 ins [2395 + 1308\_2395 + 1360]) (Fig. 4E), which led to premature termination of  $\alpha 5(\text{IV})$  chain (p.Gly799Alafs\*15). *COL4A5* intron 29 was amplified by PCR from genomic DNA in proband 4 and his mother. A hemizygous C to G variant in intron 29 at 1275 bp and a G to T variant in intron 29 at 1292 bp downstream from exon 29 (IVS29 c.2395 + 1275C > G and c.2395 + 1292G > T) were found in genomic DNA of proband 4 (Fig. 4F). Neither of the two variants were identified in his mother.

In proband 5, agarose gel electrophoresis revealed a bigger *COL4A5* mRNA transcript (fragment 1, Fig. 4G) in proband 5's skin fibroblasts (Fig. 4H). Sequencing of 10 RT-PCR products revealed a 58 bp sequence from intron 4 of *COL4A5* gene in *COL4A5* mRNA from skin fibroblasts inserted between exon 4 and exon 5 heterozygously (r276\_r.277ins [276 + 4108\_276 + 4165]) (Fig. 4G and H), which led to premature termination of  $\alpha 5(\text{IV})$  chain (p.Pro941lfs\*7). *COL4A5* intron 4 was amplified by PCR from genomic DNA in proband 5 and his daughter. A heterozygous A to G variant in intron 4 at 559 bp upstream from exon 5 (IVS4 c.277–559 A > G) was found in genomic DNA of proband 5 and her daughter (Fig. 4I).

## In silico prediction of deep intronic pathogenic variants

Table 2 shows the output of HSF, NNSPLICE, and NetGene2 for each deep intronic splice variant identified in this study. Only two out of the nine variants were correctly predicted as deleterious by all three tools and six variants were only detected by one tool; no effect was predicted for *COL4A5* intron 31 variant c.2677 + 646C > T. As this variant and the same intron variant c.2677 + 487C > A were segregated with AS in proband 2's family, and not all available in silico predictions were used, it was difficult to consider this variant as a neutral variant.

Table 2  
The output of in silico splice tools for deep intronic pathogenic variants detected in this study\*

Proband	DNA level (experimentally determined)	RNA level (experimentally determined)	In silico splice tools					
			Human Splicing Finder		NNSPLICE		NetGene2	
			Acceptor	Donor	Acceptor	Donor	Acceptor	Donor
1	<i>COL4A4</i> intron 2: c.72 – 26_72- 23delTAAT	Del exon 3 (r.72_r113del)	No significant impact on splicing signals		WT:1.00 Mut:1.00	WT:0.99 Mut:0.99	WT: 0.26 Mut: 0.26	WT:0.63 Mut:0.67
	<i>COL4A4</i> intron 23: c.1623 + 702T > A	Retention 109 nucleotide of intron 23 (r.1623_r.1624 ins [1623 + 590_1623 + 698])	Significant alteration of ESE / ESS motifs ratio		WT:0.76 Mut:0.76	WT:0.67 Mut:0.96	WT: 0 Mut: 0	WT:0.41 Mut:0.70
	<i>COL4A4</i> intron 25: c.1804-158A > G	Del exon 26 (r.1804_r1987del)	WT:42.98 Mut:70.85	-	WT: 0 Mut: 0	WT: 0 Mut: 0	WT: 0 Mut: 0	WT: 0 Mut: 0
2	<i>COL4A5</i> intron 31: c.2677 + 487 C > A	Del exon 32 (r.2678_r2767del)	Significant alteration of ESE / ESS motifs ratio		WT:0.79 Mut:0.79	WT:0.80 Mut:0.80	WT: 0.25 Mut: 0.25	WT: 0.91 Mut: 0.91
	<i>COL4A5</i> intron 31: c.2677 + 646 C > T	Del exon 32 (r.2678_r2767del)	No significant impact on splicing signals.		WT:0.79 Mut:0.79	WT:0.80 Mut:0.80	WT: 0.79 Mut: 0.79	WT: 0.80 Mut: 0.80
3	<i>COL4A5</i> intron 10: c.609 + 879A G	Retention 128 nucleotide of intron 10 (r.609_r.610 ins [609 + 751_609 + 878])	-	WT:67.2 Mut:94.34	WT:0.94 Mut:0.94	WT: 0 Mut: 0.90	WT:0.43 Mut:0.43	WT: 0 Mut: 0.41
4	<i>COL4A5</i> intron 29: c.2395 + 1275C G	Retention 53 nucleotide of intron 29 (r.2395_r.2396 ins [2395 + 1308_2395 + 1360])	Significant alteration of ESE / ESS motifs ratio		WT:0.65 Mut:0.65	WT:0.99 Mut:0.99	WT: 0 Mut: 0	WT:0.46 Mut:0.46
	<i>COL4A5</i> intron 29: c.2395 + 1292G T	Retention 53 nucleotide of intron 29 (r.2395_r.2396 ins [2395 + 1308_2395 + 1360])	No significant impact on splicing signals		WT:0.65 Mut:0.89	WT:0.99 Mut:0.99	WT: 0 Mut: 0.26	WT:0.46 Mut:0.46
5	<i>COL4A5</i> intron 4: c.277 – 559 A > G	Retention 58 nucleotide of intron 4 (r.276_r.277ins [276 + 4108_276 + 4165])	No significant impact on splicing signals		WT:0.87 Mut:0.87	WT:0.57 Mut:0.57	WT:0.71 Mut:0.77	WT: 0 Mut: 0

\*Variants were named following Human Genome Variation Society guidelines (<http://www.hgvs.org/mutnomen>). For genomic DNA positioning, numbering was based on the reference sequences (chr2:227007325–227147483 for *COL4A4* gene and chrX:108440126–108696378 for *COL4A5* gene), whereas RNA position was numbered according to the reference sequences (NM\_000092.5 for *COL4A4* mRNA and NM\_000495 for *COL4A5* mRNA) using the first coding ATG of exon 1 as the initiation codon. WT: wild-type sequence; Mut: mutant sequence; ESE: exonic splicing enhancer; ESS: exonic splicing silencer.

## Discussion

In this study, by analyzing *COL4A3–5* mRNAs from urine or skin fibroblasts, nine deep intronic pathogenic variants were identified in five unrelated Alport syndrome patients with negative NGS results. These findings indicate that our developed approach may be applied to help provide personalized evaluation and care of patients and their families. In addition, this is the first report on compound heterozygous deep intronic splicing mutations in *COL4A4* gene in an Alport syndrome patient.

Numerous studies have shown that NGS is effective in finding single nucleotide variations and small indels in exons and the flanking intronic regions<sup>16</sup>. However, some genetic events such as deep intronic variants, copy number variants, and somatic cell mosaicism may be missed by NGS<sup>17</sup>. Therefore, for a patient with clinically diagnosed or suspected AS and no pathogenic variants detected by NGS, it is necessary to further analyze *COL4A3–5* genes by mRNA sequencing, chromosome microarray analysis, droplet digital PCR or other approaches to improve genetic diagnosis<sup>18,19</sup>.

According to the literature and public databases (Human Gene Mutation Database and Leiden Open source DNA Variation Database), pathogenic splicing variants account for 14.9–24.5% in the *COL4A5* gene<sup>20,21</sup>. Approximately 70.4% (112/159) occurred at consensus splice sites, and only seven splicing variants occurred in introns at more than 100 base pairs up/downstream from exon-intron junctions. Approximately 70% (23/32) of the

pathogenic *COL4A3* splicing variants occurred at consensus splice sites and only two variants were located in introns at more than 100 base pairs upstream from the exons. No deep intronic *COL4A4* splicing variants have been reported to date. These findings indicate that deep intronic *COL4A3–5* mutations are rare. The nine novel deep intronic pathogenic variants obtained in the present study extend the mutational spectrum of AS. These findings also highlight *COL4A3–5* mRNA analysis as an effective supplementary approach for NGS in molecular diagnosis of this disease.

Previous studies have reported that GBM collagen  $\alpha3(\text{IV})$  is synthesized solely by podocytes<sup>22</sup>, and the urine podocyte detachment rate (assessed by podocin mRNA in urine pellets) is increased in AS patients<sup>23,24</sup>. Therefore, extraction of RNA directly from patient-originated urine may be a valuable approach to the analysis of all three Alport gene variants, which was demonstrated by the findings of the present study. In addition, compared with the method that extracted RNA from urine-derived podocyte-lineage cells, our developed approach for isolation of RNA directly from urine is simpler and more practical. A weak point of our approach is the requirement for patient cooperation to obtain enough fresh urine, which means that young patients who cannot rapidly drink 1000–1500 ml water are not suitable for urine mRNA analysis.

Given that the deep intronic variants identified in the present study could be detected using whole genome sequencing, and *in silico* splicing prediction tools are usually used to select variants that are predicted to have an effect on splicing in a molecular diagnostic setting<sup>25</sup>, we assessed the reliability of HSF, NNSPLICE, and NetGene2 in discriminating between neutral and pathogenic variants. Assuming that the splice outcomes obtained from one tool were consistent with transcript analysis results, eight out of the nine variants detected in this study were correctly predicted, which indicated that these tools are useful to select deep intronic variants that are likely to be worth RNA analysis. However, extensive *in silico* analysis should be compared with transcript analysis results to determine their benefit in the context of molecular diagnosis.

In summary, three novel pathogenic *COL4A4* variants and six novel pathogenic *COL4A5* splicing variants were detected in five unrelated AS patients with negative NGS test results. All identified variants were deep intronic variants. As obtaining urine is feasible and non-invasive, we suggest analyzing *COL4A3–5* mRNA from urine as the preferred method for evaluation of patients with clinically diagnosed or suspected AS with negative NGS results.

## Materials And Methods

All methods were carried out in accordance with relevant guidelines and regulations.

## Ethical Considerations

The Ethical Committee of Peking University First Hospital approved the procedures in this study.

## Patients

Patients with hematuria or hematuria and proteinuria were enrolled from August 2019 to August 2020 by pediatric nephrologists from the Department of Pediatrics, Peking University First Hospital based on fulfillment of the following two criteria: diagnosed or suspected AS and no pathogenic *COL4A3–5* variants identified by exome sequencing. Patients were diagnosed with AS if they met one of the following three criteria: 1. abnormal staining of the type IV collagen  $\alpha5$  chain in skin and/or renal specimens; 2. ultrastructural alterations in the GBM typical of AS; or 3. positive family history of hematuria and/or end stage renal disease. Informed consent was obtained from all subjects or, if subjects are under 18, from a parent and/or legal guardian. Patients were excluded if informed consent was not obtained from either themselves or their parents. In total, five patients with AS were included in this study.

## Analysis of *COL4A3–5* mRNA from urine

For AS patients with an uncertain inheritance pattern, *COL4A3–5* mRNAs from urine were analyzed. When available, RNA from parents was sequenced to assess the segregation of variants with the disease in the respective families. To obtain fresh urine, patients were asked to drink approximately 1000–1500 ml water rapidly after emptying the bladder and spontaneously void every 30–45 min. Approximately 500 ml of urine per patient was collected and allocated in 50 mL centrifuge tubes pre-treated with RNeasy (Qiagen, 145023696). Urine samples were centrifuged for 5 min (1200 rpm at 4°C), and the supernatants were carefully removed using pipettes. The urinary pellets were washed twice with ice-cold PBS supplemented with RNeasy (1 ml RNeasy per 50 ml PBS) and the samples were centrifuged for 5 min (1200 rpm at 4°C). Total RNA was isolated from urinary pellets using TRIzol reagent (Gibco, Grand Island, NY, USA) according to the manufacturer's instructions. The concentration of RNA was quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Reverse-transcription was performed using the RevertAid First Strand cDNA Synthesis Kit (TAKARA, K1622). Ten pairs of PCR primers were designed for *COL4A3* (NM\_000091.5), *COL4A4* (NM\_000092.5), and *COL4A5* (NM\_000495.5); the sequences are listed in Table 3. The 'Touchdown' PCR procedure included annealing from 64°C to 57°C, descending 1°C every two cycles, followed by annealing at 57°C for 26 cycles. The PCR amplification products were checked by 2% agarose gel electrophoresis and sequenced on an ABI 3730XL (SinoGenoMax Company Limited, China).

Table 3  
Primers for *COL4A3-5* cDNA analysis

Fragment	Forward prime (5'-3')	Reward prime (3'-5')	Product (bp)
<i>COL4A3-1</i>	GCGAGGCGAGCTTTCCAG	GGAGCACCCCTTTTGCCTTT	586
<i>COL4A3-2</i>	TACGGACTTGTCGGTGTACC	GGCCTTGATGATCCAGGACT	815
<i>COL4A3-3</i>	GGGTGAAGATGGCATTAAAGG	AGACCTGTATTTCTGGGGAC	594
<i>COL4A3-4</i>	CAGGAAGACAAGGCGCAG	GACAACCCAGTGATCCTTTTGT	666
<i>COL4A3-5</i>	CTGGGGAATGTGGAGATCC	TGGCCCTAAAATTCCCGGAT	618
<i>COL4A3-6</i>	CACTGGGTCAAAGAGGATATCC	GTCCTGGCCTTGACCTTCT	591
<i>COL4A3-7</i>	CAGGGAGATAAGGGGAGAGCC	GCTTCTCTTGAGCCTGGT	600
<i>COL4A3-8</i>	GTGCAATTATCCCTGGCCAG	CCTTTCAAACCTGGCAATCC	598
<i>COL4A3-9</i>	GGCAAGGATGGAAAACCAGG	GCACGTTCTCTTCCATGAC	687
<i>COL4A3-10</i>	CGGCTGGATTTCTCTCTGGA	ACAGCACAGATTAGAGACCCA	498
<i>COL4A4-1</i>	TGACCCAGAACACAGAACCT	CCAAGAGCTCCTCTTCTCC	564
<i>COL4A4-2</i>	CCACAATGGCTCAAGAGGTG	TCCAAATAGCCAGGATCTCC	595
<i>COL4A4-3</i>	TCCTGGTTCCTATGGATCTCC	ACCTTTTGTTC AAGCCAGC	646
<i>COL4A4-4</i>	CTGTGAGCCTGGACCCATG	ATTTCGGCTTCTGGTGT	672
<i>COL4A4-5</i>	CAACGTAACCTACCCTGGGA	CCAGGTAGCCATCATCTCC	696
<i>COL4A4-6</i>	CAGCTGGAATGAAAGGCCTC	AGGTCTCTTGCTCCATCAA	617
<i>COL4A4-7</i>	GCTCAACTGGTCTAAGAGGGT	GCTCTTCTGTGGCACCT	652
<i>COL4A4-8</i>	GGGCTAAAAGGGGAGAGAGG	CTCTCATTCCAGGGAGCCC	583
<i>COL4A4-9</i>	CAGGCATGAGAGGACCAGAA	CTGACATAGGGGCGGATCG	538
<i>COL4A4-10</i>	TGCCAGAGAAACGACAGAT	ACGTGTTGGTGAATTCGCA	499
<i>COL4A5-1</i>	GCTCTCTCCATATAAACCCCTCAA	TCCTGGCAGTGATGACATAATT	600
<i>COL4A5-2</i>	TGCAATGGAACCAAGGGAGA	TCATCACCTTTCTGACCCCTT	863
<i>COL4A5-3</i>	GCAAAGATGGAGAAAATGGCC	TGGAGGCCCTGAAATACCA	579
<i>COL4A5-4</i>	AGGAGAACAAGGAGTGAAAGGT	CCGGCTGGGTTATAGTCTGA	584
<i>COL4A5-5</i>	CATACAAGGTGTGGCAGGAA	CTATTGGCCCAGGAATCCC	596
<i>COL4A5-6</i>	GCCAGGAATAGGTGTTTCAGG	CGATGGTTCCTTTAAGTCCAGG	644
<i>COL4A5-7</i>	ACCTGGACTGAGTGGACAAC	GTTACCCTTCTGTCCAGCT	575
<i>COL4A5-8</i>	TGTAGGTGGTGGAGGTCATC	ACCAATAAGTCCCGTTCCC	613
<i>COL4A5-9</i>	GGATTCCCAGGCATGAAAGG	CCTTTAGGGGTTGCATGCTC	673
<i>COL4A5-10</i>	GCCTTTCATGTTCTGCAAC	GGGGACAATGAGACACTGAC	594

The reference sequence of *COL4A3-5* transcripts was NM\_000091.5, NM\_000092.5, and NM\_000495.5, respectively.

## Analysis of *COL4A5* mRNA from skin fibroblasts

For patients with XLAS, *COL4A5* mRNA from cultured skin fibroblasts was analyzed. Dermal fibroblasts were cultured as described previously<sup>11</sup>. Primers for *COL4A5* cDNA analyses were performed using the same primers as shown in Table 3. RT-PCR and direct sequencing followed the above protocol.

## Genomic DNA analysis

Genomic DNA was extracted from peripheral blood lymphocytes. Once abnormal *COL4A3-5* transcripts were detected, the corresponding exons with flanking intronic sequences were further analyzed using PCR and direct sequencing to identify the point variants that may cause new splice sites. PCR primers are available on request.

## In silico splice tools for identifying deep intronic pathogenic variants

To evaluate the reliability of in silico splicing prediction tools in discriminating the deep intronic pathogenic variants identified in this study, three tools including HSF (<http://www.umd.be/HSF/>), NNSPLICE ([http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html)) and NetGene2 (<http://www.cbs.dtu.dk/services/NetGene2/>) were used. NNSPLICE and NetGene2 present scores of 0–1 for the predicted site; the higher the score the more likely a variant is a splicing site.

## Declarations

## Acknowledgements

We thank the patients, the families, and the physicians who contributed to this project. This study was supported by grants from the National Key Research and Development Program of China (No. 2016YFC0901505) (the registry study of rare diseases in children) and the Beijing key laboratory of molecular diagnosis and study on pediatric genetic diseases (BZ0317). We thank Gabrielle White Wolf, PhD, from Liwen Bianji, Edanz Editing China ([www.liwenbianji.cn/ac](http://www.liwenbianji.cn/ac)), for editing the English text of a draft of this manuscript.

## Author information

Xiaoyuan Wang and Yanqin Zhang contributed equally

## Affiliations

Department of Pediatrics, Peking University First Hospital, Beijing, China

## Contributions

Our co-authors have all contributed to this manuscript. Conceptualization, F.W. and J.D.; Methodology, X.W.; Software, X.W.; Formal Analysis, F.W. and J.D.; Investigation, X.W.; Resources, X.W. and Y.Z.; Data Curation, X.W. and Y.Z.; Writing-Original Draft Preparation, X.W.; Writing-Review & Editing, Y.Z., F.W., J.D.; Visualization, Y.Z., F.W. and J.D.; Supervision, F.W. and J.D.; Project Administration, F.W. and J.D.; Funding Acquisition, F.W. and J.D.

## Competing Interests

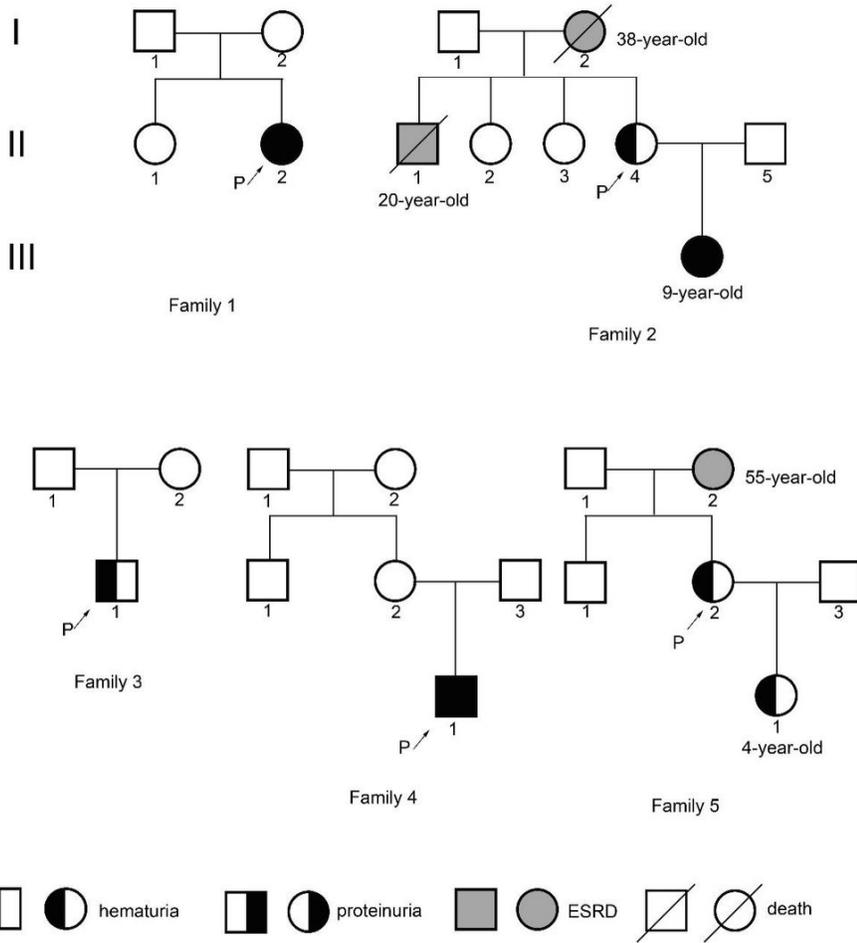
The authors declare no competing interests.

## References

1. Kruegel, J., Rubel, D. & Gross, O. Alport syndrome—insights from basic and clinical research. *Nat Rev Nephrol.***9**, 170–178 <https://doi.org/10.1038/nrneph.2012.259> (2013).
2. Mochizuki, T. *et al.* Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome. *Nature genetics.***8**, 77–81 <https://doi.org/10.1038/ng0994-77> (1994).
3. Barker, D. F. *et al.* Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science.***248**, 1224–1227 <https://doi.org/10.1126/science.2349482> (1990).
4. Nozu, K. *et al.* A review of clinical characteristics and genetic backgrounds in Alport syndrome. *Clin Exp Nephrol.***23**, 158–168 <https://doi.org/10.1007/s10157-018-1629-4> (2019).
5. Savage, J. *et al.* Expert guidelines for the management of Alport syndrome and thin basement membrane nephropathy. *J Am Soc Nephrol.***24**, 364–375 <https://doi.org/10.1681/ASN.2012020148> (2013).
6. Yamamura, T. *et al.* Comparison between conventional and comprehensive sequencing approaches for genetic diagnosis of Alport syndrome. *Molecular genetics & genomic medicine.***7**, e883 <https://doi.org/10.1002/mgg3.883> (2019).
7. Moriniere, V. *et al.* Improving mutation screening in familial hematuric nephropathies through next generation sequencing. *J Am Soc Nephrol.***25**, 2740–2751 <https://doi.org/10.1681/ASN.2013080912> (2014).
8. Plant, K. E., Boye, E., Green, P. M., Vetrie, D. & Flinter, F. A. Somatic mosaicism associated with a mild Alport syndrome phenotype. *Journal of medical genetics.***37**, 238–239 <https://doi.org/10.1136/jmg.37.3.238> (2000).
9. King, K., Flinter, F. A., Nihalani, V. & Green, P. M. Unusual deep intronic mutations in the COL4A5 gene cause X linked Alport syndrome. *Human genetics.***111**, 548–554 <https://doi.org/10.1007/s00439-002-0830-3> (2002).
10. Nagano, C. *et al.* Detection of copy number variations by pair analysis using next-generation sequencing data in inherited kidney diseases. *Clin Exp Nephrol.***22**, 881–888 <https://doi.org/10.1007/s10157-018-1534-x> (2018).
11. Wang, F., Wang, Y., Ding, J. & Yang, J. Detection of mutations in the COL4A5 gene by analyzing cDNA of skin fibroblasts. *Kidney international.***67**, 1268–1274 <https://doi.org/10.1111/j.1523-1755.2005.00204.x> (2005).

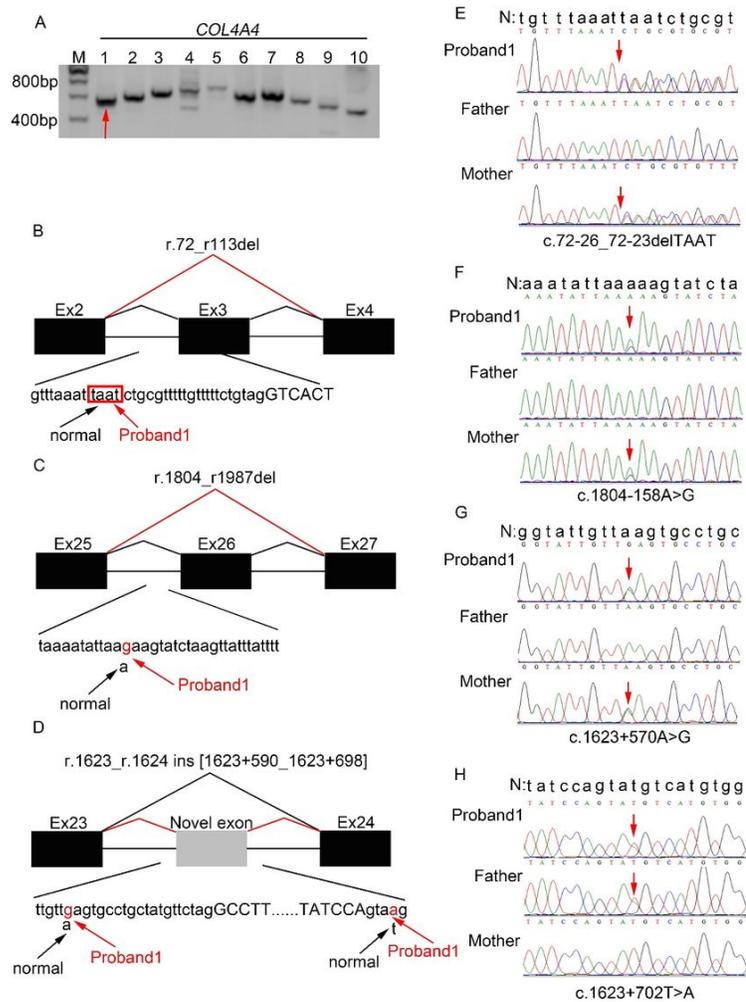
12. Nakazato, H. *et al.* Splicing mutations in the COL4A5 gene in Alport's syndrome: different mRNA expression between leukocytes and fibroblasts. *American journal of kidney diseases: the official journal of the National Kidney Foundation*.**26**, 732–739 [https://doi.org/10.1016/0272-6386\(95\)90436-0](https://doi.org/10.1016/0272-6386(95)90436-0) (1995).
13. Tazon-Vega, B. *et al.* Genetic testing for X-linked Alport syndrome by direct sequencing of COL4A5 cDNA from hair root RNA samples. *American journal of kidney diseases: the official journal of the National Kidney Foundation***50**, 257 e251-214, doi:10.1053/j.ajkd.2007.05.004 (2007).
14. Guo, C. *et al.* Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain. *Kidney international*.**44**, 1316–1321 <https://doi.org/10.1038/ki.1993.384> (1993).
15. Oka, M. *et al.* Natural history of genetically proven autosomal recessive Alport syndrome. *Pediatric nephrology*.**29**, 1535–1544 <https://doi.org/10.1007/s00467-014-2797-4> (2014).
16. Fernandez-Marmiesse, A., Gouveia, S. & Couce, M. L. NGS Technologies as a Turning Point in Rare Disease Research, Diagnosis and Treatment. *Curr Med Chem*.**25**, 404–432 <https://doi.org/10.2174/0929867324666170718101946> (2018).
17. Yohe, S. & Thyagarajan, B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med*.**141**, 1544–1557 <https://doi.org/10.5858/arpa.2016-0501-RA> (2017).
18. Tarilonte, M. *et al.* Parental Mosaicism in PAX6 Causes Intra-Familial Variability: Implications for Genetic Counseling of Congenital Aniridia and Microphthalmia. *Front Genet*.**9**, 479 <https://doi.org/10.3389/fgene.2018.00479> (2018).
19. Hoischen, A. *et al.* Array-CGH in unclear syndromic nephropathies identifies a microdeletion in Xq22.3-q23. *Pediatric nephrology*.**24**, 1673–1681 <https://doi.org/10.1007/s00467-009-1184-z> (2009).
20. Wang, F. *et al.* Skin biopsy is a practical approach for the clinical diagnosis and molecular genetic analysis of X-linked Alport's syndrome. *J Mol Diagn*.**14**, 586–593 <https://doi.org/10.1016/j.jmoldx.2012.06.005> (2012).
21. Jais, J. P. *et al.* X-linked Alport syndrome: natural history in 195 families and genotype- phenotype correlations in males. *J Am Soc Nephrol*.**11**, 649–657 (2000).
22. Abrahamson, D. R., Hudson, B. G., Stroganova, L. & Borza, D. B. & St John, P. L. Cellular origins of type IV collagen networks in developing glomeruli. *J Am Soc Nephrol*.**20**, 1471–1479 <https://doi.org/10.1681/ASN.2008101086> (2009).
23. Wickman, L. *et al.* Urine podocyte mRNAs, proteinuria, and progression in human glomerular diseases. *J Am Soc Nephrol*.**24**, 2081–2095 <https://doi.org/10.1681/ASN.2013020173> (2013).
24. Ding, F. *et al.* Accelerated podocyte detachment and progressive podocyte loss from glomeruli with age in Alport Syndrome. *Kidney international*.**92**, 1515–1525 <https://doi.org/10.1016/j.kint.2017.05.017> (2017).
25. Houdayer, C. *et al.* Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human mutation*.**33**, 1228–1238 <https://doi.org/10.1002/humu.22101> (2012).

## Figures



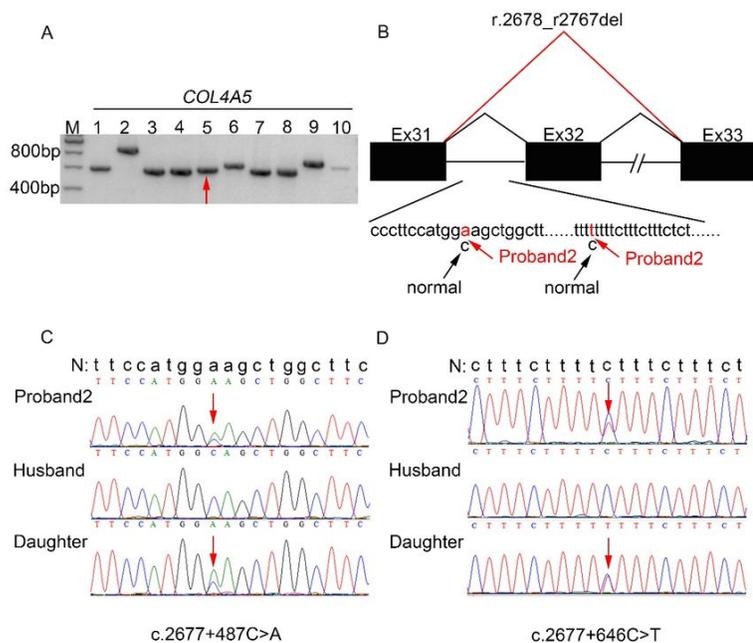
**Figure 1**

Pedigrees of the 5 families in this study. The proband is indicated by an arrow.



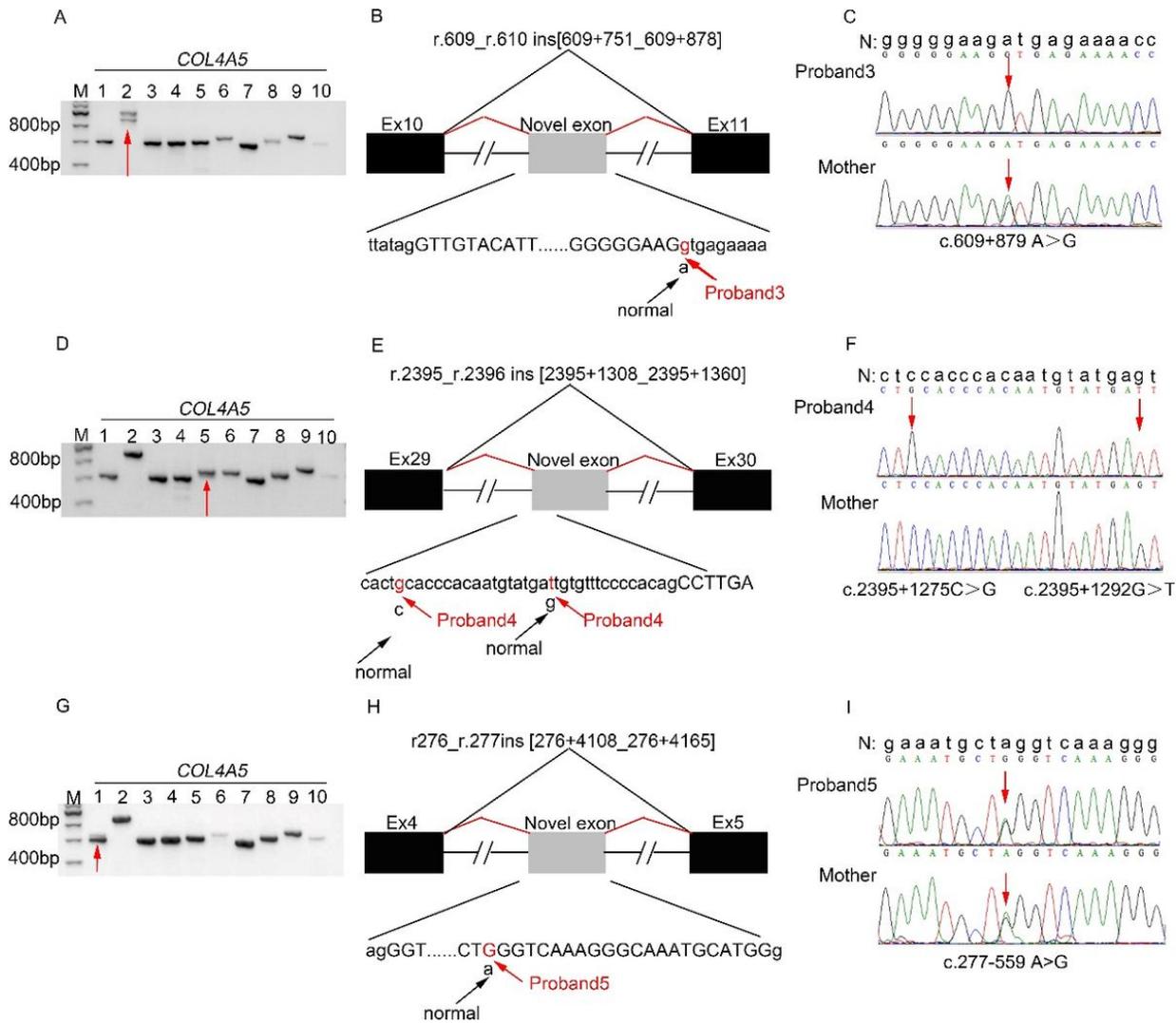
**Figure 2**

COL4A4 gene analysis in proband 1. A: Agarose-gel images of RT-PCR products for proband 1 urine. M: DNA molecular mass marker. Lane 1 to 10: 10 overlapping PCR products covering the entire COL4A4 cDNA from proband 1. The red arrow indicates a smaller sized RT-PCR product than the expected fragment. B, C, and D: Schematic representation of the aberrant COL4A4 cDNA caused by the deep intronic splice variants detected in proband 1. Exons are represented by black filled rectangles and are numbered, and the novel exon is indicated by the gray filled rectangle. Capital and lowercase letters depict exon and intron sequences, respectively. The black and red lines indicate wild-type and abnormal splicing, respectively. The black arrows indicate the wild-type nucleotide(s), the red arrows show the changed nucleotides, and the red box shows the nucleotides that should have been detected in wild-type but were deleted in the proband. E, F, G, and H: Sequencing of PCR amplified products containing the deep intronic variants for family 1. N, normal sequence. Intron sequences are depicted by lowercase letters. The red arrows indicate the breakpoint and changed nucleotide.



**Figure 3**

COL4A5 gene analysis in proband 2. A: Agarose gel images of RT-PCR products for proband 2 urine. M: DNA molecular mass marker. Lane 1 to 10: 10 overlapping PCR products covering the entire COL4A5 cDNA. The red arrow indicates a smaller sized RT-PCR product than the expected fragment. B: Schematic representation of the aberrant COL4A5 cDNA caused by the deep intronic splice variants detected in proband 2. Exons are represented by black filled rectangles and are numbered. Lowercase letters are used to depict intron sequences. The black and red lines indicate wild-type and abnormal splicing, respectively. The black arrows indicate the wild-type nucleotide(s), the red arrows show the changed nucleotide. C and D: Sequencing of PCR amplified products containing the deep intronic variants for family 2. N, normal sequence. Intron sequences are depicted by lowercase letters. The red arrows indicate the changed nucleotide.



**Figure 4**

COL4A5 gene analysis in proband 3-5. A, D, and G: Agarose gel images of RT-PCR products for skin fibroblasts of probands 3–5, respectively. M: DNA molecular mass marker. Lane 1 to 10: 10 overlapping PCR products covering the entire COL4A5 cDNA from probands 3–5. The red arrow indicates a larger sized RT-PCR product than the expected fragment. B, E, and H: Schematic representation of the aberrant COL4A5 cDNA caused by the deep intronic splice variants detected in probands 3–5, respectively. Exons are represented by black filled rectangles and are numbered, and the novel exon is indicated by the gray filled rectangle. Capital and lowercase letters are used to depict exon and intron sequences, respectively. The black and red lines indicate wild-type and abnormal splicing, respectively. The black arrows indicate the wild-type nucleotide, and the red arrows showed the changed nucleotide. C, F, and I: sequencing of PCR amplified products containing the deep intronic variants for families 3–5, respectively. N, normal sequence. Intron sequences are depicted by lowercase letters. The red arrows indicate the changed nucleotide.

## Supplementary Files

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

- [supplementaryfiles.docx](#)