Splicing predictions, minigene analyses, and ACMG-AMP clinical classification of 42 germline PALB2 splice-site variants

Alberto Valenzuela-Palomo1, Elena Bueno-Martínez1, Lara Sanoguera-Miralles1, Víctor Lorca2, Eugenia Fraile-Bethencourt1,3, Ada Esteban-Sánchez4, Susana Gómez-Barrero3, Sara Carvalho5, Jamie Allen5, Alicia García-Álvarez1, Pedro Pérez-Segura5, Leila Dorling5, Douglas F Easton5, Peter Devilee6, Maaike PG Vreeswijk6, Miguel de la Hoya7,8 and Eladio A Velasco1††

1 Splicing and Genetic Susceptibility to Cancer, Unidad de Excelencia Instituto de Biología y Genética Molecular, Consejo Superior de Investigaciones Científicas (CSIC-UvA), Valladolid, Spain
2 Molecular Oncology Laboratory, Hospital Clínico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain
3 Knight Cancer Research Building, Portland, OR, USA
4 VISAVET Health Surveillance Centre, Complutense University, Madrid, Spain
5 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
6 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

† Correspondence to: EA Velasco, Grupo de Splicing y Cáncer, Instituto de Biología y Genética Molecular (IBGM), Consejo Superior de Investigaciones Científicas (CSIC)-UvA, Sanz y Forés 3, 47003 Valladolid, Spain. E-mail: eavelsam@ibgm.uva.es

†† Joint senior authors.

Abstract

PALB2 loss-of-function variants confer high risk of developing breast cancer. Here we present a systematic functional analysis of PALB2 splice-site variants detected in approximately 113,000 women in the large-scale sequencing project Breast Cancer After Diagnostic Gene Sequencing (BRIDGES; https://bridges-research.eu/). Eighty-two PALB2 variants at the intron-exon boundaries were analyzed with MaxEntScan. Forty-two variants were selected for the subsequent splicing functional assays. For this purpose, three splicing reporter minigenes comprising exons 1–12 were constructed. The 42 potential spliceogenic variants were introduced into the minigenes by site-directed mutagenesis and assayed in MCF-7/MDA-MB-231 cells. Splicing anomalies were observed in 35 variants, 23 of which showed no traces or minimal amounts of the expected full-length transcripts of each minigene. More than 30 different variant-induced transcripts were characterized, 23 of which were predicted to truncate the PALB2 protein. The pathogenicity of all variants was interpreted according to an in-house adaptation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) variant classification scheme. Up to 23 variants were classified as pathogenic/likely pathogenic. Remarkably, three <1.2 variants (c.49–2A>T, c.108+2T>C, and c.211+1G>A) were classified as variants of unknown significance, as they produced significant amounts of either in-frame transcripts of unknown impact on the PALB2 protein function or the minigene full-length transcripts. In conclusion, we have significantly contributed to the ongoing effort of identifying spliceogenic variants in the clinically relevant PALB2 cancer susceptibility gene. Moreover, we suggest some approaches to classify the findings in accordance with the ACMG-AMP rational.

Keywords: breast cancer; susceptibility genes; PALB2; splicing; aberrant splicing; VUS; functional assay; minigene; clinical interpretation

Introduction

PALB2 [MIM #610355] was originally identified as a BRCA2-interacting protein that plays a key role in DNA repair by homologous recombination [1,2]. Monallelic germline loss-of-function variants in PALB2 confer high risk of developing breast cancer (BC) [3–5]. Moreover, germline biallelic loss-of-function variants of PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood malignancies [6]. Pathogenic PALB2 variants have been identified in 0.6–3.9% of families with a history of BC, while the absolute BC risk by 70 years of age for PALB2 carriers was estimated in the 33–58% range, and more recently in the 44–68% range, depending on the familial aggregation of BC [4,5]. Two large population-based studies have estimated the BC relative risk in carriers of PALB2 protein truncating variants to be 3.83 (2.68–5.63) and 5.02 (3.73–6.76), respectively [7,8]. Moreover, the
association of PALB2 truncating variants with estrogen receptor negative and triple negative BC is particularly strong (odds ratios = 7.35 and 10.36, respectively) [8]. PALB2 loss-of-function variants are also associated with pancreatic cancer, ovarian cancer (OC), and male BC [5,9–12]. The association of PALB2 with gastric and colorectal cancer is currently disputed [5,13,14].

There are several critical steps for gene expression, which are frequent targets of loss-of-function variants [15,16]. Splicing is a pivotal RNA processing stage whereby introns from eukaryotic genes are removed and exons are sequentially joined. Most human protein-coding genes (94%) contain introns and produce multiple mRNA isoforms via alternative splicing. This process is finely controlled by small nuclear ribonucleoproteins, protein factors, and a broad array of cis-acting elements that are responsible for the splicing efficiency [16]. There are fundamental preserved sequences that are recognized by the splicing machinery: the 5’ (donor/AG) and the 3’ (acceptor/AG) splice sites (5’ss and 3’ss, respectively), the polypyrimidine tract, and the branch point. These sequences are possible targets for spliceogenic variants that can produce anomalous transcripts and modify disease risk [17]. Actually, splicing disruption is a frequent deleterious mechanism in genetic diseases [18]. The identification and understanding of the underlying causative mechanisms are crucial in clinical practice [19] with the aim of improving genetic diagnosis, prevention strategies, and therapy.

We had previously shown that spliceogenic variants make a significant contribution to the overall pool of germ-line loss of function alleles in the BC/OC genes RAD51C [20], RAD51D [21], and BRCA2 [22–24]. The present work was carried out in the context of the Breast Cancer After Diagnostic Gene Sequencing (BRIDGES) project (https://bridges-research.eu/), where 34 known or suspected BC/OC genes were sequenced in 60,466 BC cases and 53,461 controls [8]. Our purpose was to experimentally characterize the PALB2 spliceogenic variants detected in BRIDGES subjects. Here, we used three ad hoc developed splicing reporter minigenes to test 42 PALB2 unique variants with a high prior probability of being spliceogenic, assigning a final American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) clinical classification to all of them.

Materials and methods

Ethics approval

Ethical approval for this study was obtained from the Ethics Committee of the Spanish National Research Council-CSIC (28/05/2018).

Variant and transcript annotations

We identified in BRIDGES carriers (cases and/or controls) a total of 82 unique variants at PALB2 intron/exon boundaries [8]: 3’ss: intron/exon [IVS-10_ IVS-1/2nt]; 5’ss: exon/intron [2nt/IVS+1_ IVS+10]. Variants, transcripts, and predicted protein products were described according to the Human Genome Variation Society guidelines (https://varnomen.hgvs.org), using the Ensembl reference transcript ID ENSG00000083093 (Genbank NM_024675.4). We also annotated transcripts according to a former shortened description [25].

Bioinformatics selection of PALB2 variants

We selected 42 of 82 variants at the intron-exon boundaries for minigenes analysis based on: (1) splice-site disruption at the ±1,2 (AG/AT) positions; (2) important MaxEntScan (MES) score changes (≥15%) [26–28]; (3) creation of de novo alternative splice sites (cut-off ≥3.0); (4) regardless of MES predictions, variants at other conserved positions of the acceptor (Y[11]NCAG/G) and donor (MAG[TRAGT]) consensus sequences, such as pyrimidine to purine changes or deletions at the polypyrimidine tract, substitutions of a conserved nucleotide at the intronic positions – 3C, +3R, +4A, +5G, +6T, as well as the first (G) and the last three nucleotides of the exon (M, A, G). Patient RNA was not available for any of the 42 selected variants.

Minigenome construction and mutagenesis

The PALB2 gene is composed of 13 exons, 12 of which were cloned into the pSAD v9.0 vector (patent P201231427-CSIC) [22,23]. Exon 13 was excluded given that BRIDGES did not report any potential splice-site variant in it. Exons 1–12 were cloned into three different minigenes: mgPALB2_ex1-3, mgPALB2_ex4-6, and mgPALB2_ex5-12 (Figure 1, supplementary material, Figure S1). The construction of these minigenes is detailed in Supplementary materials and methods. Basically, the inserts were generated either by PCR with Phusion High Fidelity polymerase (Fisher Scientific, Waltham, MA, USA) and the primers shown in supplementary material, Table S1, or by gene synthesis (Genewiz, South Plainfield, NJ, USA). Then, fragments were cloned into the pSAD splicing vector by two methods, either by classical restriction enzyme digestion and ligation or by overlapping extension PCR [29]. All the insert sequences with their structures are shown in supplementary material. Figure S1 and the final minigenes are outlined in Figure 1. The wild type (wt) minigenes were used as templates to generate 42 DNA variants (supplementary material, Table S1) with the QuikChange Lightning kit (Agilent, Santa Clara, CA, USA). All constructs were confirmed by sequencing (Macrogen, Madrid, Spain).

Splicing functional assays

Approximately 2 × 10^5 MCF-7 (human breast adenocarcinoma cell line) cells were transfected with 1 μg minigene using 2 μl lipofectamine LTX (Life Technologies, Carlsbad, CA, USA), as previously described [20,21]. To check the reproducibility of the splicing outcomes, MDA-MB-231 (triple-negative BC cell line) cells were transfected with the wt and mutant minigenes
carrying the following variants: c.47A>G, c.48+1G>A, c.48+2T>C, c.48+4C>T, c.2749-1G>T, c.2834+3A>G, c.3201+3_3201+4insTG, and c.3350+4A>G. RNA was purified by means of the Genematrix Universal RNA Purification Kit (EURx, Gdansk, Poland), with on-column DNAse I (EURx) digestion. Reverse transcription was carried out using 400 ng RNA and the RevertAid First Strand cDNA Synthesis Kit (Life Technologies), using the vector-specific primer RTPSPL3-RT-FW (5'-TGAGGAGTGAATTGGTCGAA-3'). Then, 40 ng cDNA was amplified with platinum-Taq DNA polymerase (Life Technologies) and the primers SD6-PSPL3_RT-FW (5'-TCCTGGACAACCTCAAAG-3') and RTpSAD-RV (patent P201231427). Samples were denatured at 94 °C for 2 min, followed by 35 cycles of 94 °C/30 s, 60 °C/30 s, and 72 °C (1 min/kb), and a final extension step at 72 °C for 5 min. RT-PCR products were sequenced by Macrogen. The expected sizes of the minigene full-length (mgFL) transcripts were 366-nt (mgPALB2_ex1-3), 2,556-nt (mgPALB2_ex4-6), and 1,847-nt (mgPALB2_ex5-12).

Figure 1. Schematic representation of the wt PALB2 minigenes. Exons are indicated by boxes; green elbow arrows indicate the expected splicing reactions in eukaryotic cells and black arrows locate specific vector RT-PCR primers. (A) Minigene with PALB2 exons 1–3 (mgPALB2_ex1-3). (B) Minigene with PALB2 exons 4–6 (mgPALB2_ex4-6). (C) Minigene with PALB2 exons 5–12 (mgPALB2_ex5-12).

ACMG-AMP clinical classification of PALB2 genetic variants

We classified all variants according to a recently proposed ACMG-AMP point system Bayesian framework (Supplementary materials and methods, and supplementary material, Table S2, Figures S2 and S3A–C) [31–33]; mgPALB2 read-outs have been incorporated into the classification system as PS3/BS3 codes with variable strength depending on the actual outcome. As most tested variants produce two or more different transcripts, we proceeded as follows: (1) we assigned a specific PS3/BS3 code strength to each individual transcript.
and (2) depending on the relative contribution of pathogen-
ic and benign codes to the overall expression, we
assigned an overall PS3/BS3 code strength to each vari-
ant. To assess its contribution to the final classification,
we classified with and without incorporating mgPALB2
data into the ACMG-AMP scheme. To assist in the clas-
ification process, we developed a PALB2 adaptation of
the PVS1 decision tree proposed by the ClinGen sequence
variant interpretation (SVI) working group [34]. We
considered that some pathogenic (PS2, PM1, PM6, PP2,
PP4, PP5) and benign (BS2, BP1, BP3, BP5, BP6) codes
are not applicable to the classification of PALB2 variants.
Splicing predictive codes PVS1 (variable strength) and
PP3/BP4 were assigned according to SpliceAI predictions
(https://spliceailookup.broadinstitute.org/) [35]. We
selected SpliceAI for ACMG-AMP classification because:
(1) we have previously used MES to select many variants under investigation and (2) SpliceAI
scores probabilities associated with specific splicing
outcomes, a feature not provided by MES, but critical
to assign specific PVS1 strengths.

Results

Bioinformatics selection of PALB2 variants

Eighty-two PALB2 variants were analyzed with MES
(supplementary material, Table S3). Forty-four of them
were predicted to have a potential damaging effect on
splicing. Variants c.2748+1G>A and c.2834+2T>C
were excluded from the functional study because their
effects on splicing were expected to be similar to
selected variants c.2748+1G>T and c.2834+1G>A,
respectively. A total of 11 and 30 variants impaired the
3’ss and the 5’ss, respectively, while one variant
was predicted to generate a de novo 5’ss
(c.48+7G>C). Seven splice-site disrupting variants
(c.49-2A>T, c.109-2A>G, c.2587-2A>G, c.2749-1G>T,
c.2997-1G>A, c.3348C>T, c.3350+4A>G) were also
estimated to create a de novo splice site. Variants
c.2748+4A>T, c.2834+6T>C, c.3113+3A>G, and
c.3201+6T>A were chosen despite their weak MES
score changes (<15%) because they affect highly conserved
positions at the 5’ss (consensus sequence EXON/
intron: MAG/gtratg).

Minigene assays

The three wt minigene constructs (mgPALB2_ex1-3,
mgPALB2_ex4-6, and mgPALB2_ex5-12) were vali-
dated in MCF-7 cells, where they produced the expected
mgFL transcripts: [V1-PALB2_ex1-3-V2, 3566-nt],
[V1-PALB2_ex4-6-V2, 2556-nt], and [V1-PALB2-
ex5_12-V2, 1847-nt], respectively.

The 42 candidate variants were introduced into the cor-
responding minigene by site-directed mutagenesis and
functionally assayed in MCF-7 cells. Thirty-five variants
(83.3%) impaired splicing, 23 of which showed a com-
plete lack or minimal amounts (c.48G>A, 0.9%;
c.3113+5G>C, 4.9%) of the mgFL transcripts (Table 1,
Figures 2, 3, supplementary material, Figure S4). Overall,
the 35 spliceogenic variants each produced one to five
anomalous transcripts. Eight variants (see Materials and
methods) were also examined in MDA-MB-231 cells,

Table 1. Splicing outcomes of PALB2 variants.

| Variant (HGVS) | Bioinformatics summary | mgFL transcripts | PTC transcripts | In-frame transcripts |
|---------------|------------------------|------------------|----------------|--------------------|
| wt mgPALB2_ex1-3 | - | 100% | | |
| c.47A>G | [-]5’ss (5.74 → −0.01) | 0.9 ± 0.1% | △(E1q17) [88.6 ± 0.1%] | △(E1q9) [5.4 ± 0.1%] |
| c.48G>A | [-]5’ss (5.74 → −3.48) | - | △(E1q17) [100%] | △(E1q9) [5.4 ± 0.1%] |
| c.47+1G>A | [-]5’ss (5.74 → −2.43) | - | △(E1q17) [100%] | △(E1q9) [5.4 ± 0.1%] |
| c.47+2T>C | [-]5’ss (5.74 → 2.88) | 94.5 ± 0.5% | △(E1q17) [5.5 ± 0.5%] | |
| c.48+16G | 5’ss (5.74 → −5.74) | 100% | | |
| c.49-2A>T | [+]5’ss (6.49) 9-nt downstream | 85.5 ± 0.3% | △(E2) [14.5 ± 0.3%] | |
| c.108+1G>A | [-]5’ss (8.8-6) 6-nt downstream | 100% | | |
| c.108+2T>C | [-]5’ss (8.8-6) 6-nt downstream | 100% | | |
| c.109-6, 109-4del | [+]5’ss (10.06 → 2.11) | - | △(E3) [41.0 ± 0.4%] | |
| c.109-2A>G | [-]5’ss (3.38) 11-nt downstream | - | △(E3) [45.7 ± 0.4%] | △(E3) [54.3 ± 0.4%] |
| c.211+1G>A | - | - | △(E3) [54.3 ± 0.4%] | |
| c.211+1G>A | - | - | △(E3) [54.3 ± 0.4%] | |
| wt mgPALB2_ex4-6 | - | 100% | | |
| c.1684+4A>G | [-]5’ss (8.88 → −7.24) | 41.7 ± 1.2% | △(E4) [58.3 ± 1.2%] | |
| c.1685+2A>C | [-]3’ss (11.15 → −3.11) | - | △(E5p13) [67.3 ± 3.6%] | △(E5p5) [15.4 ± 1.5%] |

(Continues)
where they replicated the splicing outcomes detected in MCF-7 cells.

Eighteen variants affected the classical ±1.2 positions of the 5’ or 3’ss. All these variants, except for c.108+2T>C, induced complete aberrant patterns. Variant c.108+2T>C, as well as c.48+2T>C, transforms a GT splice-donor into an atypical GC donor [36]. Remarkably, although c.48+2T>C only displayed anomalous transcripts, c.108+2T>C produced 85.5% of the mgFL transcript. In addition, other 17 spliceogenic variants affected other positions of the 5’ss, including the three last exonic nucleotides and intronic nucleotides +3, +4, +5, and +6 (Table 1).

### Transcript analysis

Fluorescent fragment analysis revealed the existence of 39 different transcripts, including six mgFL transcripts.
Five of them carrying rare variants (Table 1; supplementary material, Table S4 and Figure S4). Alternative site usage was the most frequent spliceogenic mechanism, explaining up to 22 different aberrant transcripts. Two of them, Δ(E10q41) and Δ(E10q65), derived from the activation of atypical GC donors. The second most prevalent mechanism was exon skipping, which was detected in 10 different transcripts. Finally, one

Figure 2. Splicing functional assays of variants in minigenes. (A) mgPALB2_ex1-3 and (B) mgPALB2_ex4-6. The maps of variants are shown on the left. At the bottom, fluorescent fragment analysis of transcripts generated by the wt and mutant minigenes. FAM-labeled products (blue peaks) were run with LIZ-1200 (orange peaks) as size standard. FL, minigene full-length transcript. Transcript ▼(E1q416) of the variants c.48G>A and c.48+1G>A is not shown because it is out of the size range displayed in minigene mgPALB2_ex1-3 electropherograms.
transcript contained a full intron 5 retention (▼(I5)). The identity of three transcripts (592, 710, and 888-nt long) could not be characterized. Up to 16 of the abovementioned isoforms had been previously characterized as naturally occurring events [25], including: ▼(E1q9), Δ(E1q17), Δ(E2p6), Δ(E2),...
Table 2. ACMG-AMP clinical classification of 42 PALB2 genetic variants detected in the BRIDGES cohort.

| c.HGVS | p.HGVS | ClinVar† | ACMG-AMP‡ classification | PS3/Bs3 | (mgPALB2) | PS4‡ | PM2** | PM3†‡ | PM4†‡ | PP3/BP4§ (splicing) | PP3/BP4 (protein) | BP7†† |
|--------|--------|----------|--------------------------|---------|-----------|-------|-------|-------|-------|-------------------|------------------|-------|
| c.47A>G | p.(Lys16Arg) | VUS | LP | PS3 | V5 | PM2 | P | PP3 |
| c.48G>A | p.(Lys16Arg) | VUS/LP/P | LP | PS3 | V5 | PM2 | P (PM4) | PP3 |
| c.48+1G>A | p. | LP | LP | PSV1 | M | PS3 | V5 | PM2 | P (PM4) |
| c.48+2T>C | p. | (--) | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.48+4C>T | p. | VUS | LB | BS3 | PM2 | P | BP4 |
| c.48+7G>C | p. | B/LBVUS | LB | BS3 | PM2 | P | BP4 |
| c.49-2A>T | p. | VUS/LP | VUS | PSV1 | M | PS3 | M | PM2 | P (PM4) |
| c.108+1G>A | p. | LP | P | PSV1 | M | PS3 | V5 | PM2 | P |
| c.108+2T>C | p. | LP | VUS | N/A | N/A | PM2 | P |
| c.109-6_109-4del | p. | (--) | LB | BS3 | PM2 | P | BP4 |
| c.109+2A>G | p. | LP | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.211+4G>A | p. | VUS | LB | BS3 | PM2 | P | BP4 |
| c.211-5del | p. | VUS | LB | BS3 | PM2 | P | BP4 |
| c.1684+4A>G | p. | (--) | VUS | N/A | N/A | PM2 | P |
| c.1685-2A>C | p. | VUS | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.1685-2A>G | p. | VUS | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.2513A>C | p.(Gln838Pro) | (--) | VUS | N/A | PM2 | P | (BP4_M) |
| c.2515-2A>G | p. | (--) | LP | N/A | PS3 | V5 | PM2 | P |
| c.2518+1G>T | p. | VUS | LB | BS3 | PM2 | P | BP4 |
| c.2748+1G>T | p. | LP/P | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.2748+2dup | p. | VUS | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.2748+4A>T | p. | (--) | LB | BS3 | PM2 | P | BP3 |
| c.2749+1G>T | p. | LP | LP | PSV1 | M | PS3 | V5 | PM2 | P (PM4) |
| c.2750+1G>T | p.(Val917Ala) | VUS | LB | BS3 | M | PM2 | P | n/a (BP4_M) |
| c.2834+4C>G | p.(Arg945Thr) | VUS | LB | BS3 | M | PM2 | P | BP3 | (BP4_M) |
| c.2834+1G>G | p. | LP/P | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.2834+3A>G | p. | LB/VUS | VUS | N/A | PM2 | P | BP3 |
| c.2834+5G>A | p. | VUS | VUS | N/A | N/A | PM2 | P |
| c.2834+6T>C | p. | (--) | LB | BS3 | PM2 | P | BP4 |
| c.2997+2del | p. | (--) | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.3007-1G>A | p. | VUS/LP | LP | PSV1 | M | PS3 | V5 | PM2 | P |
| c.3201+1G>A | p. | LP | LP | PSV1 | M | PS3 | V5 | PM2 | P |

(Continues)
Table 2. Continued

| cHGVS       | pHGVS  | ClinVar | ACMG-AMP classification | PVS1 | PS3/BS3 (mpgPALB2) | PS4 | PM2 | PM3 | PM4 | PP3/PP4 (splicing) | PP3/PP4 (protein) | BP7††† |
|-------------|--------|---------|--------------------------|------|--------------------|-----|-----|-----|-----|-------------------|-------------------|-------|
| c.3201+3_+4insTG | p.?   | (-) | LP                     | PS3_VS | PM2_P               | PP3 |     |     |     |                   |                   |       |
| c.3201+6T>A   | p.?   | VUS | VUS                    | N/A | PM2_P              | PM3 |     |     |     |                   |                   |       |
| c.3348C>T    | (p.(Y1116=) | VUS | VUS                    | N/A | PM2_P              | BP4 |     |     |     |                   |                   |       |
| c.3350+4A>G   | p.?   | P/LP | P                      | PS3_VS | PM2_P | PM3 |     |     |     |                   |                   |       |
| c.3350+5G>A   | p.?   | VUSLP | P                     | PS3_VS | PM2_P | PM3 |     |     |     |                   |                   |       |

‡‡‡ Not reported; HGVS, Human Genome Variation Society; LB, likely benign; LP, likely pathogenic; N/A, the pSAD read-out did not support neither a pathogenic nor a benign code strength; P, pathogenic; VUS, variant of uncertain significance.

☆☆☆ We used an ACMG-AMP point system Bayesian framework to combine all pathogenic and benign evidence. For each individual PALB2 variant under investigation, we evaluated all 16 pathogenic and 12 benign ACMG-AMP codes. The table shows only the pathogenic and benign codes that contributed to the final classification (see Supplementary materials and methods for further details).

††† We evaluated the clinical relevance of each variant per ACMG-AMP classification scheme using a Bayesian approach and incorporating the mpPALB2 splicing evidence (see Materials and methods). As shown in Table 2, this resulted in classifying 27% of the 42 PALB2 variants as pathogenically pathogenic, and 15% as possibly pathogenic. Eleven variants were classified as variants of uncertain significance (VUS). The number of VUSs rose from 26% (11 of 42) with previous experimental data in carriers (limited to nonspliceogenic variants) to 49% (12 of 24) with previous experimental data in carriers (see supplementary material, Table S5). Taken together, the data support the robustness of the mpPALB2 assay and the accuracy of the classification scheme with and without incorporating the mpPALB2 splicing evidence (supplementary material, Tables S6, S7; Figures S2, S3). Only 10969896, 2022, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/path.5839 by Csic Organización Central Om... on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License
classified as VUS (see supplementary material, Table S2, mgPALB2 incorporated classification). Note that c.3348C>T is one of very few examples in which SpliceAI failed to predict mgPALB2 read-outs accurately (supplementary material, Table S5).

Discussion

The genetic landscape of hereditary BC is characterized by a high complexity where loss-of-function variants in a minimum of eight genes (BRCA1, BRCA2, PALB2, BARD1, RAD51C, RAD51D, ATM, and CHEK2) show a significant association with BC risk in the general population, as several large-scale sequencing studies have recently reported [5,7,8,37]. Indeed, PALB2 has been firmly established as a high-risk BC susceptibility gene by these studies, with overall lifetime female BC risk above 30%. Accordingly, germline pathogenic variants in PALB2 are considered actionable findings in many clinical settings, with proposed actions ranging from intensified surveillance to prophylactic surgery [38], according to NCCN (www.nccn.org) and NICE guidelines (www.nice.org.uk/guidance). The latter highlights the clinical relevance of a robust PALB2 variant classification system. Moreover, PALB2 germline defects are not only clinically actionable, but are relatively prevalent, accounting for approximately 10% of the pathogenic variants of the eight genes mentioned above and 0.5% of all BC cases (422 pathogenic variants in 87,158 patients), underlining the role of this gene in hereditary BC. However, a relevant fraction of BC patients carries a VUS in their BC susceptibility genes (1.65% in PALB2) [7], which poses a challenge in genetic counseling, as risk estimates are based solely on personal and familial cancer history. Functional assays of VUS provide critical information for their clinical interpretation.

Aberrant splicing is a known frequent mechanism of gene inactivation associated with germline variants in BC genes [20,21,24,39]. Herein, we have focused on 82 PALB2 variants located at the splice-site boundaries detected in the BRIDGES cohort, so that we have accomplished the largest splicing functional study of PALB2 by minigenes to date. As we pointed out in previous reports, simplicity, sensitivity, robustness, or versatility are the main features of the minigene strategy, thus supporting its suitability for the preliminary characterization of potential spliceogenic variants. Furthermore, any other potential spliceogenic variant might be tested in these three constructs. Thus, a preliminary analysis of the 3,627 different ClinVar PALB2 variants (https://www.ncbi.nlm.nih.gov/clinvar/, last date accessed 16 April 2021) with splice-site predictors would select other additional 65 candidate spliceogenic variants that might be promptly checked in the three PALB2 minigenes.

Remarkably, a large proportion of tested variants (35/42, 83.3%) impaired splicing, supporting the specificity of our pre-selection approach. Moreover, the high sensitivity and resolution of fluorescent fragment analysis allowed us to detect up to 36 different aberrant transcripts. Furthermore, this methodology is appropriate to successfully distinguish transcripts with just 2-nucleotide deletions (Δ(E10p2)) from the mgFL transcript. In general, the splicing outcomes of these variants or of other ones affecting the same splice sites support the reproducibility of the minigenes when they were compared with patient RNA analysis previously described by other groups [25,40–42]. Hence, c.48G>A, c.49-2A>T, c.2748+1G>T (in this case versus c.2748+2T>G), c.2834+1G>A (versus c.2834+2T>C), c.3113+5G>C, and c.3350+5G>A yielded very similar or identical results. However, some discrepancies were observed between patient and minigene RNA assays, or even among different patient RNA reports, for variants c.3113G>A and c.3113+5G>C [6,25,42–44]. This may be due to the complexity of alternative splicing patterns in certain genomic regions as described for BRCA2 variant c.7976+5G>T [45], to technical issues, like the use of NMD inhibitors in the minigene assays that allow the detection of minor PTC transcripts, to the cell type where the tests are performed or to the methods used for characterizing transcripts, such as RT-PCR or RNA-seq, from patient RNA.

Despite that ±1,2 splice-site variants are usually considered pathogenic, some of them produced in-frame or the expected mgFL transcripts. Additionally, variants at other conserved nucleotides of the splice sites are potentially deleterious. Thus, although most spliceogenic variants of our study (18/34) affected ±1,2 nucleotides, 16 changes at other positions (the last three exonic and intronic +3, +4, +5, and +6 nucleotides) also disrupted splicing, confirming their relevance in splice-site recognition and suggesting their possible role in disease pathogenicity. In this regard, splicing impairments of +3 to +6 variants were especially difficult to predict by MES, because weak or no splicing impacts were observed for several variants, such as c.48+4C>T or c.2834+3A>G, among others. Moreover, two spliceogenic variants (c.3113+3A>G and c.3201+6T>A) did not significantly affect the MES score but they were selected because they changed conserved nucleotides of the 5′ss (see Materials and methods and Table 1). Then, we compared MES with other algorithms, such as NNSplice and SpliceAI (supplementary material, Tables S9–S11) in 38 previously assayed [+3_to_+6] variants [20,22–24,46]. The accuracy of MES and SpliceAI in the subset of +3 to +6 variants was similar (78.9 and 76.3%, respectively), outperforming NNSplice.

It is also worth mentioning two +2T>C variants (c.48+2T>C and c.108+2T>C) that convert a canonical GT donor into an atypical GC one (Figure 4). About 1% of human 5′ss are GC and are commonly associated to physiological alternative splicing [47,48]. Variants c.48+2T>C and c.108+2T>C induce different splicing impacts (0 and 86% of the mgFL transcript, respectively, Table 1). In fact, it has been previously estimated that approximately 15–18% of +2T>C substitutions are capable of generating variable amounts of canonical
transcripts (1–84%) [49]. Consequently, all these changes should be carefully interpreted because they are not invariably associated with major splicing disruptions. Likewise, two minor anomalous transcripts (Δ[E10q41] and Δ[E10q65]) are generated by using a GC-5’s.

The GC-5’s are intrinsically weak due to the change of the essential +2 nucleotide, but, in general, the other splice-site positions seem to be more conserved. The natural GC 5’s of PALB2-exon 12 and BRCA2-exon 17 as well as the GC-5’s created by c.108+2T>C in intron 2 (mgFL transcript, 86%) are identical (CAG|gcaagt) (Figure 4). These data suggest that the de novo GC of c.108+2T>C behaves like the natural ones of PALB2-exon 12 and BRCA2-exon 17 and might be regulated in a similar way. Indeed, several siRNA experiments of splicing factors suggested that BRCA2-exon 17 recognition is mediated by SC35, SF2/ASF, and Tra2β [23], while 9G8, Tra2β, and SC35 participate in the recognition of an anomalous GC used in the BTK gene [47]. Finally, the splicing outcome of variant c.3348C>T that affects the 5’s (GC) of exon 12 should also be highlighted. As calculated by MES, c.3348C>T weakens the natural GC donor (MES score: 3.1 to 1.9) and concomitantly creates a stronger GT site (MES score: 6.5). However, Δ[E12q4] and the mgFL transcript represent 26.7 and 68%, respectively, of the overall expression, indicating preferential use of the GC site and supporting a specific regulation of PALB2-exon 12 by splicing factors that enhance GC recognition.

Clinical interpretation

As PALB2 expert panel specifications of the ACMG-AMP guidelines are not yet available (https://clinicalgenome.org/, last accessed 25 November 2021), we have developed in-house specifications (Supplementary materials and methods). Incorporating mgPALB2 into the classification system makes a substantial impact on final variant classification, reducing VUSs from 28 to 11 (60% reduction), with a particularly high reclassification impact in the subgroup of 18 non-GT-AG intron variants. The latter reflects an intrinsic feature of the ACMG-AMP classification scheme: predictive codes for GT-AG and non-GT-AG intronic variants with similar splicing predictions have nonetheless very different strength (PVS1 versus PP4). As a result, experimental splicing data contribute very little to GT-AG variants with accurate splicing predictions (replacing a PVS1 code for a PS3 code with identical strength).

Incorporating mgPALB2 read-outs into the ACMG-AMP schema has been challenging. In the end, we propose solutions not necessarily supported by the ClinGen SVI. These include: (1) an approach to transform complex mgPALB2 read-outs into a PS3/BS3 code (supplementary material, Figure S3A) and (2) the arbitrary definition of a ≥90% overall pathogenic (or benign) expression threshold to support PS3 (or BS3). We think that, lacking experimental data supporting another threshold (we are not aware of experimental data demonstrating the level of PALB2 expression conferring haplo-sufficiency), 90% has the merit of being conservative. Based on this, mgPALB2 read-outs in up to nine variants were considered conflicting and, therefore, not contributing to the final ACMG-AMP classification (Table 2 and supplementary material, Table S2). Not surprisingly, all these variants ended up as VUSs. It is tempting to speculate that some of them (e.g. c.2834+5G>A, producing 59% of PTC-NMD transcripts and 41% of mgFL transcripts) could be associated with some intermediate risk level for developing BC (or other malignancies).

Our study shows the limited contribution of the ACMG-AMP evidence code PS4 to classify rare genetic variants in BC susceptibility genes, even if associated with high risk (as it is the case of PALB2). After analyzing the larger PALB2 association study reported so far (>113,000 woman in the BRIDGES cohort), only one of the 42 variants under investigation (c.108+3G>A, odds ratio = 4.57, p = 0.007) qualifies for PS4.

An overall reduction in the proportion of VUSs is a desirable feature of any variant classification system but relocating previously classified variants into the VUS category might be clinically relevant. In this regard, we highlight GT-AG variants c.49-2A>T, c.108+2T>C, and c.211+1G>A, all of them considered likely pathogenic by ClinVar submitters. We have classified these variants as VUS. It is not the mgPALB2 data that explain the difference (all three ended up as VUSs, regardless of including/excluding mgPALB2 data from our classification scheme), but it is the accuracy of SpliceAI combined with the PALB2 adaptation of the PVS1 decision tree.

We recommend being cautious when classifying GT-AG variants in high-risk BC genes such as PALB2 [5], for which prophylactic surgery might be recommended.
to healthy carriers. In a previous study [25], we already warned about certain PALB2 GT-AG variants. Based on the present data, we have reevaluated and updated these warnings (supplementary material, Table S7).

In summary, we have tested 42 PALB2 splice-site variants in minigenes that have proven to be an appropriate and straightforward strategy for the characterization of splicing outcomes of putative splicogenic variants. The subsequent application of the ACMG-AMP evidence code PS3/BS3 reduced VUSs by 60%. We ended up classifying 23 variants as pathogenic/likely pathogenic. Remarkably, these 23 variants account for approximately 15% of all presumed pathogenic variants reported in BRIDGES subjects.

Acknowledgements

PD, MPGV, DFE, MdlH, and EAV have received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement no. 634935. The laboratory of EAV is supported by grants from the Spanish Ministry of Science and Innovation, Plan Nacional de I+D+i 2013-2016, ISCIII (PI17/00227 and PI20/00225) co-funded by FEDER from Regional Development European Funds (European Union) and from the Consejería de Educación, Junta de Castilla y León, ref. CSI242P18 (actuación cofinanciada P.O. FEDER 2014-2020 de Castilla y León). The laboratory of MdlH is supported by a grant from the Spanish Ministry of Science and Innovation, Plan Nacional de I+D+i 2013-2016, ISCIII (PI20/00110) co-funded by FEDER from Regional Development European Funds (European Union). Programa Estratégico Instituto de Biología y Genética Molecular (IBGM), Escuela de Excelencia, Junta de Castilla y León (ref. CLU-2019-02). AV-P is supported by a predoctoral fellowship from the Consejería de Educación, Junta de Castilla y León (2018–2022). LS-M is supported by a predoctoral fellowship from the AECC-Scientific Foundation, Sede Provincial de Valladolid (2019–2023). AE-S is supported through the Operational Program for Youth Employment and Youth Employment Initiative (YEI), called by the Community of Madrid in 2020, and co-financed by the European Social Fund.

Author contributions statement

MdlH and EAV conceived the study. AV-P, LS-M, EB-M, SC, JA, LD, DFE, PD and MPGV were responsible for data curation. AV-P, EB-M, LS-M, VL, MdlH and EAV carried out the formal analysis. PD, DFE, MPGV, MdlH and EAV acquired funding. AV-P, LS-M, EB-M, VL, EF-B, MD, PP-S, AE-S, AGA, SG-B, MPGV, MdlH and EAV carried out the investigations. AV-P, EF-B, LS-M, EB-M, AGA and EAV were responsible for methodology. EAV supervised the study. AV-P, MdlH and EAV wrote the original draft of the manuscript. SG-B, MdlH and EAV reviewed and edited the manuscript. All authors approved the final version of the manuscript.

Data availability statement

All sequencing and fragment analysis data are available at https://digital.csic.es/handle/10261/244566

References

1. Xia B, Sheng Q, Nakanishi K, et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell 2006; 22: 719–729.
2. Hanenberg H, Andreassen PR. PALB2 (partner and localizer of BRCA2). Atlas Genet Cytogenet Oncol Haematol 2018; 22: 484–490.
3. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet 2007; 39: 165–167.
4. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med 2014; 371: 497–506.
5. Yang X, Leslie G, Doroszuk A, et al. Cancer risks associated with germline PALB2 pathogenic variants: an international study of 524 families. J Clin Oncol 2020; 38: 674–685.
6. Reid S, Schindler D, Hanenberg H, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet 2007; 39: 162–164.
7. Hu C, Hart SN, Gnaoaviru R, et al. A population-based study of genes previously implicated in PALB2. N Engl J Med 2021; 384: 440–451.
8. Breast Cancer Association Consortium, Dorling L, Carvalho S, et al. Breast cancer risk genes – association analysis in more than 113,000 women. N Engl J Med 2021; 384: 428–439.
9. Hofstatter EW, Domchek SM, Miron A, et al. PALB2 mutations in familial breast and pancreatic cancer. Fam Cancer 2011; 10: 225–231.
10. Blanco A, de la Hoya M, Osorio A, et al. Analysis of PALB2 gene in BRCA1/BRCA2 negative Spanish hereditary breast/ovarian cancer families with pancreatic cancer cases. PLoS One 2013; 8: e76538.
11. Ramus SJ, Song H, Dicks E, et al. Germline mutations in the BRRP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. J Natl Cancer Inst 2015; 107: djv214.
12. Blanco A, de la Hoya M, Balmaña J, et al. Detection of a large rearrangement in PALB2 in Spanish breast cancer families with male breast cancer. Breast Cancer Res Treat 2012; 132: 307–315.
13. Pearlman R, Frankel WL, Swanson B, et al. Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. JAMA Oncol 2017; 3: 464–471.
14. Fewings E, Larianov A, Redman J, et al. Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without CDH1 mutation: a whole-exome sequencing study. Lancet Gastroenterol Hepatol 2018; 3: 489–498.
15. de Voogt KMK, van Wijk R, van Solinge WW. Management of gene promoter mutations in molecular diagnostics. Clin Chem 2009; 55: 698–708.
16. Manning KS, Cooper TA. The roles of RNA processing in translating genotype to phenotype. Nat Rev Mol Cell Biol 2017; 18: 102–114.
17. Ward AJ, Cooper TA. The pathobiology of splicing. J Pathol 2010; 220: 152–163.
Functional and clinical classification of PALB2 splice-site variants

18. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 2007; 8: 749–761.

19. Baralle D, Lucassen A, Buratti E. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep* 2009; 10: 810–816.

20. Sanoguera-Miralles L, Valenzuela-Palomo A, Bueno-Martínez E, et al. Comprehensive functional characterization and clinical interpretation of 20 splice-site variants of the RAD51C gene. *Cancers* (Basel) 2020; 12: 3771.

21. Bueno-Martínez E, Sanoguera-Miralles L, Valenzuela-Palomo A, et al. RAD51D aberrant splicing in breast cancer: identification of splicing regulatory elements and miRNA-based evaluation of 53 DNA variants. *Cancers* (Basel) 2021; 13: 2845.

22. Acedo A, Hernández-Moro C, Curiel-García Á, et al. Functional classification of BRCA2 DNA variants by splicing assays in a large miniature with 9 exons. *Hum Mutat* 2015; 36: 210–221.

23. Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, et al. Functional classification of DNA variants by hybrid minigenes: identification of 30 splice-ogenic variants of BRCA2 exons 17 and 18. *PLoS Genet* 2017; 13: e1006691.

24. Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, et al. Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays. *J Pathol* 2019; 248: 609–420.

25. Lopez-Peralto I, Leman R, Behar R, et al. Alternative splicing and ACMG-AMP 2015-based classification of PALB2 genetic variants: an ENIGMA report. *J Med Genet* 2019; 56: 453–460.

26. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; 11: 377–394.

27. Moles-Fernández A, Duran-Lozano L, Montalban G, et al. Computational tools for splicing defect prediction in breast/ovarian cancer genes: how efficient are they at predicting RNA alterations? *Front Genet* 2018; 9: 366.

28. Houdayer C, Caux-Moncoutier V, Krieger S, et al. Splicing analysis in molecular diagnosis derived from a set of 327 combined in silico and in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 2012; 33: 1228–1238.

29. Bryskin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* 2010; 48: 463–465.

30. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012; 9: 671–675.

31. Richards S, Aziz N, Bale S, et al. 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* 2015; 17: 405–424.

32. Tavitgian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* 2018; 20: 1054–1060.

33. Tavitgian SV, Harrison SM, Boucher KM, et al. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat* 2020; 41: 1734–1737.

34. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 2018; 39: 1517–1524.

35. Jaganathan K, Kyriazopoulou Fanagiopoulou S, McRae IF, et al. Predicting splicing from primary sequence with deep learning. *Cell 2019; 176: 535–548.e24.

36. Chabananon A, Winters-Hilt S, Koonin EV, et al. Accumulation of GC donor splice signals in mammals. *Bio Direct* 2008; 3: 30.

37. Narod SA. Which genes for hereditary breast cancer? *N Engl J Med* 2021; 384: 471–473.

38. Tischkowitz M, Balmaha J, Foulkes WD, et al. Management of individuals with germ-line variants in PALB2: a clinical practice resource
61. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* 2019; 12: 3.

62. Brnich SE, Arteaga EC, Wang Y, et al. A validated functional analysis of partner and localizer of BRCA2 missense variants for use in clinical variant interpretation. *J Mol Diagn* 2021; 23: 847–864.

63. Rodrigue A, Margaillan G, Torres Gomes T, et al. A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. *Nucleic Acids Res* 2019; 47: 10662–10677.

64. Boonen RACM, Rodrigue A, Stoepker C, et al. Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. *Nat Commun* 2019; 10: 5296.

65. Wiltshire T, Ducy M, Foo TK, et al. Functional characterization of 84 PALB2 variants of uncertain significance. *Genet Med* 2020; 22: 622–632.

66. Li SC, Goto NK, Williams KA, et al. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc Natl Acad Sci U S A* 1996; 93: 6676–6681.

67. Whiffin N, Roberts AM, Minikel E, et al. Using high-resolution variant frequencies empowers clinical genome interpretation and enables investigation of genetic architecture. *Am J Hum Genet* 2019; 104: 187–190.

68. Southey MC, Goldgar DE, Wingqvist R, et al. PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. *J Med Genet* 2016; 53: 800–811.

69. Mori M, Hira A, Yoshida K, et al. Pathogenic mutations identified by a multimodality approach in 117 Japanese Fanconi anemia patients. *Haematologica* 2020; 105: 1166–1167.

70. Biesecker LG, Harrison SM, ClinGen Sequence Variant Interpretation Working Group. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med* 2018; 20: 1687–1688.

**SUPPLEMENTARY MATERIAL ONLINE**

Supplementary materials and methods

Figure S1. Insert sequences of minigenes mgPALB2_ex1-3, _ex4-6 and _ex5-12

Figure S2. Mapping PALB2 regions critical to protein function

Figure S3. (A) Proposed decision tree assigning a PS3/BS3 code strength to mgPALB2 minigene read-outs. (B) Pathogenic/Benign annotation of mgPALB2 minigene deconvolute read-outs (transcripts). (C) Pathogenic/Benign annotation of PALB2 transcripts

Figure S4. Additional splicing functional assays of PALB2 variants

Table S1. Cloning and mutagenesis primers of PALB2

Table S2. ACMG-AMP point-based classification of 42 PALB2 variants

Table S3. Bioinformatics analysis of PALB2 variants with Max Ent Score

Table S4. Transcript annotations according to PALB2 sequence NM_024675.4

Table S5. Comparative of SpliceAI prediction, mgPALB2 read-outs, and experimental splicing data in carriers

Table S6. PALB2 regions critical to protein function

Table S7. PALB2 sites/variants for which we place a warning

Table S8. Impact of mgPALB2 data on the ACMG-AMP classification of 42 PALB2 variants

Table S9. Bioinformatics predictions of +3 to +6 variants and splicing outcomes

Table S10. Summary of bioinformatics predictions

Table S11. Sensitivity and specificity for +3 to +6 variants