When splicing is not all or none: Challenges for variant classification

Jin-Huan Lin,[1,2,7] Hao Wu,[1,2,7] Xin-Ying Tang,[3] Emmanuelle Masson,[4,5] David N. Cooper,[6] Claude Férec,[4] Zhuan Liao,[1,2] Wen-Bin Zou,[1,2,*] and Jian-Min Chen[4,*]

[1]Department of Gastroenterology, Changhai Hospital, Naval Medical University, Shanghai 200433, China; [2]Shanghai Institute of Pancreatic Diseases, Shanghai 200433, China; [3]Department of Gastroenterology, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai 200433, China; [4]Univ Brest, Inserm, EFS, UMR 1078, GGB, 29200 Brest, France; [5]Service de Génétique Médicale et de Biologie de la Reproduction, CHU Brest, 29200 Brest, France; [6]Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK

[7]These authors contributed equally.

[*]**Correspondence:**

dr.wenbinzou@hotmail.com (W.-B.Z.), jian-min.chen@univ-brest.fr (J.-M.C.)

Abstract

Variants with intermediate functional effects—those that neither abolish nor fully preserve gene function—constitute an underappreciated source of genetic complexity and a grey zone for classification within the current ACMG/AMP binary pathogenic-or-benign framework. Here, we address this issue using GT>GC (+2T>C) 5′ splice site variants, approximately 15% of which retain variable amounts of wild-type transcript, as a tractable model. Drawing on disease-associated GT>GC variants and those analyzed by full-length gene splicing assays, we show that residual wild-type transcript levels form a continuum from barely detectable to near normal. Notably, GT>GC variants retaining substantial wild-type transcript (e.g., *SPINK1* c.194+2T>C, *HBB* c.315+2T>C, and *BRCA2* c.8331+2T>C) frequently present classification challenges. Our findings also indicate that, in classical Mendelian disorders caused by monoallelic or biallelic complete loss-of-function alleles, GT>GC

variants are not invariably pathogenic (e.g., *DOCK8* c.3234+2T>C). Minigene analyses of selected *CFTR* and *BAP1* variants further illustrate the inherent difficulty of predicting and experimentally quantifying residual wild-type transcript levels (e.g., *CFTR* c.2490+2T>C and *CFTR* c.4242+2T>C). Because GT>GC variants that generate wild-type transcript reflect the degeneracy and context dependence of splice-site recognition, insights from their study may extend to other splice-altering variants outside the invariant +1, −1, and −2 positions. Overall, our analysis highlights that grey-zone variants with intermediate functional effects represent a biological reality that complicates the genotype–phenotype relationship and underscores the need for classification frameworks that accommodate the full spectrum of variant effects and the true complexity of human genetic disease.

Keywords

ACMG/AMP variant classification guidelines; ClinVar; Incomplete penetrance; Intermediate functional effects; Loss-of-function (LoF) variants; Missing heritability; Residual wild-type transcript; SpliceAI; Variant effect spectrum; Variants of uncertain significance (VUS)

Introduction

Accurate classification of genetic variants remains a central challenge in human genetics.[1,2] The increasing use of a genotype-first or genomic-first approach,[3,4] enabled by exome and genome sequencing, has greatly magnified this challenge. In many cases, the mode of inheritance becomes apparent only after genetic analysis, and disorders presumed to follow a simple Mendelian pattern may instead have a complex, multifactorial, or oligogenic basis. This complexity underscores the limitations of the widely used American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) framework.[5] Its dichotomization of variants as either “pathogenic” or “benign,” although useful for clinical consistency, cannot account for the full spectrum of variant effects that contribute to human genetic disease.[3,6-9] It should be noted that the current ACMG/AMP category “variant of uncertain significance (VUS)”[5] is not a biological entity but

54 a temporary designation expected to resolve as either pathogenic or benign once sufficient evidence
55 becomes available.

56 For clarity, we focus here on variants that act through a loss-of-function (LoF) mechanism. In
57 contrast to variants at the two extremes—those causing complete or near-complete loss of function of
58 the affected allele, typically responsible for classic Mendelian disease phenotypes, and those retaining
59 complete or near-complete normal function—variants with intermediate effects are challenging to
60 interpret because they lie outside the scope of binary pathogenicity models. This challenge is
61 illustrated by the multiple partial LoF variants identified in *SPINK1* (MIM: 167790).[8] Despite their
62 clear biological reality and clinical relevance, such variants remain underexplored and are not
63 adequately addressed by any widely established classification frameworks, which offer limited
64 guidance for interpreting quantitative or intermediate functional effects and thereby risk systematic
65 over- or under-classification.

66 In this study, we explore GT>GC (+2T>C) 5′ splice site variants as a tractable model to investigate
67 how intermediate functional effects challenge classification. Although substitutions at canonical GT
68 donor or AG acceptor sites typically abolish normal splicing,[10,11] a subset of GT>GC substitutions
69 defy this expectation. Specifically, a meta-analysis of 45 disease-associated variants, together with
70 full-length gene splicing assay (FLGSA) data from 103 substitutions, showed that 15–18% retain
71 residual wild-type (WT) transcript, in some cases at levels approaching 80% of normal.[12] This
72 variability can be clinically significant. For instance, in cystic fibrosis, as little as 5% normal cystic
73 fibrosis transmembrane conductance regulator (CFTR) activity may be sufficient to prevent
74 pulmonary disease,[13,14] whereas in the hemophilias, factor VIII or IX levels above 5% may lead to
75 only mild bleeding diatheses.[15,16] Conversely, in disorders typically caused by complete LoF alleles,
76 GT>GC variants may not be invariably pathogenic if substantial WT transcript is preserved, as we
77 previously proposed.[12]

78 Using residual (leaky) WT transcript levels as a quantitative measure, this study aims to identify
79 the conceptual and methodological limitations in classifying GT>GC variants that generate substantial
80 yet incomplete amounts of WT transcript. Our analysis highlights the classification challenges posed

81 by intermediate functional effects and reinforces the need for expanded frameworks capable of
82 capturing the full spectrum of variant effects.[6,8]
83

84 # Methods

85 **Identification and selection of GT>GC variants**

86 Our 2019 study[12] served as the starting point. From the FLGSA-analyzed GT>GC variants reported
87 therein, we included all variants that both retained WT transcript and were registered in ClinVar[17]
88 (https://www.ncbi.nlm.nih.gov/clinvar/; accessed September 5, 2025) for analysis. To identify
89 additional disease-associated GT>GC variants with evidence of residual WT transcript generation
90 published since 2019, we conducted a Google-based literature search using the keyword "+2T>C,"
91 complemented by cascade reference tracking and cross-referencing within the peer-reviewed literature
92 (up to September 5, 2025). As in our previous work,[12] we focused on studies in which RNA analysis
93 was performed on material derived from affected individuals, and simple heterozygotes were excluded
94 unless residual WT transcript generation from the variant allele was explicitly demonstrated by allele-
95 specific reverse transcription-PCR (RT-PCR) in the original study. At the single-gene level, all 26
96 possible *CFTR* (MIM: 602421) GT>GC variants were subjected to initial evaluation. Finally, a *BAP1*
97 (MIM: 603089) variant, c.783+2T>C (GenBank: NM_004656.4, NG_031859.1), identified during the
98 literature search as having conflicting classifications,[18,19] was included for additional minigene splicing
99 analysis.
100

101 ***In silico* splicing predictions**

102 *In silico* predictions were obtained using SpliceAI[20] via the SpliceAI Lookup tool
103 (https://spliceailookup.broadinstitute.org/; accessed September 5, 2025) with the following settings:
104 genome version, hg38; GENCODE, basic; maximum distance, 10,000 bp; and REF & ALT scores
105 displayed. The resulting delta donor loss (ΔDL) scores, derived from comparisons between WT and
106 variant sequences, were used for analysis.

107 **Variant classification or description in public databases**

108 Selected variants were queried in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/)[17] to determine their
109 current classification status. The Clinical and Functional TRanslation of CFTR database (CFTR2;
110 http://www.cftr2.org/)[21] and the CFTR-France database (https://cftr.chu-montpellier.fr/)[22] were
111 consulted to evaluate *CFTR* variant classifications. Variant information from the Cystic Fibrosis
112 Mutation Database (CFTR1; http://www.genet.sickkids.on.ca/) was used to assist in the classification
113 of two *CFTR* variants that were not registered in either CFTR2 or CFTR-France. Our evaluation of the
114 variant classification or description data in these databases was last updated on October 28, 2025.

115

116 **Minigene splicing analysis of selected *CFTR* GT>GC variants**

117 ***Variant selection***

118 Of the 26 possible *CFTR* GT>GC variants (GenBank: NM_000492.4, NG_016465.4), we selected all
119 three with low SpliceAI DL scores (≤0.56)—c.2490+2T>C, c.3717+2T>C and c.4242+2T>C—
120 together with one variant (c.3873+2T>C) that had a high SpliceAI score (0.98) but was previously
121 reported to generate WT transcript,[23] for minigene splicing analysis.

122

123 ***Construction of pET01 and pSPL3 minigene expression vectors***

124 For each selected +2T>C variant, the corresponding WT genomic sequences cloned into the pET01
125 and pSPL3 exon trapping vectors[24] were always identical. Each *CFTR* WT insert comprised 213–217
126 bp from the 3′ end of intron N−1, the entire exon N, and 213–216 bp from the 5′ end of intron N (N =
127 variant-affected intron). Cloning of WT constructs and introduction of the selected +2T>C variants,
128 followed by Sanger confirmation, were performed by GENEWIZ Biotech Co. (Suzhou, China).

129

130 ***Cell culture, transfection, RT-PCR, and sequencing***

131 Experiments were carried out essentially as previously described.[24] Human embryonic kidney 293T
132 (HEK293T) cells were maintained in DMEM (Gibco) supplemented with 10% fetal calf serum
133 (Procell). Cells (3.5×10^5/well) were seeded in 6-well plates 24 h before transfection. Per well, 2.5 µg
134 WT or variant plasmid were transfected using HieffTrans Universal Transfection Reagent (Yeasen).

135 Forty-eight hours later, total RNA was extracted with the FastPure Cell/Tissue Total RNA Isolation
136 Kit V2 (+gDNA wiper) (Vazyme). RT was carried out using the HiScript III 1st Strand cDNA
137 Synthesis Kit (Vazyme), incorporating 2 μL 5×gDNA wiper Mix, 2 μL 10×RT Mix, 2 μL HiScript III
138 Enzyme Mix, 1 μL Oligo (dT)20VN, and 1 μg total RNA. RT-PCR was performed in a 25-μL reaction
139 mixture containing 12.5 μL 2×Taq Master Mix (Vazyme), 1 μL cDNA, and 0.4 μM each primer. The
140 primer pairs were as follows: pET01 vector—forward 5′-GAGGGATCCGCTTCCTGGCCC-3′,
141 reverse 5′-CTCCCGGGCCACCTCCAGTGCC-3′; pSPL3 vector—forward 5′-
142 TCTGAGTCACCTGGACAACC-3′, reverse 5′-ATCTCAGTGGTATTTGTGAGC-3′. The PCR
143 program comprised an initial denaturation step at 94°C for 5 min, followed by 35 cycles of
144 denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min, and a final
145 extension step at 72°C for 7 min. RT-PCR products were separated on agarose gels, excised, and
146 purified (Omega Bio-Tek Gel Extraction Kit). Sequencing primers were the same as those used for
147 RT-PCR, and reactions were performed with the BigDye Terminator v3.1 Cycle Sequencing Kit
148 (Applied Biosystems).
149

150 **Full-length and minigene splicing analysis of *BAP1* c.783+2T>C**

151 The full-length genomic *BAP1* WT insert, spanning from the translation initiation codon to the
152 termination codon (GenBank: NM_004656.4, NG_031859.1), was amplified from genomic DNA of a
153 normal control using long-range PCR. The resulting fragment was cloned into the pcDNA3.1/V5-His-
154 TOPO vector. These procedures, as well as subsequent transfection and RT–PCR, were performed
155 essentially as previously described.[12]

156 Minigene splicing analysis followed the same procedures as for *CFTR* variants, except that two
157 WT constructs (and their corresponding variant constructs) were generated. The *BAP1* Exon 9 insert
158 comprised 337 bp from intron 8, exon 9, and 302 bp from intron 9, whereas the *BAP1* Exon 8–10
159 insert comprised 197 bp from intron 7, exon 8, intron 8, exon 9, intron 9, exon 10, and 243 bp from
160 intron 10.

Results

Analytical approach

Building on our 2019 findings that some GT>GC variants retain WT transcript,[12] we analyzed three successive variant categories (Figure 1). First, we re-examined variants previously characterized by FLGSA. Second, we considered disease-associated GT>GC variants reported since 2019 that generate WT transcript. Third, we evaluated all possible GT>GC variants in *CFTR* and functionally tested four selected variants. Across these categories, we assessed residual WT transcript levels, genetic and clinical findings, SpliceAI ΔDL scores, and database classifications, where available. In addition, we conducted minigene splicing analysis of the *BAP1* c.783+2T>C variant, which has conflicting classifications.[18,19] Integrating these findings, we delineated the classification challenges posed by variants with intermediate functional effects—hereafter referred to conceptually as grey-zone variants.

Comparing disease-associated and FLGSA-analyzed GT>GC variants

In our previous study,[12] the proportion of GT>GC variants retaining WT transcript was similar between disease-associated and FLGSA-analyzed variants: 15.6% (7 of 45) versus 18.4% (19 of 103). The magnitude of residual WT transcript levels, however, differed markedly. The seven disease-associated variants yielded only low levels, from trace amounts up to 15% normal, whereas of the ten variants generating exclusively WT transcript (as judged by gel electrophoresis), eight retained levels ranging from 13.7% to 84% of normal, with only two falling below 5%.

The consistently low WT transcript levels observed for the seven disease-associated variants most likely reflect clinical ascertainment bias, as noted previously.[12] Nonetheless, of the four variants retaining 10–15% WT transcript, three—*CAV3* (MIM: 601253) c.114+2T>C (GenBank: NM_033337.3, NG_008797.2),[25] *CD40LG* (MIM: 300386) c.346+2T>C (GenBank: NM_000074.3, NG_007280.1),[26] and *DMD* (MIM: 300377) c.8027+2T>C (GenBank: NM_004006.3, NG_012232.1)[27]—were associated with milder clinical phenotypes; the remaining variant, *SPINK1* c.194+2T>C (GenBank: NM_001379610.1, NG_008356.2), is discussed below.

187 The corresponding WT constructs for the 103 FLGSA-analyzed substitutions were invariably
188 preselected to yield a single or quasi-single RT–PCR band of the expected WT transcript size,
189 consistent with their respective reference mRNA sequences.[12] We also analyzed *HBB* (MIM: 141900)
190 c.315+2T>C (GenBank: NM_000518.5, NG_059281.1),[28] despite its WT construct producing two
191 strong bands—one corresponding to the WT transcript and the other representing an aberrantly spliced
192 transcript. This variant also generated WT transcript in FLGSA,[12] bringing the total number of WT
193 transcript–generating GT>GC variants identified by FLGSA to 20.
194
195 **Detailed analysis of three FLGSA-analyzed GT>GC variants**
196 Of the 20 GT>GC variants previously shown by FLGSA[12] to generate residual WT transcript, only
197 three were found to be registered in ClinVar[17] (https://www.ncbi.nlm.nih.gov/clinvar/; assessed
198 September 5, 2025). These three variants, described below, illustrate how the presence of residual WT
199 transcript can challenge variant classification.
200
201 ***SPINK1 c.194+2T>C***
202 Among GT>GC variants, *SPINK1* c.194+2T>C (GenBank: NM_001379610.1, NG_008356.2) is
203 distinctive in that transcript expression has been assessed both *in vivo* (using RNA from an affected
204 homozygote)[29] and *in vitro* using FLGSA.[12] Both approaches yielded concordant results, with
205 approximately 10% of normal WT transcript levels from the variant allele.[30] SpliceAI predicts a ΔDL
206 score of 0.33, consistent with a substantial yet incomplete reduction in normal splicing.
207 Our in-depth analysis of *SPINK1*-related chronic pancreatitis[8] indicated that this ~10% residual
208 WT transcript places c.194+2T>C in the grey zone, conferring a well-defined genetic effect that is too
209 weak for "pathogenic" under current ACMG/AMP criteria[5] but too strong for the proposed "risk
210 allele" category.[7] We therefore suggested classifying it as "predisposing," reflecting its partial LoF
211 nature and intermediate genetic impact.[8] ClinVar lists this variant as pathogenic/likely pathogenic
212 (VCV000132142.62; accessed October 28, 2025), but convergent evidence from functional assays, *in*
213 *silico* prediction, and clinical and genetic data supports its role as proof of concept highlighting the
214 need for expanded classification frameworks.[8]

215 ***HBB c.315+2T>C***

216 This variant was identified during routine thalassemia screening in a 34-year-old individual who
217 exhibited hypochromia and microcytosis consistent with a thalassemia trait rather than β-thalassemia
218 major. The individual was compound heterozygous for *HBB* c.315+2T>C and *HBB* c.79G>A
219 (p.Glu27Lys) (GenBank: NM_000518.5, NG_059281.1).[28] The original authors reported that *HBB*
220 c.315+2T>C had a limited effect on splicing efficiency, stating that it "results in expression that is
221 approximately two-thirds of the normal allele."[28] However, this estimate was based on hemoglobin
222 quantitation and was likely confounded by the *trans*-inherited *HBB* c.79G>A (p.Glu27Lys) variant.
223 The latter likely retains substantial normal function, as most individuals homozygous for *HBB*
224 c.79G>A (p.Glu27Lys) exhibit only mild hypochromic microcytic anemia.[31]

225 Our FLGSA analysis confirmed that *HBB* c.315+2T>C allows appreciable production of WT
226 transcript.[12] Here, using ImageJ (https://imagej.net/ij/), we compared the intensity of the WT transcript
227 band derived from the variant allele with that from the WT allele in the original gel.[12] The variant
228 allele was estimated to retain approximately 30% of WT transcript output, providing a more direct and
229 reliable measure of its residual transcript level. Notably, the orthologous rabbit variant also generated
230 WT transcripts,[32,33] supporting conservation of the partial splicing effect across species. The SpliceAI
231 ΔDL score for *HBB* c.315+2T>C was 0.72.

232 *HBB* alleles in β-thalassemia have traditionally been classified as β^0 (complete loss of β-globin
233 synthesis), β^+ (moderate-to-severe reduction), or β^{++} ("silent," minimal reduction).[34] *HBB* c.315+2T>C
234 was originally described as a mild β^+ allele,[28] consistent with our FLGSA finding of significant but not
235 complete loss of WT transcript.[12] Nevertheless, it is currently annotated in ClinVar as "pathogenic"
236 (VCV000869355.2; accessed October 28, 2025)—a designation that may warrant re-evaluation in
237 view of its substantial residual WT transcript generation and the mild hematological phenotype
238 observed in the 34-year-old compound heterozygote.

239

240 ***MGP c.94+2T>C***

241 *MGP* (MIM: 154870) c.94+2T>C (GenBank: NM_000900.5, NG_023331.2) was shown by FLGSA[12]
242 to retain approximately 80% of WT transcript, whereas SpliceAI predicted a strong splice-disrupting

243 effect (ΔDL score, 0.97). Despite this discrepancy, ClinVar lists *MGP* c.94+2T>C as "likely
244 pathogenic" (VCV001987659.2; accessed October 28, 2025), based on a single clinical testing
245 submission (SCV003031993.2) with unknown affected status and no published case reports. This
246 classification appears to rest on four implicit assumptions:
247 (1) the variant alters a canonical donor splice site;
248 (2) such variants generally disrupt splicing;
249 (3) disruption of canonical sites typically abolishes protein function; and
250 (4) LoF variants in *MGP* are known to be pathogenic.
251 As this example shows, ClinVar classifications may at times rely on inferred reasoning rather than
252 direct experimental evidence.
253
254 **Disease-associated GT>GC variants reported since 2019**
255 We next examined disease-associated GT>GC variants reported since 2019, focusing on those with
256 WT transcript generation documented in samples derived from affected individuals. Among the five
257 variants identified, *SBDS* (MIM: 607444) c.258+2T>C (GenBank: NM_016038.4, NG_007277.1)[35]
258 and *SRP68* (MIM: 604858) c.184+2T>C (GenBank: NM_014230.4, NC_000017.11),[36] produced only
259 trace (~2%) or low amounts of WT transcript, with SpliceAI ΔDL scores of 0.93 and 0.37,
260 respectively. *TNXB* (MIM: 600985) c.12469+2T>C (GenBank: NM_001365276.2, NG_008337.2)
261 retained a moderate level of WT transcript by allele-specific RT–PCR,[37] consistent with its very low
262 ΔDL score of 0.01, and has been associated with moderate Ehlers–Danlos–like manifestations in
263 congenital adrenal hyperplasia.[37] The remaining two variants—*BRCA2* (MIM: 600185) c.8331+2T>C
264 (GenBank: NM_000059.4, NG_012772.3) and *DOCK8* (MIM: 611432) c.3234+2T>C (GenBank:
265 NM_203447.4, NG_017007.1)—are discussed below.
266
267 ***BRCA2 c.8331+2T>C***
268 *BRCA2* c.8331+2T>C (GenBank: NM_000059.4, NG_012772.3) is associated with a high SpliceAI
269 ΔDL score (0.93) and is listed in ClinVar as "pathogenic/likely pathogenic" (VCV000267692.71;
270 accessed October 28, 2025). However, allele-specific RT–PCR in a heterozygous carrier demonstrated

271 approximately 60% of WT transcript from the variant allele.[38] A validated *BRCA1/2* history-weighting
272 algorithm, which compares cancer histories with expectations for pathogenic versus benign variants,[39]
273 yielded a benign call.[38] Moreover, *BRCA2* appears relatively tolerant of reduced transcript dosage, and
274 one study suggested that variants producing more than 5% WT transcript should, in the absence of
275 other supporting evidence, be considered VUS.[40] While the original report reclassified c.8331+2T>C
276 from "pathogenic" to a "VUS,"[38] the combined evidence appears to point toward a benign
277 interpretation under current ACMG/AMP principles.
278
279 ***DOCK8* c.3234+2T>C**
280 Biallelic LoF variants in *DOCK8* cause autosomal recessive hyper-IgE syndrome-2 (MIM: 243700).
281 *DOCK8* c.3234+2T>C (GenBank: NM_203447.4, NG_017007.1) was reported in the homozygous
282 state in an infant with severe viral infections.[41] T cells from this individual showed a predominance of
283 WT transcript, normal DOCK8 protein accumulation, intact cellular function, and no DOCK8
284 deficiency phenotype. The clinical presentation was instead explained by a homozygous *IFNAR1*
285 (MIM: 107450) nonsense variant (c.1156G>T, p.Glu386Ter; GenBank: NM_000629.3).[41] Despite this
286 evidence for near-complete WT transcript generation, lack of disease relevance, and a modest
287 SpliceAI ΔDL score (0.49), three of the four ClinVar submissions for *DOCK8* c.3234+2T>C
288 (VCV000573664.15; accessed October 28, 2025) classified it as "likely pathogenic," while one
289 classified it as "benign."
290 *BRCA2* c.8331+2T>C and *DOCK8* c.3234+2T>C exemplify how variants predicted—or
291 assumed—to be fully disruptive can nevertheless yield intermediate or even near-complete levels of
292 WT transcript, underscoring the need for direct transcript-level evidence for accurate classification.
293 These two variants are summarized in Table 1 alongside the three FLGSA-analyzed examples
294 (*SPINK1* c.194+2T>C, *HBB* c.315+2T>C, and *MGP* c.94+2T>C).
295
296 **Locus-wide analysis of *CFTR* GT>GC variants**
297 Having examined GT>GC variants that retain WT transcript generation across multiple genes, we next
298 asked how frequent and systematic this phenomenon might be within a single locus. We selected

299 *CFTR* (GenBank: NM_000492.4, NG_016465.4), in which biallelic LoF variants cause cystic fibrosis
300 (CF; MIM: 219700),[42,43] as a model system because it harbors 26 GT donor splice sites (Table S1), is
301 among the most extensively studied disease-related genes, and benefits from expert-curated databases
302 CFTR2[21] and CFTR-France[22] that provide authoritative variant annotations.

303 Building upon our previous[12,30] and current findings, we observed that although some GT>GC
304 variants retaining WT transcript are associated with high SpliceAI ΔDL scores, most have low scores.
305 We therefore used SpliceAI ΔDL scores as the starting point for this analysis, complemented by cross-
306 examination of variant classifications in CFTR2 (http://www.cftr2.org/; accessed October 28, 2025)
307 and CFTR-France (https://cftr.chu-montpellier.fr/; accessed October 28, 2025) (Figure 2A). Notably,
308 all 16 *CFTR* GT>GC variants that we grouped as "pathogenic" had SpliceAI ΔDL scores greater than
309 0.80. These included 13 designated as "CF-causing" in CFTR2, one absent from CFTR2 but listed as
310 "likely pathogenic" in CFTR-France (c.3468+2T>C), and two absent from both CFTR2 and CFTR-
311 France but classified by us as "pathogenic" based on clinical and genetic data in CFTR1
312 (http://www.genet.sickkids.on.ca/; accessed October 28, 2025) (c.2908+2T>C and c.2988+2T>C).
313 Eight *CFTR* GT>GC variants were not registered in these three databases, of which seven also scored
314 above 0.80 and one had a score of 0.56 (c.3717+2T>C). The remaining two *CFTR* GT>GC variants—
315 one listed as a VUS in CFTR-France (c.2490+2T>C) and one reported in CFTR1 in two individuals
316 with suspected CF (c.4242+2T>C)—had SpliceAI ΔDL scores below 0.56 (Figure 2A,B). Table S1
317 provides full details of all 26 *CFTR* GT>GC variants, including the rationale for our classifying two as
318 "pathogenic."

319 Assuming that the three variants with relatively low SpliceAI ΔDL scores (c.2490+2T>C, 0.35;
320 c.3717+2T>C, 0.56; and c.4242+2T>C, 0.15) generate WT transcript, at least 11.5% of the 26 *CFTR*
321 GT>GC variants would retain WT transcript. This proportion is broadly consistent with the 15–18%
322 frequency observed across genes in our previous study.[12]

323 Among these three *CFTR* variants with relatively low SpliceAI ΔDL scores, only c.4242+2T>C
324 had previously undergone functional analysis. Using an expression minigene (EMG) system
325 comprising the full *CFTR* coding sequence and intronic regions tailored to the variant under study,
326 generation of residual WT transcript from the c.4242+2T>C allele was confirmed by Sanger

327 sequencing, and mature CFTR protein was detected in transfected cells by immunoblotting.[23] The only
328 other *CFTR* GT>GC variant reported to yield WT transcripts is c.3873+2T>C, although this was
329 mentioned only briefly and without accompanying experimental data.[23] Notably, c.3873+2T>C is
330 classified as "CF-causing" in both CFTR2 and CFTR-France, consistent with its high SpliceAI ΔDL
331 score (0.98) (Figure 2B).

332 We tested the three variants with low SpliceAI ΔDL scores, together with c.3873+2T>C, in
333 minigene assays using both pET01 and pSPL3 vectors (Figure 3; see also Figure S1). Findings for
334 each variant are detailed below.

335

336 ***c.2490+2T>C (intron 14)***

337 WT constructs in both minigene vectors generated multiple RT–PCR products, including the WT
338 transcript. Importantly, WT transcript from the variant allele was detected in the pET01 vector (Figure
339 3A), consistent with its modest SpliceAI ΔDL score (0.35). Although this variant is not classified by
340 CFTR2, it is listed in CFTR-France as a VUS, having been reported in two asymptomatic
341 homozygotes. In CFTR1, it was described in the heterozygous state in the partner of an individual
342 affected by congenital bilateral absence of the vas deferens (Figure 2B). Furthermore, ClinVar lists
343 this variant under "conflicting classifications of pathogenicity" (VCV000370078.23; accessed October
344 28, 2025), based on six clinical-testing submissions ("pathogenic" × 3, "likely pathogenic" × 2, and
345 "uncertain significance" × 1). Only one submission specified a known affected status—homozygosity
346 for c.2490+2T>C, labeled "pathogenic," in an individual with fetal cystic hygroma
347 (SCV001737053.1). Another submission, designated "likely pathogenic" (SCV001363801.5),
348 explicitly noted that "no occurrence of c.2490+2T>C in individuals affected with cystic fibrosis and
349 no experimental evidence demonstrating its impact on protein function have been reported." Taken
350 together, the clinical, genetic, computational, and experimental evidence indicate that c.2490+2T>C is
351 unlikely to represent a pathogenic *CFTR* variant in the context of classic CF.

352 ***c.3717+2T>C (intron 22)***

353 No WT transcript from the variant allele was detected in either minigene vector (Figure 3B), despite
354 its intermediate SpliceAI ΔDL score (0.56). This variant is not listed in CFTR2, CFTR-France,
355 CFTR1, or ClinVar (Figure 2B).

356

357 ***c.3873+2T>C (intron 23)***

358 No WT transcript from the variant allele was observed in either vector (Figure 3C); this finding is
359 consistent with its high SpliceAI ΔDL score (0.98), its designation as "CF-causing" in both CFTR2
360 and CFTR-France, and its "pathogenic" classification in ClinVar (VCV000053828.18; accessed
361 October 28, 2025) (Figure 2B). Moreover, consistent with its description in CFTR1, this variant is
362 among the *CFTR* alleles recurrently identified in individuals with clinically diagnosed CF in
363 Norway.[44] Taken together, c.3873+2T>C represents a pathogenic *CFTR* variant causing classic CF.
364 The level of WT transcript generated from the c.3873+2T>C allele, as previously reported,[23] is
365 presumably minimal.

366

367 ***c.4242+2T>C (intron 26)***

368 No WT transcript from the variant allele was detected in either minigene vector (Figure 3D), in
369 contrast to the earlier EMG report of residual WT transcript and protein production[23] and to its very
370 low SpliceAI ΔDL score (0.15). Notably, this variant is not listed in CFTR2 or CFTR-France but
371 appears in CFTR1 as present in two individuals with suspected cystic fibrosis (Figure 2B). ClinVar
372 lists c.4242+2T>C as "pathogenic/likely pathogenic" (VCV000035884.8; accessed October 28, 2025),
373 based on three submissions ("pathogenic" × 1 and "likely pathogenic" × 2), all from clinical genetic
374 testing laboratories. The collection method was described as "clinical testing" for two submissions
375 (SCV002630543.2 and SCV004675881.2), both with unknown affected status, and as "curation" for
376 the remaining submission (SCV000052191.3). Taken together, this variant may best be classified as a
377 VUS at present.

378 In summary, locus-wide analysis of *CFTR* confirms that a subset of GT>GC variants retain WT
379 transcript. Functional testing of selected *CFTR* GT>GC variants highlights both inter-assay variability
380 and the inherent limitations of minigene systems in fully recapitulating native splicing.[24]
381
382 **Challenges of splicing assays: the case of *BAP1* c.783+2T>C**
383 As illustrated by *CFTR* variants, *in vitro* splicing assays may fail to yield clear or consistent results,
384 complicating variant classification. This challenge is further exemplified by *BAP1* c.783+2T>C
385 (GenBank: NM_004656.4, NG_031859.1). Because heterozygous germline LoF variants in *BAP1*
386 underlie tumor predisposition syndrome 1 (TPDS1; MIM: 614327) and *BAP1* c.783+2T>C affects the
387 canonical intron 9 donor splice site, this variant was initially classified in ClinVar as "likely
388 pathogenic," as noted by Goldberg et al.[18] These authors reclassified it as a VUS based on the
389 following considerations:
390 (1) the variant was identified in seven families (six Jewish Iraqi and one non-Jewish) with
391 nonclassical *BAP1*-related TPDS1 tumor types;
392 (2) the variant exhibited low penetrance;
393 (3) the variant had a relatively high allele frequency in Jewish Iraqi individuals (3.6%);
394 (4) the variant had a low SpliceAI score (0.08); and
395 (5) no aberrantly spliced transcripts were detected by RT-PCR using RNA from blood cells of
396 two probands.
397 A subsequent study[19] identified a methodological limitation: the forward primer used by Goldberg
398 et al.[18] was located within exon 9, rendering it incapable of detecting exon 9 skipping. Using a primer
399 in exon 8, Sculco et al.[19] analyzed *BAP1* c.783+2T>C heterozygotes and observed an alternative
400 splicing pattern consistent with exon 9 skipping. These authors proposed reinstating a "likely
401 pathogenic" classification while acknowledging that the effect of *BAP1* c.783+2T>C on generating
402 aberrant transcripts might be incomplete.
403 To further clarify this issue, we performed additional splicing assays (Figure 4; Figure S2). We
404 first attempted FLGSA, as the *BAP1* genomic sequence from start to stop codon (<8 kb) falls within
405 the upper size limit of the insert cloned into the pcDNA3.1/V5-His-TOPO vector routinely used in this

406 assay.[12] However, the WT construct, when transfected into HEK293T cells, failed to yield clear RT-
407 PCR bands (Figure 4A). Notably, in our previous study, only 33 of 119 (28%) WT full-length
408 constructs produced a single or quasi-single band of the expected size after transfection.[12] We
409 therefore turned to minigene assays using pET01 and pSPL3 vectors. WT constructs spanning
410 genomic sequence from exon 8 to exon 10 (including exon 8, intron 8, exon 9, intron 9, and exon 10)
411 with short flanking sequences from introns 7 and 10 produced multiple RT-PCR bands, none
412 corresponding to the WT transcript defined by NM_004656.4; the corresponding variant constructs
413 also failed to generate a normally spliced transcript, although the banding patterns differed from those
414 of the WT (Figure 4B). In contrast, shorter constructs containing exon 9 with short flanking intronic
415 sequences generated the expected WT band from the WT allele, whereas only transcripts lacking exon
416 9 were produced from the variant allele (Figure 4C; Figure S2).
417 ClinVar currently lists c.783+2T>C under "conflicting classifications of pathogenicity"
418 (VCV000422670.29; accessed October 28, 2025), with five submissions of "uncertain significance"
419 and one of "likely benign." In light of these inconclusive findings, allele-specific transcript analyses in
420 heterozygous individuals—if possible in the presence of a *cis*-linked coding variant outside exon 9—
421 would be most informative. Alternatively, targeted knock-in cell models could provide additional
422 insight into the splicing effect of *BAP1* c.783+2T>C.
423

424 ## Discussion

425 Our analysis yielded several insights into the impact of GT>GC 5′ splice-site variants. Consistent with
426 findings from FLGSA,[12] we show that disease-associated GT>GC variants can generate a wide range
427 of residual WT transcript levels, from barely detectable to nearly normal. Moreover, our findings
428 confirm our earlier postulate that "GT>GC variants in human disease genes may not invariably be
429 pathogenic,"[12] as exemplified by the *DOCK8* c.3234+2T>C variant.
430 Second, GT>GC variants that generate substantial yet incomplete WT transcript levels often fall
431 within a functional grey zone, complicating their classification (Figure 5). Such indeterminacy
432 underscores the limitations of binary frameworks and has important clinical implications. A clear

example is *SPINK1* c.194+2T>C, which retains approximately 10 % of normal WT transcript. In the simple heterozygous state, its effect is too weak to be classified as "pathogenic" yet too strong to be regarded as a mere "risk" allele. However, in homozygous, compound heterozygous, or *trans*-heterozygous states—or in the presence of environmental modifiers—the same variant confers a substantially greater pathogenic effect.[8] This exemplifies how variants with intermediate functional effects contribute to the genetic complexity of human disease. To address this challenge, we have proposed expanding the ACMG/AMP variant classification framework by incorporating additional categories, including a "predisposing" category.[6,8]

Third, our findings underscore the limited resolution of current *in silico* tools, such as SpliceAI,[20] for predicting whether GT>GC variants generate WT transcript and, if so, to what extent. Whereas most WT transcript–retaining variants exhibit low SpliceAI ΔDL scores, some have high values (>0.80), and transcript levels show no consistent correlation with predicted scores. This limitation partly reflects the intrinsic ability of GC to function as a natural donor site in approximately 1% of U2-type introns.[45] More broadly, it highlights the difficulty of predicting splicing outcomes that are not all-or-none, leaving many variants within a functional grey zone that current algorithms cannot readily model or quantify.

Fourth, our results highlight the methodological challenges inherent in experimentally characterizing splice-altering variants. RNA from disease-relevant tissues provides the most reliable readout, yet such material is rarely available, and surrogate sources such as blood cells or lymphoblastoid lines are informative only if the gene of interest is expressed. In the absence of RNA from affected individuals, cell-based assays—FLGSA,[12] EMG,[23] midigene,[46] or minigene[47] systems—are often used, but each has intrinsic limitations. For example, not only is FLGSA impractical for large genes, but also not all successfully constructed full-length WT constructs yield the expected WT transcript. The most frequently used minigene assays are performed in a highly artificial background that lacks the natural genomic context required for proper splicing regulation.[48,49] Such loss of context is particularly problematic for variants with intermediate effects, for which classification depends not only on detecting aberrant transcripts but also on quantifying residual WT transcript from the variant allele. Even when RNA from affected individuals is available, interpretation in heterozygotes can be

461 confounded by WT transcript derived from the unaffected allele. For this reason, in both our
462 previous[12] and current studies, analyses of disease-associated GT>GC variants were focused primarily
463 on homozygotes and hemizygotes. RNA-seq, although increasingly applied in clinical contexts,[50-52]
464 faces similar challenges of tissue accessibility and allele-specific quantification. Collectively, these
465 methodological obstacles underscore the need for refined experimental systems, such as those based
466 on transactivation or transdifferentiation.[53]

467 GT>GC variants that retain residual WT transcript reflect the degeneracy and context dependence
468 of splice-site recognition. As such, findings from this study can likely be extrapolated to many other
469 splice-altering variants—excluding those at the invariant +1, −1, and −2 positions—because *cis*-acting
470 motifs are degenerate, and deep intronic variants can generate novel splice sites that compete with
471 physiological ones. Although a comprehensive survey is beyond the scope of this study, several
472 observations support this view. In SpliceVarDB, approximately 50% of variants could not be assigned
473 unambiguously to "splice-altering" or "non–splice-altering" categories but instead fell into a "low-
474 frequency splice-altering" group, corresponding to weak or indeterminate evidence of
475 spliceogenicity,[54] suggesting that grey-zone variants are common. Among 12 FLGSA-analyzed
476 *SPINK1* coding variants showing aberrant splicing, nine (including presumed missense and
477 synonymous variants) also generated residual WT transcript.[55] Likewise, deep intronic variants such as
478 *CFTR* c.1584+18672A>G (GenBank: NM_000492.4, NG_016465.4)[56] and *CLRN1* c.254−643G>T
479 (GenBank: NM_174878.3, NG_009168.1)[57] have been reported to yield both aberrant and WT
480 transcripts. Collectively, these examples indicate that residual WT transcript generation is a recurrent
481 and underappreciated phenomenon across genes and variant types, producing outcomes along a
482 continuum rather than at binary extremes (Figure 5). Because residual WT transcript cannot always be
483 readily assessed in simple or compound heterozygotes, and splice-altering variants account for an
484 estimated 10–30% of all disease-associated alleles,[11,54,58] we infer that variants generating residual WT
485 transcript may represent a significant source of variable penetrance and expressivity.

486 Nonetheless, the boundaries of the functional grey zone (Figure 5) likely vary across genes,
487 depending on functional tolerance thresholds and modes of inheritance, such that an identical level of
488 residual WT transcript may be clinically inconsequential in one gene–disease context yet functionally

489 significant in another. Although developed here in the context of LoF variants for conceptual
490 simplicity, the grey-zone model may also extend to gain-of-function or dominant-negative variants,[59]
491 in which disease-associated alleles within a given gene are likewise unlikely to exert uniform effects.
492 Notably, this conceptual framework was established primarily for genes in which complete or near-
493 complete LoF alleles cause classical Mendelian diseases. As we have previously opined,[8] some
494 disease-associated genes may not necessarily harbor clearly "pathogenic" variants.

495 Taken together, our study establishes GT>GC variants as a tractable model for addressing the
496 classification challenges posed by splice-altering variants with intermediate functional consequences.
497 By using residual WT transcript as a measurable parameter—and integrating clinical and genetic data,
498 RNA analyses of material derived from affected individuals, full-length and minigene assays, and
499 computational predictions—we show that residual WT transcript generation is a recurrent and
500 biologically relevant phenomenon that frequently complicates classification. Gradual functional
501 effects are also observed for other variant types, as exemplified by missense variants in *BRCA1* (MIM:
502 113705)[60] and *SOD1* (MIM: 147450),[61] both analyzed using multiplexed assays of variant effect
503 (MAVE). These observations underscore the need to move beyond binary pathogenic/benign
504 frameworks toward models that accommodate a continuum of effects,[6,8] as efforts focused solely on
505 refining ACMG/AMP criteria within a binary framework (as reflected in recent studies[62-65]) are
506 unlikely to resolve the fundamental problem amplified by the field's transition from a "phenotype-
507 first" to a "genotype-first" approach.

508 To address this challenge, new conceptual, computational, and experimental strategies are
509 emerging. As highlighted by Whitcomb,[66] our recently proposed "pathogenic–predisposing–risk–
510 benign" framework[8] represents an important step toward resolving the variant classification challenge.
511 The integration of machine-learning models trained on large-scale population-based datasets[9] with
512 functional data generated from MAVE[67,68] holds great promise for defining the full spectrum of variant
513 effects revealed through analyses of increasingly large cohorts.[69-71] Ultimately, grey-zone variants may
514 constitute a substantial portion of those currently classified as VUS or excluded for exceeding the
515 allele-frequency thresholds set for high-penetrance disease variants. Such variants could also account
516 for part of the "missing heritability" observed in both rare and complex diseases.[72-74] Recognizing the

517 widespread presence of grey-zone variants and their contribution to genetic complexity will be
518 essential for realizing the full potential of precision medicine.[75]

519 Our study has several limitations. First, the number of GT>GC 5′ splice site variants known to
520 generate substantial amounts of WT transcript remains limited. Second, the level of WT transcript
521 generated from certain GT>GC variants, as determined by FLGSA or RT–PCR using material derived
522 from affected individuals, may differ from that occurring under true *in vivo* conditions within the
523 pathologically relevant tissue. Third, a potential alternative impact of GT>GC variants—namely, a
524 reduction in overall splicing efficiency without producing aberrant transcripts—could not be
525 adequately assessed in this analysis. Likewise, GT>GC variants that generate both WT and aberrant
526 transcripts, but in which the aberrant transcripts are degraded by nonsense-mediated decay (NMD),[76]
527 could not be evaluated. These two scenarios could lead to underestimation of the true extent of partial
528 LoF GT>GC variants. Fourth, while we highlighted the classification challenges posed by the variants
529 discussed in detail, we did not systematically attempt to specify how each variant should be classified.
530 This is primarily because, as we have previously noted, "defining precise thresholds between different
531 variant categories remains challenging and must be tailored to each gene and disease context,"[8] and
532 because the available functional analytic data are not conclusive or fully reliable in some cases, and
533 the associated phenotypes were not always well defined. Despite these limitations, our integrative
534 analysis of GT>GC variants provides a coherent framework for understanding how residual WT
535 transcript generation shapes the functional continuum of splice-altering variants and informs their
536 broader biological and clinical interpretation.

537 In conclusion, viewing GT>GC variants through the lens of residual WT transcript generation
538 offers valuable insight into how partial LoF splice-altering variants—and, more broadly, variants with
539 intermediate functional effects—can contribute to genetic complexity, complicate both *in silico*
540 prediction and experimental analysis, and challenge variant classification (Figure 5).

Data and code availability

All data supporting this study are provided within the article and its supplementary materials. This study did not generate any new datasets or code.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (82000611 (to J.-H.L.)) and the Shanghai Pujiang Program (2020PJD061 (to J.-H.L.)). Support for this study also came from the Institut National de la Santé et de la Recherche Médicale (INSERM), France. D.N.C. wishes to acknowledge financial support from Qiagen Inc through a License Agreement with Cardiff University. The funding bodies did not play any role in the study design, collection, analysis, and interpretation of data or the writing of the article and the decision to submit it for publication.

Author contributions

Conceptualization, J.M.C.; study design, J.H.L., H.W., W.B.Z. and J.M.C.; methodology, J.H.L., H.W., W.B.Z. and J.M.C.; experiments, J.H.L., H.W., X.Y.T. and W.B.Z.; literature search and variant collation, E.M. and J.M.C.; writing—original draft preparation, J.M.C.; writing—review and editing, J.H.L., H.W., X.Y.T., E.M., D.N.C., C.F., Z.L., W.B.Z. and J.M.C.; project administration, W.B.Z. and J.M.C.; funding acquisition, J.H.L. and J.M.C. All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Declaration of interests

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 5 in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Fowler, D.M., and Rehm, H.L. (2024). Will variants of uncertain significance still exist in 2030? Am J Hum Genet *111*, 5-10. 10.1016/j.ajhg.2023.11.005.
2. Forrest, I.S., Huang, K.L., Eggington, J.M., Chung, W.K., Jordan, D.M., and Do, R. (2025). Using large-scale population-based data to improve disease risk assessment of clinical variants. Nat Genet *57*, 1588-1597. 10.1038/s41588-025-02212-3.
3. Wright, C.F., Sharp, L.N., Jackson, L., Murray, A., Ware, J.S., MacArthur, D.G., Rehm, H.L., Patel, K.A., and Weedon, M.N. (2024). Guidance for estimating penetrance of monogenic disease-causing variants in population cohorts. Nat Genet *56*, 1772-1779. 10.1038/s41588-024-01842-3.
4. Torene, R.I., Murphy, K.M., Brandt, T., Kelly, M.A., Willard, H.F., and Retterer, K. (2025). A scalable approach for genomic-first rare disorder detection in a healthcare-based population. Am J Hum Genet. 10.1016/j.ajhg.2025.09.010.
5. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med *17*, 405-424. 10.1038/gim.2015.30.
6. Masson, E., Zou, W.B., Génin, E., Cooper, D.N., Le Gac, G., Fichou, Y., Pu, N., Rebours, V., Férec, C., Liao, Z., and Chen, J.M. (2022). Expanding ACMG variant classification guidelines into a general framework. Hum Genomics *16*, 31. 10.1186/s40246-022-00407-x.
7. Schmidt, R.J., Steeves, M., Bayrak-Toydemir, P., Benson, K.A., Coe, B.P., Conlin, L.K., Ganapathi, M., Garcia, J., Gollob, M.H., Jobanputra, V., et al. (2024). Recommendations for risk allele evidence curation, classification, and reporting from the ClinGen Low Penetrance/Risk Allele Working Group. Genet Med *26*, 101036. 10.1016/j.gim.2023.101036.
8. Wang, Y.C., Masson, E., Wang, Q.W., Genin, E., Le Gac, G., Fichou, Y., Cooper, D.N., Liao, Z., Ferec, C., Zou, W.B., and Chen, J.M. (2025). *SPINK1*-related chronic pancreatitis: A model that encapsulates the spectrum of variant effects, genetic complexity, and classificatory challenges. Am J Hum Genet *112*, 2043-2066. 10.1016/j.ajhg.2025.07.013.
9. Forrest, I.S., Vy, H.M.T., Rocheleau, G., Jordan, D.M., Petrazzini, B.O., Nadkarni, G.N., Cho, J.H., Ganapathi, M., Huang, K.L., Chung, W.K., and Do, R. (2025). Machine learning-based penetrance of genetic variants. Science *389*, eadm7066. 10.1126/science.adm7066.
10. Walker, L.C., Hoya, M., Wiggins, G.A.R., Lindy, A., Vincent, L.M., Parsons, M.T., Canson, D.M., Bis-Brewer, D., Cass, A., Tchourbanov, A., et al. (2023). 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 *110*, 1046-1067. 10.1016/j.ajhg.2023.06.002.

608 11. Sullivan, P.J., Quinn, J.M.W., Ajuyah, P., Pinese, M., Davis, R.L., and Cowley, M.J. (2025). Data-
609 driven insights to inform splice-altering variant assessment. Am J Hum Genet *112*, 764-778.
610 10.1016/j.ajhg.2025.02.012.
611 12. Lin, J.H., Tang, X.Y., Boulling, A., Zou, W.B., Masson, E., Fichou, Y., Raud, L., Le Tertre, M.,
612 Deng, S.J., Berlivet, I., et al. (2019). First estimate of the scale of canonical 5' splice site
613 GT>GC variants capable of generating wild-type transcripts. Hum Mutat *40*, 1856-1873.
614 10.1002/humu.23821.
615 13. Ramalho, A.S., Beck, S., Meyer, M., Penque, D., Cutting, G.R., and Amaral, M.D. (2002). Five
616 percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates
617 the severity of pulmonary disease in cystic fibrosis. Am J Respir Cell Mol Biol *27*, 619-627.
618 10.1165/rcmb.2001-0004OC.
619 14. Raraigh, K.S., Han, S.T., Davis, E., Evans, T.A., Pellicore, M.J., McCague, A.F., Joynt, A.T., Lu, Z.,
620 Atalar, M., Sharma, N., et al. (2018). Functional assays are essential for interpretation of
621 missense variants associated with variable expressivity. Am J Hum Genet *102*, 1062-1077.
622 10.1016/j.ajhg.2018.04.003.
623 15. Scalet, D., Maestri, I., Branchini, A., Bernardi, F., Pinotti, M., and Balestra, D. (2019). Disease-
624 causing variants of the conserved +2T of 5' splice sites can be rescued by engineered
625 U1snRNAs. Hum Mutat *40*, 48-52. 10.1002/humu.23680.
626 16. Den Uijl, I.E., Mauser Bunschoten, E.P., Roosendaal, G., Schutgens, R.E., Biesma, D.H.,
627 Grobbee, D.E., and Fischer, K. (2011). Clinical severity of haemophilia A: does the
628 classification of the 1950s still stand? Haemophilia *17*, 849-853. 10.1111/j.1365-
629 2516.2011.02539.x.
630 17. Landrum, M.J., Chitipiralla, S., Kaur, K., Brown, G., Chen, C., Hart, J., Hoffman, D., Jang, W.,
631 Liu, C., Maddipatla, Z., et al. (2025). ClinVar: updates to support classifications of both
632 germline and somatic variants. Nucleic Acids Res *53*, D1313-D1321. 10.1093/nar/gkae1090.
633 18. Goldberg, Y., Laitman, Y., Ben David, M., Bazak, L., Lidzbarsky, G., Salmon, L.B., Shkedi-Rafid,
634 S., Barshack, I., Avivi, C., Darawshe, M., et al. (2021). Re-evaluating the pathogenicity of the
635 c.783+2T>C *BAP1* germline variant. Hum Mutat *42*, 592-599. 10.1002/humu.24189.
636 19. Sculco, M., La Vecchia, M., Aspesi, A., Clavenna, M.G., Salvo, M., Borgonovi, G., Pittaro, A.,
637 Witel, G., Napoli, F., Listi, A., et al. (2022). Diagnostics of *BAP1*-tumor predisposition
638 syndrome by a multitesting approach: A ten-year-long experience. Diagnostics (Basel) *12*,
639 1710. 10.3390/diagnostics12071710.
640 20. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D.,
641 Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting splicing
642 from primary sequence with deep learning. Cell *176*, 535-548. 10.1016/j.cell.2018.12.015.
643 21. Castellani, C., and CFTR2 team. (2013). CFTR2: How will it help care? Paediatr Respir Rev 14
644 Suppl 1, 2-5. 10.1016/j.prrv.2013.01.006.
645 22. Claustres, M., Theze, C., des Georges, M., Baux, D., Girodon, E., Bienvenu, T., Audrézet, M.P.,
646 Dugueperoux, I., Férec, C., Lalau, G., et al. (2017). CFTR-France, a national relational patient
647 database for sharing genetic and phenotypic data associated with rare *CFTR* variants. Hum
648 Mutat *38*, 1297-1315. 10.1002/humu.23276.
649 23. Joynt, A.T., Evans, T.A., Pellicore, M.J., Davis-Marcisak, E.F., Aksit, M.A., Eastman, A.C., Patel,
650 S.U., Paul, K.C., Osorio, D.L., Bowling, A.D., et al. (2020). Evaluation of both exonic and
651 intronic variants for effects on RNA splicing allows for accurate assessment of the
652 effectiveness of precision therapies. PLoS Genet *16*, e1009100.
653 10.1371/journal.pgen.1009100.
654 24. Lin, J.H., Wu, H., Zou, W.B., Masson, E., Fichou, Y., Le Gac, G., Cooper, D.N., Férec, C., Liao, Z.,
655 and Chen, J.M. (2021). Splicing outcomes of 5' splice site GT>GC variants that generate wild-
656 type transcripts differ significantly between full-length and minigene splicing assays. Front
657 Genet *12*, 701652. 10.3389/fgene.2021.701652.
658 25. Muller, J.S., Piko, H., Schoser, B.G., Schlotter-Weigel, B., Reilich, P., Gurster, S., Born, C.,
659 Karcagi, V., Pongratz, D., Lochmuller, H., and Walter, M.C. (2006). Novel splice site mutation

660 in the caveolin-3 gene leading to autosomal recessive limb girdle muscular dystrophy.
661 Neuromuscul Disord *16*, 432-436. 10.1016/j.nmd.2006.04.006.
662 26. Seyama, K., Nonoyama, S., Gangsaas, I., Hollenbaugh, D., Pabst, H.F., Aruffo, A., and Ochs,
663 H.D. (1998). Mutations of the CD40 ligand gene and its effect on CD40 ligand expression in
664 patients with X-linked hyper IgM syndrome. Blood *92*, 2421-2434.
665 27. Bartolo, C., Papp, A.C., Snyder, P.J., Sedra, M.S., Burghes, A.H., Hall, C.D., Mendell, J.R., and
666 Prior, T.W. (1996). A novel splice site mutation in a Becker muscular dystrophy patient. J Med
667 Genet *33*, 324-327. 10.1136/jmg.33.4.324.
668 28. Frischknecht, H., Dutly, F., Walker, L., Nakamura-Garrett, L.M., Eng, B., and Waye, J.S. (2009).
669 Three new beta-thalassemia mutations with varying degrees of severity. Hemoglobin *33*,
670 220-225. 10.1080/03630260903089060.
671 29. Kume, K., Masamune, A., Kikuta, K., and Shimosegawa, T. (2006). [-215G>A; IVS3+2T>C]
672 mutation in the *SPINK1* gene causes exon 3 skipping and loss of the trypsin binding site. Gut
673 *55*, 1214. 10.1136/gut.2006.095752.
674 30. Chen, J.M., Lin, J.H., Masson, E., Liao, Z., Férec, C., Cooper, D.N., and Hayden, M. (2020). The
675 experimentally obtained functional impact assessments of 5' splice site GT>GC variants differ
676 markedly from those predicted. Curr Genomics *21*, 56-66.
677 10.2174/1389202921666200210141701.
678 31. Jayasree, D., Shaji, R.V., George, B., Mathews, V., Srivastava, A., and Edison, E.S. (2016).
679 Clinical, hematological and molecular analysis of homozygous Hb E (*HBB*: c.79G > A) in the
680 Indian population. Hemoglobin *40*, 16-19. 10.3109/03630269.2015.1086880.
681 32. Aebi, M., Hornig, H., Padgett, R.A., Reiser, J., and Weissmann, C. (1986). Sequence
682 requirements for splicing of higher eukaryotic nuclear pre-mRNA. Cell *47*, 555-565.
683 10.1016/0092-8674(86)90620-3.
684 33. Aebi, M., Hornig, H., and Weissmann, C. (1987). 5' cleavage site in eukaryotic pre-mRNA
685 splicing is determined by the overall 5' splice region, not by the conserved 5' GU. Cell *50*,
686 237-246. 10.1016/0092-8674(87)90219-4.
687 34. Tesio, N., and Bauer, D.E. (2023). Molecular basis and genetic modifiers of thalassemia.
688 Hematol Oncol Clin North Am *37*, 273-299. 10.1016/j.hoc.2022.12.001.
689 35. Peretto, L., Tonetto, E., Maestri, I., Bezzerri, V., Valli, R., Cipolli, M., Pinotti, M., and Balestra,
690 D. (2023). Counteracting the common Shwachman-Diamond Syndrome-causing *SBDS*
691 c.258+2T>C mutation by RNA therapeutics and base/prime editing. Int J Mol Sci *24*, 4024.
692 10.3390/ijms24044024.
693 36. Schmaltz-Panneau, B., Pagnier, A., Clauin, S., Buratti, J., Marty, C., Fenneteau, O., Dieterich,
694 K., Beaupain, B., Donadieu, J., Plo, I., and Bellanne-Chantelot, C. (2021). Identification of
695 biallelic germline variants of *SRP68* in a sporadic case with severe congenital neutropenia.
696 Haematologica *106*, 1216-1219. 10.3324/haematol.2020.247825.
697 37. Lao, Q., Mallappa, A., Rueda Faucz, F., Joyal, E., Veeraraghavan, P., Chen, W., and Merke, D.P.
698 (2021). A *TNXB* splice donor site variant as a cause of hypermobility type Ehlers-Danlos
699 syndrome in patients with congenital adrenal hyperplasia. Mol Genet Genomic Med *9*,
700 e1556. 10.1002/mgg3.1556.
701 38. Nix, P., Mundt, E., Coffee, B., Goossen, E., Warf, B.M., Brown, K., Bowles, K., and Roa, B.
702 (2022). Interpretation of *BRCA2* splicing variants: A case series of challenging variant
703 interpretations and the importance of functional RNA analysis. Fam Cancer *21*, 7-19.
704 10.1007/s10689-020-00224-y.
705 39. Pruss, D., Morris, B., Hughes, E., Eggington, J.M., Esterling, L., Robinson, B.S., van Kan, A.,
706 Fernandes, P.H., Roa, B.B., Gutin, A., et al. (2014). Development and validation of a new
707 algorithm for the reclassification of genetic variants identified in the *BRCA1* and *BRCA2*
708 genes. Breast Cancer Res Treat *147*, 119-132. 10.1007/s10549-014-3065-9.
709 40. Tubeuf, H., Caputo, S.M., Sullivan, T., Rondeaux, J., Krieger, S., Caux-Moncoutier, V.,
710 Hauchard, J., Castelain, G., Fievet, A., Meulemans, L., et al. (2020). Calibration of

711 pathogenicity due to variant-induced leaky splicing defects by using *BRCA2* exon 3 as a model
712 system. Cancer Res *80*, 3593-3605. 10.1158/0008-5472.CAN-20-0895.
713 41. Huynh, A., Gray, P.E., Sullivan, A., Mackie, J., Guerin, A., Rao, G., Pathmanandavel, K., Mina,
714 E.D., Hollway, G., Hobbs, M., et al. (2024). A Novel case of *IFNAR1* deficiency identified a
715 common canonical splice site variant in *DOCK8* in Western Polynesia: The importance of
716 validating variants of unknown significance in under-represented ancestries. J Clin Immunol
717 *44*, 170. 10.1007/s10875-024-01774-x.
718 42. Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J.,
719 Lok, S., Plavsic, N., Chou, J.L., and et al. (1989). Identification of the cystic fibrosis gene:
720 cloning and characterization of complementary DNA. Science *245*, 1066-1073.
721 10.1126/science.2475911.
722 43. Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A.,
723 Buchwald, M., and Tsui, L.C. (1989). Identification of the cystic fibrosis gene: genetic analysis.
724 Science *245*, 1073-1080. 10.1126/science.2570460.
725 44. Lundman, E., Gaup, H.J., Bakkeheim, E., Olafsdottir, E.J., Rootwelt, T., Storrosten, O.T., and
726 Pettersen, R.D. (2016). Implementation of newborn screening for cystic fibrosis in Norway.
727 Results from the first three years. J Cyst Fibros *15*, 318-324. 10.1016/j.jcf.2015.12.017.
728 45. Burset, M., Seledtsov, I.A., and Solovyev, V.V. (2000). Analysis of canonical and non-canonical
729 splice sites in mammalian genomes. Nucleic Acids Res *28*, 4364-4375.
730 10.1093/nar/28.21.4364.
731 46. Sangermano, R., Khan, M., Cornelis, S.S., Richelle, V., Albert, S., Garanto, A., Elmelik, D.,
732 Qamar, R., Lugtenberg, D., van den Born, L.I., et al. (2018). *ABCA4* midigenes reveal the full
733 splice spectrum of all reported noncanonical splice site variants in Stargardt disease. Genome
734 Res *28*, 100-110. 10.1101/gr.226621.117.
735 47. Buisine, M.P., Bellanne-Chantelot, C., Calmels, N., Vache, C., Besnard, T., Cogne, B., Vitobello,
736 A., Piton, A., Martins, A., Gaildrat, P., et al. (2025). RNA-based diagnostic studies in genetics:
737 Review and guidance from a multidisciplinary French network. Eur J Hum Genet *33*, 1219-
738 1227. 10.1038/s41431-025-01881-2.
739 48. Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-
740 binding proteins. Nat Rev Genet *15*, 689-701. 10.1038/nrg3778.
741 49. Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing kinetics and coordination
742 revealed by direct nascent RNA sequencing through nanopores. Mol Cell *77*, 985-998.
743 10.1016/j.molcel.2019.11.017.
744 50. Oh, R.Y., AlMail, A., Cheerie, D., Guirguis, G., Hou, H., Yuki, K.E., Haque, B.,
745 Thiruvahindrapuram, B., Marshall, C.R., Mendoza-Londono, R., et al. (2024). A systematic
746 assessment of the impact of rare canonical splice site variants on splicing using functional
747 and in silico methods. HGG Adv *5*, 100299. 10.1016/j.xhgg.2024.100299.
748 51. Zhao, S., Macakova, K., Sinson, J.C., Dai, H., Rosenfeld, J., Zapata, G.E., Li, S., Ward, P.A.,
749 Wang, C., Qu, C., et al. (2025). Clinical validation of RNA sequencing for Mendelian disorder
750 diagnostics. Am J Hum Genet *112*, 779-792. 10.1016/j.ajhg.2025.02.006.
751 52. Sha, Y., Ortiz, J.B., Bristow, S.L., Loranger, K., Meng, L., Zhao, X., Xia, F., Parmar, S., ElNaggar,
752 A.C., and Xu, W. (2025). Refining the interpretation of variants of uncertain significance in
753 hereditary cancer screening through integrated RNA sequencing. Genet Med Open *3*,
754 101914. 10.1016/j.gimo.2024.101914.
755 53. Nicolas-Martinez, E.C., Robinson, O., Pflueger, C., Gardner, A., Corbett, M.A., Ritchie, T.,
756 Kroes, T., van Eyk, C.L., Scheffer, I.E., Hildebrand, M.S., et al. (2024). RNA variant assessment
757 using transactivation and transdifferentiation. Am J Hum Genet *111*, 1673-1699.
758 10.1016/j.ajhg.2024.06.018.
759 54. Sullivan, P.J., Quinn, J.M.W., Wu, W., Pinese, M., and Cowley, M.J. (2024). SpliceVarDB: A
760 comprehensive database of experimentally validated human splicing variants. Am J Hum
761 Genet *111*, 2164-2175. 10.1016/j.ajhg.2024.08.002.

55. Wu, H., Lin, J.H., Tang, X.Y., Marenne, G., Zou, W.B., Schutz, S., Masson, E., Génin, E., Fichou, Y., Le Gac, G., et al. (2024). Combining full-length gene assay and SpliceAI to interpret the splicing impact of all possible *SPINK1* coding variants. Hum Genomics *18*, 21. 10.1186/s40246-024-00586-9.
56. Costantino, L., Rusconi, D., Solda, G., Seia, M., Paracchini, V., Porcaro, L., Asselta, R., Colombo, C., and Duga, S. (2013). Fine characterization of the recurrent c.1584+18672A>G deep-intronic mutation in the cystic fibrosis transmembrane conductance regulator gene. Am J Respir Cell Mol Biol *48*, 619-625. 10.1165/rcmb.2012-0371OC.
57. Elasal, M.A., Khateb, S., Panneman, D.M., Roosing, S., Cremers, F.P.M., Banin, E., Sharon, D., and Sarma, A.S. (2024). A leaky deep intronic splice variant in *CLRN1* is associated with non-syndromic retinitis pigmentosa. Genes (Basel) *15*, 1363. 10.3390/genes15111363.
58. Stenson, P.D., Mort, M., Ball, E.V., Chapman, M., Evans, K., Azevedo, L., Hayden, M., Heywood, S., Millar, D.S., Phillips, A.D., and Cooper, D.N. (2020). The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Hum Genet *139*, 1197-1207. 10.1007/s00439-020-02199-3.
59. Gerasimavicius, L., Livesey, B.J., and Marsh, J.A. (2022). Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. Nat Commun *13*, 3895. 10.1038/s41467-022-31686-6.
60. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of *BRCA1* variants with saturation genome editing. Nature *562*, 217-222. 10.1038/s41586-018-0461-z.
61. Axakova, A., Ding, M., Cote, A.G., Subramaniam, R., Senguttuvan, V., Zhang, H., Weile, J., Douville, S.V., Gebbia, M., Al-Chalabi, A., et al. (2025). Landscapes of missense variant impact for human superoxide dismutase 1. Am J Hum Genet *112*, 2295-2315. 10.1016/j.ajhg.2025.08.019.
62. Richardson, M.E., Bishop, M.F.H., Holdren, M.A., de la Hoya, M., Spurdle, A.B., Tavtigian, S.V., Brannan, T., Young, C.C., Zec, L., Hiraki, S., et al. (2025). Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline *PALB2* sequence variants. Am J Hum Genet *112*, 2266-2280. 10.1016/j.ajhg.2025.08.020.
63. Rowlands, C.F., Allen, S., Garrett, A., Durkie, M., Burghel, G.J., Robinson, R., Callaway, A., Field, J., Frugtniet, B., Palmer-Smith, S., et al. (2025). Availability of benign missense variant "truthsets" for validation of functional assays: Current status and a systematic approach. Am J Hum Genet *112*, 2281-2294. 10.1016/j.ajhg.2025.08.014.
64. Liu, S., Feng, X., Wu, Y., and Bu, F. Calibration and refinement of ACMG/AMP criteria for variant classification with BayesQuantify. J Med Genet, 2025 Oct 2022:jmg-2025-110863. doi: 110810.111136/jmg-112025-110863. Online ahead of print. 10.1136/jmg-2025-110863.
65. Yuan, X., Xia, X., Li, J., and Zhao, G. (2025). HCSeeker: A classification tool for human genetic variant hot and cold spots designed for PM1 and benign criteria in the ACMG-AMP guideline. Genet Med, 101591. 10.1016/j.gim.2025.101591.
66. Whitcomb, D.C. (2025). Adoption of a “Pathogenic, Predisposing, Risk and Benign” classification of DNA variants for precision medicine – a step forward. SMART-MD Journal of Precision Medicine *2*, e.i-e.ii. https://doi.org/10.69734/h7x6zm57.
67. Rubin, A.F., Stone, J., Bianchi, A.H., Capodanno, B.J., Da, E.Y., Dias, M., Esposito, D., Frazer, J., Fu, Y., Grindstaff, S.B., et al. (2025). MaveDB 2024: A curated community database with over seven million variant effects from multiplexed functional assays. Genome Biol *26*, 13. 10.1186/s13059-025-03476-y.
68. Allen, S., Garrett, A., Rowlands, C.F., Durkie, M., Burghel, G.J., Robinson, R., Callaway, A., Field, J., Frugtniet, B., Palmer-Smith, S., et al. (2025). Validating data from multiplex assays of variant effect: A CanVIG-UK national survey of NHS clinical scientists. Am J Hum Genet *112*, 1479-1488. 10.1016/j.ajhg.2025.04.006.

69. Heyne, H.O., Karjalainen, J., Karczewski, K.J., Lemmela, S.M., Zhou, W., FinnGen, Havulinna, A.S., Kurki, M., Rehm, H.L., Palotie, A., and Daly, M.J. (2023). Mono- and biallelic variant effects on disease at biobank scale. Nature *613*, 519-525. 10.1038/s41586-022-05420-7.
70. Lee, D.S.M., Cardone, K.M., Zhang, D.Y., Tsao, N.L., Abramowitz, S., Sharma, P., DePaolo, J.S., Conery, M., Aragam, K.G., Biddinger, K., et al. (2025). Common-variant and rare-variant genetic architecture of heart failure across the allele-frequency spectrum. Nat Genet *57*, 829-838. 10.1038/s41588-025-02140-2.
71. Garcia Hernandez, S., de la Higuera Romero, L., Fernandez, A., Luisa Pena Pena, M., Mora-Ayestaran, N., Basurte-Elorz, M.T., Larranaga-Moreira, J.M., Cardenas Reyes, I., Villacorta, E., Valverde-Gomez, M., et al. (2025). Redefining the genetic architecture of hypertrophic cardiomyopathy: Role of Intermediate-effect variants. Circulation *152*, 1060-1075. 10.1161/CIRCULATIONAHA.125.074529.
72. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753. 10.1038/nature08494.
73. Maroilley, T., and Tarailo-Graovac, M. (2019). Uncovering missing heritability in rare diseases. Genes (Basel) *10*, 275. 10.3390/genes10040275.
74. Génin, E. (2020). Missing heritability of complex diseases: case solved? Hum Genet *139*, 103-113. 10.1007/s00439-019-02034-4.
75. Denny, J.C., and Collins, F.S. (2021). Precision medicine in 2030-seven ways to transform healthcare. Cell *184*, 1415-1419. 10.1016/j.cell.2021.01.015.
76. Torene, R.I., Guillen Sacoto, M.J., Millan, F., Zhang, Z., McGee, S., Oetjens, M., Heise, E., Chong, K., Sidlow, R., O'Grady, L., et al. (2024). Systematic analysis of variants escaping nonsense-mediated decay uncovers candidate Mendelian diseases. Am J Hum Genet *111*, 70-81. 10.1016/j.ajhg.2023.11.007.

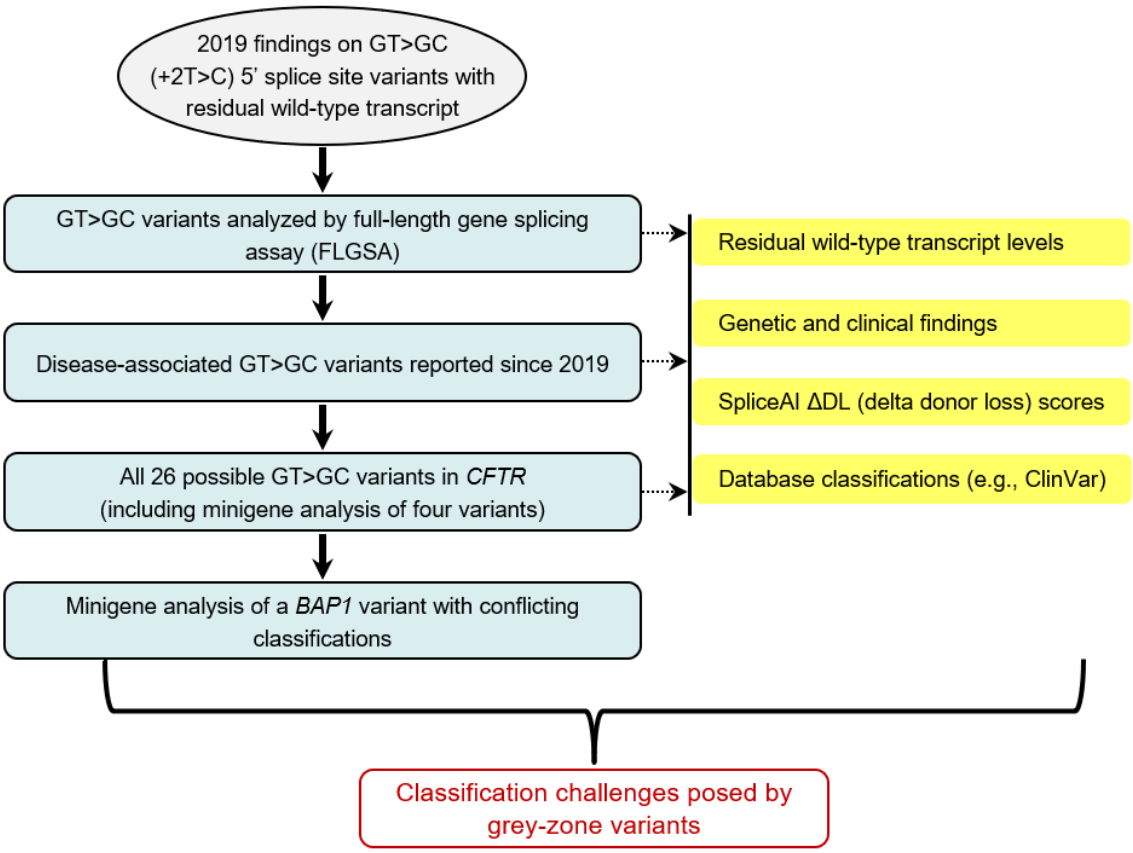
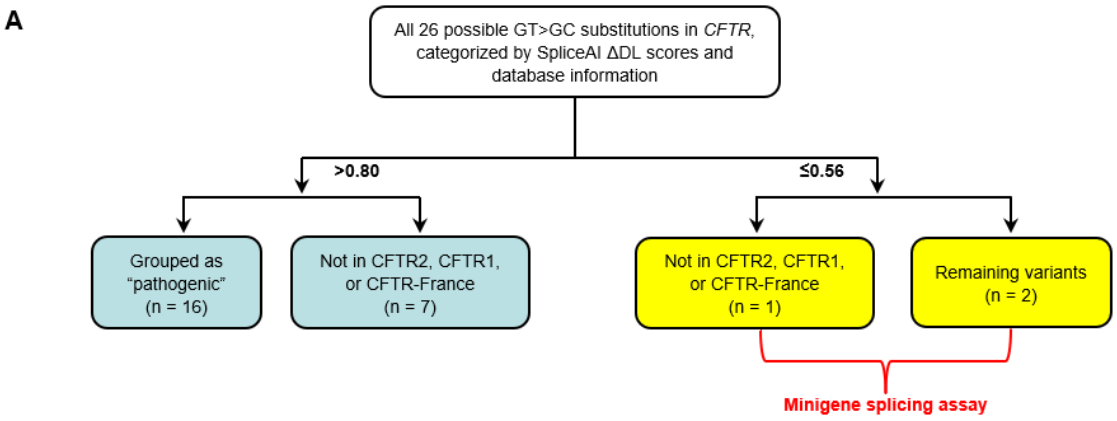


Figure 1. Scheme of the analytical approach

GT>GC variants were analyzed in three successive categories, together with minigene analysis of a *BAP1* variant with conflicting classifications. Within each category, variants were examined for four features (yellow shading). The red box highlights the central issue addressed in this study—the classification challenges posed by grey-zone variants that retain substantial yet incomplete wild-type transcript generation.



B

CFTR variants selected for functional analysis using minigene splicing assays

Variant	SpliceAI ΔDL score	Database classification and reporting				Residual wild-type transcript	
		CFTR2	CFTR-France	CFTR1	ClinVar	Previous study	This study
c.2490+2T>C	0.35	–	VUS (reported in two asymptomatic homozygotes)	Identified in heterozygous state in the partner of a CBAVD patient	CCP	–	Yes (pET01); No (pSPL3)
c.3717+2T>C	0.56	–	–	–	–	–	No in both assays
c.3873+2T>C	0.98	CF-causing	CF-causing	Commonly reported in Norwegian CF patients	P	Yes	No in both assays
c.4242+2T>C	0.15	–	–	Reported in two individuals with suspected CF diagnosis	P/LP	Yes	No in both assays

Figure 2. Analysis of GT>GC substitutions in *CFTR*

(A) Flowchart summarizing the 26 possible *CFTR* GT>GC substitutions categorized by SpliceAI ΔDL scores and database information (see Table S1 for details). Variants with low scores (≤0.56) were selected for minigene splicing assays.

(B) Four variants analyzed by minigene splicing assays. In "Database classification and reporting," note that, unlike CFTR2 and CFTR-France, CFTR1 does not classify registered variants. In "Residual wild-type transcript," "Yes" and "No" indicate whether the variant allele generated wild-type transcript. Results from this study are shown separately for the two assay systems (pET01 and pSPL3; see Figure 3), with "No in both assays" indicating absence of wild-type transcript from the variant allele in both systems. "Previous study" refers to Joynt et al.[23]

Abbreviations: CBAVD, congenital bilateral absence of the vas deferens; CF, cystic fibrosis; CCP, conflicting classifications of pathogenicity; ΔDL, delta donor loss; P, pathogenic; P/LP, pathogenic/likely pathogenic; VUS, variant of uncertain significance.

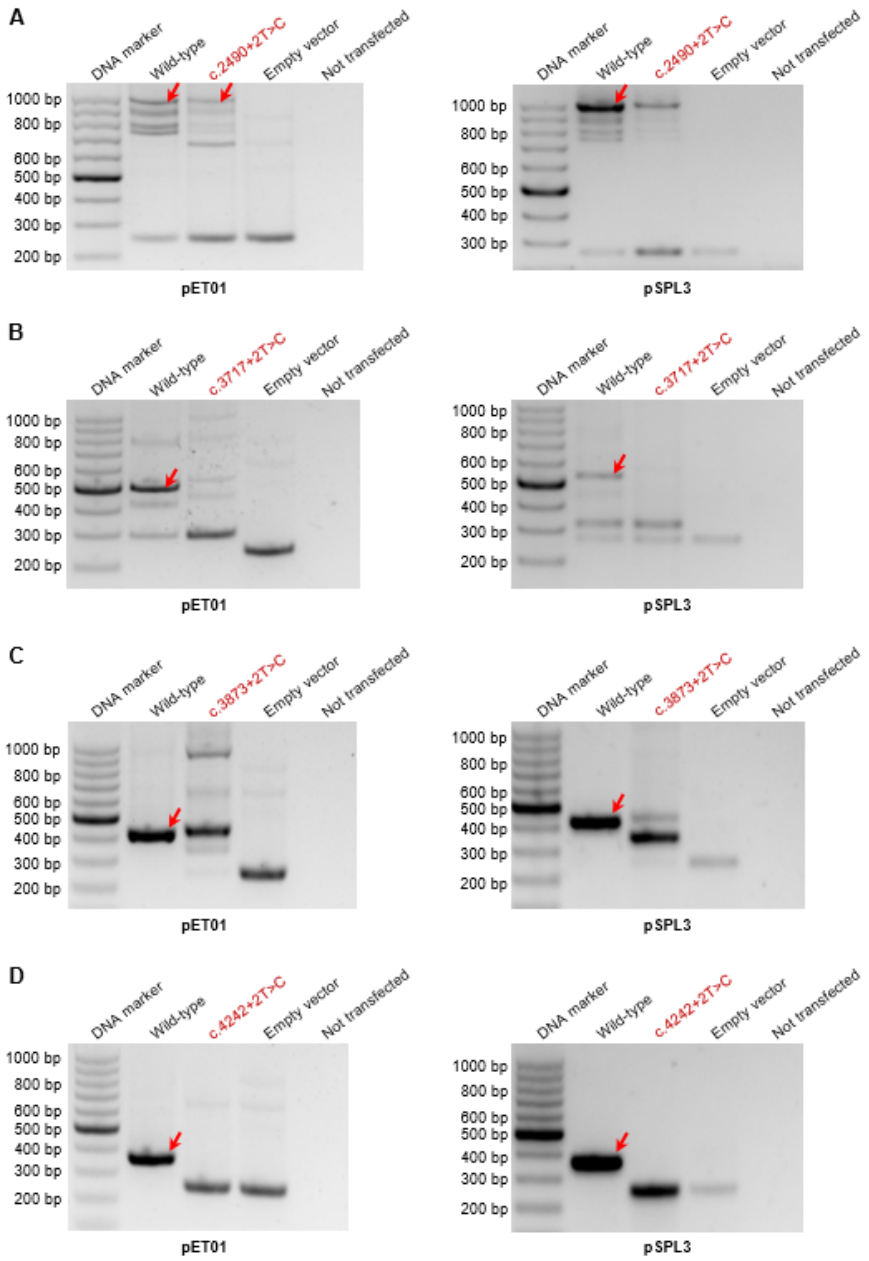
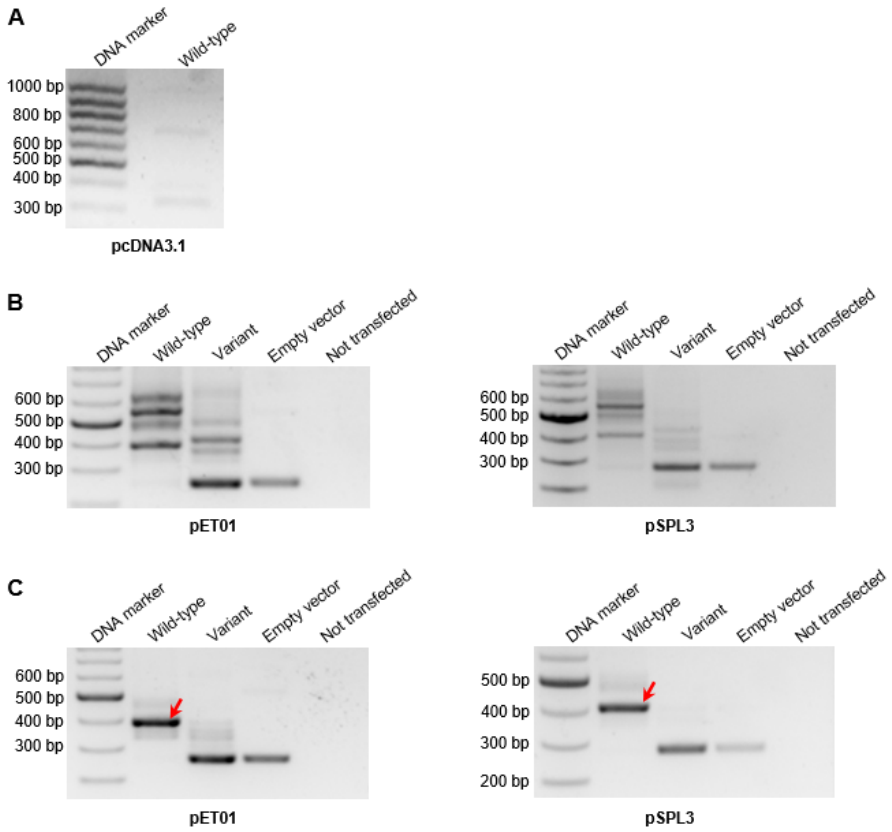


Figure 3. Minigene splicing analysis of four selected *CFTR* GT>GC substitutions using pET01 and pSPL3 vectors

Each variant was analyzed with its preceding exon and short flanking intronic sequences cloned into the respective reporter vectors. Arrows indicate bands corresponding to the normally spliced transcript, in accordance with the *CFTR* mRNA reference sequence NM_000492.4. Additional bands are annotated in Figure S1.



871
872

873 **Figure 4. Splicing analysis of the *BAP1* c.783+2T>C variant in different assay contexts**

874 **(A)** Full-length gene splicing assay (FLGSA). The wild-type (WT) construct failed to generate clear
875 reverse transcription–PCR (RT–PCR) bands.

876 **(B)** Minigene analysis of exons 8–10 (including exon 8, intron 8, exon 9, intron 9, and exon 10)
877 with short flanking sequences from introns 7 and 10 cloned into the pET01 and pSPL3 vectors. No
878 normally spliced transcripts were detected, in accordance with the *BAP1* mRNA reference sequence
879 NM_004656.4, even from the WT constructs. For band annotations, see Figure S2.

880 **(C)** Minigene analysis of exon 9 with short flanking intronic sequences cloned into the pET01 and
881 pSPL3 vectors. Arrows indicate bands corresponding to the normally spliced transcript, in accordance
882 with the *BAP1* mRNA reference sequence NM_004656.4. Additional bands are annotated in Figure S2.
883 *BAP1* genomic reference sequence: NG_031859.1.

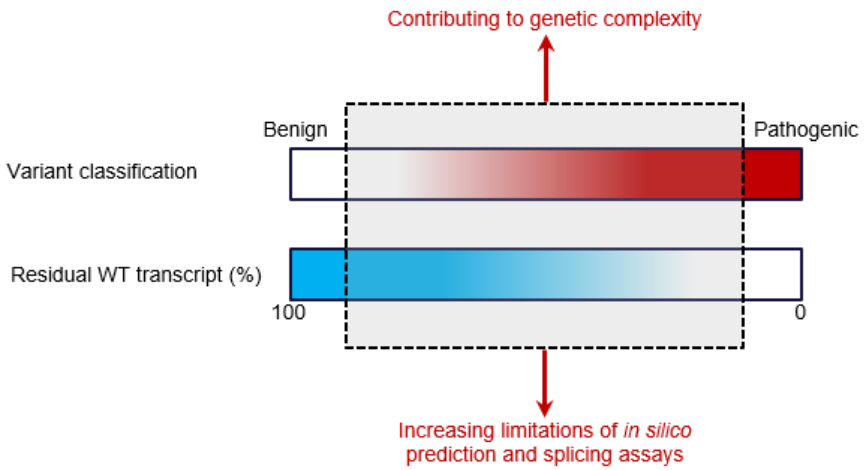


Figure 5. Conceptual model of the functional grey zone in variant classification

Residual wild-type (WT) transcript levels and variant classification lie along continua rather than binary states. Variants with intermediate WT transcript levels often fall within a functional grey zone, where *in silico* prediction and experimental analysis become increasingly difficult. Such grey-zone variants may act through variant–variant (both intra- and intergenic) or variant–environment interactions, thereby contributing to the genetic complexity of human disease. Although illustrated here with splice-altering variants, this conceptual model also applies to other variant types and disease mechanisms shaped by graded levels of functional impact.

895 **Table 1. Five GT>GC variants with residual WT transcript that present classification challenges**

Variant[a]	Reference mRNA and genomic sequences	hg38 coordinate	Residual WT transcript[b]	Assessment method	SpliceAI ΔDL score	Clinical/genetic findings	ClinVar classification
SPINK1 c.194+2T>C	NM_001379610.1, NG_008356.2	chr5-147828020-A-G	10%	RT–PCR of material from an affected homozygote;[29] FLGSA[12]	0.33	Well-defined genetic effect, too weak for pathogenic classification[8]	Pathogenic/likely pathogenic
HBB c.315+2T>C	NM_000518.5, NG_059281.1	chr11-5226575-A-G	30%	FLGSA[12]	0.72	Originally described as a mild β^+-thalassemia allele[28]	Pathogenic
MGP c.94+2T>C	NM_000900.5, NG_023331.2	chr12-14884211-A-G	80%	FLGSA[12]	0.97	Not reported in affected individuals	Likely pathogenic
BRCA2 c.8331+2T>C	NM_000059.4, NG_012772.3	chr13-32363535-T-C	60%	Allele-specific RT–PCR of affected individual–derived material[38]	0.93	*BRCA2* history-weighting algorithm[39] indicated a benign profile;[38] *BRCA2* exhibits tolerance to reduced transcript dosage[40]	Pathogenic/likely pathogenic
DOCK8 c.3234+2T>C	NM_203447.4, NG_017007.1	chr9-399261-T-C	Predominant WT	RT–PCR and protein analyses of T cells from an affected homozygote[41]	0.49	Disease explained by an unlinked homozygous *IFNAR1* nonsense variant[41]	Conflicting classifications of pathogenicity

896 Abbreviations: ΔDL, delta donor loss; FLGSA, full-length gene splicing assay; RT-PCR, reverse transcription-PCR; WT, wild-type.

897 [a]Variants are arranged in order of mention in the main text.

898 [b]Percentage of WT transcript generated from the variant allele relative to that from the WT allele.