scientific reports

Check for updates

OPEN

Identifying six single nucleotide variants in the *COL17A1* gene that alter RNA splicing: database analysis and minigene assays

Yingfei Shao[1] & Ran Zhang[2]✉

Collagen type XVII alpha 1 chain (*COL17A1*) is a protein in the collagen family crucial for maintaining the integrity of skin and epithelial tissues. It is also vital for enamel formation and plays a significant role in the differentiation of ameloblasts. Many studies have indicated that single nucleotide variants (SNVs) can disrupt normal splicing process of the pre-mRNA by altering various splicing regulatory signals. This study aimed to explore the potential impact of SNVs in *COL17A1* geneon splicing events, with the ultimate aim of improving the prediction of disease prognosis. Here, we analyzed 703 SNVs including 446 exonic variants and 257 intronic variants in the *COL17A1* gene using bioinformatics tools and identified candidate variants that may induce splicing alterations via minigene assays. Our study identified that, among eight candidate variants, six variants (c.1139 C > T, c.1834G > A, c.3198 C > T, c.202 + 6T > G, c.1222 + 4 A > G, c.3071-5G > A) induced splicing alterations by interfering with the recognition of classical splice sites or disrupting the ratio of exonic splicing enhancers/exonic splicing silencers, or both. This study emphasizes the necessity of assessing the effects of SNVs on at the mRNA level, aiding accurate characterization of *COL17A1* variants and enabling the development of personalized treatment options.

Keywords *COL17A1*, Amelogenesis imperfecta, Junctional epidermolysis Bullosa, Minigene analysis, Single nucleotide variants

Collagen type XVII alpha 1 chain (*COL17A1*) is expressed in skin, oesophagus and reproductive organs according to Human Protein Atlas. It is located in the basal membrane area of the skin, and connects the skin epidermis and dermis as a hemidesmosome component in humans. It also plays a role in the process of enamel formation and contributes to the differentiation of ameloblasts[1]. *COL17A1* gene participates in the interaction between enamel epithelium and underlying mesenchymal through the epithelial-mesenchymal junction[2]. *COL17A1*−/− mice lack Tomes' processes in ameloblasts, exhibit malformed ameloblasts and reduced production of enamel matrix during the secretory stage and prolonged calcification in the amelogenesis maturation stage[1].

Biallelic mutations in the *COL17A1* gene (ENSG00000065618) cause the recessively inherited, genetically heterogeneous mucocutaneous blistering condition junctional epidermolysis bullosa (JEB), with amelogenesis imperfecta (AI). Of note, most mutations in the database are only analyzed at the genomic level, so exonic single-nucleotide variants (SNVs) are generally classified as missense or nonsense, even synonymous variants. However, SNVs may lead to aberrant splicing, especially variants located in the donor splice region (the last 3 bases of the exon and 3–6 nt of intronic sequence adjacent to the exon) and the acceptor splice region (the first base of the exon and from 3 to 40 nt upstream from the exon boundary)[3]. The impact of SNVs in *COL17A1* on splicing regulation may be underestimated due to the lack of studies on pre-mRNA splicing. In this study, we explored the effect of SNVs in the *COL17A1* gene on pre-mRNA splicing by bioinformatics tools and minigene technology.

Materials and methods

Variant collection

DNA variant numbering was grounded in the complementary DNA (cDNA) sequence for *COL17A1* (NM_000494.4). Missense and nonsense variants were collected from the Human Gene Mutation Database (February 2024), while synonymous variants and intronic variants (located 3–40 nt upstream of a canonical

[1]Wenzhou Medical University Renji College, Wenzhou, China. [2]Department of Nephrology, The Affiliated Qingdao Municipal Hospital of Qingdao University, Qingdao, China. ✉email: 18562788709@163.com

splice acceptor site or 3–6 nt downstream of a canonical splice donor site) were collected from ClinVar (February 2024).

Bioinformatics analyses and screening criteria

Each of collected variants was analyzed to predict the effects on premRNAs splicing through online bioinformatics software. Firstly, the variants predicted by SpliceAI (https://spliceailookup.broadinstitute.org/, accessed on 2 February 2024) scoring less than 0.5 were excluded. Further, we used BDGP (http://www.fruitfly.org) and MaxEntScan (http://hollywood.mit.edu/burgelab/maxent/Xmaxent.html) to analyze the potential effects of variations on canonical splice sites and to predict the generation and/or activation of novel sites, the variants with BDGP score reduction or the absolute value of the MaxEntScan score > 15% were selected to continue the analyses. Moreover, the Human Splice Finder (http://www.umd.be/HSF3) was performed to investigate potential impact of exonic variants on splicing regulatory sequences, including Exonic Splicing Enhancers (ESEs) broken and/or new Exonic Splicing Silencers (ESSs) creation, and identify the putative effect of variants on splicing regulatory motifs.

Minigene construction and mutagenesis

This study received approval from the Ethics Committee at Qingdao Municipal Hospital, affiliated with Qingdao University, and informed consent was obtained from all participants. All experiments were performed in accordance with relevant guidelines and regulations. Genomic DNA was extracted from the blood samples of healthy volunteers using the GenElute Blood Genomic DNA Extraction Kit (Sigma, NA2010). The pSPL3 exon trapping vector was used minigene splicing assay in vitro (Supplementary Fig. S1a). We amplify by PCR reactions the target wild-type (WT) fragments containing the target exons and shortened flanking introns with specific primers (Supplementary Table S1). The primers were engineered to include *Xho*I and *Nhe*I restriction sites (*Xho*I: CCGC^CTCGAG; *Nhe*I: CTAG^CTAGC), and were designed by Primer Premier 5 and Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast). The PCR amplification reaction was performed as follows: in 50 µL volume, 3 µL of DNA, 25 µL of 2×Tap Plus MasterMix (Vazyme, China), 1 µL of each primer (10 µM), and 20 µL of ddH_2O; the initial denaturation at 95 ℃ for 5 min, followed by 3 cycles of 94 ℃ for 30 s, 62 ℃ for 30 s, and 72 ℃ for 90 s, then 5 cycles of 94 ℃ for 30 s, 60 ℃ for 30 s, and 72 ℃ for 90 s; and finally 10 cycles of 94 ℃ for 30 s, 58 ℃ for 30 s, and 72 ℃ for 90 s. PCR products were subsequently purified using Gel Extraction Kit (Cwbio, China). Both the purified PCR products and the pSPL3 exon trapping vector were subjected to digestion by the respective enzymes *Xho*I and *Nhe*I (Takara, Japan). Then, the PCR products and the pSPL3 splicing vector were connected at 16 ℃ using 1 µL of T4 DNA Ligase (Takara, Japan) and transformed into DH5α competent *E. coli* cells to construct the WT minigene. Two-hundred microliters of bacteria suspension was spread evenly on the solid ampicillin-Luria-Bertani agar plates and coated for 16 h at 37 ℃. Monoclonal colonies were screened and sequenced using the forward primers. Sequence alignment analysis was conducted using Snapgene software. Finally, plasmid DNA from positive monoclonal colonies was extracted with Pure Plasmid Mini Kit (Cwbio, China).

Mutagenesis primers were designed by Snapgene and Primer BLAST (Supplementary Table S2). Variants of interest were introduced into WT plasmids with QuikChange II Site Directed Mutagenesis Kit (Stratagene, La Jolla, CA, United States) according to the manufacturer's instruction. Primer extension and PCR amplification reactions are as follows: in 50 µL volume, 200 ng WT plasmid, 25 µL of PrimeSTAR HS (Premix) (Takara, Japan), 1 µL of each primer (10 µM), and make up the remaining volume with ddH_2O; the first step is denaturation at 98 ℃ for 10 s, followed by 30 cycles, denaturation at 98 ℃ for 10 s, annealing at 62 – 55 ℃ for 10 s, elongation at 72 ℃ for 2 min, and finally extension at 72 ℃ for 5 min. In order to confirm the existence of target mutations, all minigenes were further transformed into DH5α competent *E. coli* cells. The positive monoclonal colonies was extracted with Pure Plasmid Mini Kit (Cwbio, China).

All constructed WT and variant minigenes were confirmed through direct sequencing for ensuring obtained target variants (Supplementary Table S4 and Supplementary Fig. S2).

Analysis of minigenes

Human embryonic kidney 293T (HEK293T) and HeLa cells were purchased from Procell (Wuhan, China), and cultured in DMEM with high glucose levels (4.5 g/L), enriched with 10% fetal bovine serum, along with penicillin (100 U/L) and streptomycin (100 mg/L). The cells were maintained at 37 ℃ in an incubator with a 5% CO_2 atmosphere. One day prior to transfection, the cells were seeded on 12-well plate in an antibiotic-free medium to achieve 70–80% confluence. Each group of plasmids (empty pSPL3-control, pSPL3-WT and pSPL3-Mutation, 2 µg) were transfected to HEK239T and Hela cells using Lipofectamine 2000 (Invitrogen, United States) following the manufacturer's instructions.

Total RNA was extracted with RNA-easy Isolation Reagent (Vazyme Biotech Co.,Ltd, China) after 48 h incubation. Then cDNA was synthesized by reverse transcription PCR (RT-PCR) using PrimeScript 1st Strand cDNA Synthesis kit (Takara, Japan), and amplified by PCR. The PCR amplification reaction of cDNAs was executed with vector-specific primers: SD6 (the forward primer: 5′-TCTGAG TCACCTGGACAACC-3′) and SA2 (the reverse primer: 5′-ATCTCAGTGGTATTTGTGAGC-3′). The PCR amplification reaction was performed as follows: in 30 µL volume, 3 µL of cDNA, 10 µL of 2×Tap Plus MasterMix (Vazyme, China), 1 µL of each primer (10 µM), and 15 µL of ddH_2O. Thermal conditions were denaturation at 95 °C for 3 min, 5 cycles of 95 °C for 20 s, 62 °C for 30 s, and 72 °C for 30 s; 5 cycles of 95 °C for 30 s, 60 °C for 20 s, and 72 °C for 30 s; 10 cycles of 95 °C for 30 s, 58 °C for 20 s, and 72 °C for 30 s; 10 cycles of 95 °C for 30 s, 55 °C for 20 s, and 72 °C for 30 s; and followed by a final elongation step at 72 °C for 10 min. The amplified products were separated by 1.5% agarose gel electrophoresis. Each band was quantified by ImageJ software and analyzed by Sanger sequencing (Sangon Biotech (Shanghai) Co., Ltd.). The result of sequences were aligned with the reference *COL17A1* sequence from

GenBank using Snapgene software. Original TA Cloning Kit was used toseparate each PCR product from band contained multiple PCR products, and the relative expression of each PCR product was quantified by real-time qPCR. Total RNA was extracted from cells using RNA-easy Isolation Reagent (Vazyme Biotech Co.,Ltd, China). Then, RNA was converted to cDNA using Evo M-MLV RT Mix Kit (Accurate biotechnology Co.,Ltd, China). The quantification of gene expression was determined with SYBR Green Premix Pro Taq HS qPCR Kit (Accurate biotechnology Co.,Ltd, China). Sequences and location for qPCR primers were shown in Supplementary Table S3 and Supplementary Fig. S4. QuantStudio 3 Real-Time Fluorescent Quantitative PCR Instrument was used for the qPCR reaction. The qPCR reaction system: in 10 μL volume, 2×SYBR Green Pro Taq HS Premix 5μL, upstream and downstream primers (10 μM) 0.2μL, cDNA template 2μL, ddH2O 2.6μL. After mixing the reaction system thoroughly, it was aliquoted into centrifuge tubes, with three replicates for each sample. The reaction conditions were set as follows: 95 ℃ for 30 s; 40 cycles of 95 ℃ for 5 s, 60 ℃ for 30 s; 95 ℃ for 1 s, 60 ℃ for 20 s, 95 ℃ for 1 s. Fluorescence data were collected and analyzed. All PCR reactions were performed in triplicate. Different splicing patterns compared to the wild-type minigene pattern were considered to result in incorrect splicing. The experiments were conducted three times in HEK239T and HeLa cells.

Statistical analysis

The percentage of aberrant band (%) = (aberrant band/all bands) × 100. The qPCR analysis utilized the 2^−ΔΔCt method to calculate relative gene expression levels, and normalized to the geometric mean of GAPDH genes. Statistical analysis was analyzed using the unpaired Student's t-test by GraphPad Prism (Version 6.02, GraphPad Software, USA). Error bars represent SEM ($n = 3$). $P < 0.05$ was considered statistically significant. All experiments were repeated at least three times to ensure reproducibility.

Results

A total of 703 SNVs (446 exonic variants and 257 intronic variants) in the *COL17A1* gene were collected and analyzed. After preliminary analysis by bioinformatics software, 11 SNVs were collected, among which, 3 SNVs previously proven to have an effect on pre-mRNA splicing were excluded[4–6]. Finally, 8 SNVs were enrolled as follows: c.1139 C > T, c.1467G > A, c.1834G > A, c.3198 C > T, c.202 + 6T > G, c.1222 + 4 A > G, c.2335 + 3 A > G, c.3071-5G > A (Table 1 and Supplementary Fig. S1). Minigene splicing assays verified that six of the eight variants altered normal pre-mRNA splicing (Table 2). All RT-PCR products were sequenced by Sanger sequencing (Supplementary Fig. S3).

Variant c.202 + 6T > G led to skipping of exon 4

Minigene assays showed that RT-PCR products generated by minigenes with variant c.202 + 6T > G and WT were different (Fig. 1a and g*(1)*). The band of WT demonstrated one fragment of 368 bp that contains exon 4, whereas the mutant c.202 + 6T > G generated a 263 bp fragment with skipping of exon 4.

Variant c.1139 c > t gave rise to partial skipping of exon 14

The RT-PCR products of minigene with variant c.1139C > T consisted of a single aberrant band of 421 bp (Fig. 1b and g*(2)*), which indicated that variant c.1139C > T created a novel splice donor site in exon 14 (Supplementary Table S4), which led to a 4-base deletion at the 3' end of exon 14.

Variant c.1222 + 4 A > G caused a significant skipping of exon 15

The results of minigene assays of confirmed the variant c.1222 + 4 A > G produced two different fragments of 344 bp and 263 bp (Fig. 1c), the larger fragment containing exon 15, SD and SA of the pSPL3, and the smaller skipping exon 15, which made up 19.7% and 80.21%, respectively ($p < 0.001$) (Fig. 1g*(3)*).

Variant	Amino acid	Affected motifs	RNA effect	Frameshift	Protein effect
c.1139 C > T	p.Ala380Val	Canonical splice site and ESEs/ESSs	r.1138_1141del; Partial deletion (4 bp) of exon 14	p.Ala380Leu fs*8	Truncated protein in non-collagenous 16 C domains
c.1834G > A	p.Gly612Arg	Canonical splice site	r.1831_1834del; Partial deletion (4 bp) of exon 22	p.Val611Glu fs*71	Truncated protein in collagenous 15 domains
c.3198 C > T	p.Ser1066=	Canonical splice site and ESEs/ESSs	r.3193_3208del; Partial deletion (16 bp) of exon 46	p.Val1065Leu fs*35	Truncated protein in non-collagenous 9 domains
c.202 + 6T > G	Intronic mutation	Canonical splice site	r.98_202del; Skipping of exon 4	–	Absence of 35 amino acids in non-collagenous 16 C domains
c.1222 + 4 A > G	Intronic mutation	Canonical splice site	r.1142_1222del; Skipping of exon 15	–	Absence of 27 amino acids in non-collagenous 16 C domains
c.3071−5G > A	Intronic mutation	Canonical splice site	r.3070_3071ins3071−3_3071−1; Partial retention (3 bp) of intron 45/r.3071_3208del; Skipping of exon 46	–	Insertion of 1 amino acid/ Absence of 46 amino acids in non-collagenous 10 domains

Table 1. Bioinformatics analysis of variants screened in *COL17A1*. *DS* donor splice site, *AS* acceptor splice site, *ESE* exonic splicing enhancer, *ESS* exonic splicing silencer, *WT* wild type, *MUT* mutation, *NA* not applicable. [a]Location of eight variants: "+" indicates the distance to the 5′ end of the exon/intron, and "−" indicates the distance to the 3′ end of the exon/intron. [b]ESE/ESS motifs ratio.

Variant	Amino acid	Exon/intron	Location in exon/intron[a]	Splice AI	BDGP (WT → MUT score)	HSF	MaxEntScan	
							△MaxEnt Donor site	△MaxEnt Acceptor site
c.1139 C > T	p.Ala380Val	Exon 14	−3	1.00	5'DS: 0.28 → 0 (100%)	−6[b];New DS	956.79%	/
c.1467G > A	p.Ala489=	Exon 18	+2	0.80	3'AS: 0.41 → 0 (100%)	New AS	/	229.77%
c.1834G > A	p.Gly612Arg	Exon 22	−1	0.92	5'DS: 0.91 → 0 (100%)	Broken WT DS	−42.65%	
c.3198 C > T	p.Ser1066=	Exon 46	−11	0.65	NA	−5[b];New DS	157.42%	/
c.202 + 6T > G	NA	Intron 4	+6	0.50	5'DS: 0.99 → 0.82 (17.17%)	NA	−35.71	/
c.1222 + 4 A > G	NA	Intron 15	+4	0.64	5'DS: 1 → 0.76 (24%)	Broken WT DS	NA	
c.2335 + 3 A > G	NA	Intron 31	+3	0.52	5'DS: 0.9 → 0.57 (36.67%)	Broken WT DS	−38.18%	/
c.3071−5G > A	NA	Intron 45	−5	0.93	3'AS: 0.41 → 0(100%)	New AS	/	175.11%

Table 2. Splicing outcomes for variants with impact on splicing. *COL17A1* reference sequence: NM_000494.4, NP_000485.3.

Variant c.1834G > A generated different splice transcript

Agarose gel electrophoresis of minigene splicing experiment showed that the RT-PCR product of the variant c.1834G > A minigene made of a single band. However, sequencing results verified "a single band" contained two fragments actually (Fig. 1d). Original TA Cloning Kit was further used to separate each PCR product (Supplementary Fig. S3), which confirmed that there were two distinct products of 353 bp and 349 bp, the larger band including a complete exon 22 with variant c.1834G > A, the smaller one containing a 4-bp deletion at the 3′ end of exon 22. This result suggested the mutation activating a cryptic donor splice site located 4 nt upstream of a canonical splice donor site (Supplementary Table S4). Then, real-time qPCR was used to quantify the relative expression of 353 bp and 349 bp product, which were almost equally expressed (Fig. 1h).

Variant c.3071-5G > A and c.3198 c > t resulted in abnormal splice transcripts of exon 46

Similarly, original TA Cloning Kit verified that the RT-PCR product of mutant c.3071-5G > A minigene contained three fragment of 404 bp, 401 bp and 263 bp (Fig. 1e, Supplementary Fig. S3), which indicated variant c.3071-5G > A had multiple effects on the splicing of exon 46. Firstly, it producted one fragment containing exon 46, SD and SA of the pSPL3. Additionally, it created a novel acceptor splice site resulting in the insertion of a 3-bp of the 3′ end of intron 45, and disturbing the canonical splice acceptor site causing skipping of exon46 (Supplementary Table S4). The real-time qPCR experiment showed that the majority of the products were normal, with a comparable amount of abnormal splicing products (Fig. 1h).

The RT-PCR product from variant c.3198C > T consisted of a single aberrant band of 385 bp (Fig. 1e and g*(4)*). Sequencing results indicated that variant c.3198C > T triggered a cryptic splice site leading to a 16-bp deletion at the 3′ end of exon 46 (Supplementary Fig. S3, Supplementary Table S4).

Variant c.1467G > A and c.2335 + 3 A > G did not alter pre-mRNA splicing

Minigene assays showed that variant c.1467G > A produced a band of complete exon 18 with variant c.1467G > A, and c.2335 + 3 A > G generated the same transcripts as the WT (Fig. 1f).

Discussion

Pre-mRNA splicing involves connecting exon sequences and simultaneously excising noncoding intron sequences to produce mature mRNA within the nucleus. During the pre-mRNA splicing process, the cis-elements necessary for splicing or its regulation (including the 5′ splice site, 3′ splice site, branch point sequence, and intronic or exonic enhancer and silencer elements) play a crucial role in maintaining normal splicing. Mutations in these elements have led to the development of many human diseases, such as Alport syndrome, Gitelman syndrome, Bartter syndrome, Cystinosis[7–11]. The impact of these mutations on splicing patterns can be verified through minigene splicing assays or by analyzing RNA samples derived from patients. Currently, a growing number of SNVs have been confirmed to impact RNA splicing. Approximately 8.6% of human pathogenic variants documented in the HGMD (released in 2021.4) are known to affect pre-mRNA splicing. However, the actual number of splicing variants is probably underestimated. Thus, it is essential to integrate RNA-level analyses into routine genetic disease research. The minigene assay, used to confirm whether potential splicing variants affect splicing patterns, has been widely validated for its reliability and effectiveness[12,13].

Mutations affecting pre-mRNA splicing

Variants positioned close to canonical splice sites can disrupt exon definition by reducing the specificity or accuracy of splice site selection[14].

Firstly, the 5′ splice site (5'ss) consists of a 9-bp consensus sequence (CAG/GUAAGU) found at exon/intron boundaries. The 5′ end of U1 snRNA recognizes the 5'ss through base pairing[15]. Not all base pairs at different 5′ss positions are equally important, with their splicing contribution generally correlating with their conservation. The most conserved positions are the first two intronic nucleotides (+1 and +2), determining the 5'ss subtype. The next most conserved positions are −1G and +5G, forming strong G-C base pairs with U1 due to three hydrogen bonds. In our study, variant c.1834G > A is located at the last nucleotide of the exon 22, which weakened recognition of natural 5′ss, concurrently activated cryptic 5′ss, upstream of the natural DS, leading

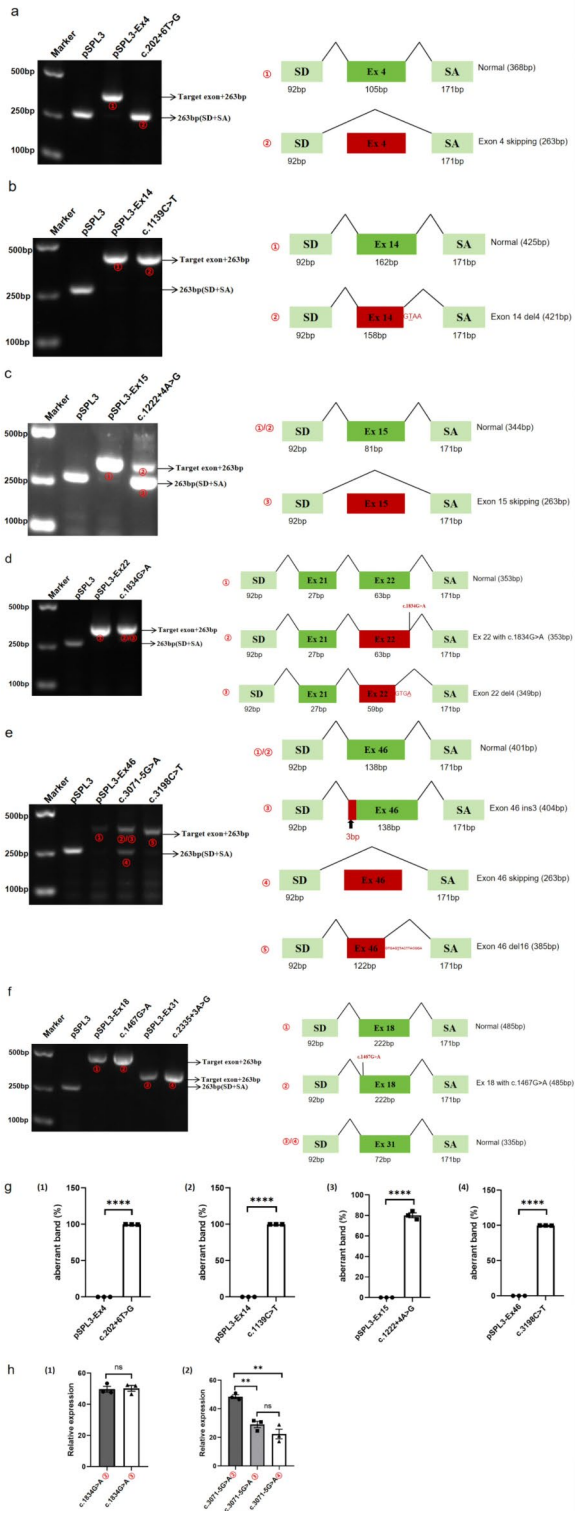

Fig. 1. Agarose gel electrophoresis and statistical analysis of RT-PCR and qPCR expressed from the *COL17A1* minigenes products. The percentage of aberrant band (%) = (aberrant band/all bands) ×100. Error bars represent SEM ($n = 3$). unpaired Student's t-test. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; ns: no significance; "_": mutation site.

to the production of two transcripts with nearly equal expression levels. The use of both alternative 5′ss may depend on their sequences. Consensus nucleotides − 2 A, +3 A, +4 A, and + 6U also contribute to the strength of the 5′ss through two hydrogen bonds each. The variants c.202 + 6T > G and c.1222 + 4 A > G were + 6U and + 4 A, respectively, leading to the disruption of the normal DS. While − 3 C forms a C-G base pair with U1 but is less

conserved and contributes less to splicing. However, in this study, the c.1139 C > T mutation at − 3 C still affected splicing, leading to the creation of a new DS, suggesting that other mechanisms may also influence splicing.

Furthermore, the 3′ splice site (3′ss) is comprised of the branch point site (BPS), the polypyrimidine tract, and the 3′-AG dinucleotides[16]. The mechanism by which mutations cause aberrant 3′ss splicing is more complex. As predicted, the splicing outcome caused by the c.3071-5G > A mutation at the 3'ss is also more diverse. This mutation produced a normal transcript, while simultaneously created a new "AG" dinucleotide competing with the natural one, additionally generated a transcript with complete skipping of the exon 46. The expression of the normal transcript is much higher than that of two abnormal transcripts. This finding suggests a preference for recognizing downstream "AG" over bypassing upstream "AG", possibly related to the preference for G at positions following AGs in competitive scenarios between closely spaced AGs.

Equally important, numerous regulatory elements within the exon, such as ESEs and ESSs, either promote or inhibit the recognition of adjacent splice sites by recruiting various protein factors[17]. In this study, variants c.1139 C > T and c.3198 C > T were identified to alter pre-mRNA splicing, which were predicted to make significant alteration of ESE/ESS motifs ratio (− 6, − 5) in silico analysis. The two variants lead to the disruption of multiple ESEs and the creation of multiple ESSs, thus detrimentally affecting splicing.

Splicing alterations and pathogenic mechanisms

The *COL17A1* gene comprises 56 exons that encode a 1497-amino acid protein, forming a homotrimer made up of three alpha (α1) chains, each with a molecular mass of 180 kDa. Each chain features a globular amino-terminal intracellular endodomain, a short transmembrane domain, and a rod-shaped carboxy-terminal extracellular ectodomain (Supplementary Fig. S5). The ectodomain contains 15 collagenous sequences (COL1–COL15) with repeating Gly-X-Y tripeptides, which assemble into the typical collagen triple helices within the homotrimer. These collagenous sequences are flanked by 16 non-collagenous sequences (NC1–NC16)[18]. Variants c.202 + 6T > G, c.1139 C > T, c.1222 + 4 A > G all led to absence of amino acids in NC16C, which is an intracellular domain that may be involved in signal transduction and cell adhesion. The c.1834G > A mutation occurs in the extracellular domain of COL15, disrupting the repeating Gly-X-Y triplet structure and leading to protein truncation. Variants c.3071-5G > A and c.3198 C > T are located within the NC domain of the extracellular region. Specifically, c.3071-5G > A resulted in the insertion of 1 amino acid and the absence of 46 amino acids in the NC10 domain, while c.3198 C > T produced a truncated protein in the NC9 domain. The extracellular domain plays a crucial role in binding to the skin matrix, and mutations in this region can lead to abnormal binding to the skin matrix, thereby causing disease.

RNA splicing, as a crucial process in human gene expression, not only produces diverse mRNA isoforms but also influences protein biosynthesis. Therefore, we advocate for increased focus on their pathogenicity and potential therapeutic strategies. Antisense oligonucleotide (ASO) therapies targeting exon skipping have demonstrated promising efficacy in Junctional Epidermolysis Bullosa[19,20], which highlighted the potential of ASO-mediated splice modulation for treating. The exploration and application of ASOs hold significant promise for the future.

In addition, this study's limitations include that minigene analysis may not fully capture the complexities of regulatory mechanisms in vivo. It is recommended to analyze RNA samples from relevant tissues of patients to better elucidate the pathogenicity of variants.

In conclusion, our results demonstrated that the influence of SNVs in the *COL17A1* gene on splicing using bioinformatics tools and minigene assays. Previously presumed missense variants (c.1139 C > T, c.1834G > A), synonymous mutation (c.3198 C > T), and three intronic variants (c.202 + 6T > G, c.1222 + 4 A > G, c.3071-5G > A) were confirmed to alter pre-mRNA splicing. Thus, we recommend considering splicing alterations as part of the pathogenic mechanism of these variants, which has implications not only for diagnosis but also for therapeutic strategies.

Data availability

The datasets generated, used, and analysed during this study are available from the corresponding author on reasonable request, also available in the GenBank repository [GenBank accession numbers: PQ815856, PQ815857, PQ815858, PQ815859, PQ815860, PQ815861, PQ815862].

Received: 12 August 2024; Accepted: 24 March 2025
Published online: 03 April 2025

References

1. Asaka, T. et al. Type XVII collagen is a key player in tooth enamel formation. *Am. J. Pathol.* **174**, 91–100 (2009).
2. Danpanichkul, P. et al. Global epidemiology of alcohol-associated liver disease in adolescents and young adults. *Aliment Pharmacol. Ther.* apt.18101 (2024).
3. Walker, L. C. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: recommendations from the ClinGen SVI splicing subgroup. *Am. J. Hum. Genet.* **6**, 1046–1067 (2023).
4. Cifuentes, L. et al. A case of junctional epidermolysis Bullosa with prurigo-like lesions and reduction of collagen XVII and Filaggrin: correspondence. *Br. J. Dermatol.* **169**, 195–198 (2013).
5. Dinckan, N. et al. Whole-Exome sequencing identifies novel variants for tooth agenesis. *J. Dent. Res.* **97**, 49–59 (2018).
6. Jonsson, F. et al. Mutations in collagen, type XVII, alpha 1 (*COL17A1*) cause epithelial recurrent erosion dystrophy (ERED). *Hum. Mutat.* **36**, 463–473 (2015).
7. Wong, M. S., Kinney, J. B. & Krainer, A. R. Quantitative activity profile and context dependence of all human 5′ splice sites. *Mol. Cell.* **71**, 1012–1026e3 (2018).
8. Li, C. et al. Functional analysis of the CTNS gene exonic variants predicted to affect splicing. *Clin. Genet.* **105**, 323–328 (2024).
9. Okada, E. Aberrant splicing caused by exonic single nucleotide variants positioned 2nd or 3rd to the last nucleotide in the COL4A5 gene. *Clin. Exp. Nephrol.* **27**, 218–226 (2023).

10. Shi, X. et al. Minigene splicing assays reveal new insights into exonic variants of the SLC12A3 gene in Gitelman syndrome. *Mol. Genet. Genom. Med.* **11**, e2128 (2023).
11. Xin, Q. et al. Twelve exonic variants in the SLC12A1 and CLCNKB genes alter RNA splicing in a minigene assay. *Front. Genet.* **25**, 961384 (2022).
12. Blakes, A. J. M. et al. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 genomes project. *Genome Med.* **14**, 79 (2022).
13. Truty, R. et al. Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *Am. J. Hum. Genet.* **108**, 696–708 (2021).
14. Aoto, Y. et al. Last nucleotide substitutions of COL4A5 exons cause aberrant splicing. *Kidney Int. Rep.* **7**, 108–116 (2022).
15. Roca, X., Krainer, A. R. & Eperon, I. C. Pick one, but be quick: 5′ splice sites and the problems of too many choices. *Genes Dev.* **27**, 129–144 (2013).
16. Smith, C. W., Chu, T. T. & Nadal-Ginard, B. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell. Biol.* **13**, 4939–4952 (1993).
17. Zhang, R. et al. Identification of seven exonic variants in the *SLC4A1*, *ATP6V1B1*, and *ATP6V0A4* genes that alter RNA splicing by minigene assay. *Hum. Mutat.* **42**, 1153–1164 (2021).
18. Giudice, G., Emery, D. & Diaz, L. Cloning and primary structural analysis of the bullous pemphigoid autoantigen BP180. *J. Investig. Dermatol.* **99**, 243–250 (1992).
19. Ablinger, M. et al. Personalized development of antisense oligonucleotides for exon skipping restores type XVII collagen expression in junctional epidermolysis bullosa. *IJMS.* **22**, 3326 (2021).
20. Hainzl, S. et al. Splicing modulation via antisense oligonucleotides in recessive dystrophic epidermolysis Bullosa. *IJMS.* **25**, 761 (2024).

Author contributions
R.Z. conceived and designed and the experiments. Y.F.S. conducted experiments, conducted in silico prediction of variants and the data analysis. Y.F.S. wrote the manuscript. R.Z. revised the manuscript. All authors read and approved the final manuscript.

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

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95851-9.

Correspondence and requests for materials should be addressed to R.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.