



# Splicing Outcomes of 5' Splice Site GT>GC Variants That Generate Wild-Type Transcripts Differ Significantly Between Full-Length and Minigene Splicing Assays

Jin-Huan Lin<sup>1,2†</sup>, Hao Wu<sup>1,2†</sup>, Wen-Bin Zou<sup>1,2†</sup>, Emmanuelle Masson<sup>3,4</sup>, Yann Fichou<sup>3,5</sup>, Gerald Le Gac<sup>3,4,5</sup>, David N. Cooper<sup>6</sup>, Claude Férec<sup>3,4</sup>, Zhuan Liao<sup>1,2\*</sup> and Jian-Min Chen<sup>3\*</sup>

## OPEN ACCESS

### Edited by:

Stephen J. Bush,  
University of Oxford, United Kingdom

### Reviewed by:

Alex Star Nord,  
University of California, Davis,  
United States  
Ludovica Celli,  
Institute of Molecular Genetics, Italian  
National Research Council, Italy

### \*Correspondence:

Zhuan Liao  
liao zhuan@smmu.edu.cn  
Jian-Min Chen  
jian-min.chen@univ-brest.fr

†These authors share first authorship

### Specialty section:

This article was submitted to  
Human and Medical Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 28 April 2021

Accepted: 13 July 2021

Published: 05 August 2021

### Citation:

Lin J-H, Wu H, Zou W-B,  
Masson E, Fichou Y, Le Gac G,  
Cooper DN, Férec C, Liao Z and  
Chen J-M (2021) Splicing Outcomes  
of 5' Splice Site GT>GC Variants That  
Generate Wild-Type Transcripts Differ  
Significantly Between Full-Length  
and Minigene Splicing Assays.  
*Front. Genet.* 12:701652.  
doi: 10.3389/fgene.2021.701652

<sup>1</sup> Department of Gastroenterology, Changhai Hospital, Second Military Medical University, Shanghai, China, <sup>2</sup> Shanghai Institute of Pancreatic Diseases, Shanghai, China, <sup>3</sup> Univ Brest, Inserm, EFS, UMR 1078, GGB, Brest, France, <sup>4</sup> Service de Génétique Médicale et de Biologie de la Reproduction, CHRU Brest, Brest, France, <sup>5</sup> Laboratory of Excellence GR-Ex, Paris, France, <sup>6</sup> Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, United Kingdom

Combining data derived from a meta-analysis of human disease-associated 5' splice site GT>GC (i.e., +2T>C) variants and a cell culture-based full-length gene splicing assay (FLGSA) of forward engineered +2T>C substitutions, we recently estimated that ~15–18% of +2T>C variants can generate up to 84% wild-type transcripts relative to their wild-type counterparts. Herein, we analyzed the splicing outcomes of 20 +2T>C variants that generate some wild-type transcripts in two minigene assays. We found a high discordance rate in terms of the generation of wild-type transcripts, not only between FLGSA and the minigene assays but also between the different minigene assays. In the pET01 context, all 20 wild-type minigene constructs generated the expected wild-type transcripts; of the 20 corresponding variant minigene constructs, 14 (70%) generated wild-type transcripts. In the pSPL3 context, only 18 of the 20 wild-type minigene constructs generated the expected wild-type transcripts whereas 8 of the 18 (44%) corresponding variant minigene constructs generated wild-type transcripts. Thus, in the context of a particular type of variant, we raise awareness of the limitations of minigene splicing assays and emphasize the importance of sequence context in regulating splicing. Whether or not our findings apply to other types of splice-altering variant remains to be investigated.

**Keywords:** aberrant transcript, full-length gene splicing assay, genetic variant, minigene splicing assay, splice site, SpliceAI

## INTRODUCTION

In principle, both coding and intronic variants within a gene have the potential to affect splicing (Cooper et al., 2009; Scotti and Swanson, 2016; Anna and Monika, 2018; Truty et al., 2021). Genetic variants occurring within the 5' splice site GT dinucleotide, whenever found in disease-causing or disease-predisposing genes, have generally been classified as pathogenic (Mount et al., 2019;

Stenson et al., 2020). However, a problem is posed by 5' splice site GT>GC variants (henceforth simply termed +2T>C variants) due to the fact that in the human genome, a small but nevertheless significant minority (~1%) of introns contain the 5' splice site GC dinucleotide (Burset et al., 2000, 2001; Abril et al., 2005; Sheth et al., 2006; Parada et al., 2014). Recently, combining data derived from a meta-analysis of human inherited disease-associated +2T>C variants and a cell culture-based Full-Length Gene Splicing Assay (FLGSA) of forward engineered +2T>C substitutions, we estimated that ~15–18% of +2T>C variants can generate up to 84% wild-type transcripts relative to their wild-type counterparts (Lin et al., 2019). This finding was corroborated by a re-analysis (Chen et al., 2020) of the saturation genome editing data on 12 *BRCA1* +2T>C substitutions (Findlay et al., 2018).

Our aforementioned findings have two direct clinical implications. Firstly, many +2T>C variants in human disease genes that have been capable of generating some wild-type transcripts are likely to have gone largely unreported; this represents a significant deficiency in terms of our understanding of genotype-phenotype relationships and tailored treatment options given that even the minor retention of wild-type transcripts derived from a variant allele might significantly impact disease expression and severity (Ramalho et al., 2002; Den Uijl et al., 2011; Raraigh et al., 2018; Lin et al., 2019; Scalet et al., 2019; Joynt et al., 2020). In this regard, it is pertinent to mention that *CFTR* c.3873+2T>C and c.4242+2T>C transitions (Joynt et al., 2020) and *SRP68* c.184+2T>C (Schmaltz-Panneau et al., 2021) are among the most recently reported examples of disease-causing +2T>C variants that generated some wild-type transcripts. Secondly, +2T>C variants in human disease genes may not invariably be pathogenic, a notion that has received support from at least two recent publications, which reclassified *BRCA2* c.8331+2T>C (Nix et al., 2021) and *BAP1* c.783+2T>C (Goldberg et al., 2021) as variants of unknown significance.

Another important finding arising from our study was that none of the widely used splicing prediction tools were capable of reliably distinguishing those +2T>C variants that generated wild-type transcripts from those that did not (Lin et al., 2019). The root of this problem is twofold: apart from the use of GC instead of GT as the 5' splice site dinucleotide in ~1% of introns, these prediction tools only take into consideration short local DNA sequence motifs (Chen et al., 2020). The recently developed deep learning-based tool, SpliceAI (Jaganathan et al., 2019), performed somewhat better in this regard but was still far from perfect (Chen et al., 2020). These observations underscored the importance of experimentally determining the splicing outcomes of +2T>C variants in a clinical as well as a basic research setting. Whilst RNA analysis, using pathophysiologically relevant tissues, provides the most accurate and reliable mRNA phenotyping information on human splicing variants, this is often not possible if appropriate tissue samples are not available (Aicher et al., 2020). RNA analysis using either patient blood cells or immortalized lymphoblastoid cells represents an alternative option, providing that the gene of interest is normally expressed in these cells (Wai et al., 2020). In case of the non-feasibility of both approaches, a cell culture-based

minigene splicing assay has often been devised (for some most recent examples, see Damasio et al., 2021; Hao et al., 2021; Kim et al., 2021; Kortum et al., 2021; Le Tertre et al., 2021; Morbidoni et al., 2021; Qian et al., 2021; Saint-Martin et al., 2021; Torrado et al., 2021).

Our FLGSA assay (focused on genes whose genomic sizes were < 8 kb) (Lin et al., 2019, 2020) cannot be readily used for large genes for various practical and/or technical reasons. Genome editing (Findlay et al., 2018) is a promising trend but its wide application is still some way from becoming reality. Thus, the minigene splicing assay will for the time being remain the mainstream approach for functionally characterizing potential splice-altering variants. However, an inherent drawback of the minigene splicing assay is the lack of the wider genomic sequence context of the gene under study (Zou et al., 2016; Lin et al., 2019, 2020; Tang et al., 2019). This could lead to inaccurate results and incorrect conclusions being drawn owing to the complexity of the splicing code (Fu and Ares, 2014; Drexler et al., 2020), as exemplified by the contrasting findings from the study of the *SPINK1* c.194G>A variant in a minigene assay (Beer and Sahin-Toth, 2014) and our own FLGSA assay (Wu et al., 2017). Herein, we explored whether the splicing outcomes of 20 +2T>C variants that have been previously shown to generate some wild-type transcripts by means of FLGSA and/or patient RNA analysis (Lin et al., 2019) could be replicated in two minigene assays.

## MATERIALS AND METHODS

### +2T>C Variants Included for Minigene Splicing Assay and Variant Nomenclature

A total of 26 +2T>C variants were previously shown to generate some wild-type transcripts by means of FLGSA and/or patient RNA analyses (Lin et al., 2019). Of these, six variants that occurred within the first intron of their respective genes could not be readily analyzed by the minigene assay and hence were excluded from further consideration.

All the remaining 20 variants were included in the current analysis (Table 1). Of these, six had been originally reported to be both naturally occurring and disease causing. These six pathogenic variants included the five variants that had previously been demonstrated to generate some wild-type transcripts by means of patient RNA analysis (i.e., *CD3E* IVS7+2T>C, *CD40LG* IVS3+2T>C, *DMD* IVS54+2T>C, *PLP1* IVS5+2T>C and *SPINK1* IVS3+2T>C) plus *HBB* IVS2+2T>C. Although the latter *HBB* IVS2+2T>C variant had no accompanying patient RNA data, it was suggested to have had a limited impact on splicing due to its associated hematological phenotype that was milder than would have been expected from a null allele (Frischknecht et al., 2009); its orthologous counterpart in the rabbit *Hbb* gene has been experimentally shown to generate wild-type transcripts (Aebi et al., 1986, 1987); and the human variant was also shown to generate wild-type transcripts in a FLGSA assay (Lin et al., 2019). Of the five pathogenic variants subjected to patient RNA analysis, only *SPINK1* IVS3+2T>C was also analyzed by FLGSA; the findings from the patient RNA analysis

(Kume et al., 2006) and FLGSA assay (Zou et al., 2016; Lin et al., 2019) were remarkably similar.

The remaining 14 variants included in this study were not known to be disease causing at the time (Lin et al., 2019). They represent forward engineered +2T>C substitutions, all being found to generate wild-type transcripts by means of FLGSA (Lin et al., 2019).

For ease of description and to be consistent with our previous publications (Lin et al., 2019, 2020; Chen et al., 2020), all included +2T>C variants were described in accordance with the traditional IVS (intervening sequence or intron) nomenclature (Table 1). Their respective chromosome locations, hg38 coordinates, reference alleles in hg38 and HGVS nomenclature (den Dunnen et al., 2016) are, however, provided in Supplementary Table 1.

### Construction of pET01 and pSPL3 Wild-Type Minigene Expression Vectors by Means of In-Fusion Cloning

For a given +2T>C variant, the corresponding wild-type genomic sequences cloned into the pET01 and pSPL3 exon trapping vectors were always identical. Of the 20 wild-type inserts, 18 comprised 63–330 bp sequence from the 3' end of N-1 intron, the entire exon N and 65–328 bp sequence from

the 5' end of intron N (N is the number of the variant-affected intron) (see upper panel in Figure 1A). The other two wild-type inserts (for *PSMC5* IVS10+2T>C and *SELENOS* IVS5+2>C, respectively) instead comprised 50–116 bp sequence from intron N-2, entire exon N-1, entire intron N-1, entire exon N and 96–294 bp sequence from intron N (lower panel in Figure 1A); this was done primarily due to the small size (<100 bp) of the respective intron N-1 in these two cases. See Supplementary Table 2 for the sequences of all inserts.

Two pairs of 5' *Xho*I-harboring and 3' *Bam*HI-harboring primers, one for in-fusion cloning into the pET01 trapping vector and the other for in-fusion cloning into the pSPL3 exon trapping vector, were designed to amplify each insert (Figure 1B). Primer sequences are provided in Supplementary Table 3. PCR was performed in a 25 µL reaction mixture containing 0.5 U KAPA HiFi HotStart DNA Polymerase (Kapa Biosystems), 0.75 µL KAPA dNTP Mix (300 µM final), 5 µL 5× KAPA HiFi Buffer, 50 ng DNA (from a healthy Chinese subject), and 0.3 µM forward and reverse primers. The PCR program comprised an initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 98°C for 20 s, annealing at 66°C for 15 s, extension at 72°C for 1 min, and a final extension at 72°C for 5 min.

PCR products of the expected size were purified with the Cloning Enhancer kit (TaKaRa). The purified products were

**TABLE 1** | Results from minigene splicing analyses of 20 +2T>C variants that were previously reported to generate wild-type transcripts.

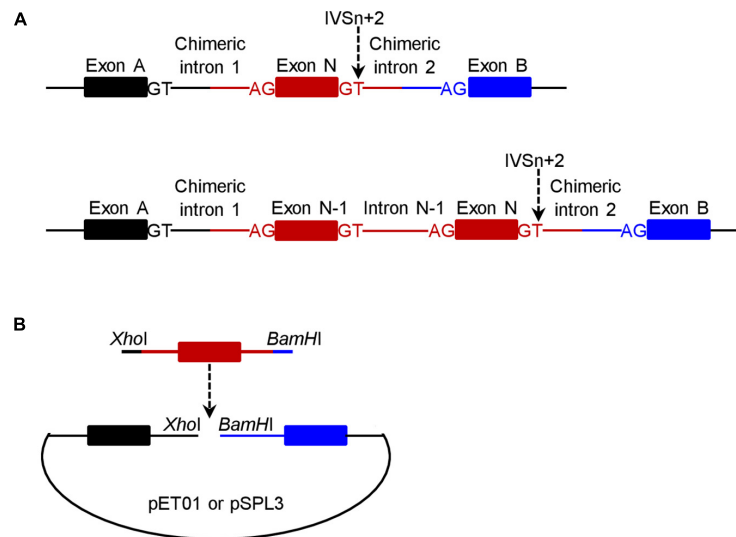
Gene	mRNA reference	Variant <sup>a</sup>	Summary of previous data		Current results <sup>c</sup>	
			Analytical method employed	Expression level of the wild-type transcript <sup>b</sup>	pET01	pSPL3
<i>CD3E</i>	NM_000733.3	IVS7+2T>C	Patient RNA analysis	1–5%	–	–
<i>CD40LG</i>	NM_000074.2	IVS3+2T>C	Patient RNA analysis	15%	+	+
<i>DBI</i>	NM_001079862.2	IVS2+2T>C	FLGSA <sup>d</sup>	Not quantified	+	–
<i>DMD</i>	NM_004006.2	IVS54+2T>C	Patient RNA analysis	10%	+	+
<i>DNAJC19</i>	NM_145261.3	IVS5+2T>C	FLGSA	42%	+	×
<i>FOLR3</i>	NM_000804.3	IVS4+2T>C	FLGSA	Not quantified	+	+
<i>HBB</i>	NM_000518.5	IVS2+2T>C	FLGSA	Not quantified	+	–
<i>IFNL2</i>	NM_172138.1	IVS5+2T>C	FLGSA	5%	+	+
<i>IL10</i>	NM_000572.3	IVS3+2T>C	FLGSA	Not quantified	+	+
<i>MGP</i>	NM_000900.4	IVS2+2T>C	FLGSA	80%	–	–
<i>PLP1</i>	NM_000533.4	IVS5+2T>C	Patient RNA analysis	8%	–	–
<i>PSMC5</i>	NM_001199163.1	IVS6+2T>C	FLGSA	56%	–	–
		IVS8+2T>C	FLGSA	56%	–	–
		IVS10+2T>C	FLGSA	46%	+	–
<i>RPL11</i>	NM_000975.5	IVS2+2T>C	FLGSA	Not quantified	+	+
		IVS3+2T>C	FLGSA	Not quantified	+	+
<i>RPS27</i>	NM_001030.4	IVS2+2T>C	FLGSA	63%	–	–
		IVS3+2T>C	FLGSA	Not quantified	+	×
<i>SELENOS</i>	NM_203472.2	IVS5+2T>C	FLGSA	14%	+	+
<i>SPINK1</i>	NM_003122.3	IVS3+2T>C	Patient RNA analysis/FLGSA	10%	+	–

<sup>a</sup>In accordance with the traditional IVS (intervening sequence or intron) nomenclature as previously described (Lin et al., 2019, 2020). See Supplementary Table 1 for additional information including specifically the HGVS nomenclature (den Dunnen et al., 2016).

<sup>b</sup>Expression level of the wild-type transcripts generated from the variant allele relative to that (set as 1) generated from the wild-type allele (see Lin et al., 2019; Chen et al., 2020).

<sup>c</sup>+, the variant generated wild-type transcripts; –, the variant did not generate wild-type transcripts; ×, the variant was not analyzed by the minigene assay owing to the non-expression of the wild-type transcripts from the corresponding wild-type minigene vectors.

<sup>d</sup>FLGSA, full-length gene splicing assay.



**FIGURE 1** | Schematic illustrations of the minigene expression constructs in the context of wild-type inserts. **(A)** Illustration of the target gene sequence (in red) inserted into the pET01 and pSPL3 exon trapping vectors. The number of the +2T>C variant-affected intron is N and the IVS<sub>n</sub>+2T site is indicated by a downward pointing arrow. For most variants ( $n = 18$ ), the insert comprised a single exon (i.e., exon N) plus flanking intronic sequences on both sides (upper panel). For two variants, the insert comprised two exons (exon N-1 and exon N) plus flanking and intervening intronic sequences (lower panel). The two exons (exon A and exon B) located within the vector and the canonical splice donor GT and acceptor AG sites defining the two chimeric introns (upper panel) or the two chimeric introns and intron N-1 (lower panel) are also denoted in the figure. **(B)** Illustration of how a wild-type minigene expression vector was constructed. The insert (in red) was PCR amplified with 5' *XhoI*-harboring and 3' *BamHI*-harboring in-fusion primers with respect to the pET01 or pSPL3 vector. The resulting PCR products were inserted into their respective linearized vectors by means of in-fusion cloning.

then cloned into the *XhoI* and *BamHI* restriction sites of the linearized pET01 or pSPL3 vector with the In-Fusion HD Cloning kit (TaKaRa) according to the manufacturer's instructions. Transformation was performed using Stellar Competent Cells (TaKaRa). Transformed cells were spread onto LB agar plates with 50  $\mu\text{g}/\text{mL}$  ampicillin and incubated at 37°C overnight. Plasmid constructs containing inserts were confirmed by Sanger sequencing.

Primers were designed by our laboratory in the Changhai Hospital. Primer synthesis, insert amplification, in-fusion cloning and verification of the inserted fragments were all performed by GENEWIZ, Beijing, China.

### Generation of pET01 and pSPL3 +2T>C Variant Minigene Expression Vectors by Means of Site-Directed Mutagenesis

+2T>C variants were introduced into their respective wild-type minigene expression constructs by means of the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Mutagenesis, transformation, plasmid preparation and validation of the introduced variants were performed as previously described (Lin et al., 2019). Sequences of the mutagenesis primers are provided in **Supplementary Table 4**.

### Cell Culture, Transfection, RNA Extraction, and Reverse Transcription

These were performed as previously described (Lin et al., 2019).

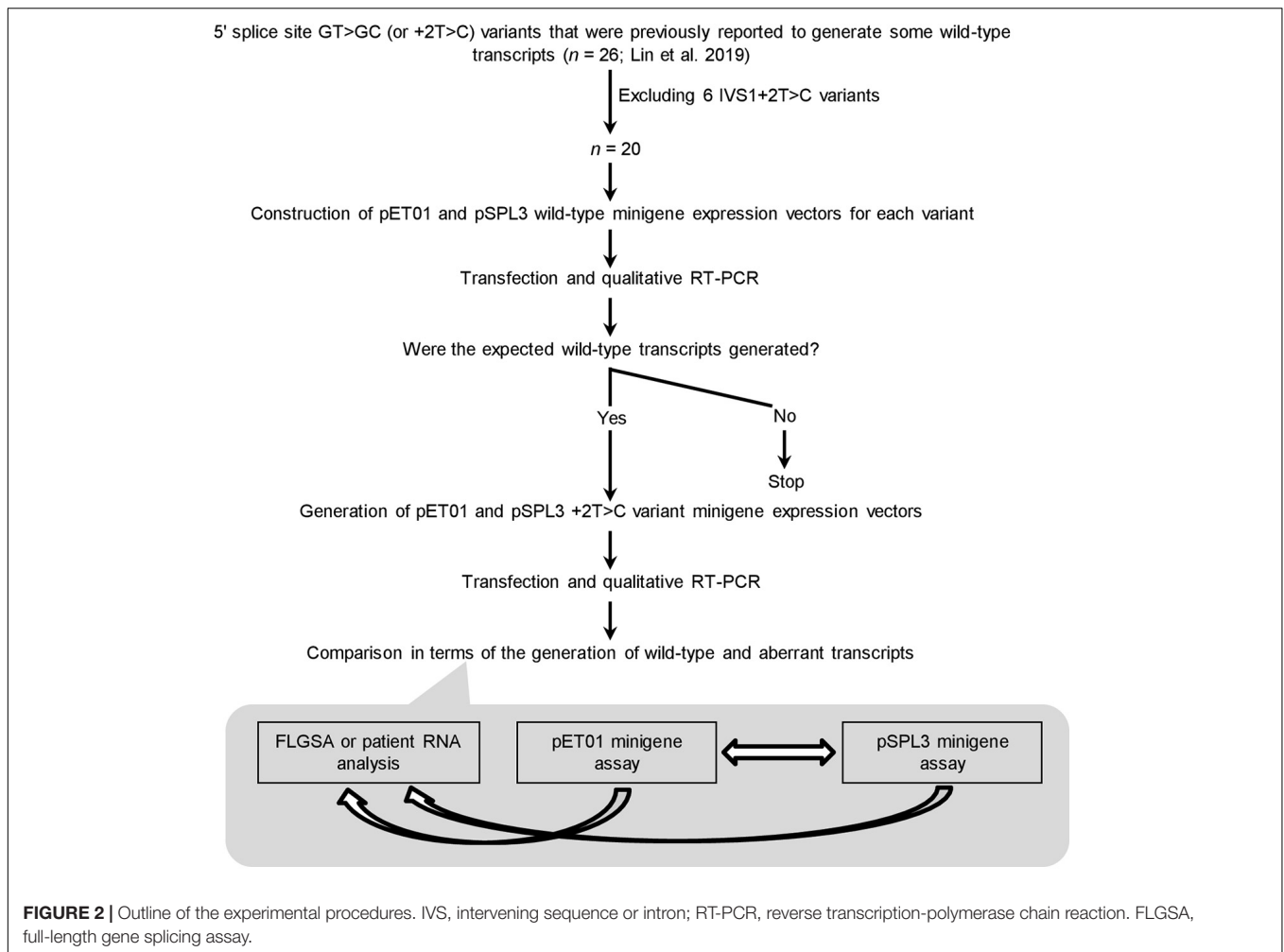
### Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Analysis

RT-PCR was performed in a 25- $\mu\text{L}$  reaction mixture containing 12.5  $\mu\text{L}$  HotStarTaq Master Mix (Qiagen), 1  $\mu\text{L}$  cDNA, and 0.4  $\mu\text{M}$  each primer (5'-GAGGGATCCGCTTCTCGGCC-3' (forward) and 5'-CTCCGGGCCACCTCCAGTGCC-3' (reverse) for pET01 expression vectors (both primers are located within the pET01 vector sequence); 5'-TCTGAGTCACCTGGACAACC-3' (forward) and 5'-ATCTCAGTGGTATTTGTGAGC-3' (reverse) for pSPL3 expression vectors (both primers are located within the pSPL3 vector sequence)). The PCR program had an initial denaturation step at 95°C for 15 min, followed by 30 cycles of denaturation at 94°C for 45 s, annealing at 58°C for 45 s, extension at 72°C for 1 min/kb, and a final extension step at 72°C for 10 min. RT-PCR products of a single band were cleaned by ExoSAP-IT (Affymetrix). In the case of multiple bands, the bands were excised from the agarose gel and then purified by QIAquick Gel Extraction Kit (Qiagen). Sequencing primers were those used for the RT-PCR analyses and sequencing was performed using the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems).

## RESULTS

### Rationale of Experimental Protocol

The experimental procedures adopted in this study are summarized in **Figure 2**. Before presenting the results obtained



in some detail, we would like to summarize our protocol in terms of its four components.

First, the 20 +2T>C variants included for minigene assay analysis represent the totality of the variants that were previously found to generate wild-type transcripts as assessed by FLGSA and/or patient RNA analyses excluding all IVS1+2T>C ones (Lin et al., 2019). (NB. Intronic variants located near the first or last exon of the gene cannot be readily evaluated by a minigene assay without special adaptation (Chen et al., 2018; Raud et al., 2019; Tang et al., 2019).) The accuracy and reliability of the FLGSA-obtained functional assessment of the +2T>C variants have been extensively addressed in our previous publications (Lin et al., 2019, 2020; Chen et al., 2020).

Second, for each +2T>C variant under study, the corresponding wild-type genomic sequences inserted into the two minigene vectors, pET01 and pSPL3, were always identical.

Third, the generation (or not) of wild-type transcripts—determined by qualitative RT-PCR analyses as previously described (Lin et al., 2019)—was used as the basis for comparison.

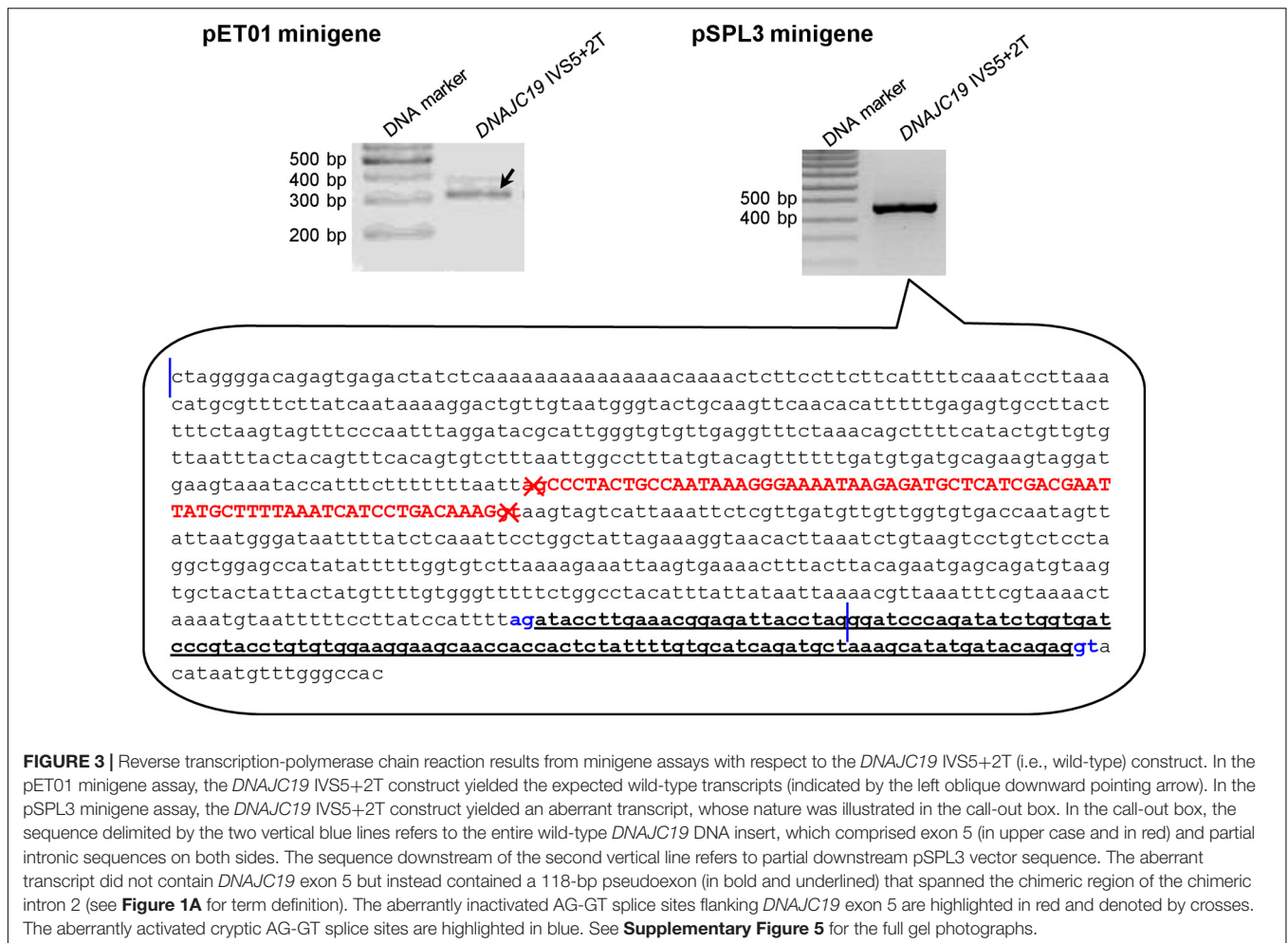
Fourth, a wild-type transcript refers to the product containing precisely Exon A, Exon N and Exon B or Exon A, Exon N-1, Exon N and Exon B as depicted in **Figure 1A**. The authenticity of all

wild-type transcripts was confirmed by Sanger sequencing. Most aberrant transcripts were also Sanger sequenced.

## Generation (or Not) of Wild-Type Transcripts

### Two Exceptions to the Rule That Wild-Type Minigene Constructs Invariably Express the Expected Wild-Type Transcripts

Of the 40 wild-type minigene constructs (20 in the pET01 context and 20 in the pSPL3 context), only two did not express the expected wild-type transcripts (**Supplementary Figures 1–20**), both of them in the pSPL3 context. Specifically, the pSPL3 *DNAJC19* IVS5+2T minigene construct expressed a transcript lacking *DNAJC19* exon 5 but containing instead a 118-bp pseudoexon (**Figure 3**). The pSPL3 *RPS27* IVS3+2T minigene construct expressed a transcript with *RPS27* exon 3 being skipped (**Supplementary Figure 18**). These two wild-type minigene constructs were thus not mutated to their corresponding variant versions. In other words, *DNAJC19* IVS5+2T>C and pSPL3 *RPS27* IVS3+2T>C were not analyzed by the minigene assay in the pSPL3 context.



#### Four +2T>C Variants Exhibited Discordance Between the Two Minigene Assays

Four variants, *DBI* IVS2+2T>C (**Figure 4**), *HBB* IVS2+2>C (**Supplementary Figure 7**), *PSMC5* IVS10+2T>C (**Supplementary Figure 14**) and *SPINK1* IVS3+2T>C (**Supplementary Figure 20**), generated wild-type transcripts in the pET01 minigene assay but not in the pSPL3 minigene assay.

#### Six +2T>C Variants Failed to Generate Wild-Type Transcripts in Both Minigene Assays

The six variants are *CD3E* IVS7+2T>C (**Supplementary Figure 1**), *MGP* IVS2+2T>C (**Supplementary Figure 10**), *PLP1* IVS5+2T>C (**Supplementary Figure 11**), *PSMC5* IVS6+2T>C (**Supplementary Figure 12**), *PSMC5* IVS8+2T>C (**Supplementary Figure 13**) and *RPS27* IVS2+2T>C (**Figure 5**).

#### Eight Variants Generated Wild-Type Transcripts in Both Minigene Assays

These eight variants are *CD40LG* IVS3+2T>C (**Supplementary Figure 2**), *DMD* IVS54+2T>C (**Supplementary Figure 4**), *FOLR3* IVS4+2T>C (**Figure 6**), *IFNL2* IVS5+2T>C (**Supplementary Figure 8**), *IL10* IVS3+2T>C (**Supplementary Figure 9**), *RPL11* IVS2+2T>C (**Supplementary Figure 15**),

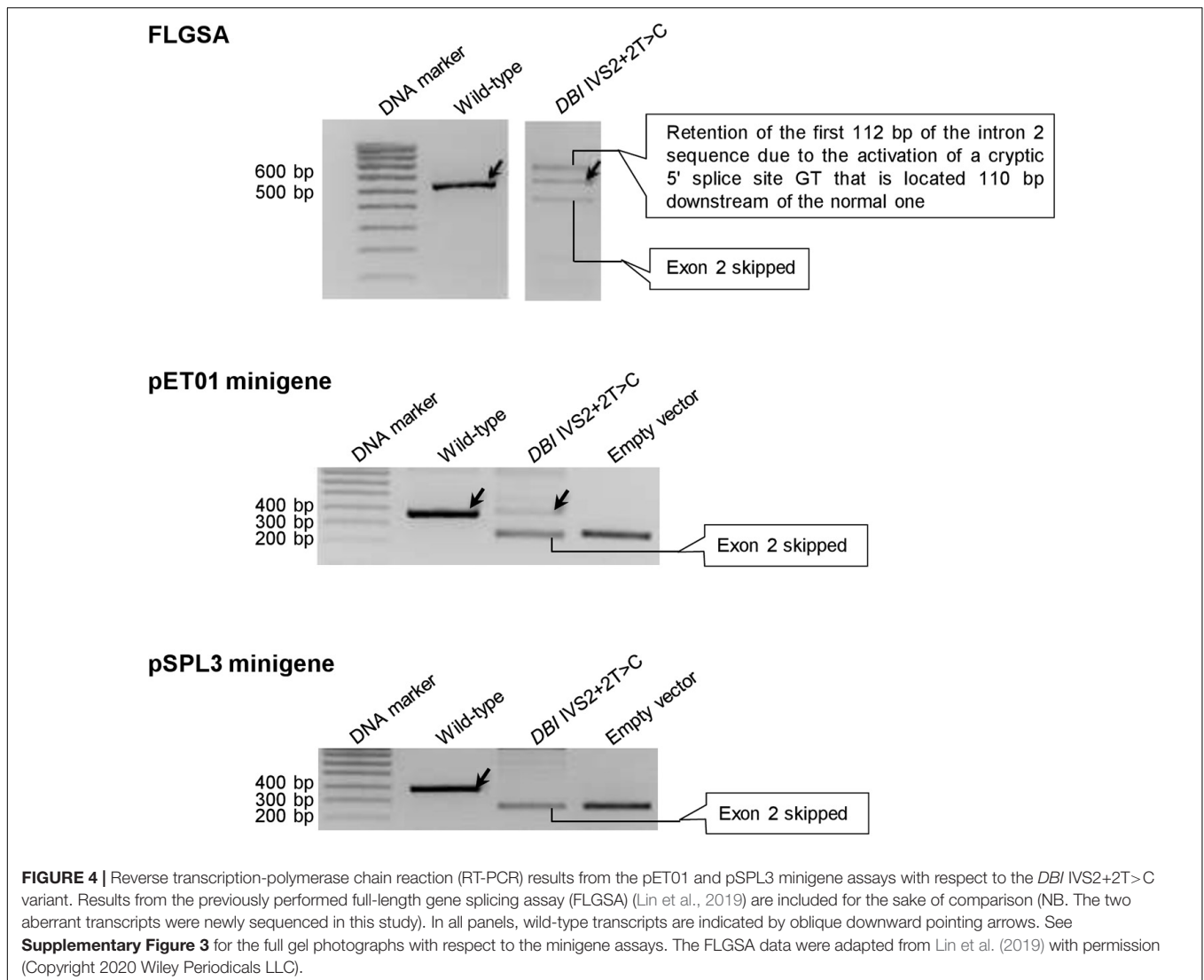
*RPL11* IVS3+2T>C (**Supplementary Figure 16**) and *SELENOS* IVS5+2T>C (**Supplementary Figure 19**).

#### Synthesis

The above findings are summarized in **Table 1**. In short, in the pET01 context, all 20 wild-type minigene constructs generated the expected wild-type transcripts; of the 20 corresponding variant minigene constructs, 14 (70%) generated wild-type transcripts. In the pSPL3 context, only 18 of the 20 wild-type minigene constructs generated the expected wild-type transcripts; of the 18 corresponding variant minigene constructs, only 8 (44%) generated wild-type transcripts.

## DISCUSSION

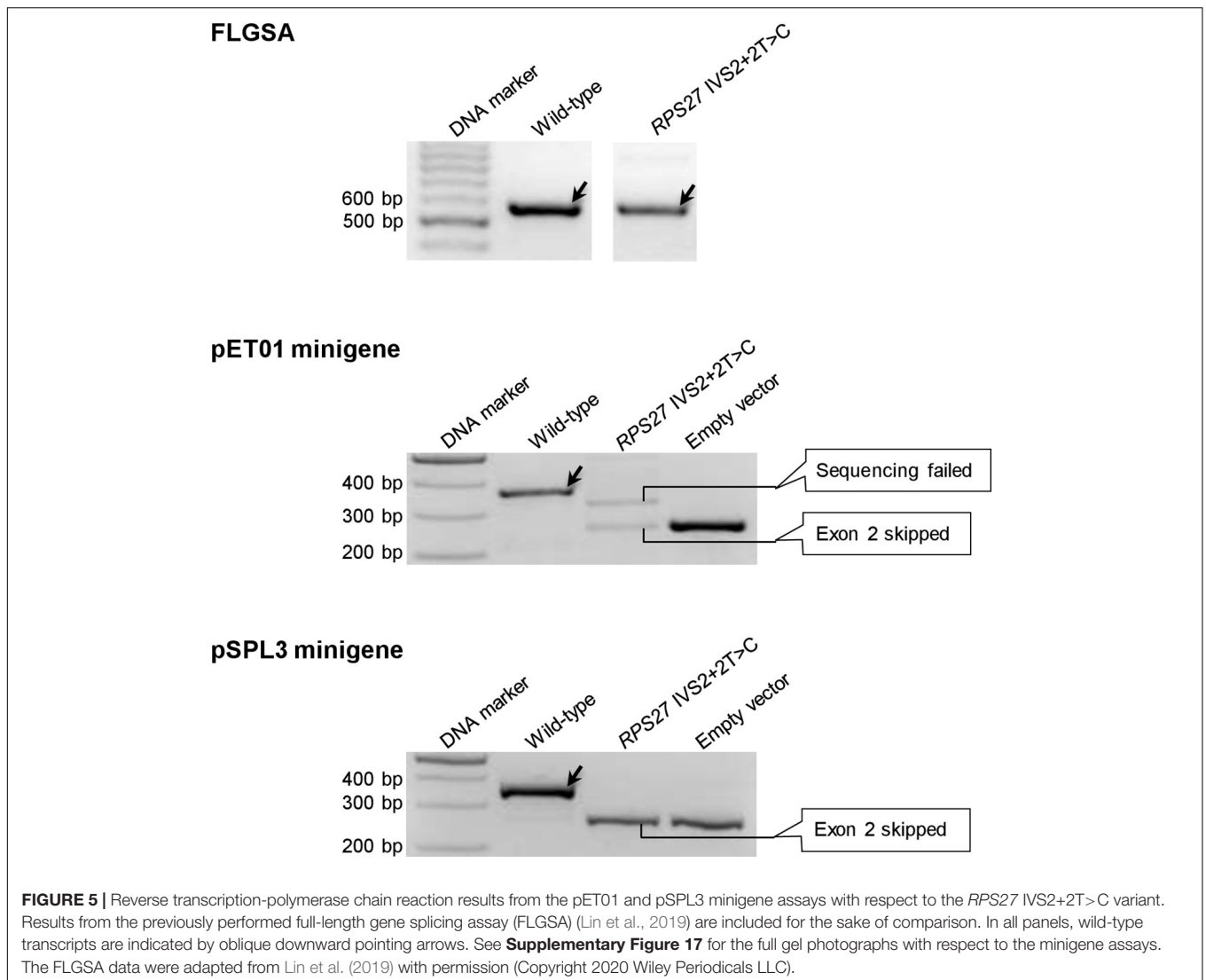
In this study, we set out to systematically analyze the splicing outcomes of 20 +2T>C variants that had been previously shown to generate varying levels of wild-type transcripts by means of FLGSA and/or patient RNA analyses (Lin et al., 2019), in two minigene systems. We found a fairly high level of discordance between the different systems in terms of the generation of wild-type transcripts (**Table 1**). First and foremost,



30% ( $n = 6$ ) of the 20 +2T>C variants analyzed in the pET01 minigene assay and 56% ( $n = 10$ ) of the 18 +2T>C variants analyzed in the pSPL3 minigene assay failed to generate wild-type transcripts. It would thus appear that the minigene assays have a tendency to exaggerate the negative effect of the +2T>C variants on splicing. Whether this is bound up with the artificiality of the minigene structure or simply represents a chance finding, remains to be established. In line with our own findings, the aforementioned reclassified *BRCA2* c.8331+2T>C variant (Nix et al., 2021) had been previously found to generate no wild-type transcripts at all by means of a minigene assay (Fraile-Bethencourt et al., 2017). However, using exonic tag-SNP analysis of transcripts expressed in Epstein-Barr virus-immortalized lymphoblastoid cells from a heterozygous *BRCA2* c.8331+2T>C carrier, Gelli et al. (2019) demonstrated that wild-type transcripts were derived from both the wild-type and c.8331+2T>C alleles, although they did not specify the relative levels of wild-type transcript emanating from the wild-type and variant alleles. More recently, using exonic tag-SNP analysis of

transcripts expressed in blood cells from a *BRCA2* c.8331+2T>C heterozygote, Nix et al. (2021) demonstrated that 62 and 38% of the wild-type transcripts were derived from the wild-type and variant alleles, respectively. For the purpose of comparison, *BRCA2* c.68-7T>A, which causes an ~20% functional loss of the variant allele, has been firmly established to be nonpathogenic (Colombo et al., 2018) whilst analysis of a neutral leaky variant (c.231T>G) has served to demonstrate that a reduction of ~60% of full-length *BRCA2* transcripts from the mutant allele does not give rise to any measurable increase in cancer risk (Tubeuf et al., 2020).

Significant discordance was also apparent between the two minigene systems, into which identical inserts were cloned for each variant under study, in terms of the results obtained (**Table 1**). Moreover, even in the cases that showed concordance in terms of the generation (or not) of wild-type transcripts, the splicing outcomes may have differed in terms of the nature of the aberrant transcripts and/or relative levels of the wild-type transcripts. Take, for example, the *FOLR3* IVS4+2T>C variant

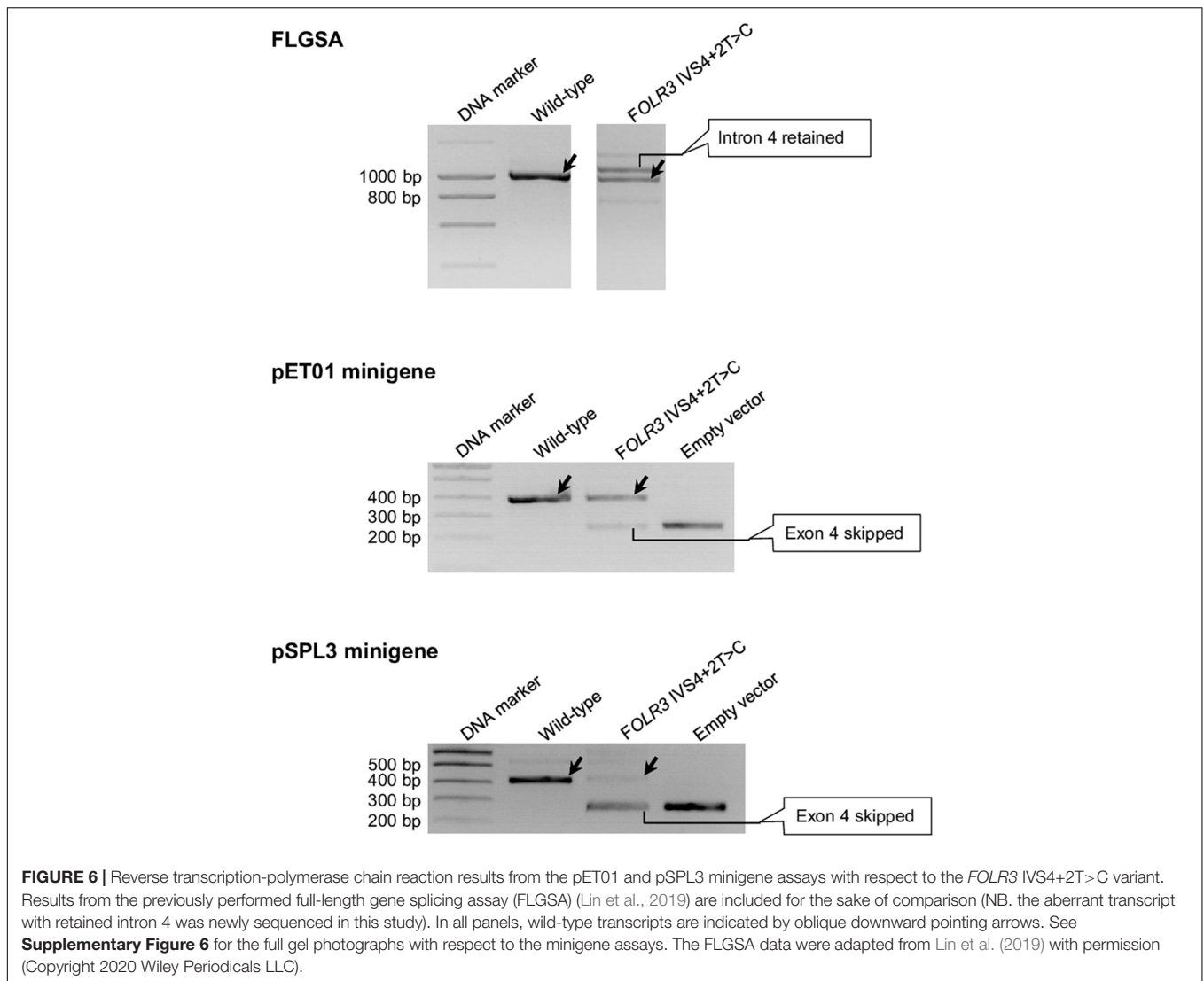


that generated wild-type transcripts in all three systems: the aberrant transcripts generated from FLGSA were different from those generated from the two minigene assays; moreover, the relative levels of the wild-type transcript were markedly different between the two minigene assays. Specifically, the FLGSA-derived aberrant transcript had retained intron 4 whereas the minigene-derived aberrant transcript had skipped exon 4; further, the level of the pET01-derived wild-type transcripts was much higher than that of the pSPL3-derived wild-type transcripts, as indicated by the relative intensities of the wild-type and aberrant transcript bands (**Figure 6**).

All the above mentioned discordant findings could be attributed primarily (if not solely) to differences in the underlying sequence contexts because our previous FLGSA and the current minigene assays were all performed under the same experimental conditions and employing the same procedures. As such, the high level of discordant findings between the different systems used should not be regarded as surprising given that (i) the sequence determinants for the 5' splice site go beyond the best studied

9-bp consensus sequence motif (see Lin et al., 2019 and references therein) and (ii) splicing is a complicated as well as a coordinated process across different introns (Fu and Ares, 2014; Drexler et al., 2020). In this context, it is pertinent to cite a previous study, in which two splicing reporter minigenes were found to exhibit very different sensitivities in relation to the effects of 13 *MLH1* variants on exon 10 skipping; it was the one that most closely approximated the pattern of exon 10 skipping *in vivo* (in the context of the wild-type *MLH1* exon 10 minigene construct) that was used for the final analysis (Soukariéh et al., 2016). Taken together with our current findings, this indicates that it is most unlikely that a universal splicing reporter minigene could ever be developed that would be suitable for the analysis of all splicing variants. In other words, different exon trapping vectors carrying a particular wild-type target gene insert might need to be tested in advance with a view to selecting one empirically that most closely resembled the normal expression pattern of the gene designated for functional analysis. Alternatively, a midigene splicing assay (Sangermano et al., 2018) might be considered





with a view to increasing the natural sequence context of the variant under study.

Although we have provided experimental evidence that genomic sequence context has influenced the splicing outcome of +2T>C variants capable of generating wild-type transcripts, it is beyond our current ability to discern precisely how and why these differences originated. For illustrative purposes, let us take the two pSPL3 wild-type minigene constructs that did not generate the expected wild-type transcripts. First, the pSPL3 *DNAJC19* IVS5+2T minigene construct expressed a transcript lacking *DNAJC19* exon 5 and containing instead a 118-bp pseudoexon. As shown in **Figure 3**, this was due to the inactivation of the physiological GT-AG splice sites defining *DNAJC19* exon 5 and the concurrent activation of cryptic splice sites located within the chimeric intron 2. However, based upon comparisons with the 3' splice site consensus sequence (CAG|G) and 5' splice site consensus sequence (MAG|GTRAGT where M is A or C and R is A or G)<sup>1</sup>, we could not draw any meaningful conclusions

about the alternative use of the *DNAJC19* exon 5-defining GT-AG splice sites and the aberrantly activated cryptic GT-AG splice sites observed in the minigene construct. Second, the pSPL3 *RPS27* IVS3+2T minigene construct expressed an *RPS27* transcript skipping exon 3 (**Supplementary Figure 18**). One might argue that this could somehow be associated with alternative splicing. However, although *RPS27* has three alternative transcripts,<sup>2</sup> exon 3 (in the context of NM\_001030.6) is common to all three. Even if *RPS27* exon 3 was differentially used by the three transcripts, the fact of its being skipped only in the pSPL3 context points to differences in sequence that extend beyond the gene inserts.

There is one final point to make. Although we have provided experimental evidence that points to limitations in the minigene-based analysis of splicing, our findings should not be interpreted as a challenge to the current preeminence of the minigene splicing

<sup>1</sup><https://science.umd.edu/labs/mount/RNAinfo/consensus.html>

<sup>2</sup><https://www.ncbi.nlm.nih.gov/gene/6232>

assay which is one of the most widely used analytical tools employed for the interpretation of potentially pathogenic variants. Indeed, all methods for assessing splicing have their advantages and inconveniences.

Our study has its limitations. For example, we used only one cell line for transfection and subsequent RT-PCR analysis. It would be interesting to see whether the same results were obtained employing another cell line. Here it may nonetheless be pertinent to mention that in our previous study, we have analyzed 10 +2T>C substitutions that generated wild-type transcripts and 10 +2T>C substitutions that did not generate wild-type transcripts in HEK293T cells for FLGSA in HeLa cells; we observed entirely consistent findings in the two cell lines in terms of the generation of wild-type transcripts or not (Lin et al., 2019). Moreover, in common with our previous studies (Lin et al., 2019, 2020), our findings were based on qualitative RT-PCR/gel analysis in terms of the absence or presence of the wild-type transcripts. Repeating the experiments using another more precise method would strengthen our findings. However, we believe that since an aberrant transcript band was always generated by the variant minigene expression vectors (or in other words, the aberrant transcripts always served as an internal control for gene expression), our findings should be highly reliable.

## CONCLUSION

Our study provides experimental evidence that +2T>C variants capable of generating some wild-type transcripts exhibit remarkable differences not only between minigene and full-length gene splicing assays but also between different minigene assays. Our results therefore bring fresh glimpses of the limitations that are inherent to minigene splicing assays and emphasize the role of sequence context in regulating splicing of a particular variant type. Whether our findings also apply to other types of splice-altering variant remains to be investigated.

## REFERENCES

- Abril, J. F., Castelo, R., and Guigo, R. (2005). Comparison of splice sites in mammals and chicken. *Genome Res.* 15, 111–119. doi: 10.1101/gr.3108805
- Aebi, M., Hornig, H., and Weissmann, C. (1987). 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* 50, 237–246. doi: 10.1016/0092-8674(87)90219-4
- Aebi, M., Hornig, H., Padgett, R. A., Reiser, J., and Weissmann, C. (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* 47, 555–565. doi: 10.1016/0092-8674(86)90620-3
- Aicher, J. K., Jewell, P., Vaquero-Garcia, J., Barash, Y., and Bhoj, E. J. (2020). Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet. Med.* 22, 1181–1190. doi: 10.1038/s41436-020-0780-y
- Anna, A., and Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* 59, 253–268. doi: 10.1007/s13353-018-0444-7

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

J-HL, HW, and W-BZ designed the study, performed the experiments, and assisted in writing the manuscript. EM, YF, GLG, DNC, and CF analyzed the data and revised the manuscript with important intellectual input. ZL contributed to study design, obtained funding, supervised the experiments, and revised the manuscript with important intellectual input. J-MC conceived and coordinated the study and drafted the manuscript. All authors approved the final manuscript.

## FUNDING

This research was funded by the National Natural Science Foundation of China (82000611 to J-HL and 81800569 to HW), the Shanghai Pujiang Program (2020PJD061 to J-HL and 18PJD057 to HW), the Scientific Innovation Program of Shanghai Municipal Education Committee, China (201901070007E00052 to ZL), and the Institut National de la Santé et de la Recherche Médicale (INSERM), France. The funding bodies did not play any role in the study design, collection, analysis and interpretation of data or in the writing of the article and the decision to submit it for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.701652/full#supplementary-material>

- Beer, S., and Sahin-Toth, M. (2014). Exonic variants affecting pre-mRNA splicing add to genetic burden in chronic pancreatitis. *Gut* 63, 860–861. doi: 10.1136/gutjnl-2013-305981
- Burset, M., Seledtsov, I. A., and Solovyev, V. V. (2000). Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.* 28, 4364–4375. doi: 10.1093/nar/28.21.4364
- Burset, M., Seledtsov, I. A., and Solovyev, V. V. (2001). SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res.* 29, 255–259. doi: 10.1093/nar/29.1.255
- Chen, J. M., Lin, J. H., Masson, E., Liao, Z., Férec, C., Cooper, D. N., et al. (2020). The experimentally obtained functional impact assessments of 5' splice site GT>GC variants differ markedly from those predicted. *Curr. Genomics* 21, 56–66. doi: 10.2174/1389202921666200210141701
- Chen, Y., Huang, L., Jiao, X., Riazuddin, S., Riazuddin, S. A., and Fielding Hetmancik, J. (2018). A novel *LRAT* mutation affecting splicing in a family with early onset retinitis pigmentosa. *Hum. Genomics* 12:35.
- Colombo, M., Lopez-Perolio, I., Meeks, H. D., Caleca, L., Parsons, M. T., Li, H., et al. (2018). The *BRCA2* c.68-7T >A variant is not pathogenic: a model for clinical calibration of spliceogenicity. *Hum. Mutat.* 39, 729–741.

- Cooper, T. A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell* 136, 777–793.
- Damasio, J., Santos, D., Sara Morais, S., Bras, J., Guerreiro, R., Sardoeira, A., et al. (2021). Congenital ataxia due to novel variant in *ATP8A2*. *Clin. Genet.* 100, 79–83. doi: 10.1111/cge.13954
- den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., et al. (2016). HGVS recommendations for the description of sequence variants: 2016 update. *Hum. Mutat.* 37, 564–569. doi: 10.1002/humu.22981
- Den Uijl, I. E., Mauser Bunschoten, E. P., Roosendaal, G., Schutgens, R. E., Biesma, D. H., Grobbee, D. E., et al. (2011). Clinical severity of haemophilia A: does the classification of the 1950s still stand? *Haemophilia* 17, 849–853. doi: 10.1111/j.1365-2516.2011.02539.x
- Drexler, H. L., Choquet, K., and Churchman, L. S. (2020). Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* 77, 985–998. doi: 10.1016/j.molcel.2019.11.017
- Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., et al. (2018). Accurate classification of *BRCA1* variants with saturation genome editing. *Nature* 562, 217–222. doi: 10.1038/s41586-018-0461-z
- Fraille-Bethencourt, E., Diez-Gomez, B., Velasquez-Zapata, V., Acedo, A., Sanz, D. J., and Velasco, E. A. (2017). Functional classification of DNA variants by hybrid minigenes: identification of 30 spliceogenic variants of *BRCA2* exons 17 and 18. *PLoS Genet.* 13:e1006691. doi: 10.1371/journal.pgen.1006691
- Frischknecht, H., Dutly, F., Walker, L., Nakamura-Garrett, L. M., Eng, B., and Wayne, J. S. (2009). Three new beta-thalassemia mutations with varying degrees of severity. *Hemoglobin* 33, 220–225. doi: 10.1080/03630260903089060
- Fu, X. D., and Ares, M. Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701. doi: 10.1038/nrg3778
- Gelli, E., Colombo, M., Pinto, A. M., De Vecchi, G., Foglia, C., Amitrano, S., et al. (2019). Usefulness and limitations of comprehensive characterization of mRNA splicing profiles in the definition of the clinical relevance of *BRCA1/2* variants of uncertain significance. *Cancers (Basel)* 11:295. doi: 10.3390/cancers11030295
- Goldberg, Y., Laitman, Y., Ben David, M., Bazak, L., Lidzbarsky, G., Salmon, L. B., et al. (2021). Re-evaluating the pathogenicity of the c.783+2T>C *BAP1* germline variant. *Hum. Mutat.* 42, 592–599. doi: 10.1002/humu.24189
- Hao, Z., Jin, D. Y., Chen, X., Schurgers, L. J., Stafford, D. W., and Tie, J. K. (2021). Gamma-Glutamyl carboxylase mutations differentially affect the biological function of vitamin K-dependent proteins. *Blood* 137, 533–543. doi: 10.1182/blood.2020006329
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., et al. (2019). Predicting splicing from primary sequence with deep learning. *Cell* 176, 535–548.e24.
- Joynt, A. T., Evans, T. A., Pellicore, M. J., Davis-Marcisak, E. F., Aksit, M. A., Eastman, A. C., et al. (2020). Evaluation of both exonic and intronic variants for effects on RNA splicing allows for accurate assessment of the effectiveness of precision therapies. *PLoS Genet.* 16:e1009100. doi: 10.1371/journal.pgen.1009100
- Kim, B. J., Jeon, H. W., Jeon, W., Han, J. H., Oh, J., Yi, N., et al. (2021). Rising of *LOXHD1* as a signature causative gene of down-sloping hearing loss in people in their teens and 20s. *J. Med. Genet.* doi: 10.1136/jmedgenet-2020-107594 [Epub ahead of print].
- Kortum, F., Kieninger, S., Mazzola, P., Kohl, S., Wissinger, B., Prokisch, H., et al. (2021). X-linked retinitis pigmentosa caused by non-canonical splice site variants in *RPGR*. *Int. J. Mol. Sci.* 22:850. doi: 10.3390/ijms22020850
- Kume, K., Masamune, A., Kikuta, K., and Shimosegawa, T. (2006). [-215G>A; IVS3+2T>C] mutation in the *SPINK1* gene causes exon 3 skipping and loss of the trypsin binding site. *Gut* 55:1214. doi: 10.1136/gut.2006.095752
- Le Tertre, M., Ka, C., Raud, L., Berlivet, I., Gourlaouen, I., Richard, G., et al. (2021). Splicing analysis of *SLC40A1* missense variations and contribution to hemochromatosis type 4 phenotypes. *Blood Cells Mol. Dis.* 87:102527. doi: 10.1016/j.bcmd.2020.102527
- Lin, J. H., Masson, E., Boulling, A., Hayden, M., Cooper, D. N., Férec, C., et al. (2020). 5' splice site GC>GT and GT>GC variants differ markedly in terms of their functionality and pathogenicity. *Hum. Mutat.* 41, 1358–1364. doi: 10.1002/humu.24029
- Lin, J. H., Tang, X. Y., Boulling, A., Zou, W. B., Masson, E., Fichou, Y., et al. (2019). First estimate of the scale of canonical 5' splice site GT>GC variants capable of generating wild-type transcripts. *Hum. Mutat.* 40, 1856–1873. doi: 10.1002/humu.23821
- Morbidity, V., Baschiera, E., Forzan, M., Fumini, V., Ali, D. S., Giorgi, G., et al. (2021). Hybrid minigene assay: an efficient tool to characterize mRNA splicing profiles of *NFI* variants. *Cancers (Basel)* 13:999. doi: 10.3390/cancers13050999
- Mount, S. M., Avsec, Z., Carmel, L., Casadio, R., Celik, M. H., Chen, K., et al. (2019). Assessing predictions of the impact of variants on splicing in *CAG15*. *Hum. Mutat.* 40, 1215–1224. doi: 10.1002/humu.23869
- Nix, P., Mundt, E., Coffee, B., Goossen, E., Warf, B. M., Brown, K., et al. (2021). Interpretation of *BRCA2* splicing variants: a case series of challenging variant interpretations and the importance of functional RNA analysis. *Fam. Cancer* doi: 10.1007/s10689-020-00224-y [Epub ahead of print].
- Parada, G. E., Munita, R., Cerda, C. A., and Gysling, K. (2014). A comprehensive survey of non-canonical splice sites in the human transcriptome. *Nucleic Acids Res.* 42, 10564–10578. doi: 10.1093/nar/gku744
- Qian, X., Wang, J., Wang, M., Igelman, A. D., Jones, K. D., Li, Y., et al. (2021). Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. *Front. Genet.* 12:647400. doi: 10.3389/fgene.2021.647400
- Ramalho, A. S., Beck, S., Meyer, M., Penque, D., Cutting, G. R., and Amaral, M. D. (2002). Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. *Am. J. Respir Cell Mol. Biol.* 27, 619–627. doi: 10.1165/rcmb.2001-0004oc
- Raiaigh, K. S., Han, S. T., Davis, E., Evans, T. A., Pellicore, M. J., McCague, A. F., et al. (2018). Functional assays are essential for interpretation of missense variants associated with variable expressivity. *Am. J. Hum. Genet.* 102, 1062–1077. doi: 10.1016/j.ajhg.2018.04.003
- Raud, L., Ka, C., Gourlaouen, I., Callebaut, I., Férec, C., Le Gac, G., et al. (2019). Functional analysis of novel *RHD* variants: splicing disruption is likely to be a common mechanism of variant D phenotype. *Transfusion* 59, 1367–1375. doi: 10.1111/trf.15210
- Saint-Martin, C., Cauchois-Le Miere, M., Rex, E., Soukarieh, O., Arnoux, J. B., Buratti, J., et al. (2021). Functional characterization of *ABCC8* variants of unknown significance based on bioinformatics predictions, splicing assays, and protein analyses: benefits for the accurate diagnosis of congenital hyperinsulinism. *Hum. Mutat.* 42, 408–420. doi: 10.1002/humu.24164
- Sangermano, R., Khan, M., Cornelis, S. S., Richelle, V., Albert, S., Garanto, A., et al. (2018). *ABCA4* midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in stargardt disease. *Genome Res* 28, 100–110. doi: 10.1101/gr.226621.117
- Scalet, D., Maestri, I., Branchini, A., Bernardi, F., Pinotti, M., and Balestra, D. (2019). Disease-causing variants of the conserved +2T of 5' splice sites can be rescued by engineered U1snRNAs. *Hum. Mutat.* 40, 48–52. doi: 10.1002/humu.23680
- Schmaltz-Panneau, B., Pagnier, A., Clauin, S., Buratti, J., Marty, C., Fenneteau, O., et al. (2021). Identification of biallelic germline variants of *SRP68* in a sporadic case with severe congenital neutropenia. *Haematologica* 106, 1216–1219. doi: 10.3324/haematol.2020.247825
- Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32. doi: 10.1038/nrg.2015.3
- Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R., and Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* 34, 3955–3967. doi: 10.1093/nar/gkl556
- Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frebourg, T., et al. (2016). Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. *PLoS Genet.* 12:e1005756. doi: 10.1371/journal.pgen.1005756
- Stenson, P. D., Mort, M., Ball, E. V., Chapman, M., Evans, K., Azevedo, L., et al. (2020). The human gene mutation database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum. Genet.* 139, 1197–1207. doi: 10.1007/s00439-020-02199-3
- Tang, X. Y., Lin, J. H., Zou, W. B., Masson, E., Boulling, A., Deng, S. J., et al. (2019). Toward a clinical diagnostic pipeline for *SPINK1* intronic variants. *Hum. Genomics* 13:8.
- Torrado, M., Fernandez, G., Ganoza, C. A., Maneiro, E., Garcia, D., Sonicheva-Paterston, N., et al. (2021). A cryptic splice-altering *KCNQ1* variant in trans with R259L leading to Jervell and Lange-Nielsen syndrome. *NPJ Genom. Med.* 6:21.
- Truty, R., Ouyang, K., Rojahn, S., Garcia, S., Colavin, A., Hamlington, B., et al. (2021). 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. doi: 10.1016/j.ajhg.2021.03.006
- Tubeuf, H., Caputo, S. M., Sullivan, T., Rondeaux, J., Krieger, S., Caux-Moncoutier, V., et al. (2020). Calibration of pathogenicity due to variant-induced leaky splicing defects by using *BRCA2* exon 3 as a model system. *Cancer Res.* 80, 3593–3605. doi: 10.1158/0008-5472.can-20-0895
- Wai, H. A., Lord, J., Lyon, M., Gunning, A., Kelly, H., Cibir, P., et al. (2020). Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet. Med.* 22, 1005–1014. doi: 10.1038/s41436-020-0766-9
- Wu, H., Boulling, A., Cooper, D. N., Li, Z. S., Liao, Z., Chen, J. M., et al. (2017). *In vitro* and *in silico* evidence against a significant effect of the *SPINK1* c.194G>A variant on pre-mRNA splicing. *Gut* 66, 2195–2196. doi: 10.1136/gutjnl-2017-313948
- Zou, W. B., Boulling, A., Masson, E., Cooper, D. N., Liao, Z., Li, Z. S., et al. (2016). Clarifying the clinical relevance of *SPINK1* intronic variants in chronic pancreatitis. *Gut* 65, 884–886. doi: 10.1136/gutjnl-2015-311168
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lin, Wu, Zou, Masson, Fichou, Le Gac, Cooper, Férec, Liao and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.