When splicing is not all or none: Implications 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 exemplify the degeneracy and context dependence of splice-site recognition, insights from their study 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
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,[17,18] was included for additional minigene splicing
99 analysis.
100

101 ***In silico* splicing predictions**

102 *In silico* predictions were obtained using SpliceAI[19] 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/)[20] 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 (Table S1), we selected all three with low SpliceAI DL
119 scores (≤0.56)—c.2490+2T>C, c.3717+2T>C and c.4242+2T>C—together with one variant
120 (c.3873+2T>C) that had a high SpliceAI score (0.98) but was previously reported to generate WT
121 transcript,[23] for minigene splicing analysis.

122 NM_000492.4 and NG_016465.4 were used as the mRNA and genomic reference sequences of
123 *CFTR*, respectively.

124

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

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

131

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

133 Experiments were carried out essentially as previously described.[24] Human embryonic kidney 293T
134 (HEK293T) cells were maintained in DMEM (Gibco) supplemented with 10% fetal calf serum

135 (Procell). Cells (3.5 × 10^5/well) were seeded in 6-well plates 24 h before transfection. Per well, 2.5 μg
136 WT or variant plasmid were transfected using HieffTrans Universal Transfection Reagent (Yeasen).
137 Forty-eight hours later, total RNA was extracted with the FastPure Cell/Tissue Total RNA Isolation
138 Kit V2 (+gDNA wiper) (Vazyme). RT was carried out using the HiScript III 1st Strand cDNA
139 Synthesis Kit (Vazyme), incorporating 2 μL 5×gDNA wiper Mix, 2 μL 10×RT Mix, 2 μL HiScript III
140 Enzyme Mix, 1 μL Oligo (dT)20VN, and 1 μg total RNA. RT-PCR was performed in a 25-μL reaction
141 mixture containing 12.5 μL 2×Taq Master Mix (Vazyme), 1 μL cDNA, and 0.4 μM each primer. The
142 primer pairs were as follows: pET01 vector—forward 5′-GAGGGATCCGCTTCCTGGCCC-3′,
143 reverse 5′-CTCCCGGGCCACCTCCAGTGCC-3′; pSPL3 vector—forward 5′-
144 TCTGAGTCACCTGGACAACC-3′, reverse 5′-ATCTCAGTGGTATTTGTGAGC-3′. The PCR
145 program comprised an initial denaturation step at 94°C for 5 min, followed by 35 cycles of
146 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
147 extension step at 72°C for 7 min. RT-PCR products were separated on agarose gels, excised, and
148 purified (Omega Bio-Tek Gel Extraction Kit). Sequencing primers were the same as those used for
149 RT-PCR, and reactions were performed with the BigDye Terminator v3.1 Cycle Sequencing Kit
150 (Applied Biosystems).
151
152 **Full-length and minigene splicing analysis of *BAP1* c.783+2T>C**
153 The full-length genomic *BAP1* WT insert, spanning from the translation initiation codon to the
154 termination codon, was amplified from genomic DNA of a normal control using long-range PCR. The
155 resulting fragment was cloned into the pcDNA3.1/V5-His-TOPO vector. These procedures, as well as
156 subsequent transfection and RT–PCR, were performed essentially as previously described.[12]
157 Minigene splicing analysis followed the same procedures as for *CFTR* variants, except that two
158 WT constructs (and their corresponding variant constructs) were generated. The *BAP1* Exon 9 insert
159 comprised 337 bp from intron 8, exon 9, and 302 bp from intron 9, whereas the *BAP1* Exon 8–10
160 insert comprised 197 bp from intron 7, exon 8, intron 8, exon 9, intron 9, exon 10, and 243 bp from
161 intron 10.

NM_004656.4 and NG_031859.1 were used as the mRNA and genomic reference sequences of *BAP1*, respectively.

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.[17,18] 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,

189 NG_012232.1)[27]—were associated with milder clinical phenotypes; the remaining variant, *SPINK1*
190 c.194+2T>C (GenBank: NM_001379610.1, NG_008356.2), is discussed below.
191 The corresponding WT constructs for the 103 FLGSA-analyzed substitutions were invariably
192 preselected to yield a single or quasi-single RT–PCR band of the expected WT transcript size,
193 consistent with their respective reference mRNA sequences.[12] We also analyzed *HBB* (MIM: 141900)
194 c.315+2T>C (GenBank: NM_000518.5, NG_059281.1),[28] despite its WT construct producing two
195 strong bands—one corresponding to the WT transcript and the other representing an aberrantly spliced
196 transcript. This variant also generated WT transcript in FLGSA,[12] bringing the total number of WT
197 transcript–generating GT>GC variants identified by FLGSA to 20.
198
199 **Detailed analysis of three FLGSA-analyzed GT>GC variants**
200 Of the 20 GT>GC variants previously shown by FLGSA[12] to generate residual WT transcript, only
201 three were found to be registered in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/; assessed
202 September 5, 2025). These three variants, described below, illustrate how the presence of residual WT
203 transcript can challenge variant classification.
204
205 ***SPINK1 c.194+2T>C***
206 Among GT>GC variants, *SPINK1* c.194+2T>C is distinctive in that transcript expression has been
207 assessed both *in vivo* (using RNA from an affected homozygote)[29] and *in vitro* using FLGSA.[12] Both
208 approaches yielded concordant results, with approximately 10% of normal WT transcript levels from
209 the variant allele.[30] SpliceAI predicts a ΔDL score of 0.33, consistent with a substantial yet incomplete
210 reduction in normal splicing.
211 Our in-depth analysis of *SPINK1*-related chronic pancreatitis[8] indicated that this ~10% residual
212 WT transcript places c.194+2T>C in the grey zone, conferring a well-defined genetic effect that is too
213 weak for "pathogenic" under current ACMG/AMP criteria[5] but too strong for the proposed "risk
214 allele" category.[7] We therefore suggested classifying it as "predisposing," reflecting its partial LoF
215 nature and intermediate genetic impact.[8] ClinVar lists this variant as pathogenic/likely pathogenic
216 (VCV000132142.62; accessed October 28, 2025), but convergent evidence from functional assays, *in*

silico prediction, and clinical and genetic data supports its role as proof of concept highlighting the need for expanded classification frameworks.[8]

HBB c.315+2T>C

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

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

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

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

246 *MGP* (MIM: 154870) c.94+2T>C (GenBank: NM_000900.5, NG_023331.2) was shown by FLGSA[12]
247 to retain approximately 80% of WT transcript, whereas SpliceAI predicted a strong splice-disrupting
248 effect (ΔDL score, 0.97). Despite this discrepancy, ClinVar lists *MGP* c.94+2T>C as "likely
249 pathogenic," based on a single clinical testing submission (SCV003031993.2; accessed October 28,
250 2025) with unknown affected status and no published case reports. This classification appears to rest
251 on four implicit assumptions:

252 (1) the variant alters a canonical donor splice site;
253 (2) such variants generally disrupt splicing;
254 (3) disruption of canonical sites typically abolishes protein function; and
255 (4) LoF variants in *MGP* are known to be pathogenic.

256 As this example shows, ClinVar classifications may at times rely on inferred reasoning rather than
257 direct experimental evidence.

258

259 **Disease-associated GT>GC variants reported since 2019**

260 We next examined disease-associated GT>GC variants reported since 2019, focusing on those with
261 WT transcript generation documented in samples derived from affected individuals. Among the five
262 variants identified, *SBDS* (MIM: 607444) c.258+2T>C (GenBank: NM_016038.4, NG_007277.1)[35]
263 and *SRP68* (MIM: 604858) c.184+2T>C (GenBank: NM_014230.4, NG_137597.1),[36] produced only
264 trace (~2%) or low amounts of WT transcript, with SpliceAI ΔDL scores of 0.93 and 0.37,
265 respectively. *TNXB* (MIM: 600985) c.12469+2T>C (GenBank: NM_001365276.2, NG_008337.2)
266 retained a moderate level of WT transcript by allele-specific RT–PCR,[37] consistent with its very low
267 ΔDL score of 0.01, and has been associated with moderate Ehlers–Danlos–like manifestations in
268 congenital adrenal hyperplasia.[37] The remaining two variants—*BRCA2* (MIM: 600185) c.8331+2T>C
269 (GenBank: NM_000059.4, NG_012772.3) and *DOCK8* (MIM: 611432) c.3234+2T>C (GenBank:
270 NM_203447.4, NG_017007.1)—are discussed below.

271 ***BRCA2 c.8331+2T>C***

272 *BRCA2* c.8331+2T>C is associated with a high SpliceAI ΔDL score (0.93) and is listed in ClinVar as
273 "pathogenic/likely pathogenic" (VCV000267692.71; accessed October 28, 2025). However, allele-
274 specific RT–PCR in a heterozygous carrier demonstrated approximately 60% of WT transcript from
275 the variant allele.[38] A validated *BRCA1/2* history-weighting algorithm, which compares cancer
276 histories with expectations for pathogenic versus benign variants,[39] yielded a benign call.[38] Moreover,
277 *BRCA2* appears relatively tolerant of reduced transcript dosage, and one study suggested that variants
278 producing more than 5% WT transcript should, in the absence of other supporting evidence, be
279 considered VUS.[40] While the original report reclassified c.8331+2T>C from "pathogenic" to a
280 "VUS,"[38] the combined evidence appears to point toward a benign interpretation under current
281 ACMG/AMP principles.

282

283 ***DOCK8* c.3234+2T>C**

284 Biallelic LoF variants in *DOCK8* cause autosomal recessive hyper-IgE syndrome-2 (MIM: 243700).
285 *DOCK8* c.3234+2T>C was reported in the homozygous state in an infant with severe viral
286 infections.[41] T cells from this individual showed a predominance of WT transcript, normal DOCK8
287 protein accumulation, intact cellular function, and no DOCK8 deficiency phenotype. The clinical
288 presentation was instead explained by a homozygous *IFNAR1* (MIM: 107450) nonsense variant
289 (c.1156G>T, p.Glu386Ter; GenBank: NM_000629.3).[41] Despite this evidence for near-complete WT
290 transcript generation, lack of disease relevance, and a modest SpliceAI ΔDL score (0.49), three of the
291 four ClinVar submissions for *DOCK8* c.3234+2T>C (VCV000573664.15; accessed October 28, 2025)
292 classified it as "likely pathogenic," while one classified it as "benign."

293 *BRCA2* c.8331+2T>C and *DOCK8* c.3234+2T>C exemplify how variants predicted—or
294 assumed—to be fully disruptive can nevertheless yield intermediate or even near-complete levels of
295 WT transcript, underscoring the need for direct transcript-level evidence for accurate classification.
296 These two variants are summarized in Table 1 alongside the three FLGSA-analyzed examples
297 (*SPINK1* c.194+2T>C, *HBB* c.315+2T>C, and *MGP* c.94+2T>C).

298 **Locus-wide analysis of *CFTR* GT>GC variants**

299 Having examined GT>GC variants that retain WT transcript generation across different genes, we next
300 asked how frequent and systematic this phenomenon might be within a single locus. We selected
301 *CFTR* (GenBank: NM_000492.4, NG_016465.4), in which biallelic LoF variants cause cystic fibrosis
302 (CF; MIM: 219700),[42,43] as a model system because it harbors 26 GT donor splice sites, is among the
303 most extensively studied disease-related genes, and benefits from expert-curated databases CFTR2[21]
304 and CFTR-France[22] that provide authoritative variant annotations.

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

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

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

326 WT transcript from the c.4242+2T>C allele was confirmed by Sanger sequencing, and mature CFTR
327 protein was detected in transfected cells by immunoblotting.[23] The only other *CFTR* GT>GC variant
328 reported to yield WT transcripts is c.3873+2T>C, although this was mentioned only briefly and
329 without accompanying experimental data.[23] Notably, c.3873+2T>C is classified as "CF-causing" in
330 both CFTR2 and CFTR-France, consistent with its high SpliceAI ΔDL score (0.98) (Figure 2B).
331 We tested the three low-scoring variants together with c.3873+2T>C in minigene assays using
332 both pET01 and pSPL3 backbones (Figure 3; see also Figure S1). Findings for each variant are
333 detailed below.
334
335 ***c.2490+2T>C (intron 14)***
336 WT constructs in both minigene backbones generated multiple RT–PCR products, including the WT
337 transcript. Importantly, WT transcript from the variant allele was detected in the pET01 backbone
338 (Figure 3A), consistent with its modest SpliceAI ΔDL score (0.35). Although this variant is not
339 classified by CFTR2, it is listed in CFTR-France as a VUS, having been reported in two asymptomatic
340 homozygotes. In CFTR1, it was described in the heterozygous state in the partner of an individual
341 affected by congenital bilateral absence of the vas deferens (Figure 2B). Furthermore, ClinVar lists
342 this variant under "conflicting classifications of pathogenicity" (VCV000370078.23; accessed October
343 28, 2025), based on six clinical-testing submissions ("pathogenic" × 3, "likely pathogenic" × 2, and
344 "uncertain significance" × 1). Only one submission specified a known affected status—homozygosity
345 for c.2490+2T>C, labeled "pathogenic," in an individual with fetal cystic hygroma
346 (SCV001737053.1). Another submission, designated "likely pathogenic" (SCV001363801.5),
347 explicitly noted that "no occurrence of c.2490+2T>C in individuals affected with cystic fibrosis and
348 no experimental evidence demonstrating its impact on protein function have been reported." Taken
349 together, the clinical, genetic, computational, and experimental evidence indicate that c.2490+2T>C is
350 unlikely to represent a pathogenic *CFTR* variant in the context of classic CF.

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

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

355

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

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

365

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

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

377 In summary, locus-wide analysis of *CFTR* confirms that a subset of GT>GC variants retain WT
378 transcript. Functional testing of selected *CFTR* GT>GC variants highlights both inter-assay variability
379 and the inherent limitations of minigene systems in fully recapitulating native splicing.[24]
380

381 **Challenges of splicing assays: the case of *BAP1* c.783+2T>C**

382 As illustrated by *CFTR* variants, *in vitro* splicing assays may fail to yield clear or consistent results,
383 complicating variant classification. This challenge is further exemplified by *BAP1* c.783+2T>C.
384 Because heterozygous germline LoF variants in *BAP1* underlie tumor predisposition syndrome 1
385 (TPDS1; MIM: 614327) and *BAP1* c.783+2T>C affects the canonical intron 9 donor splice site
386 (GenBank: NM_004656.4, NG_031859.1), this variant was initially classified in ClinVar as "likely
387 pathogenic," as noted by Goldberg et al.[17] These authors reclassified it as a VUS based on the
388 following considerations:

389 (1) the variant was identified in seven families (six Jewish Iraqi and one non-Jewish) with
390 nonclassical *BAP1*-related TPDS1 tumor types;
391 (2) the variant exhibited low penetrance;
392 (3) the variant had a relatively high allele frequency in Jewish Iraqi individuals (3.6%);
393 (4) the variant had a low SpliceAI score (0.08); and
394 (5) no aberrantly spliced transcripts were detected by RT-PCR using RNA from blood cells of
395 two probands.

396 A subsequent study[18] identified a methodological limitation: the forward primer used by Goldberg
397 et al.[17] was located within exon 9, rendering it incapable of detecting exon 9 skipping. Using a primer
398 in exon 8, Sculco et al.[18] analyzed *BAP1* c.783+2T>C heterozygotes and observed an alternative
399 splicing pattern consistent with exon 9 skipping. These authors proposed reinstating a "likely
400 pathogenic" classification while acknowledging that the effect of *BAP1* c.783+2T>C on generating
401 aberrant transcripts might be incomplete.

402 To further clarify this issue, we performed additional splicing assays (Figure 4; Figure S2). We
403 first attempted FLGSA, as the *BAP1* genomic sequence from start to stop codon (<8 kb) falls within
404 the pcDNA3.1/V5-His-TOPO insert size we routinely used in this system.[12] However, the WT

405 construct failed to yield clear RT-PCR bands (Figure 4A). Notably, in our previous study only 33 of
406 119 (28%) WT full-length constructs produced a single or quasi-single band of the expected size after
407 transfection.[12] We therefore turned to minigene assays using pET01 and pSPL3. WT constructs
408 spanning exons 8–10 with flanking introns produced multiple RT-PCR bands, none corresponding to
409 the WT transcript defined by NM_004656.4, nor did the variant constructs (Figure 4B). In contrast,
410 shorter constructs containing exon 9 with flanking intronic sequences generated the expected WT band
411 from the WT allele, whereas only transcripts lacking exon 9 were produced from the variant allele
412 (Figure 4C; Figure S2).

413 ClinVar currently lists c.783+2T>C under "conflicting classifications of pathogenicity"
414 (VCV000422670.29; accessed October 28, 2025), with five submissions of "uncertain significance"
415 and one of "likely benign." In light of these inconclusive findings, allele-specific transcript analyses in
416 heterozygous individuals—if possible in the presence of a *cis*-linked coding variant outside exon 9—
417 would be most informative. Alternatively, targeted knock-in cell models could provide additional
418 insight into the splicing effect of *BAP1* c.783+2T>C.

419

420 Discussion

421 Our analysis yielded several insights into the impact of GT>GC 5′ splice-site variants. Consistent with
422 findings from FLGSA,[12] we show that disease-associated GT>GC variants can generate a wide range
423 of residual WT transcript levels, from barely detectable to nearly normal. Moreover, our findings
424 confirm our earlier postulate that "GT>GC variants in human disease genes may not invariably be
425 pathogenic,"[12] as exemplified by the *DOCK8* c.3234+2T>C variant.

426 Second, GT>GC variants that generate substantial yet incomplete WT transcript levels often fall
427 within a functional grey zone, complicating their classification (Figure 5). Such indeterminacy
428 underscores the limitations of binary frameworks and has important clinical implications. A clear
429 example is *SPINK1* c.194+2T>C, which retains approximately 10 % of normal WT transcript. In the
430 simple heterozygous state, its effect is too weak to be classified as "pathogenic" yet too strong to be
431 regarded as a mere "risk" allele. However, in homozygous, compound heterozygous, or *trans*-

432 heterozygous states—or in the presence of environmental modifiers—the same variant confers a
433 substantially greater pathogenic effect.[8] This exemplifies how variants with intermediate functional
434 effects contribute to the genetic complexity of human disease. To address this challenge, we have
435 proposed expanding the ACMG/AMP variant classification framework by incorporating additional
436 categories, including a "predisposing" category.[6,8]

437 Third, our findings underscore the limited resolution of current *in silico* tools, such as SpliceAI,[19]
438 for predicting whether GT>GC variants generate WT transcript and, if so, to what extent. Whereas
439 most WT transcript–retaining variants exhibit low SpliceAI ΔDL scores, some have high values
440 (>0.80), and transcript levels show no consistent correlation with predicted scores. This limitation
441 partly reflects the intrinsic ability of GC to function as a natural donor site in approximately 1% of U2-
442 type introns.[45] More broadly, it highlights the difficulty of predicting splicing outcomes that are not
443 all-or-none, leaving many variants within a functional grey zone that current algorithms cannot readily
444 model or quantify.

445 Fourth, our results highlight the methodological challenges inherent in experimentally
446 characterizing splice-altering variants. RNA from disease-relevant tissues provides the most reliable
447 readout, yet such material is rarely available, and surrogate sources such as blood cells or
448 lymphoblastoid lines are informative only if the gene of interest is expressed. In the absence of RNA
449 from affected individuals, cell-based assays—FLGSA,[12] EMG,[23] midigene,[46] or minigene[47] systems—
450 are often used, but each has intrinsic limitations. For example, not only is FLGSA impractical for large
451 genes, but also not all successfully constructed full-length WT constructs yield the expected WT
452 transcript. The most frequently used minigene assays are performed in a highly artificial background
453 that lacks the natural genomic context required for proper splicing regulation.[48,49] Such loss of context
454 is particularly problematic for variants with intermediate effects, for which classification depends not
455 only on detecting aberrant transcripts but also on quantifying residual WT transcript from the variant
456 allele. Even when RNA from affected individuals is available, interpretation in heterozygotes can be
457 confounded by WT transcript derived from the unaffected allele. For this reason, in both our
458 previous[12] and current studies, analyses of disease-associated GT>GC variants were focused primarily
459 on homozygotes and hemizygotes. RNA-seq, although increasingly applied in clinical contexts,[50-52]

460 faces similar challenges of tissue accessibility and allele-specific quantification. Collectively, these
461 methodological obstacles underscore the need for refined experimental systems, such as those based
462 on transactivation or transdifferentiation.[53]

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

482 Nonetheless, the boundaries of the functional grey zone (Figure 5) likely vary across genes,
483 depending on functional tolerance thresholds and modes of inheritance, such that an identical level of
484 residual WT transcript may be clinically inconsequential in one gene–disease context yet functionally
485 significant in another. Although developed here in the context of LoF variants for conceptual
486 simplicity, the grey-zone model may also extend to gain-of-function or dominant-negative variants,[59]
487 in which disease-associated alleles within a given gene are likewise unlikely to exert uniform effects.

488 Notably, this conceptual framework was established primarily for genes in which complete or near-
489 complete LoF alleles cause classical Mendelian diseases. As we have previously opined,[8] some
490 disease-associated genes may not necessarily harbor clearly "pathogenic" variants.

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

504 To address this challenge, new conceptual, computational, and experimental strategies are
505 emerging. As noted by Whitcomb,[66] our recently proposed "pathogenic–predisposing–risk–benign"
506 framework[8] offers a practical means of addressing the variant classification challenge. The use of
507 machine-learning models trained on large-scale population-based datasets,[9] together with functional
508 data from MAVE,[67,68] holds promise for delineating the full spectrum of variant effects revealed
509 through analyses of increasingly large cohorts.[69-71] Ultimately, grey-zone variants may constitute a
510 substantial fraction of those currently classified as VUS or filtered out for exceeding the allele-
511 frequency threshold set for high-penetrance disease variants and could account for part of the "missing
512 heritability" in both rare and complex diseases.[72-74] Recognizing the broad prevalence of grey-zone
513 variants and their contribution to genetic complexity will be an important step toward realizing the full
514 potential of precision medicine.[75]

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

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

535

536 Data and code availability

537 All data supporting this study are provided within the article and its supplementary materials. This
538 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.
11. Sullivan, P.J., Quinn, J.M.W., Ajuyah, P., Pinese, M., Davis, R.L., and Cowley, M.J. (2025). Data-driven insights to inform splice-altering variant assessment. Am J Hum Genet *112*, 764-778. 10.1016/j.ajhg.2025.02.012.
12. Lin, J.H., Tang, X.Y., Boulling, A., Zou, W.B., Masson, E., Fichou, Y., Raud, L., Le Tertre, M., Deng, S.J., Berlivet, I., 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. 10.1002/humu.23821.
13. 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. 10.1165/rcmb.2001-0004OC.
14. Raraigh, K.S., Han, S.T., Davis, E., Evans, T.A., Pellicore, M.J., McCague, A.F., Joynt, A.T., Lu, Z., Atalar, M., Sharma, N., et al. (2018). Functional assays are essential for interpretation of

615 missense variants associated with variable expressivity. Am J Hum Genet *102*, 1062-1077.
616 10.1016/j.ajhg.2018.04.003.
617 15. Scalet, D., Maestri, I., Branchini, A., Bernardi, F., Pinotti, M., and Balestra, D. (2019). Disease-
618 causing variants of the conserved +2T of 5' splice sites can be rescued by engineered
619 U1snRNAs. Hum Mutat *40*, 48-52. 10.1002/humu.23680.
620 16. Den Uijl, I.E., Mauser Bunschoten, E.P., Roosendaal, G., Schutgens, R.E., Biesma, D.H.,
621 Grobbee, D.E., and Fischer, K. (2011). Clinical severity of haemophilia A: does the
622 classification of the 1950s still stand? Haemophilia *17*, 849-853. 10.1111/j.1365-
623 2516.2011.02539.x.
624 17. Goldberg, Y., Laitman, Y., Ben David, M., Bazak, L., Lidzbarsky, G., Salmon, L.B., Shkedi-Rafid,
625 S., Barshack, I., Avivi, C., Darawshe, M., et al. (2021). Re-evaluating the pathogenicity of the
626 c.783+2T>C *BAP1* germline variant. Hum Mutat *42*, 592-599. 10.1002/humu.24189.
627 18. Sculco, M., La Vecchia, M., Aspesi, A., Clavenna, M.G., Salvo, M., Borgonovi, G., Pittaro, A.,
628 Witel, G., Napoli, F., Listi, A., et al. (2022). Diagnostics of *BAP1*-tumor predisposition
629 syndrome by a multitesting approach: A ten-year-long experience. Diagnostics (Basel) *12*,
630 1710. 10.3390/diagnostics12071710.
631 19. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D.,
632 Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting splicing
633 from primary sequence with deep learning. Cell *176*, 535-548. 10.1016/j.cell.2018.12.015.
634 20. Landrum, M.J., Chitipiralla, S., Kaur, K., Brown, G., Chen, C., Hart, J., Hoffman, D., Jang, W.,
635 Liu, C., Maddipatla, Z., et al. (2025). ClinVar: updates to support classifications of both
636 germline and somatic variants. Nucleic Acids Res *53*, D1313-D1321. 10.1093/nar/gkae1090.
637 21. Castellani, C., and CFTR2 team. (2013). CFTR2: How will it help care? Paediatr Respir Rev 14
638 Suppl 1, 2-5. 10.1016/j.prrv.2013.01.006.
639 22. Claustres, M., Theze, C., des Georges, M., Baux, D., Girodon, E., Bienvenu, T., Audrézet, M.P.,
640 Dugueperoux, I., Férec, C., Lalau, G., et al. (2017). CFTR-France, a national relational patient
641 database for sharing genetic and phenotypic data associated with rare *CFTR* variants. Hum
642 Mutat *38*, 1297-1315. 10.1002/humu.23276.
643 23. Joynt, A.T., Evans, T.A., Pellicore, M.J., Davis-Marcisak, E.F., Aksit, M.A., Eastman, A.C., Patel,
644 S.U., Paul, K.C., Osorio, D.L., Bowling, A.D., et al. (2020). Evaluation of both exonic and
645 intronic variants for effects on RNA splicing allows for accurate assessment of the
646 effectiveness of precision therapies. PLoS Genet *16*, e1009100.
647 10.1371/journal.pgen.1009100.
648 24. Lin, J.H., Wu, H., Zou, W.B., Masson, E., Fichou, Y., Le Gac, G., Cooper, D.N., Férec, C., Liao, Z.,
649 and Chen, J.M. (2021). Splicing outcomes of 5' splice site GT>GC variants that generate wild-
650 type transcripts differ significantly between full-length and minigene splicing assays. Front
651 Genet *12*, 701652. 10.3389/fgene.2021.701652.
652 25. Muller, J.S., Piko, H., Schoser, B.G., Schlotter-Weigel, B., Reilich, P., Gurster, S., Born, C.,
653 Karcagi, V., Pongratz, D., Lochmuller, H., and Walter, M.C. (2006). Novel splice site mutation
654 in the caveolin-3 gene leading to autosomal recessive limb girdle muscular dystrophy.
655 Neuromuscul Disord *16*, 432-436. 10.1016/j.nmd.2006.04.006.
656 26. Seyama, K., Nonoyama, S., Gangsaas, I., Hollenbaugh, D., Pabst, H.F., Aruffo, A., and Ochs,
657 H.D. (1998). Mutations of the CD40 ligand gene and its effect on CD40 ligand expression in
658 patients with X-linked hyper IgM syndrome. Blood *92*, 2421-2434.
659 27. Bartolo, C., Papp, A.C., Snyder, P.J., Sedra, M.S., Burghes, A.H., Hall, C.D., Mendell, J.R., and
660 Prior, T.W. (1996). A novel splice site mutation in a Becker muscular dystrophy patient. J Med
661 Genet *33*, 324-327. 10.1136/jmg.33.4.324.
662 28. Frischknecht, H., Dutly, F., Walker, L., Nakamura-Garrett, L.M., Eng, B., and Waye, J.S. (2009).
663 Three new beta-thalassemia mutations with varying degrees of severity. Hemoglobin *33*,
664 220-225. 10.1080/03630260903089060.

29. 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. 10.1136/gut.2006.095752.
30. Chen, J.M., Lin, J.H., Masson, E., Liao, Z., Férec, C., Cooper, D.N., and Hayden, M. (2020). The experimentally obtained functional impact assessments of 5' splice site GT>GC variants differ markedly from those predicted. Curr Genomics *21*, 56-66. 10.2174/1389202921666200210141701.
31. Jayasree, D., Shaji, R.V., George, B., Mathews, V., Srivastava, A., and Edison, E.S. (2016). Clinical, hematological and molecular analysis of homozygous Hb E (*HBB*: c.79G > A) in the Indian population. Hemoglobin *40*, 16-19. 10.3109/03630269.2015.1086880.
32. 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. 10.1016/0092-8674(86)90620-3.
33. 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. 10.1016/0092-8674(87)90219-4.
34. Tesio, N., and Bauer, D.E. (2023). Molecular basis and genetic modifiers of thalassemia. Hematol Oncol Clin North Am *37*, 273-299. 10.1016/j.hoc.2022.12.001.
35. Peretto, L., Tonetto, E., Maestri, I., Bezzerri, V., Valli, R., Cipolli, M., Pinotti, M., and Balestra, D. (2023). Counteracting the common Shwachman-Diamond Syndrome-causing *SBDS* c.258+2T>C mutation by RNA therapeutics and base/prime editing. Int J Mol Sci *24*, 4024. 10.3390/ijms24044024.
36. Schmaltz-Panneau, B., Pagnier, A., Clauin, S., Buratti, J., Marty, C., Fenneteau, O., Dieterich, K., Beaupain, B., Donadieu, J., Plo, I., and Bellanne-Chantelot, C. (2021). Identification of biallelic germline variants of *SRP68* in a sporadic case with severe congenital neutropenia. Haematologica *106*, 1216-1219. 10.3324/haematol.2020.247825.
37. Lao, Q., Mallappa, A., Rueda Faucz, F., Joyal, E., Veeraraghavan, P., Chen, W., and Merke, D.P. (2021). A *TNXB* splice donor site variant as a cause of hypermobility type Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. Mol Genet Genomic Med *9*, e1556. 10.1002/mgg3.1556.
38. Nix, P., Mundt, E., Coffee, B., Goossen, E., Warf, B.M., Brown, K., Bowles, K., and Roa, B. (2022). Interpretation of *BRCA2* splicing variants: A case series of challenging variant interpretations and the importance of functional RNA analysis. Fam Cancer *21*, 7-19. 10.1007/s10689-020-00224-y.
39. Pruss, D., Morris, B., Hughes, E., Eggington, J.M., Esterling, L., Robinson, B.S., van Kan, A., Fernandes, P.H., Roa, B.B., Gutin, A., et al. (2014). Development and validation of a new algorithm for the reclassification of genetic variants identified in the *BRCA1* and *BRCA2* genes. Breast Cancer Res Treat *147*, 119-132. 10.1007/s10549-014-3065-9.
40. Tubeuf, H., Caputo, S.M., Sullivan, T., Rondeaux, J., Krieger, S., Caux-Moncoutier, V., Hauchard, J., Castelain, G., Fievet, A., Meulemans, L., 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. 10.1158/0008-5472.CAN-20-0895.
41. Huynh, A., Gray, P.E., Sullivan, A., Mackie, J., Guerin, A., Rao, G., Pathmanandavel, K., Mina, E.D., Hollway, G., Hobbs, M., et al. (2024). A Novel case of *IFNAR1* deficiency identified a common canonical splice site variant in *DOCK8* in Western Polynesia: The importance of validating variants of unknown significance in under-represented ancestries. J Clin Immunol *44*, 170. 10.1007/s10875-024-01774-x.
42. Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., and et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science *245*, 1066-1073. 10.1126/science.2475911.

716 43. Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A.,
717 Buchwald, M., and Tsui, L.C. (1989). Identification of the cystic fibrosis gene: genetic analysis.
718 Science *245*, 1073-1080. 10.1126/science.2570460.
719 44. Lundman, E., Gaup, H.J., Bakkeheim, E., Olafsdottir, E.J., Rootwelt, T., Storrosten, O.T., and
720 Pettersen, R.D. (2016). Implementation of newborn screening for cystic fibrosis in Norway.
721 Results from the first three years. J Cyst Fibros *15*, 318-324. 10.1016/j.jcf.2015.12.017.
722 45. Burset, M., Seledtsov, I.A., and Solovyev, V.V. (2000). Analysis of canonical and non-canonical
723 splice sites in mammalian genomes. Nucleic Acids Res *28*, 4364-4375.
724 10.1093/nar/28.21.4364.
725 46. Sangermano, R., Khan, M., Cornelis, S.S., Richelle, V., Albert, S., Garanto, A., Elmelik, D.,
726 Qamar, R., Lugtenberg, D., van den Born, L.I., et al. (2018). *ABCA4* midigenes reveal the full
727 splice spectrum of all reported noncanonical splice site variants in Stargardt disease. Genome
728 Res *28*, 100-110. 10.1101/gr.226621.117.
729 47. Buisine, M.P., Bellanne-Chantelot, C., Calmels, N., Vache, C., Besnard, T., Cogne, B., Vitobello,
730 A., Piton, A., Martins, A., Gaildrat, P., et al. (2025). RNA-based diagnostic studies in genetics:
731 Review and guidance from a multidisciplinary French network. Eur J Hum Genet *33*, 1219-
732 1227. 10.1038/s41431-025-01881-2.
733 48. Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-
734 binding proteins. Nat Rev Genet *15*, 689-701. 10.1038/nrg3778.
735 49. Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing kinetics and coordination
736 revealed by direct nascent RNA sequencing through nanopores. Mol Cell *77*, 985-998.
737 10.1016/j.molcel.2019.11.017.
738 50. Oh, R.Y., AlMail, A., Cheerie, D., Guirguis, G., Hou, H., Yuki, K.E., Haque, B.,
739 Thiruvahindrapuram, B., Marshall, C.R., Mendoza-Londono, R., et al. (2024). A systematic
740 assessment of the impact of rare canonical splice site variants on splicing using functional
741 and in silico methods. HGG Adv *5*, 100299. 10.1016/j.xhgg.2024.100299.
742 51. Zhao, S., Macakova, K., Sinson, J.C., Dai, H., Rosenfeld, J., Zapata, G.E., Li, S., Ward, P.A.,
743 Wang, C., Qu, C., et al. (2025). Clinical validation of RNA sequencing for Mendelian disorder
744 diagnostics. Am J Hum Genet *112*, 779-792. 10.1016/j.ajhg.2025.02.006.
745 52. Sha, Y., Ortiz, J.B., Bristow, S.L., Loranger, K., Meng, L., Zhao, X., Xia, F., Parmar, S., ElNaggar,
746 A.C., and Xu, W. (2025). Refining the interpretation of variants of uncertain significance in
747 hereditary cancer screening through integrated RNA sequencing. Genet Med Open *3*,
748 101914. 10.1016/j.gimo.2024.101914.
749 53. Nicolas-Martinez, E.C., Robinson, O., Pflueger, C., Gardner, A., Corbett, M.A., Ritchie, T.,
750 Kroes, T., van Eyk, C.L., Scheffer, I.E., Hildebrand, M.S., et al. (2024). RNA variant assessment
751 using transactivation and transdifferentiation. Am J Hum Genet *111*, 1673-1699.
752 10.1016/j.ajhg.2024.06.018.
753 54. Sullivan, P.J., Quinn, J.M.W., Wu, W., Pinese, M., and Cowley, M.J. (2024). SpliceVarDB: A
754 comprehensive database of experimentally validated human splicing variants. Am J Hum
755 Genet *111*, 2164-2175. 10.1016/j.ajhg.2024.08.002.
756 55. Wu, H., Lin, J.H., Tang, X.Y., Marenne, G., Zou, W.B., Schutz, S., Masson, E., Génin, E., Fichou,
757 Y., Le Gac, G., et al. (2024). Combining full-length gene assay and SpliceAI to interpret the
758 splicing impact of all possible *SPINK1* coding variants. Hum Genomics *18*, 21.
759 10.1186/s40246-024-00586-9.
760 56. Costantino, L., Rusconi, D., Solda, G., Seia, M., Paracchini, V., Porcaro, L., Asselta, R.,
761 Colombo, C., and Duga, S. (2013). Fine characterization of the recurrent c.1584+18672A>G
762 deep-intronic mutation in the cystic fibrosis transmembrane conductance regulator gene.
763 Am J Respir Cell Mol Biol *48*, 619-625. 10.1165/rcmb.2012-0371OC.
764 57. Elasal, M.A., Khateb, S., Panneman, D.M., Roosing, S., Cremers, F.P.M., Banin, E., Sharon, D.,
765 and Sarma, A.S. (2024). A leaky deep intronic splice variant in *CLRN1* is associated with non-
766 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.

818 72. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy,
819 M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability
820 of complex diseases. Nature *461*, 747-753. 10.1038/nature08494.
821 73. Maroilley, T., and Tarailo-Graovac, M. (2019). Uncovering missing heritability in rare diseases.
822 Genes (Basel) *10*, 275. 10.3390/genes10040275.
823 74. Génin, E. (2020). Missing heritability of complex diseases: case solved? Hum Genet *139*, 103-
824 113. 10.1007/s00439-019-02034-4.
825 75. Denny, J.C., and Collins, F.S. (2021). Precision medicine in 2030-seven ways to transform
826 healthcare. Cell *184*, 1415-1419. 10.1016/j.cell.2021.01.015.
827 76. Torene, R.I., Guillen Sacoto, M.J., Millan, F., Zhang, Z., McGee, S., Oetjens, M., Heise, E.,
828 Chong, K., Sidlow, R., O'Grady, L., et al. (2024). Systematic analysis of variants escaping
829 nonsense-mediated decay uncovers candidate Mendelian diseases. Am J Hum Genet *111*, 70-
830 81. 10.1016/j.ajhg.2023.11.007.

831

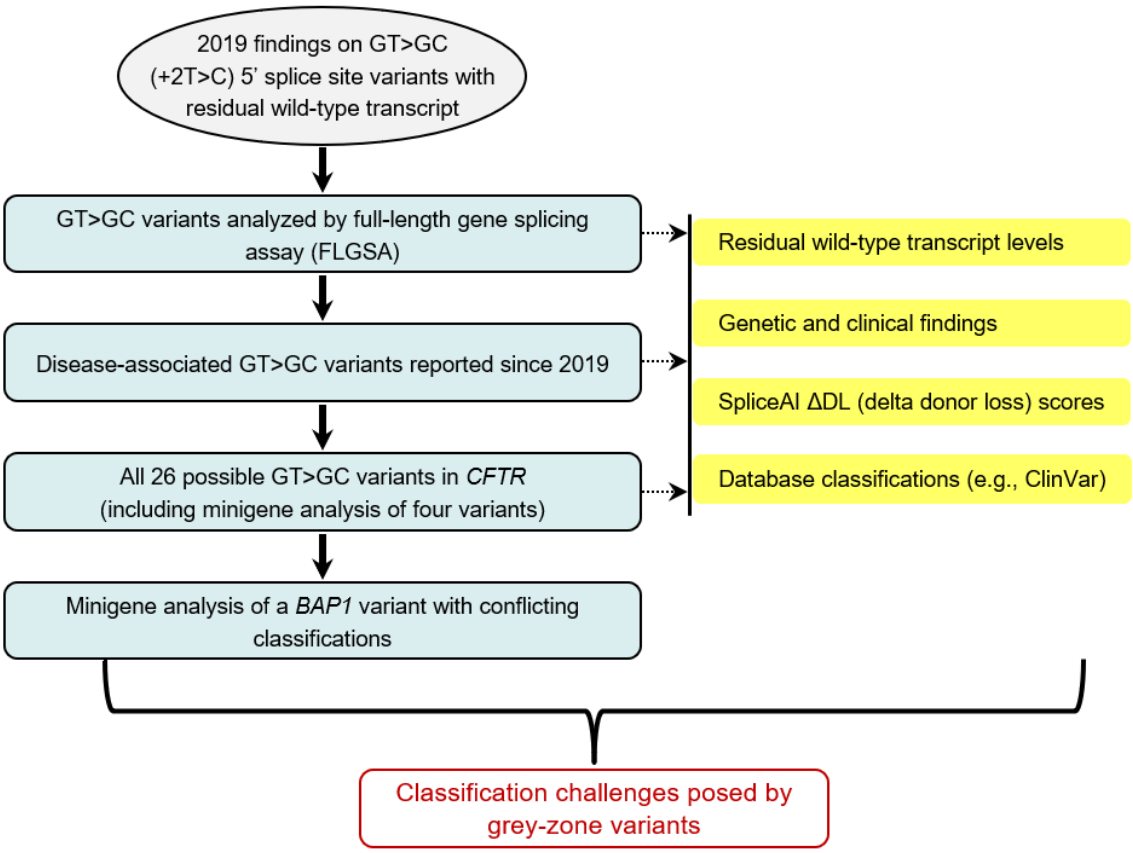
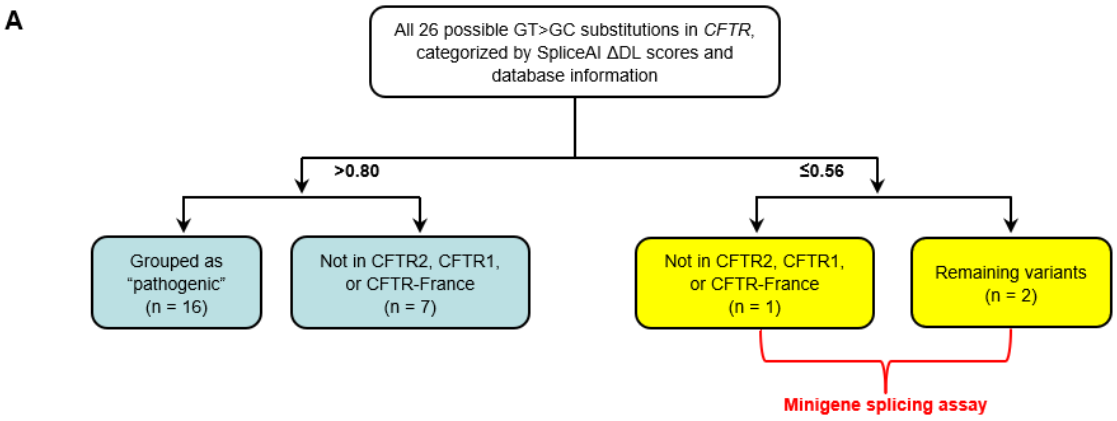


Figure 1. Scheme of the analytical approach

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.

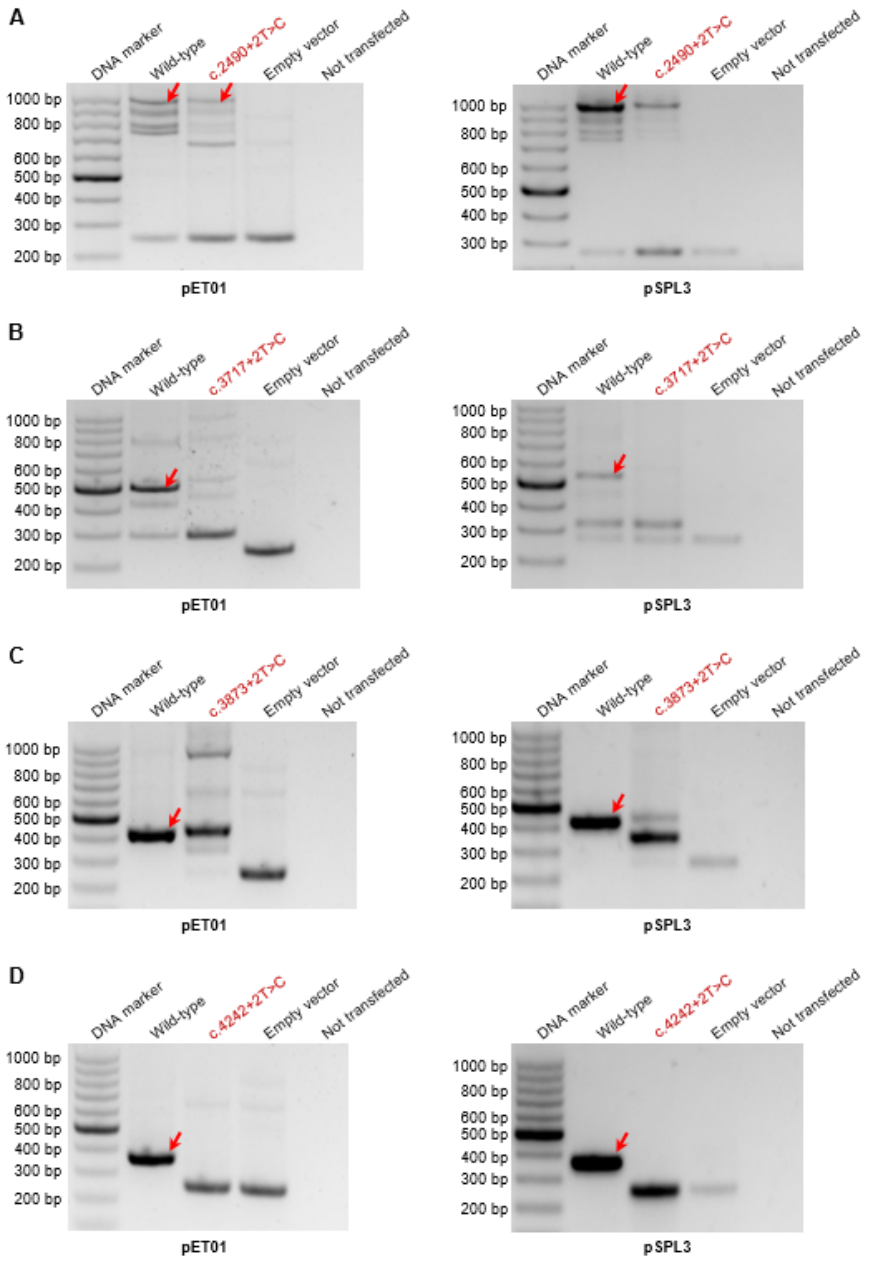


Figure 3. Minigene splicing analysis of four selected *CFTR* GT>GC substitutions in pET01 and pSPL3 backbones

Each variant was analyzed with its preceding exon and short flanking intronic regions 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.

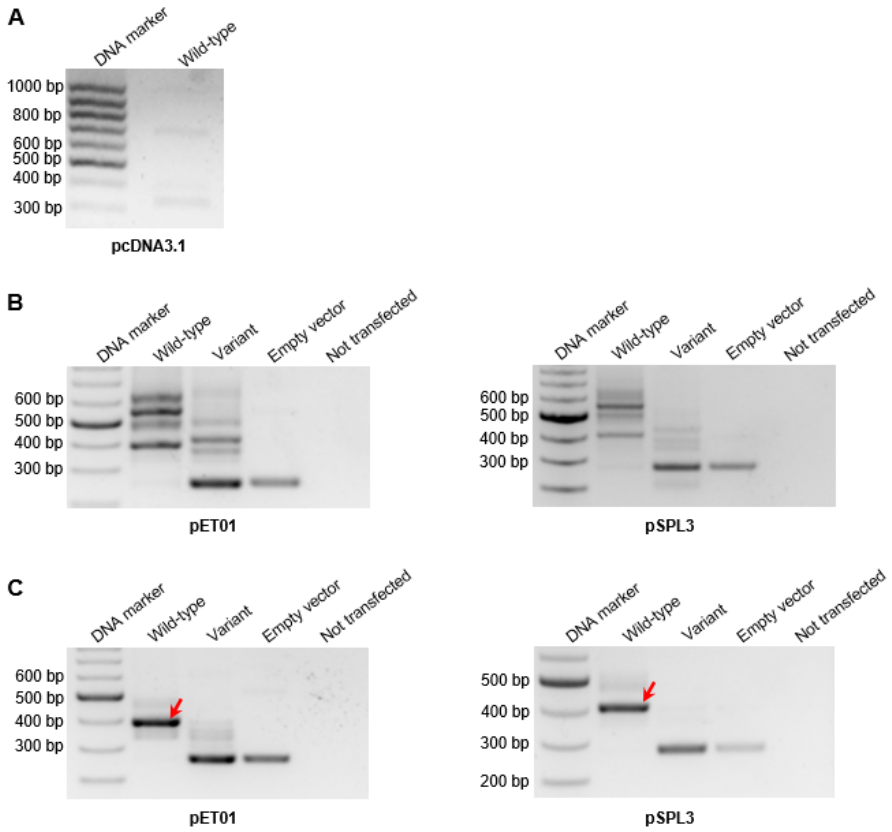


Figure 4. Splicing analysis of the *BAP1* c.783+2T>C variant in different assay contexts

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

(B) Minigene analysis of exons 8–10 with flanking intronic sequences in pET01 and pSPL3 backbones. No normally spliced transcripts were detected, even from the WT constructs. For band annotations, see Figure S2.

(C) Minigene analysis of exon 9 with flanking intronic sequences in pET01 and pSPL3 backbones. Arrows indicate bands corresponding to the normally spliced transcript. Additional bands are annotated in Figure S2.

GenBank *BAP1* mRNA and genomic reference sequences: NM_004656.4 and NG_031859.1.

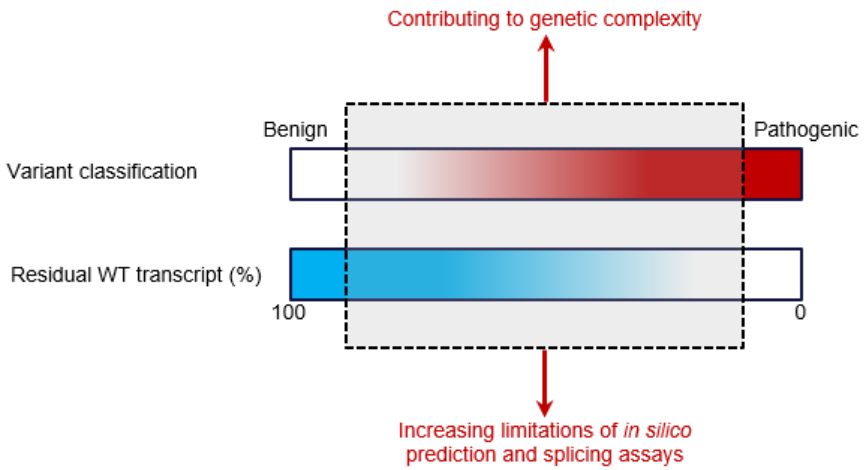


Figure 5. Conceptual model of the functional grey zone in variant classification

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.

888 **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

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

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

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