Identification of eight exonic variants in the SLC4A1, ATP6V1B1 and ATP6V0A4 gene that alter RNA splicing by minigene assay

Ruixiao Zhang[1], Zeqing Chen[2], Qijing Song[3], Sai Wang[1], Zhiying Liu[1], Xiangzhong Zhao[4], Xiaomeng Shi[1], Wencong Guo[4], Yanhua Lang[1], Irene Bottillo[5], and Leping Shao[1]

[1]Qingdao Municipal Hospital Group
[2]Fudan University
[3]People's Hospital of Jimo District
[4]The Affiliated Hospital of Qingdao University
[5]IRCCS-Casa Sollievo della Sofferenza Hospital

February 7, 2021

Abstract

Primary distal renal tubular acidosis (dRTA) is a rare tubular disease associated with variants in SLC4A1, ATP6V0A4, ATP6V1B1, FOXI1 or WDR72 genes. Currently, there is growing evidence that all types of exonic variants can alter splicing regulatory elements, affecting the pre-mRNA splicing process. This study was to determine the consequences of variants associated with dRTA on pre-mRNA splicing combined with predictive bioinformatics tools and minigene assay. As a result, among the 15 candidate variants, 8 variants distributed in SLC4A1 (c.1765C>T, p.Arg589Cys), ATP6V1B1(c.368G>T, p.Gly123Val; c.370C>T, p.Arg124Trp; c.484G>T, p.Glu162* and c.1102G>A, p.Glu368Lys) and ATP6V0A4 genes (c.322C>T, p.Gln108*; c.1571C>T, p.Pro524Leu and c.1572G>A, p.Pro524Pro) were identified to result in whole or part of exon skipping by either disruption of ESEs and generation of ESSs, or interference with the recognition of the classic splicing site, or both. To our knowledge, this is the first study on pre-mRNA splicing of exonic variants in the dRTA-related genes. These results highlight the importance of assessing the effects of exonic variants at the mRNA level and suggest that minigene analysis is an effective tool for evaluating the effects of splicing on variants in vitro

Introduction

Precursor messenger RNAs (Pre-mRNAs) composed of exons and introns are transcribed from the protein-coding genes of most eukaryotic cells[1]. Exons, consisting of two terminal untranslated regions and protein-coding regions, are interrupted by non-coding insertion segments called introns. Mature translatable mRNAs are generated by pre-mRNAs splicing, involving the removal of introns and the connection of exons, which is an important process of gene expression. Splicing is catalyzed by the spliceosome, a large ribonucleoprotein (RNP) complex constituted by 5 small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6, and many other associated non-snRNP proteins, that can recognize a large number of splicing signals in pre-mRNA and stimulate intron removal[2,3].

The splicing signals include the conserved splice sites (5' splice donor site (5'ss), and 3' splice acceptor site (3'ss)), the branch site, polypyrimidine track, exonic/intronic splicing enhancers (ESEs/ISEs) and silencers (ESSs/ISSs), as well as other regulatory components or the RNA secondary structure[4,5]. All these factors cooperate with splicing factors and the spliceosome, to accurately remove introns and join exons. In the past, only intronic mutations affecting donor or acceptor splice sites (DS or AS) were considered to potentially alter transcriptional processing. Currently, given the importance of core and auxiliary splicing signals in the pre-mRNA splicing process, it has been suggested that up to 50% of exonic variants including all types

(missense, nonsense, and small insertions or deletions) that alter splicing regulatory elements, may disrupt the splicing pathway and cause various diseases in human[6,7].

Primary distal renal tubular acidosis (dRTA) is a rare tubular disease characterized by impaired tubular secretion of hydrogen ions in the distal nephron, resulting in normal serum anion gap metabolic acidosis that often progresses to hypokalemia, nephrocalcinosis, and nephrolithiasis[8]. The vast majority of primary dRTA patients are associated with variants in *SLC4A1* , *ATP6V0A4* and *ATP6V1B1* genes, transmitted as either an autosomal dominant (AD) or recessive (AR) trait. The*SLC4A1* gene, located on chromosome 17q21, contains 20 exons and encodes a membrane protein the chloride–bicarbonate exchanger composed of 911 amino acids, also known as AE1 or band 3 protein, which is primarily responsible for the reabsorption of HCO_3^- coupled with the urinary excretion of chloride[9]; the gene *ATP6V1B1* is found on the 2p13 chromosome with a total of 14 exons, which codifies the B1 subunit of the V-ATPase in the α-intercalated cells of the renal collecting duct and endolymphatic sac epithelia[10,11]; the *ATP6V0A4* gene is located on chromosome 7q33-34 and includes 23 exons, of which 20 encode transmembrane A4 subunits, which may be involved in H^+translocation or transport and/or assembly of the H^+-ATPase[12]. In addition, recent literatures reported that variants in*FOXI1* [13], *WDR72* [14]and*ATP6V1C2*gene[15] were found in primary dRTA patients.

DNA sequence analysis has been now widely used in the clinical diagnosis of related genetic diseases, resulting in the identification of thousands of rare variations in affected and unaffected individuals. The number of these variants is expected to increase rapidly with the application of high-throughput sequencing technology based on large-scale parallel sequencing. So far, a total of 201 different dRTA-related gene variants have been described in the Human Gene Mutation Database (HGMD, Professional 2019.4), including 29 in *SLC4A1* , 73 in*ATP6V1B1* , 94 in *ATP6V0A4* , 2 in *FOXI1* and 3 in*WDR72* . Among them, missense/nonsense variants account for about 50% (101/201) of all variants. However, most variants analysis was performed to assess the effects on mRNA and protein at the genome level, but only in a few cases at both DNA and RNA levels[16].

There are few studies on the effect of splicing in exonic variants of pathogenic genes associated with primary dRTA. The purpose of this study was to gain insight into the effects of the previously described dRTA-related pathogenic variants including missense and nonsense variants on pre-mRNA splicing using a minigene technology.

Materials and Methods

Variant Nomenclature

DNA variant numbering is based on the cDNA sequence for *SLC4A1* (GenBank NM_000342.3), *ATP6V1B1* (GenBank NM_001692.3), *ATP6V0A4*(GenBank NM_020632.2), *FOXI1* (GenBank NM_-012188.4) and*WDR72* (GenBank NM_182758.3). The variant nomenclature was described according to the guidelines of the Human Genome Variation Society (*http://varnomen.hgvs.org*), with numbering starting at the first position of the translation start codon.

Variant Collection and In Silico Analyses

All point variants in *SLC4A1* , *ATP6V1B1* , *ATP6V0A4* ,*FOXI1* , *WDR72* and *ATP6V1C2* gene were collected from the Human Gene Mutation Database (April, 2019) and literatures[15,17,18], except c.409C>T (p.Pro137Ser) and c.904C>T (p.Arg302Trp) in *ATP6V1B1* that was identified in our previous studies[19].

Bioinformatics analysis was carried out to predict the splicing effect of each of the variants. Online software Human Splice Finder (version 3.1, http://www. umd.be/HSF3) was performed to investigate the presence of potential splicing regulatory sequences (ESEs and ESSs), and identify the putative effect of variants on splicing regulatory motifs. Analysis of BDGP (*http://www.fruitfly.org*) was employed to determinate the potential effects of variation on classic 5' donor or 3' acceptor consensus sites and to predict the generation and/or activation of new sites. Default thresholds were performed for all analyses. Variants that satisfy the

following criteria will be selected for further minigene splicing assay: elimination of enhancers or creation of silencers; and close to the 5' or 3' ends of exons.

Minigene Constructions

To investigate the effect on the splicing process of the candidate variants, *in vitro* analysis was performed using a minigene splicing assay based on the pSPL3 exon trapping vector shown at Figure 1 A. Minigene constructions were described as previously reported[20,21]. Briefly, for each variation, the fragments with the wild-type (WT) alleles involving interested exon, flanked by approximately 50-200 nucleotides of upstream intronic sequence and downstream intronic sequence, were cloned into the splicing vector pSPL3 using specific primers linking the XhoI and NheI restriction enzyme sites (XhoI: TGGAGCˆTCGAG; NheI: AATTTGˆCTAGC). Primers were designed for each target fragment using PP5 (SUP Table. 1). The variant sequence change was introduced in the WT plasmid through the GeneArt Site-Directed Mutagenesis PLUS System (Thermo Fisher Scientific, Massachusetts, USA) as instructed by the manufacturer. Mutagenesis primers were demonstrated at SUP Table. 2. All constructed vectors were further transformed into E. coli DH5α competent cells (Takara, Shiga, Japan), followed by screening by conventional sanger sequencing.

Transfection of HEK 293 T Cells

Human epithelial kidney 293 T (HEK 293 T) cells were cultured in DMEM medium containing 10% fetal bovine serum (FBS), penicillin (100 U/L), and streptomycin (100 mg/L) at 37 °C in a 5% CO_2. One day before transfection, cells were transferred to 12-well culture plate to grow to approximately 70% to 80% confluence in an antibiotic free medium. Cells were then transfected with 2 μg plasmid DNA in each group of minigenes (empty pSPL3-control, pSPL3-WT and pSPL3-Mutation (MUT)) using OPTI-MEM® IMedium and Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions.

RT-PCR of Minigenes

After 24 h incubation, total RNA was extracted using TRIzol reagent (Invitrogen, USA) and used for RT-PCR to confirm the splicing patterns. First-strand cDNA was synthesized from 2 μg of total RNA by random-primed reverse transcription with Superscript II Reverse Transcriptase (Invitrogen Corporation, Carlsbad, CA). To evaluate the pattern of transcripts from the transfected minigenes, the following vector-specific primers were used for RT-PCR amplification: a forward primer SD6 (5'-TCTGAGTCACCTGGACAACC-3') and a reverse primer SA2 (5'-ATCTCAGTGGTATTTGTGAGC-3').

The PCR amplification reaction was performed as follows: in 50 μl volume, 2 μl of cDNA, 10 μL of 5 × PrimerSTAR Buffer (TaKaRa, Japan), 1 μM of each primer, 0.8 μlM dNTPs, and 0.5 μL PrimerSTAR HS DNA Polymerase (TaKaRa, Japan) in a 9700 (Applied Biosystem, Foster City, CA, USA) thermal cycler. Thermal conditions were 33 cycles of 94 °C for 30 seconds, 58 °C for 30 seconds, and 72 °C for 90 seconds, and followed by a final elongation step at 72 °C for 10 minutes. The PCR products were separated by electrophoresis on a 1.5% agarose gel, and each band signal was quantified by software Image J. The target DNA bands were purified using a Gel Extraction Kit (CWBIO, China), and then all transcripts were analyzed by sequencing.

Statistical analysis

The percentage of exon exclusion (%) was calculated as (lower band / [lower band + upper band]) × 100. Statistical analysis was performed using SPSS software. The results were analyzed using the two-tailed Student's t-test or One-way ANOVA test by GraphPad Prism (Version 6.02, GraphPad Software, USA). Error bars represent SEM (n=3). A P value < 0.05 was considered statistically significant.

3 Results

A total of 113 different variants, including 20 of *SLC4A1* , 39 of*ATP6V1B1* , 48 of *ATP6V0A4* , 2 of FOXI1, 3 of WDR72, 1 of ATP6V1C2 gene, were collected for analysis by the bioinformatics software. Potential splicing variants within three bases of the 5 'or 3' end of the exon, or predicted that affect splicing regulatory elements (elimination of enhancers or creation of silencers) were selected for further minigene

assay to determine whether such variants affect splicing. Some selected variants had combined effects that both generate ESS and broke ESE (the total number of ESEs disruption and ESSs generation is more than 6 for SLC4A1, 7 for ATP6V1B1 and ATP6V0A4). From the results of these in silico analyses, the distribution of these candidate variants enrolled in this study was as follows: 3 in*SLC4A1*(c.1437C>G, c.1564G>A and c.1765C>T), 6 in *ATP6V1B1*(c.368G>T, C.370C>T, c.481G>A, c.484G>T, c.1102G>A and c.1251C>G) and 5 in *ATP6V0A4* (c.52C>T, c.322C>T, c.1571C>T, C.1572G>A, c.2190C>G and c.2257C>T) as shown in Table 1, and Figure 1 B-D. Different control minigenes were generated comprising*SLC4A1* WT sequences of exons 13 (pSPL3-*SLC4A1* Ex13) and 14 (pSPL3-*SLC4A1* Ex14), *ATP6V1B1* of exons 5 (pSPL3-*ATP6V1B1* Ex5), 6 (pSPL3-*ATP6V1B1* Ex6), 11 (pSPL3-*ATP6V1B1* Ex11), and 13 (pSPL3-*ATP6V1B1* Ex13), and*ATP6V0A4* of exons 3 (pSPL3-*ATP6V0A4* Ex3), 6 (pSPL3-*ATP6V0A4* Ex6), 15 (pSPL3-*ATP6V0A4* Ex15), and 20 (pSPL3-*ATP6V0A4* Ex20) respectively. With the corresponding WT minigenes as the template, all candidate variants minigenes successfully generated through site-directed mutagenesis, except c.1251C>G in *ATP6V1B1* and c.2257C>T in*ATP6V0A4* was not introduced into the pSPL3-WT minigene due to limitation of experimental technique.

3.1 Splicing outcome of sequences variations of *SLC4A1*

3.1.1 Variant c.1765C>T (p.Arg589Cys) induced skipping of exon 14 compared with the WT plasmids.

Variant c.1765C>T (p.Arg589Cys), located at nucleotide position -36 from the 3' end of exon 14, was predicted that the variant not only eliminates four ESEs (**C** GCAAGT, G**C** GCAA, TG**C** GCA, TGCTG**C**) but also generate two overlapping ESSs (CTG**T** GC, TGCTG**T**) by bioinformatics analysis with HSF (affected nucleotide is bold, shown at Table 1). To investigate the effect of c.1765C>T on pre-mRNA splicing, the WT (c.1765C) and mutant (c.1765C>T) minigene were created and were transfected separately into 293T cells. The result of the minigene assay revealed that two different electrophoresis bands were detected in both of WT and mutant minigene: a larger band of 400 bp and a smaller band of 263 bp (Figure 2A). Sequencing analysis showed that the larger splicing product contained exon 14 of *SLC4A1* , whereas the smaller did not contain exon 14. The absence of exon 14 resulting in the loss of 174 bp does not alter the open reading frame. Analysis of cDNA extracted from the transfected HEK293T revealed that the amounts of the exon 14-skipping transcript of c.1765C>T were significantly increased with those of the WT plasmid (83.03% versus 43.93%, $P<0.05$), suggesting that the variant c.1765C>T disturbed the normal splicing in vitro (Figure 2D).

3.1.2 Variants of c.1437C>G (p.Cys479Trp) and c.1564G>A (p.Glu522Lys) in exon 13 did not alter pre-mRNA splicing

Variant c.1437C>G (p.Cys479Trp), located at nucleotide position +6 from the 5' end of exon 13, was predicted to lead to the disruption of two ESEs and the generation of four new ESS sites using bioinformatic tool HSF. Presumed missense variant c.1564G>A (p.Glu522Lys) is located 133 bp upstream of exon 13 (Table 1). The results of HSF analysis indicated that this variant inactivated four potential ESE sites and generated two potential ESSs. However, analysis of the minigenes containing variant c.1437C>G and c.1564G>A in exon 13 of *SLC4A1* resulted in RT-PCR products that matched in size with those generated by the respective WT constructs (Figure 2A). This was further confirmed by sequencing analysis. Therefore, these variants did not affect pre-mRNA splicing.

3.2 Splicing outcome of sequences variations of ATP6V1B1

3.2.1 Variants c.368G>T (p.Gly123Val) andc.370C>T (p.Arg124Trp) led to exon 5 skipping compared with the WT plasmids.

Presumed missense variant c.368G>T (p.Gly123Val) is caused by the first nucleotide substitution in exon 5 of*ATP6V1B1* . Bioinformatic analysis with BDGP demonstrated that this variant drastically reduces the score of the WT 3' splice site from 0.96 to 0.76 (Table 1). Variant c.370C>T (p.Arg124Trp) result from substitutions at the +3 nucleotide of exon 5 and was demonstrated that this variant marginally reduces the score of the WT 3' splice site from 0.96 to 0.94 with BDGP. Additionally, analysis of HSF to two variants

predicted that they both generated three ESSs. To determine the splicing effect of variants c.368G>T and c.370C>T, a minigene containing exon 5 and flanking intronic sequences was used. The result of RT-PCR analysis showed that the WT lane demonstrated 2 different fragments of 263 bp and 368 bp, respectively (Figure 2B). Direct sequencing showed that the larger fragment contained exon 5 flanked by two exons of the pSPL3 vector, while the smaller one included only the 3' and 5' pSPL3 exons. In contrast, the mutant lane of c.368G>T revealed one unique fragment of 263 bp corresponding to skipping of exon 9 in the mRNA; whereas, the result of variant c.370C>T revealed two different electrophoresis bands whose sizes correspond to the WT products. Analysis of cDNA prepared from HEK293T showed that there was a significant increase of exon 5-skipping transcript of c.370C>T compared with the WT plasmids (28.19% versus 4.78%, $P<0.05$, Figure 2D). Deletion of exon 5 in the mutant transcript would result in the abnormal connection between exons 3 and 4, leading to the loss of 26 amino acids at the protein level without changing the open reading frame. Taken together, both of variants c.368G>T and c.370C>T broken proper recognition of the acceptor splicing site, resulting in exon 6 skipping of *ATP6V1B1* .

3.2.2 Nonsense variant c.484G>T (p.Glu162*) led to the skipping of exon 6 .

Variant c.484G>T of the internal position of exon 6 in*ATP6V1B1* alters a GAG codon for Glu to a premature TAG stop codon ((p.Glu162*), which is predicted to produce a truncated and nonfunctional protein (Table 1). However, the *in silico* analysis by HSF software indicated that c.484G>T not only disrupts five ESEs (CCCGAG**G** A, CCGAG**G** , CCGAG**G** AG, CGAG**G** A, GAG**G** AG), but also creates two ESSs (G**T** AGAT, **T** AGATG). Besides, HSF also predicted that this variant would cause the activation of a cryptic donor site (CGAG**T** AGAT). To further clarify the impact of variant c.484G>T on the splicing process, we inserted either the WT or c.484G>T-mutated exon 6, along with the nearby intron sequences, into the vector pSPL3, and analyzed the spicing pattern in HEK293T. The WT minigene resulted in two different transcripts corresponding to a mature mRNA and a truncated mRNA (1.22%) by sequencing analysis. The mutated minigene, also, produced two mature transcripts: a larger one corresponded to the single band produced by the WT, and a shorter one with the splicing of exon 6 (Figure 2B). Because exon 6 has 140 nucleotides, loss of exon 6 would disrupt the reading frame of the *ATP6V1B1* transcript. Quantitative analysis of the RT-PCR products revealed that the rate of transcript lacking exon 6 to the full transcripts was 60.35% ($P<0.05$, Figure 2D). Taking these findings together, the splicing pattern of the mutant minigene indicated that nonsense variant c.484G>T induces exon 6 skipping.

In addition, although variant c.481G>A (p.Glu161Lys) close to c.484G>T was also predicted by the software HSF to destroy five ESEs and produce two ESSs, it did not affect splicing through the minigene analysis ($P>0.05$).

3.2.3 Variant c.1102G>A (p.Glu368Lys) resulted in partial skipping of exon 11.

Variant c.1102G>A identified at nucleotide position 43 of exon 11 in *ATP6V1B1* were predicted by HSF to broke seven ESEs (CACA**G** A, CACA**G** AG, ACA**G** AGGG, CA**G** AGG, CA**G** AGGG, **G** AGGGA) and create one ESS site (A**A** AGGG) corresponding to a new hnRNPA1 binding site (Table 1). In order to verify whether this variant affected mRNA splicing, minigene splicing experiments *in vitro* were also carried. As a result, two products of 346 bp and 236 bp were both found from the RT-PCR products of the WT minigene and mutant (Figure 2B). Direct sequencing of two products showed that the larger amplicon was the transcript containing exon 11, and the smaller was the transcript excluding exon 11. Quantitative analysis of two products showed that c.1102G>A altered weakly splicing resulting in a 9.51% of the aberrant transcript that broke the open reading frame compared with WT of 2.60% ($P<0.05$, Figure 2D). In addition, the skipping of exon 11 in the mutant transcript would result in the loss of 83 nucleotides that disrupt the reading frame.

3.3 Splicing outcome of sequences variations of*ATP6V0A4*.

3.3.1 Variant c.322C>T (p.Gln108*) resulted in skipping of exon 6.

Variant c.322C>T, as a nonsense variant (p.Gln108*) located at nucleotide position +31 from the 5' end of

exon 6, is predicted to generate a truncated and nonfunctional protein (Table 1). The software HSF demonstrated that c.322C>T not only destroyed six ESEs (**C** AGGAA, TA**C** AGGA, TA**C** AGGAA, TTA**C** AG, GTTA**C** A, AGTTA**C**), but also engenders three ESSs (TA**T** AGG, TTA**T** AG, AGTTA**T** A). Consequently, two fragments of 389 bp and 263 bp were detected from that of the mutant vector by minigene assay, respectively (Figure 2C). Direct sequencing results showed that the larger fragment contained *ATP6V0A4* exon 6 flanked by two exons of the pSPL3 vector, while the smaller one included only the 3' and 5' pSPL3 exons (24.22%). However, the WT vector also revealed a smaller fragment (1.78%) of 263 bp corresponding to incorrect skipping of exon 6 in *ATP6V0A4* expect for a mature transcript (Figure 2D). The skipping of exon 6 in the mutant transcript would result in the loss of 42 amino acids (residues 98-139) at the protein level without altering the open reading frame. Therefore, variant c.322C>T (p.Gln108*) causes partial exon 9 skipping because of ESEs destruction and ESSs generation.

3.3.2Missense variant c.1571C>T (p.Pro524Leu) and synonymous variant c.1572G>A (p.Pro524Pro) resulted in skipping of exon 15.

Missense variant c.1571C>T (p.Pro524Leu) caused by substitutions at the -2 nucleotide of exon 15 downstream of the 3' splice site, was demonstrated that this variant reduces the score of the WT donor splice site of intron 15 from 0.8 to 0.68 with BDGP (Table 1). Synonymous variant c.1572G>A (p.Pro524Pro) affected the G the last nucleotide substitution in exon 15. Bioinformatic analysis with BDGP demonstrated that the score of the donor site of intron 15 is 0.8, whereas it could not be analyzed after variation. Additionally, analysis of this variant with HSF predicted to break the WT donor sites (CC**G** GTAATA), most probably affecting splicing. Taken together, to examine the splicing effect of two variants, we also used a minigene containing exon 15 and surrounding intronic sequences. RT-PCR analysis results showed the splicing products produced by the mutant and WT minigenes were different. The WT lane demonstrated one fragment of 357 bp that contains *ATP6V0A4* exon 15, whereas both of mutant c.1571C>T and c.1572G>A generated two different fragments of 263 bp and 357 bp, respectively (Figure 2C). Direct sequencing of all products showed that the larger amplicons were the exons-included transcripts and the smaller amplicons are the exons-excluded transcripts. Analysis of cDNA prepared from HEK293T revealed that there was a significant increase of exon 15-skipping in c.1571C>T and c.1572G>A with the control plasmid (3.74% versus 0 and 84.48% versus 0, respectively, Figure 2D). Besides, the skipping of exon 15 in the mutant transcript would result in the alteration of the open reading frame due to the loss of 94 nucleotides.

Discussion

Normal pre-mRNA splicing depends on accurate recognition of exons and introns, so many variations that break the correct definition of introns and exons will lead to defects in pre-mRNA splicing, causing or modifying the severity of many genetic diseases[4]. It is currently clear that these variations affect splicing not only by disrupting classic splice sites (DS or AS), but also by altering other important RNA elements (splicing enhancers or silencers binding sites) or by creating new splice sites[4,5,22,23].

Traditionally, exonic variants outside the conserved splice site dinucleotide have been classified as missense, nonsense or synonymous variants when gene analysis is performed only at the DNA level. Currently, it is widely believed that RNA analysis should be performed to assess the pathogenicity of sequence variations, especially splicing abnormalities caused by variants affecting the splicing regulatory sequences. In many cases, however, RNA samples from the affected individual are unavailable, and mRNAs are difficult to detect due to the activation of the nonsense-mediated mRNA decay pathway[24]. Fortunately, a functional splicing assay based on the minigene assay was confirmed as an effective, reliable and relatively simple tool to functionally assay potential splicing[25], which has been extensively validated in our previous studies[20,21,26].

Single-base substitutions in the *SLC4A1* , *ATP6V1B1* ,*ATP6V0A4* , *FOXI1* , *WDR72* and *ATP6V1C2* gene, many of which are predicted to lead to missense, nonsense or synonymous variants, are often identified in patients with primary dRTA. Herein, we assumed that the pathogenic effect of some of these variants is the alteration of pre-mRNA splicing. Using bioinformatics tools, a total of 15 single-base variants in *SLC4A1* , *ATP6V1B1* and*ATP6V0A4* gene were selected and tested with a minigene assay. These variants

were predicted to modify splicing regulatory sequences (disruption of ESEs or creation of ESSs), generate new splice sites or induce a significant reduction of splice site strength. Minigene plasmids ligated to either WT or mutant genomic sequences were constructed. The constructs were transfected into HEK293T cells, and the mRNA was analyzed by RT-PCR and DNA sequencing. Consequently, 8 of them were found to cause exon skipping.

Transcript analysis by quantitative PCR of four variants of c.1765C>T in *SLC4A1* , c.481G>T and c.1102G>A in *ATP6V1B1* , and c.322C>T in*ATP6V0A4* gene, located in the internal position of the exon, have shown that they affect pre-mRNA splicing by causing a significant imbalance in the proportion of ESEs/ESSs. Many splicing regulatory sequences, including ESEs and ESS, can promote or inhibit the recognition of surrounding splicing sites by recruiting different protein factors, thus coordinating to regulate the correct splicing of exons[27]. Therefore, an improper ESE/ ESS ratio will prevent correct preRNA splicing. Variant c.1765C>T (p.Arg589Cys) in *SLC4A1*gene was reported to cause dRTA in the form of autosomal dominant inheritance[28]. In the previous study, the Arg509 was considered as a mutation hotspot, which is located in the intracellular domain between the 6th and the 7th transmembrane regions of the AE1 protein and at the apex of the cytoplasm of the 6th transmembrane region, and two other forms of the variant in the same site (p.Arg589His, p.Arg589Ser) have been confirmed to be pathogenic[29,30]. Variants at p. Arg589 site may lead to failure to completely transported to the cell membrane of AE1 protein, or the defective loading of the advanced structure after transferring to the cell membrane, resulting in structural and functional abnormalities[28-30]. In this study, online software HSF was used to predict that c.1765C>T could cause a significant imbalance in the proportion of ESEs/ESSs. Further in vitro minigene analysis revealed that this variant produced part of transcripts lacking exon 14, which did not change the open reading frame. We hypothesized that it not only generated a transcript containing c.1765C>T but also produced a transcript lacking exon 14, resulting in abnormal morphogenesis of the encoded protein and leading to a phenotype of dRTA.

Nonsense variant, a single-nucleotide substitution, is generally considered to generate a premature stop codon. However, many studies have found that nonsense variants can occasionally alter exonic elements and affect the splicing pathway, which can be demonstrated by the loss of correlation between phenotype and genotype[30,31]. The result of Zhu's research has elucidated the detailed molecular mechanisms of exon skipping induced by nonsense mutations in the*DMD*gene[30]. In this context, variant c.484G>T in *ATP6V1B1* was predicted as nonsense mutation, p.Glu162*, and the minigene splicing assay for c.484G>T showed that it prevents exon 6 inclusion with a subsequent frameshift and premature termination at the 22th codon in exon 7 in the mRNA probably by affecting a functional ESE site and/or creating a functional ESS site. The bioinformatics analysis predicted the loss of five ESE and the gain of two ESS motifs with HSF. Moreover, missense variant c.1102G>A (p.Glu368Lys) in *ATP6V1B1*also generated a part of aberrant transcripts that lack exon 11, resulting in a subsequent frameshift and premature termination at the 25th codon in exon 13. We hypothesize that this is probably due to the disruption of seven functional ESE sites and the generation of a functional ESS site, which binds hnRNPA1. Similar to c.484G>T in *ATP6V1B1,* variant c.322C>T in *ATP6V0A4* was initially defined as nonsense variant (p.Gln108*) leading to an unable mRNA due to the nonsense-mediated mRNA decay. Our results showed that it influenced related ESEs and ESSs motifs and disturbed the normal splicing in vitro causing exon 6 skipping. Subsequently, the ligation of exons 5 and 7 would result in a lack of 42 amino acids and the production of a truncated protein. As a consequence, this mutated subunit A4 of apical H^+-ATP protein would destroy the first topological domain of cytoplasmic and may reduce or abolish the transport activity of H^+-ATPase.

Minigene assay allows us to categorize three exonic variants associated with dRTA, c.368G>T and c.370C>T in*ATP6V1B1* , c.1571C>T and c.1572G>A in*ATP6V0A4* , as splicing mutation. These variants located at or near to classical splice sites (DS or AS), may affect normal splicing either by decreasing the recognition efficiency of the 5'ss DS or 3'ss AS or by creating new splice sites. Sequence variations that occur in the conserved GT and AG dinucleotides flanking exons will cause splicing aberration, because the invariant GT and AG are required for the spliceosome to recognize 5's and 3'ss, respectively[1,2,25]. Variants c.368G>T and c.370C>T in *ATP6V1B1* involving the first and +3 nucleotide of exon 5, respectively, were initially

reported as missense mutations p.Gly123Val and p.Arg124Trp, respectively. Analysis of variants c.368G>T and c.370C>T in BDGP showed a reduction in the score of the 3' splice site. The minigene analysis of these two mutations revealed that they certainly caused the skipping of the exon 5.

Variants c.1571C>T and c.1572G>A in*ATP6V0A4* , located at the -2 and the last nucleotide of exon 15, were identified as missense (p.Pro524Leu) and synonymous variant (p.Pro524Pro), respectively. Such substitution usually reduces the recognition of classic splicing sites. Our BDGP analysis further showed that c.1571C>T resulted in the reduction of recognition strength of donor splicing site, while c.1572G>A did not. But HSF predicted that c.1572G>A destroyed the WT donor sites. Consequently, the minigene analysis revealed that these two variants certainly altered normal splicing by increasing approximately 3.74% and 84.48% exon 15 exclusion compared with WT, respectively. The skipping of exon 15 results in a loss of 94 nucleotides with a subsequent frameshift and premature termination at the 13th in exon 16. Significantly, synonymous substitutions are generally considered to be benign from a protein perspective. But from the mRNA's point of view, these substitutions may lead to abnormal splicing. Just as synonymous variant c.1572G>A of the*ATP6V0A4* in this study produced the aberrant transcript, the five synonymous variants were reclassified as splicing mutations using a minigene assay by Yohann Jourdy, which highlight that it is necessary to investigate the splicing impact of all synonymous substitutions to evaluate their clinical significance[2].

Among 8 variants, the PCR product of c.368G>T in*ATP6V1B1* was a unique transcript lacking exon 5, however, c.1765C>T in *SLC4A1* , c.370C>T, c.481G>T and c.1102G>A in *ATP6V1B1* , and c.322C>T, c.1571C>T and c.1572G>A in *ATP6V0A4* gene, still produced one transcript with the size of the WT product except that one RT-PCR product was the transcript missing corresponding exon, so they probably have a double damaging effect: while the some part of the transcript is deleterious due to the skipping of the exon, which would cause loss of part of some important domains or premature termination of codons due to frameshift, the remaining mRNA is damaging due to the mutated amino acid change, which could result in non-functional protein, truncated polypeptide or nonsense mediated mRNA decay.

In addition, variants of c.1437C>G and c.1564G>A in *SLC4A1* , c.481G>A in*ATP6V1B1* and c.52C>T, c.2190C>G in*ATP6V0A4* , were found to influence surrounding ESEs and ESSs motifs by the assessment of HSF or to cause the change of recognition strength of splicing site. However, results of our minigenes assay demonstrated that these variants did not affect pre-mRNA splicing, which was inconsistent with the predicted results of software, indicating that the results of the online software had a certain high sensitivity and low specificity. Also, we did not introduce these variants into cDNA and detect the cell surface expression and ion transport activity after AE1 or H^+-ATPase mutations, which require further study.

In conclusion, we have performed an extensive analysis of exonic variants in *SLC4A1* , *ATP6V1B1* and *ATP6V0A4* associated with dRTA using bioinformatics tools and mini-genes. This study allowed to reclassify 8 previously presumed missense, nonsense or synonymous variants as splicing variations, which should be taken their pathogenicity into account. These variants either disrupt ESEs and produce ESSs, or interfere with the recognition of the splicing site of 3 ' AS and a 5' DS, resulting in exon skipping. This study emphasized the importance of assessing the effect of exon point mutations at the mRNA level in the dRTA, especially under the condition of the failure to obtain the patients' RNA or kidney specimens, minigene assay may be a valuable tool.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests

Funding

This study was supported by grants from the Natural Science Foundation of China (81873594).

Acknowledgements

Not applicable.

Reference

[1] Smith SA, Lynch KW. Cell-based splicing of minigenes[J]. Methods in Molecular Biology, 2014, 1126: 243-255.

[2] Jourdy Y, Fretigny M, Nougier C, et al. Splicing analysis of 26 F8 nucleotide variations using a minigene assay[J]. Haemophilia, 2019, 25(2): 306-315.

[3] Matera AG, Wang Z. A day in the life of the spliceosome[J]. Nature Reviews: Molecular Cell Biology, 2014, 15(2): 108-121.

[4] De Conti L, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing[J]. Wiley Interdiscip Rev RNA, 2013, 4(1): 49-60.

[5] Dufner-Almeida LG, do Carmo RT, Masotti C, et al. Understanding human DNA variants affecting pre-mRNA splicing in the NGS era[J]. Advances in Genetics, 2019, 103: 39-90.

[6] Lopez-Bigas N, Audit B, Ouzounis C, et al. Are splicing mutations the most frequent cause of hereditary disease?[J]. FEBS Letters, 2005, 579(9): 1900-1903.

[7] Warf MB, Berglund JA. Role of RNA structure in regulating pre-mRNA splicing[J]. Trends in Biochemical Sciences, 2010, 35(3): 169-178.

[8] Park E, Cho MH, Hyun HS, et al. Genotype-Phenotype Analysis in Pediatric Patients with Distal Renal Tubular Acidosis[J]. Kidney and Blood Pressure Research, 2018, 43(2): 513-521.

[9] Tanner MJ. The structure and function of band 3 (AE1): recent developments (review)[J]. Molecular Membrane Biology, 1997, 14(4): 155-165.

[10] Smith AN, Skaug J, Choate KA, et al. Mutations in ATP6N1B, encoding a new kidney vacuolar proton pump 116-kD subunit, cause recessive distal renal tubular acidosis with preserved hearing[J]. Nature Genetics, 2000, 26(1): 71-75.

[11] Yang Q, Li G, Singh SK, et al. Vacuolar H+ -ATPase B1 subunit mutations that cause inherited distal renal tubular acidosis affect proton pump assembly and trafficking in inner medullary collecting duct cells[J]. Journal of the American Society of Nephrology, 2006, 17(7): 1858-1866.

[12] Stehberger PA, Schulz N, Finberg KE, et al. Localization and regulation of the ATP6V0A4 (a4) vacuolar H+-ATPase subunit defective in an inherited form of distal renal tubular acidosis[J]. Journal of the American Society of Nephrology, 2003, 14(12): 3027-3038.

[13] Enerback S, Nilsson D, Edwards N, et al. Acidosis and Deafness in Patients with Recessive Mutations in FOXI1[J]. 2018, 29(3): 1041-1048.

[14] Rungroj N, Nettuwakul C, Sawasdee N, et al. Distal renal tubular acidosis caused by tryptophan-aspartate repeat domain 72 (WDR72) mutations[J]. Clinical Genetics, 2018, 94(5): 409-418.

[15] Jobst-Schwan T, Klambt V, Tarsio M, et al. Whole exome sequencing identified ATP6V1C2 as a novel candidate gene for recessive distal renal tubular acidosis[J]. Kidney International, 2020, 97(3): 567-579.

[16] Zhao X, Lu J, Gao Y, et al. Novel compound heterozygous ATP6V1B1 mutations in a Chinese child patient with primary distal renal tubular acidosis: a case report[J]. 2018, 19(1): 364.

[17] Palazzo V, Provenzano A, Becherucci F, et al. The genetic and clinical spectrum of a large cohort of patients with distal renal tubular acidosis[J]. Kidney International, 2017, 91(5): 1243-1255.

[18] Chen L, Wang HL, Zhu YB, et al. Screening and function discussion of a hereditary renal tubular acidosis family pathogenic gene[J]. 2020, 11(3): 159.

[19] Zhang R, Wang C, Lang Y, et al. Five Novel Mutations in Chinese Children with Primary Distal Renal Tubular Acidosis[J]. Genet Test Mol Biomarkers, 2018, 22(10): 599-606.

[20] Wang S, Wang Y, Wang J, et al. Six Exonic Variants in the SLC5A2 Gene Cause Exon Skipping in a Minigene Assay[J]. Front Genet, 2020, 11: 585064.

[21] Zhang R, Wang J, Wang Q, et al. Identification of a novel TSC2 c.3610G > A, p.G1204R mutation contribute to aberrant splicing in a patient with classical tuberous sclerosis complex: a case report[J]. 2018, 19(1): 173.

[22] Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing[J]. Nat Rev Genet, 2002, 3(4): 285-298.

[23] Auclair J, Busine MP, Navarro C, et al. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing[J]. Human Mutation, 2006, 27(2): 145-154.

[24] Thery JC, Krieger S, Gaildrat P, et al. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes[J]. European Journal of Human Genetics, 2011, 19(10): 1052-1058.

[25] Fraile-Bethencourt E, Valenzuela-Palomo A, Diez-Gomez B, et al. Identification of Eight Spliceogenic Variants in BRCA2 Exon 16 by Minigene Assays[J]. Front Genet, 2018, 9: 188.

[26] Zhao X, Cui L, Lang Y, et al. A recurrent deletion in the SLC5A2 gene including the intron 7 branch site responsible for familial renal glucosuria[J]. Scientific Reports, 2016, 6: 33920.

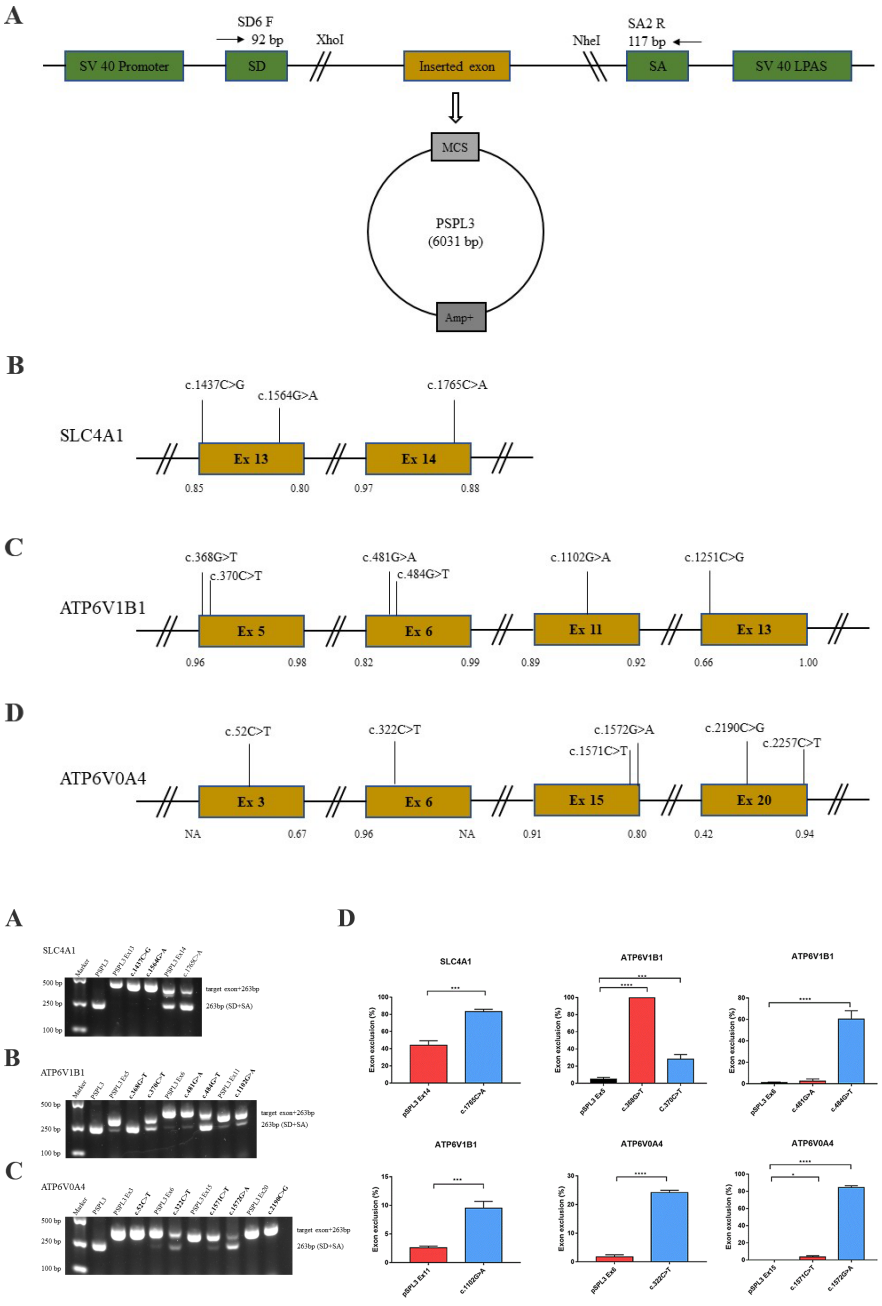
[27] Suarez-Artiles L, Perdomo-Ramirez A, Ramos-Trujillo E, et al. Splicing Analysis of Exonic OCRL Mutations Causing Lowe Syndrome or Dent-2 Disease[J]. 2018, 9(1).

[28] Sritippayawan S, Kirdpon S, Vasuvattakul S, et al. A de novo R589C mutation of anion exchanger 1 causing distal renal tubular acidosis[J]. Pediatric Nephrology, 2003, 18(7): 644-648.

[29] Karet FE, Gainza FJ, Gyory AZ, et al. Mutations in the chloride-bicarbonate exchanger gene AE1 cause autosomal dominant but not autosomal recessive distal renal tubular acidosis[J]. Proceedings of the National Academy of Sciences of the United States of America, 1998, 95(11): 6337-6342.

[30] Jarolim P, Shayakul C, Prabakaran D, et al. Autosomal dominant distal renal tubular acidosis is associated in three families with heterozygosity for the R589H mutation in the AE1 (band 3) Cl-/HCO3- exchanger[J]. Journal of Biological Chemistry, 1998, 273(11): 6380-6388.

[31] Littink KW, Pott JW, Collin RW, et al. A novel nonsense mutation in CEP290 induces exon skipping and leads to a relatively mild retinal phenotype[J]. Investigative Ophthalmology and Visual Science, 2010, 51(7): 3646-3652.



Hosted file

`Table 1.pdf` available at https://authorea.com/users/394318/articles/507714-identification-of-eight-exonic-variants-in-the-slc4a1-atp6v1b1-and-atp6v0a4-gene-that-alter-rna-splicing-by-minigene-assay