



UNIVERSITAT_{DE}
BARCELONA

Enhancing variant classification in hereditary breast and ovarian cancer predisposition genes through splicing analysis and new tumour evidence

Joanna Domènech Vivó

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

**ENHANCING VARIANT CLASSIFICATION
IN HEREDITARY BREAST AND OVARIAN
CANCER GENES THROUGH SPLICING
ANALYSIS AND NEW TUMOUR EVIDENCE**

Joanna Domènech Vivó
2024



About the cover:

The cover of this thesis features an olive tree, which symbolically weaves together the central themes of the research. The tree represents both the family pedigrees central to the research of this thesis about hereditary breast and ovarian cancer and the decision trees used in the classification of germline variants. Furthermore, it pays homage to my family roots in Camarasa.

The chosen colour palette reflects the nuanced process of classifying variants of uncertain significance. Blue represents variants that are closer to being benign, while orange symbolises variants leaning toward pathogenicity.

Cover design by Zeleste Juárez-Martín and Joanna Domènech-Vivó.



UNIVERSITAT DE
BARCELONA



Enhancing variant classification in hereditary breast and ovarian cancer predisposition genes through splicing analysis and new tumour evidence

Doctoral thesis submitted by **Joanna Domènech Vivó** for the degree of Doctor by the Universitat de Barcelona

This work was developed under the supervision of **Dr. Sara Gutiérrez-Enríquez** and **Dr. Orland Diez Gibert** at the Vall d'Hebron Institute of Oncology (VHIO)

Biomedicine Doctoral Program

Faculty of Biology, Universitat de Barcelona (UB)

Tutor: **Dr. Alex Sánchez Pla**

A handwritten signature in blue ink, reading "Joanna Domènech Vivó".

Joanna Domènech Vivó

Barcelona, 2024

This thesis was supported by the Spanish Instituto de Salud Carlos III funding, an initiative of the Spanish Ministry of Economy and Innovation, partially supported by European Regional Development FEDER Funds (grants PI19/01303 and PI22/01200), the Secretariat for Universities and Research of the Ministry of Business and Knowledge (AGAUR-FEDER, 2021SGR01112; AGAUR 2020 FI-B 00584) and Health Department (PERIS SLT028/23/000310), of the Government of Catalonia and the European Social Fund, and the ERA-NET (RAD51predict, ERAPERMED2019-215).

ABBREVIATIONS

A

ACC · Acceptor site
ACMG/AMP · American College of Medical Genetics and Genomics/Association for Molecular Pathology
AE · Allele-specific
AG · Acceptor gain
AJCC · American Joint Committee on Cancer
AL · Acceptor loss
AUC · Area under the curve

B

BAC · Bacterial artificial chromosome
BC · Breast cancer
BCCR · Breast cancer cluster region
BER · Base excision repair
BOADICEA · Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm
BPP · Branch Point Prediction
BRCA1 · BRest CAncer gene 1
BRCA2 · BRest CAncer gene 2
BRCT · BRCA1 C terminus
BRRM · Bilateral risk reduction mastectomy
Bx · Biopsy

C

C.E · Capillary electrophoresis
CC · Coiled Coil
CCC · Clear cell carcinoma
ChAM · Chromatin Association Motif
CI · Confidence interval
ClinGen · The Clinical Genome Resource
CON · Conflicting
CRRM · Contralateral risk reducing mastectomy

D

DBD · DNA binding domain
DBS · DNA double-strand break
DCIS · Ductal carcinoma *In situ*

DDR · DNA damage response
Del · Deletion
Delins · Deletion-insertion variant
DG · Donor gain
DL · Donor loss
DON · Donor site
DR-GFP · Direct repeat -green fluorescent protein
DS · Delta Score

E

EC · Endometroid carcinoma
ENIGMA · Evidence-based Network for the Interpretation of Germline Mutant Alleles
EOC · Epithelial ovarian cancer
ER · Oestrogen receptor
ESE · Exonic splicing enhancer
ESMO · European Society for Medical Oncology
ESS · Exonic splicing silencer

F

FA · Fanconi Anaemia
FDA · Food and Drug Administration
FFPE · Formalin-Fixed Paraffin-Embedded
FIGO · International Federation of Gynaecology and Obstetrics
FL · Full-length
FN · False negatives
FP · False positives
FUNC · Functional

G

gBRCA1/2 · Germline *BRCA1/2* mutation
gPALB2 · Germline *PALB2* mutation
GWAS · Genome-wide association study
GIS · Genomic instability score
gsLOH · Gene-specific loss of heterozygosity

H

HBOC · Hereditary breast and ovarian cancer

HDR · Homology-Directed Repair
 HER2 · Human epidermal growth factor receptor 2
 HGMD · Human Gene Mutation Database
 HGS · High-grade serous OC
 HNPCC · Hereditary non-polyposis colorectal cancer
 hnRNPs · Heterogeneous nuclear ribonucleoproteins
 HRD · Homologous recombination deficiency
 HRP · Homologous recombination proficiency
 HRR · Homologous recombination repair

I

IARC · International Agency for Research on Cancer
 IDC · Invasive ductal carcinoma
 IF · Immunofluorescence
 IHC · Immunohistochemistry
 ILC · Invasive lobular carcinoma
 INT · Intermediate
 IMP · Impaired
 ISE · Intronic splicing enhancer
 ISS · Intronic splicing silencer

K

KBM · KEAP1-binding motif

L

LB · Likely benign
 LCLs · Lymphoblastoid cell lines
 LGS · Low grade serous OC
 LoF · Loss of function
 LOH · Loss of heterozygosity
 LP · Likely pathogenic
 LR · Likelihood ratio
 LST · Large-scale transition

M

M · Moderate

MAF · Minor allele frequency
 MBD · MRG15-binding domain
 MC · Mucinous carcinoma
 MCC · Matthew's correlation coefficient
 MES · MaxEntScan
 mESC · Mouse embryonic stem cell
 MMR · DNA mismatch-repair
 MRI · Magnetic resonance imaging
 mRNA · messenger RNA

N

NA · Not available
 NE · Non-evaluable
 NER · Nucleotide excision repair
 NES · Nuclear Export Signal
 NLS · Nuclear Localization Sequence
 NMD · Nonsense-mediated decay
 NOS · Not otherwise specified carcinoma
 NPV · Negative predictive value
 NR · Not reported
 NST · No special type carcinoma
 nt · Nucleotide

O

OC · Ovarian cancer
 OCCR · Ovarian cancer cluster region
 OR · Odds ratio

P

p · p-value
 P · Supporting
 PARP · Poly (ADP-ribose) polymerase
 PARPi · PARP inhibitor
 PBMCs · Peripheral blood mononuclear cells
 PCR · Polymerase chain reaction
 PPV · Positive predictive value
 PR · Progesterone receptor
 PRS · Polygenic risk score
 PSA · Prostate specific antigen
 PSI · Percent Spliced-In
 PTC · Premature termination codon
 PWM · Position weight matrix

R

RING · Really Interesting New Gene
ROC · Receiver operating characteristic
RRSO · Salpingo-oophorectomy
RT-PCR · Reverse transcription PCR

S

S · Strong
SD · Standard deviation
SCD · SQ/TQ cluster domain
SEOM · Sociedad Española de Oncología Médica
SF · Splicing fraction
SNP · Single nucleotide polymorphism
snRNPs · Small nuclear ribonucleoproteins
SPiCE · Splicing Prediction in Consensus Elements
SR proteins · Serine arginine-rich protein
SRE · Splicing regulatory element
SS · Splicing site
SSBs · DNA single-strand breaks
ssDNA · Single-strand DNA
SSF · SpliceSiteFinder-like
SVI · Sequence Variant Interpretation
Sx · Surgical specimen

T

TAI · Telomeric allelic imbalance

tBRCA1/2 · Tumour mutations in *BRCA1/2*
TN BC · Triple negative breast cancer
TN · True negatives
TNM · Tumour-node-metastasis
TP · True positives

U

UTRs · Untranslated regions

V

VAF · Variant allele frequency
VCEP · Variant curation expert panel
VHIO · Vall d'Hebron Institute of Oncology
VUS · Variant of unknown significance

W

WHO · World Health Organization
WGS · Whole-genome sequencing
WT · Wild type

4

4-OHT · 4-Hydroxytamoxifen

INDEX

Index

ABSTRACT	1
INTRODUCTION	7
1. Hereditary Breast and Ovarian Cancer.....	9
1.1. Epidemiology	9
1.2. Breast Cancer Subtypes	11
1.3. Ovarian Cancer Subtypes	13
1.4. Genetic Susceptibility to HBOC	15
1.5. Patients' Clinical Management	24
2. Variant Classification.....	30
2.1. Variants of Unknown Significance.....	31
2.2. Variant Classification Guidelines.....	33
2.3. ENIGMA Consortium	38
2.4. ENGIMA as Variant Curation Expert Panel (VCEP)	41
3. mRNA Splicing in HBOC	43
3.1. Splicing Elements	43
3.2. Alternative Splicing.....	46
3.3. Aberrant Splicing and Disease	47
3.4. <i>In silico</i> Splicing Assessment	52
3.5. <i>In vitro</i> Splicing Analyses	54
3.6. Splicing Observed Impact for Variant Classification.....	58
4. High-risk HBOC Genes: Carcinogenesis, Structure and Functions	61
4.1. <i>BRCA1</i>	64
4.2. <i>BRCA2</i>	66
4.3. <i>PALB2</i>	66
4.4. Homologous Recombination Repair (HRR)	67
4.5. Alterations in HRR Genes and Homologous Recombination Repair Deficiency (HRD)	69
5. Rationale for the Thesis	72
HYPOTHESIS AND OBJECTIVES.....	75

Hypothesis	77
Objectives	79
MATERIALS AND METHODS.....	81
Part 1.....	83
Datasets.....	83
<i>In Silico</i> Variant Annotation and Analysis for All Datasets	84
Statistical Analysis.....	85
Experimental RNA Analysis in Patient’s Data Set	86
Ethical Considerations.....	88
Part 2.....	89
ENIGMA Proposal	89
Annotation and <i>In Silico</i> Analysis.....	89
Patients’ mRNA Analysis	90
Minigene Splicing Assay.....	94
Mouse Embryonic Stem Cells (mESCs)-Based Functional Assay.....	95
Variant Classification.....	98
Statistical Analysis.....	107
Part 3.....	109
Research Cohorts	109
Functional HRD: RAD51 Test	111
Genetic and Genomic HRD: Myriad’s MyChoice® HRD.....	112
Variant Classification.....	113
Statistical Analysis.....	115
ACMG/AMP LRs Evidence Strength Alignment	117
RESULTS.....	119
Part 1.....	121
SpliceAI Optimally Predicts Deep Intronic Splice-Altering Variants but with Less Sensitivity Those Affecting Splicing by Altering Regulatory Elements .	121
Splicing Regulatory Elements Balance Is Similar between Pseudoexons and Canonical Exons	125

Inclusion of SRE Landscape in the <i>In Silico</i> Detection of Deep Intronic Splice-Altering Variants	130
Experimental Analysis of Hereditary Cancer Gene Variants	130
Part 2.....	137
<i>BRCA1</i> Exon 18 Variants	137
<i>In Silico</i> Identification of Potentially Splicing-Altering Variants	138
mRNA Splicing Profiling Characterization in Patient Samples	141
RNA Splicing Evaluation through Minigene Reporter Assay.....	148
Functional Tolerability Range of Exon 18 Skipping.....	151
Observed Splicing Impact Specially Benefits the Classification of Intronic/Synonymous Variants in the Absence of Additional Evidence	161
Experimental RNA Splicing Evidence Reduces Variant Uncertainty and Is Essential for a Correct Interpretation of Variants Under ENIGMA-Adapted ACMG/AMP Guidelines.....	165
The Splicing and Functional Ambiguity behind the c.5152+7A>G Variant	171
Part 3.....	175
Study Populations	177
Prevalence of functional and genomic HRD in the Cohort of Unreviewed Variants	179
Clinical Characteristics of Pathogenic and Benign Cohorts	189
Prevalence of Functional HRD in Pathogenic and Benign Cohorts	190
RAD51% Performance in Predicting Germline Variant Status	197
RAD51% LR Estimates for Tumour-Phenotype Evidence in Germline Variant Classification	198
VUS Classification by Adding RAD51% Tumour Evidence	203
DISCUSSION	207
Part 1.....	209
Part 2.....	215
Part 3.....	227
INTEGRATED DISCUSSION	237
The Future in HBOC Genetic Testing: Whole Genome Sequencing and Need for Guidance on Deep Intronic Regions.....	237

SpliceAI in Variant Classification: Can *In Silico* Evidence Be Equally
Considered for All Variants? 240

Relevance of mRNA Experimental Testing in Variant Classification..... 245

Potential Usage and Limitations of Tumour Information for Germline Variant
Classification 249

The Relevance of ENIGMA Collaborative Efforts 253

CONCLUSIONS 259

 Part 1 261

 Part 2 261

 Part 3 262

REFERENCES 265

APPENDIX I 315

APPENDIX II 445

APPENDIX III 453

ACKNOWLEDGEMENTS 457

Table Index

Table 1. 2014 FIGO staging system and corresponding TMN for ovarian cancer.	14
Table 2. Risk of BC overall associated with protein-truncating variants in susceptibility genes from the population-based case-control BRIDGES study.	22
Table 3. SEOM's selection criteria for germline HBOC testing.	25
Table 4. IARC variant classification system and corresponding clinical recommendations.	35
Table 5. ACMG/AMP rules for combining criteria to classify sequence variants.	37
Table 6. Formulas for performance evaluation of in silico tools.	85
Table 7. PVS1(RNA) and BP7(RNA) specifications for variant classification.	103
Table 8. Likelihood ratios alignment with ACMG/AMP qualitative rules strengths.	118
Table 9. Performance of ΔSpliceAI for literature dataset with 233 deep intronic variants using the optimized threshold of 0.05.	122
Table 10. Performance of ΔSpliceAI (4999bp) optimized threshold for SREs disrupting exonic variants.	124
Table 11. Performance of ΔSpliceAI alone and sequentially combined with ESRseq tool with 233 intronic variants from the literature dataset (133 altering and 100 non altering splicing).	125
Table 12. Deep intronic variants with a spliceogenic effect detected in hereditary cancer genes.	132
Table 13. Performance of ΔSpliceAI alone and the proposed pipeline for the experimental dataset.	136
Table 14. Variants assessed in the mESC-based functional assay.	153
Table 15. Summary of splicing and functional results for the c.5152+7A>G variant.	171
Table 16. Germline and tumour characteristics of the unreviewed cohort cases.	184
Table 17. Patients' characteristics in pathogenic and benign cohorts.	190
Table 18. Estimated likelihood ratios for predicting BRCA1/2 and PALB2 mutation status by tumour HRR status by RAD51%.	201
Table 19. Proposed utilization of RAD51% evidence for variant classification.	202
Table 20. Final classification of VUS by adding RAD51% tumour evidence.	205
Table 21. ΔSpliceAI accuracy based on variant type	219
Table 22. Likelihood ratio analysis of the ΔSpliceAI proposed thresholds from (A)Walker et al. (2023) [170] and (B) those obtained in Part 1 for 233 deep intronic variants (Supplementary Table 1).	243
Table 23. Likelihood ratio analysis of the ΔSpliceAI proposed thresholds from Walker et al. (2023) [170] for 1161 exonic variants affecting SREs (Supplementary Table 2).	244
Table 24. Likelihood ratio analysis of the ΔSpliceAI proposed thresholds from Walker et al. (2023) [170] for 166 variants in exon 18 of BRCA1 (Supplementary Table 9).	244
Table 25. Gene expression in whole blood and lymphoblastoid cell lines.	248

Figure Index

Figure 1. Distribution of worldwide new cases and deaths for cancer in 2022 for (A) both sexes and (B) females.	10
Figure 2. Aetiology of breast and ovarian cancer.	10
Figure 3. Breast cancer subtypes.	13
Figure 4. Relative risk of BC and allele frequencies of high-, moderate-, and low-risk genetic variants.	15
Figure 5. Pedigree of the family of Mrs. Z described by Dr. Broca in the “Traité des tumeurs” in 1866.	16
Figure 6. Role of HBOC and other hereditary cancer syndrome susceptibility genes in genome stability pathways.	18
Figure 7. Clinical validity classifications of candidate susceptibility genes associated with syndromic or familial BC and OC.	20
Figure 8. Estimated absolute risk of BC associated with protein-truncating variants in eight genes compared with the general population risk.	23
Figure 9. PARP inhibitor’s function in BRCA1/2 deficient cells is based on synthetic lethality.	30
Figure 10. Proportion of BRCA1 and BRCA2 variants reported in ClinVar database according to their clinical significance.	32
Figure 11. Flowchart of the procedure for classification of sequence variants of unknown clinical significance described by Goldgar et al. (2004) [106].	34
Figure 12. The ACMG/AMP evidence framework.	36
Figure 13. Point system ranges to assign variant classes according to Tavtigian et al. (2020)[119].	38
Figure 14. ENIGMA working groups.	40
Figure 15. Timeline for migration from “external” expert panel to operation as a ClinGen Variant Curation Expert Panel following the ClinGen FDA-recognized protocol.	42
Figure 16. Schematic representation of cis and trans splicing elements.	45
Figure 17. Alternative splicing events.	47
Figure 18. Schematic representation of effects of splicing disruption on mRNA.	48
Figure 19. Graphical representation of in vitro splicing effect of 346 variants in HBOC and Lynch Syndrome risk genes.	50
Figure 20. Splicing effects caused by deep intronic variants.	51
Figure 21. The principles of minigene assay.	57
Figure 22. PVS1 evidence assignment of gene-specific codes to splice donor/acceptor $\pm 1,2$ dinucleotide variants according to ClinGen SVI Splicing Subgroup.	59
Figure 23. ENIGMA VCEP recommendations for applying PVS1/BP7 (RNA) evidence.	61
Figure 24. BRCA1, BRCA2, and PALB2 exon and domain maps.	63
Figure 25. Distributions of missense variants in BRCA1 and BRCA2.	65
Figure 26. DNA double-strand break (DSBs) repair by homologous recombination (HR).	69

Figure 27. HRD tumour detection approaches.	72
Figure 28. Schematic representation of the pSPL3mBK2-BRCA1-ex18_19 minigene.	94
Figure 29. Modification of in silico codes by splicing observed impact in different variant types (synonymous/intronic and missense/in-frame).	101
Figure 30. ΔSpliceAI scores of spliceogenic and not spliceogenic variants from the literature dataset, considering predictions of cryptic sites in a window of 50bp or 4999bp.	123
Figure 31. SREs abundance in different genomic regions, using the normalized SRE area calculated with ESRseq scores.	127
Figure 32. Comparison of SREs abundance in different genomic regions, using the Normalized SRE Area calculated using ESRseq scores.	128
Figure 33. Comparison of the absolute difference of normalized SRE area from 100 nucleotides before and after each deep intronic variant compiled from the literature	129
Figure 34. Spliceogenic variants characterization in patients' RNA.	135
Figure 35. Graphical representation of the 166 variants analysed in Part 2.	137
Figure 36. Splicing regulatory elements (SREs) distribution among exon 18 and intronic adjacent regions.	140
Figure 37. Representative Sanger and capillary-electrophoresis (C.E) results of BRCA1 splicing events detected in patients' samples.	141
Figure 38. Semi-quantitative mRNA splicing profiling by capillary electrophoresis of fluorescent amplicons in controls (n=46), patients (n=51 samples, 32 unique variants).	144
Figure 39. Allele-specific assays to assess the contribution of mutant allele to full-length (FL) expression.	147
Figure 40. Semi-quantitative mRNA splicing profiling by capillary electrophoresis of fluorescent amplicons in minigenes (n=166).	149
Figure 41. Correlation analysis of full-length (FL) expression in patients and minigenes.	151
Figure 42. Mouse-embryonic stem cells (mESC) mRNA splicing profiling.	154
Figure 43. Homology direct repair (HDR) activity of BRCA1 exon 18 variants relative to wild-type BRCA1 capacity in mESC.	155
Figure 44. Representative images of complementation phenotypes of BRCA1 exon 18 variants in mESC.	157
Figure 45. Linear regression analysis between BRCA1 exon 18 skipping levels (A) or FL (B) (X-axis) and HDR activity (Y-axis).	158
Figure 46. Correlation analysis between functional scores from saturation genome editing in HAP1 cells and HDR outcomes in mESCs.	159
Figure 47. Splicing profile correlation with saturation genome editing functional assay.	161
Figure 48. Sankey plot showing in silico and experimental splicing evidence usage for variant reclassification depending on coding effect for 164 variants.	163
Figure 49. Overview of the classification for 164 out of 166 variants before and after this study by considering in silico, variant coding effect, position functional, allele frequency, multifactorial clinical data, and mRNA experimental evidence generated in this project.	168

Figure 50. Overview of 166 variant classifications before and after this study by considering in silico, variant coding effect, position functional, allele frequency, multifactorial clinical data, and mRNA splicing experimental evidence generated in this project.	170
Figure 51. Sashimi plots from RT-PCRseq experiments showing BRCA1 exon 18 splicing profiles.	172
Figure 52. Flow diagram for RAD51 test.	177
Figure 53. γH2AX score distribution in breast and ovarian tumour samples from patients carrying germline pathogenic, benign and unreviewed variants.	178
Figure 54. HRR status by RAD51% and GIS in 38 tumours from germline carriers of unreviewed variants.	179
Figure 55. Correlation between RAD51% and GIS tumour results in all samples (A), in surgical specimens (B) and in biopsies (C).	181
Figure 56. Flow diagram of the cohorts used to validate the RAD51 assay for germline variant classification.	188
Figure 57. RAD51% score distribution in breast and ovarian tumour samples from patients carrying germline pathogenic, benign and VUS classified variants.	191
Figure 58. Summary of clinical and molecular features of the tumours from germline carriers of benign (A) and pathogenic (B) variants.	193
Figure 59. RAD51% score distribution among tumours from patients carrying germline pathogenic variants in BRCA1, BRCA2 and PALB2.	194
Figure 60. RAD51% score distribution among tumours from germline carriers of premature termination codon, missense/in-frame deletions or insertions, and splicing pathogenic variants.	195
Figure 61. RAD51% score distribution across tumours from patients carrying distinct pathogenic variants in BRCA1(A), BRCA2(B) and PALB2(C).	196
Figure 62. Receiver-operating characteristic curves (ROC) of RAD51% performance in predicting germline variant status.	198
Figure 63. Proposed code adaptation for in-frame splicing events targeting critical BRCA1 domains based on mESC HDR capacity and splicing profiles.	222
Figure 64. Process for calibrating splicing prediction score thresholds for in silico tools according to the ClinGen SVI splicing subgroup recommendations.	241
Figure 65. Number of unique BRCA1 exon 18 analysed variants.	255

ABSTRACT

The clinical management of hereditary breast and ovarian cancer (HBOC) families relies on identifying germline pathogenic variants in cancer susceptibility genes, such as *BRCA1* and *BRCA2*. However, genetic testing often yields an inconclusive result due to detection of only neutral variants or to the identification of variants of unknown significance (VUS), whose effects on gene function and associated cancer risk are unknown. While carriers of pathogenic variants benefit from options available for early cancer detection, preventive measures and targeted therapies, VUSs carriers are generally managed without regard to the result of genetic testing, thus limiting the options for these patients. The primary aim of this thesis is to increase the number of patients receiving a genetic diagnosis of HBOC by improving the accuracy of classifying germline genetic variants. To achieve this, we focused on improving the detection of potentially splicing-altering variants in poorly understood deep intronic regions (**Part 1**), evaluating the relevance of mRNA splicing experimental validation for variant classification (**Part 2**) and exploring the potential use of tumour information in the classification of germline variants (**Part 3**).

Deep intronic regions may contain variants that affect RNA splicing by introducing “pseudoexons” or other changes, which could potentially explain cases where the genetic cause of HBOC remains unknown. Current computational *in silico* tools to predict spliceogenic variants leading to pseudoexons have limited efficiency. In **Part 1** of this thesis, we have assessed the performance of the SpliceAI tool combined with ESRseq scores to identify spliceogenic deep intronic variants by affecting cryptic sites or splicing regulatory elements (SREs) using literature and experimental datasets. Our results with 233 published deep intronic variants showed that SpliceAI, with a 0.05 threshold, predicts spliceogenic deep intronic variants affecting cryptic

splice sites, but is less effective in detecting those affecting SREs. Next, we characterised the SREs profiles using ESRseq, showing that pseudoexons are significantly enriched in SRE-enhancers compared to adjacent intronic regions. This indicates that intronic regions with a high potential to be included as pseudoexons can be systematically identified throughout the HBOC genes. Although the combination of SpliceAI with ESRseq scores (considering Δ ESRseq and SRE landscape) showed higher sensitivity, the global performance did not improve because of the higher number of false positives. Both tools combined were tested in a tumour RNA dataset with 207 intronic variants disrupting splicing, showing a sensitivity of 86%. Following the pipeline, five spliceogenic deep intronic variants were experimentally identified from 33 variants in HBOC genes. Overall, our results provided a framework to detect deep intronic variants disrupting splicing.

A possible pathogenic mechanism for a subset of VUS is the disruption of normal mRNA splicing. However, experimental validation is generally restricted to variants affecting dinucleotide splice site positions. Consequently, variants outside these regions are not consistently evaluated across laboratories, leading to conflicting and uncertain interpretations. Additionally, variants leading to in-frame isoforms affecting clinically relevant domains are challenging to classify due to the limited knowledge about in-frame encoded proteins. One clear example of both issues are the variants located in the in-frame exon 18 (legacy number) of *BRCA1*. The results of a comprehensive study conducted within the ENIGMA consortium to identify spliceogenic variants mapping to this exon and characterize the severity of their impact are presented in **Part 2** of this thesis. Patient RNA, splicing reporter minigene, and mouse embryonic stem cell (mESC)-based functional assays were combined to determine the impact of 166

variants on splicing and estimate the tolerability functional range of exon 18 skipping ($\Delta 18$). Splicing, functional and clinical data were integrated for variant interpretation according to the recently ClinGen-approved *BRCA1/BRCA2*-specified ACMG/AMP (American College of Medical Genetics and Genomics/Association for Molecular-Pathology) classification guidelines.

Our results demonstrated that exon $\Delta 18$ is not a rescue event and that per-allele levels up to 34%, combined with at least 59% of the full-length (FL) isoform, are tolerable. Furthermore, we show that applying the recently developed *BRCA1/2* specifications significantly reduced the rate of VUS in this exon from 45.7% to 8.5%, being experimental mRNA testing essential in solving uncertainty for approximately 10% of variants

Lastly, despite the existence of multiple types of evidence for classifying germline variants, additional information is still required to classify many of them. Accordingly, we explored the potential use of tumour homologous recombination repair (HRR) status by RAD51 foci immunofluorescence detection (RAD51 test) as a predictor of pathogenicity or benignity for *BRCA1/2* and *PALB2* germline variants.

HRR status by RAD51 test was successfully assessed in a total of 148 primary untreated breast and ovarian tumours, from Vall d'Hebron Hospital, ENIGMA and EraPerMed RAD51predict collaborators. Samples were split into three cohorts depending on patients' germline status: pathogenic (91), benign (47) and VUS (8). Our results demonstrated that the RAD51 test could distinguish pathogenic from benign variants with high sensitivity (91.5%) and specificity (70.2%) (AUC 0.91).

Likelihood ratios (LR) were calculated for the association of RAD51 test outcomes (HRR deficiency-HRD or proficiency-HRP), with *BRCA1*, *BRCA2* and

PALB2 variant pathogenicity considering patients' clinical features. Then LR estimates were aligned to ACMG/AMP code strengths. Overall, RAD51-HRD status provided moderate evidence for pathogenicity if detected in primary breast tumours diagnosed before age 50, whereas RAD51-HRP provided informative moderate evidence against pathogenicity if detected in primary breast tumour independently of age at diagnosis. Lastly, we applied our recommendations to classify the eight VUS and about 38% (3/8) of variants were classified as likely benign, thus demonstrating the utility of the RAD51 test in variant classification.

INTRODUCTION

1. Hereditary Breast and Ovarian Cancer

1.1. Epidemiology

According to estimates from the International Agency for Research on Cancer (IARC) in 2022, there were close to 20 million new cases of cancer alongside 9.7 million deaths from cancer [1]. Female breast cancer (BC) ranked as the second most prevalent cancer worldwide, with approximately 2.3 million new cases, accounting for 11.6% of all cancer diagnoses (**Figure 1A**). Furthermore, it was the fourth leading cause of cancer-related deaths globally, representing 6.9% of all cancer deaths (**Figure 1A**). Amongst women, it was the most frequently diagnosed cancer, and the leading cause of cancer deaths globally (**Figure 1B**). Ovarian cancer (OC) was the eighth most frequent cancer in this population, with approximately 324,603 new cases, representing 3.4 % of all female cancer diagnoses. It is also ranked as the eighth leading cause of cancer-related deaths, with 206,956 deaths (4.8% of female cancer-related deaths) (**Figure 1B**).

While most cases are sporadic and related to risk factors such as age, diet, smoking, ionizing radiation and hormonal replacement therapy [2–9], approximately 5–10% of BCs and about 20% of OCs are hereditary [10,11] (**Figure 2**). This means that individuals inherit a genetic pathogenic variant that confers increased susceptibility to these cancer types. Additionally, about 15–20% of BC cases are familial, referring to affected women who have one or more first- or second-degree relatives with the disease but have no identifiable pathogenic variant in a gene known to cause a hereditary predisposition to cancer [12] (**Figure 2**). These cases may be caused by the interaction of low-penetrance genes, gene-environment interactions, or both [13].

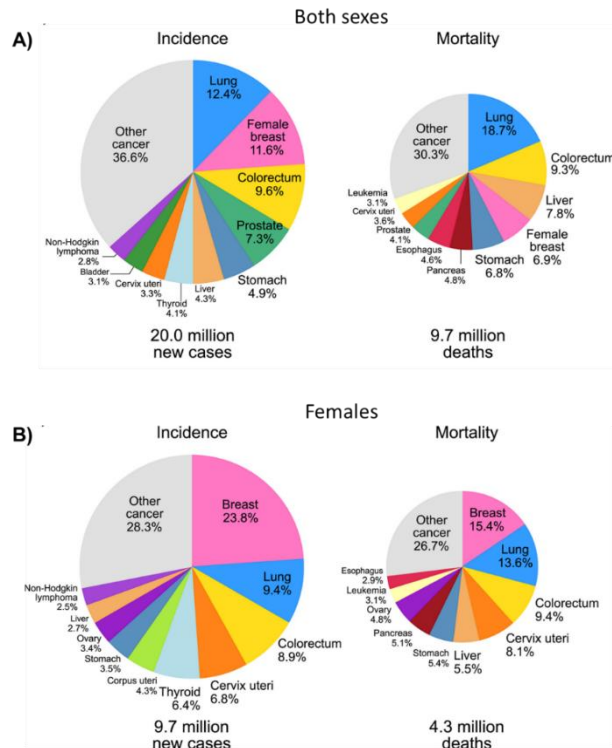


Figure 1. Distribution of worldwide new cases and deaths for cancer in 2022 for (A) both sexes and (B) females.

The area of the pie chart reflects the proportion of the total number of cases or deaths; *nonmelanoma skin cancers (excluding basal cell carcinoma)* are included in the other category. Adapted from Bray et al. (2024)[1].

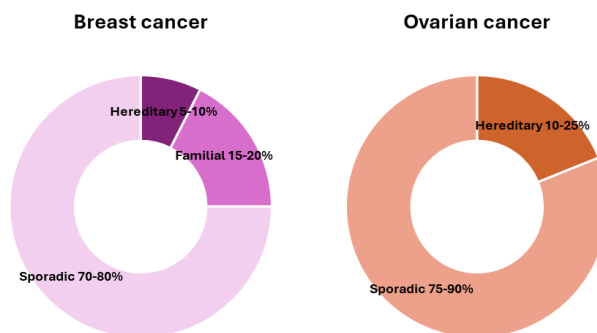


Figure 2. Aetiology of breast and ovarian cancer.

1.2. Breast Cancer Subtypes

Breast cancer comprises a heterogeneous group of tumours that displays marked variation in clinical presentation, morphology, molecular features, biological behaviour, and response to therapy.

At the morphological level, the World Health Organization (WHO) divided breast carcinomas into several subtypes [14]. Invasive carcinomas of no special type (NST) or not otherwise specified (NOS) (formerly invasive ductal carcinoma, IDC) are the most common (70-75% of cases), followed by invasive lobular carcinomas (ILC) (10-14%), and the other carcinomas of a special type [15]. The NST or NOS carcinomas do not fit into a specific histotype, whereas specific types are at least 90% pure (i.e. not mixed histology or <10% of another subtype). BCs are also classified according to the degree of differentiation into G1, G2 and G3 using the Bloom–Richardson–Scarff classification in the Elston–Ellis modification [16]. This involves the semiquantitative evaluation of three important biology-dependent morphological features: i) degree of tubule or gland formation, ii) nuclear pleomorphism, and iii) mitotic count.

Breast cancer stages are defined according to the anatomical extent of the primary tumour (T), the status of regional draining lymph nodes (N) and the presence or absence of distant metastasis (M) following the TNM system. The 7th edition of the American Joint Committee on Cancer (AJCC) staging system for breast cancer used nine stages (0, IA, IB, IIA, IIB, IIIA, IIIB, IIIC, and IV) based on different combinations of T, N, and M status [17].

Routinely determined elements of the pathomorphological examination are often insufficient to predict the clinical course of breast cancer, which makes it difficult to make appropriate therapeutic decisions. The diverse clinical

outcomes of cancers with similar morphological characteristics are attributable to different genetic profiles. The intrinsic molecular subtypes are based on mRNA gene expression levels described by Perou et al. (2000) [18]. In 2009, Parker et al. developed a 50-gene signature for subtype assignment (PAM50), that could reliably classify BC into the main intrinsic subtypes with 93% accuracy [19]. The intrinsic molecular subtypes include Luminal-A, Luminal-B, HER2-enriched, Claudin-low and Basal-like (**Figure 3**) [15]. Nonetheless, for prognostication and treatment decision making, breast tumours are grouped into surrogate intrinsic subtypes [20]. These, are based on easily accessible histological and immunohistochemical (IHC) markers such as the expression of oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and the proliferation marker Ki67 (**Figure 3**) [20–24]. Specifically, surrogate subtypes include:

- Luminal A-like: ER-positive (ER+) and/or PR-positive (PR+), HER2-negative (HER2-) and Ki-67-low (<14%).
- Luminal B-like:
 - ER+ and/or PR+, HER2- and Ki-67-high (Luminal B-like HER2-).
 - ER+ and/or PR+, HER2-positive (HER2+) (Luminal B-like HER2+).
- HER2-enriched (non-luminal): ER-negative (ER-), PR-negative (PR-) and HER2+.
- Triple-negative (TN): ER-, PR- and HER2-.

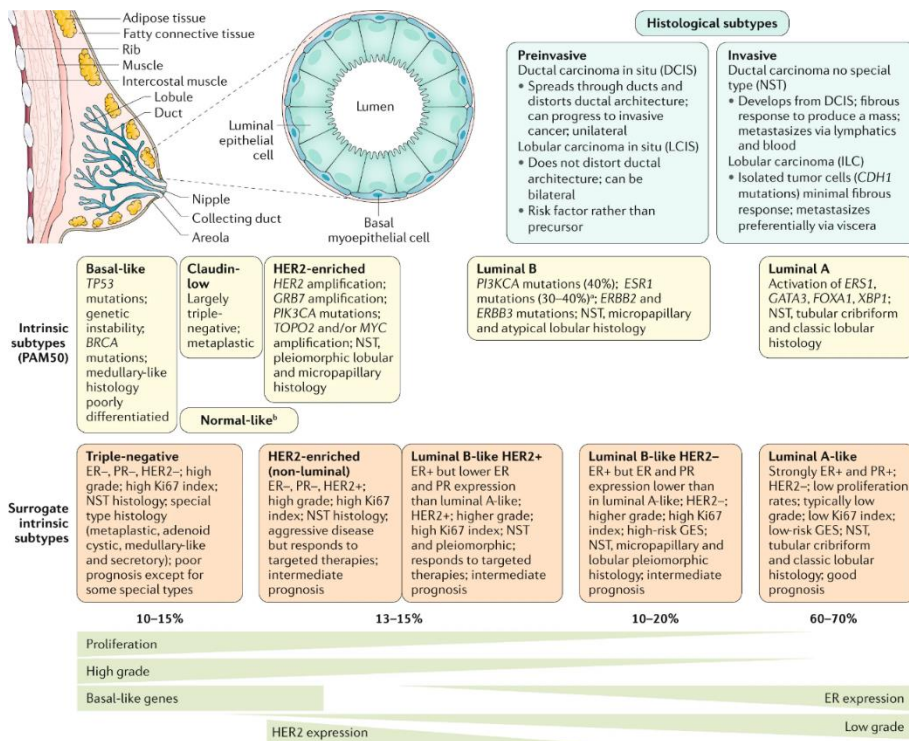


Figure 3. Breast cancer subtypes.

+: positive; -: negative; GES: gene expression signature; ^aESR1: mutations induced by aromatase inhibitor targeted therapy; ^b Artefact, expression of normal breast components due to low tumour cellularity; NST: ductal carcinoma no special type. Taken from Harbeck et al. (2019) [15].

1.3. Ovarian Cancer Subtypes

Epithelial ovarian cancer (EOC) accounts for about 95% of OC cases, while non epithelial tumours represent up to 5% of cases, and include germ cell cancer, sex cord-stromal cancer, small cell carcinoma and ovarian sarcoma [6,25].

EOC includes a heterogeneous group of tumours with different types of histology, grades and molecular and microenvironmental features, all of which contribute to treatment response and outcome. According to the WHO, EOCs are classified into five main subtypes: High-grade serous (HGS, 70% of cases), low-grade serous (LGS, 10%), endometrioid carcinoma (EC, 10%), clear cell carcinoma (CCC, 5%), and mucinous carcinoma (MC, 3%) [26].

Tumours are also classified according to the grade of differentiation of cells, as previously mentioned for BC (grades 1 to 3). However, OC tumour staging is based on the International Federation of Gynaecology and Obstetrics (FIGO) staging system (**Table 1**).

Table 1. 2014 FIGO staging system and corresponding TMN for ovarian cancer.

FIGO stage	Description	Corresponding TNM stage
I	Tumour confined to ovaries or fallopian tubes	T1
IA	Tumour limited to one ovary (with ovarian capsule intact) or fallopian tube; no tumour on ovarian or fallopian tube surface; no malignant cells in the ascites or peritoneal washings	T1a
IB	Tumour limited to both ovaries (with ovarian capsules intact) or fallopian tubes; no tumour on ovarian or fallopian tube surface; no malignant cells in the ascites or peritoneal washings	T1b
IC	Tumour limited to one or both ovaries or fallopian tubes, with any of the following C substages: IC1: surgical spill intraoperatively IC2: capsule ruptured before surgery or tumour on the ovarian or fallopian tube surface IC3: malignant cells in the ascites or peritoneal washings	T1c
II	Tumour involves one or both ovaries or the fallopian tubes with pelvic extension below the pelvic brim or primary peritoneal cancer	T2
IIA	Extension and/or implants of tumour on uterus and/or fallopian tubes and/or ovaries	T2a
IIB	Extension of tumour to other pelvic intraperitoneal tissues	T2b
III	Tumour involves one or both ovaries, or the fallopian tubes, or primary peritoneal cancer with cytologically or histologically confirmed spread to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes	T3
IIIA	Metastasis to the retroperitoneal lymph nodes with or without microscopic peritoneal involvement beyond the pelvis	T1, T2, T3aN1
	IIIA1: positive retroperitoneal lymph nodes only (pathologically proven)	
	IIIA1(i): metastasis up to 10 mm in greatest dimension	T3a/T3aN1
	IIIA1(ii): metastasis >10 mm in greatest dimension	
	IIIA2: microscopic extrapelvic (above the pelvic brim) peritoneal involvement with or without positive retroperitoneal lymph nodes	T3a/T3aN1
IIIB	Macroscopic peritoneal metastasis beyond the pelvis up to 2 cm in greatest dimension, with or without metastasis to the retroperitoneal lymph nodes	T3b/T3bN1
IIIC	Macroscopic peritoneal metastasis beyond the pelvis >2 cm in greatest dimension, with or without metastasis to the retroperitoneal lymph nodes (includes extension of tumour to capsule of liver and spleen without parenchymal involvement of either organ)	T3c/T3cN1
IV	Distant metastasis excluding peritoneal metastases	Any T, any N or M1
IVA	Pleural effusion with positive cytology	
IVB	Parenchymal metastases and metastases to extra-abdominal organs (including inguinal lymph nodes and lymph nodes outside of the abdominal cavity)	

FIGO: International Federation of Gynaecology and Obstetrics, TNM: TNM classification of malignant tumours. Adapted from Prat et al. (2015) [27].

1.4. Genetic Susceptibility to HBOC

The National Cancer Institute defines hereditary breast and ovarian cancer syndrome (HBOC) as an inherited disorder in which the risk of BC (especially before the age of 50-years) and OC is higher than normal. More specifically, the European Society for Medical Oncology (ESMO) notes that HBOC is clinically defined by family history criteria, and molecularly defined by the identification of germline pathogenic variants in clinically validated genes [28].

The genetic variants associated with HBOC susceptibility are divided into three groups depending on their frequency and conferred risk: i) high-penetrance variants are very rare in the population (with a minor allele frequency of <0.005) and are associated with a very high risk of breast cancer (relative risk of carriers versus non-carriers of 5 to >20), ii) moderate-penetrance variants are rare (with a minor allele frequency of 0.005 – 0.01) and confer a relative risk of 2–4, and iii) low-penetrance variants are common in the population (minor allele frequency >0.05) and conferred risk of breast cancer of less than 1.5-fold (**Figure 4**) [29].

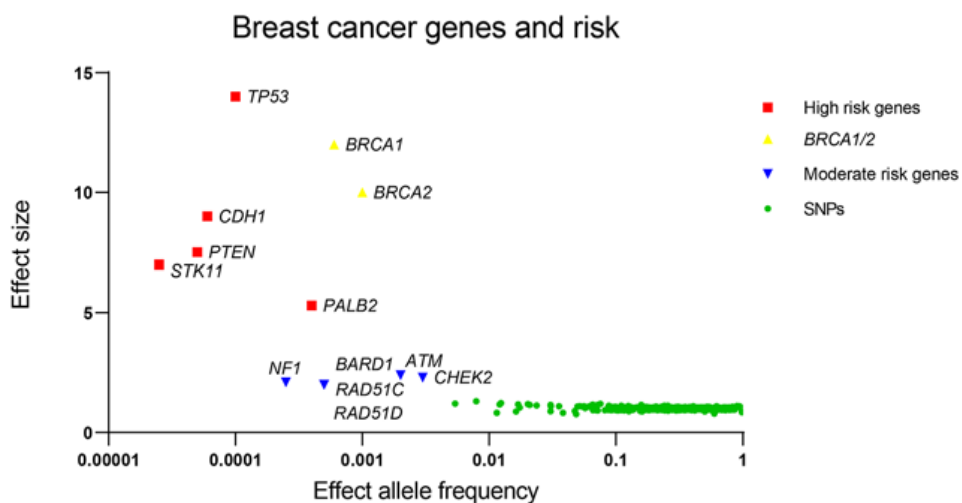


Figure 4. Relative risk of BC and allele frequencies of high-, moderate-, and low-risk genetic variants. Taken from Woodward et al. (2021) [30].

The first report of familial predisposition to breast cancer dates back to the 19th century and was made by the physician Pierre Paul Broca, who described a familial cluster of breast cancer cases affecting multiple women across four generations (**Figure 5**) [31].

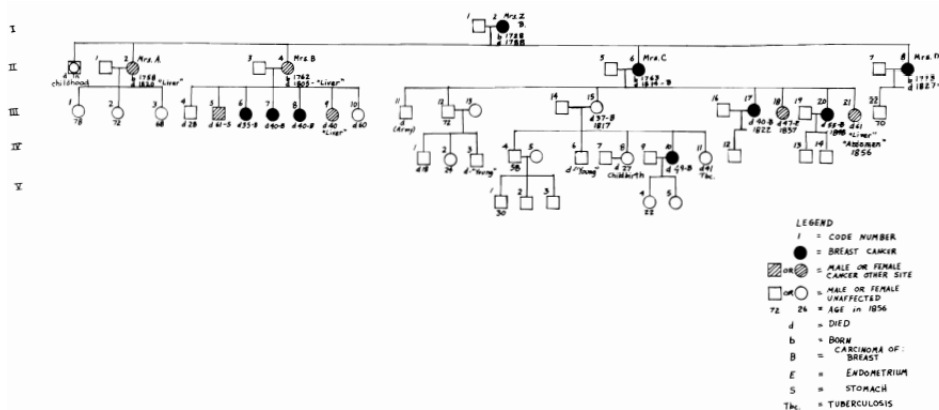


Figure 5. Pedigree of the family of Mrs. Z described by Dr. Broca in the “Traité des tumeurs” in 1866.
Taken from Broca. (1866) [31].

However, it was not until the 1990s that linkage analyses in multiple-cases families allowed the identification of *BRCA1* and *BRCA2* as high-risk genes for HBOC [32–34]. Both are tumour suppressor genes which encode proteins primarily involved in the homologous recombination repair (HRR) pathway (**Figure 6**) [35]. It has been estimated that women carrying loss of function (LoF) variants in *BRCA1* have a cumulative risk by 80 years of age of developing BC of 55-72%, and 44% of OC [36,37]. Likewise, *BRCA2* carriers have a cumulative risk of 44-69% of developing BC and 17% of OC [36,37].

A recently published weighted meta-analysis carried out for three population-base case-control studies, revealed that the prevalence of pathogenic variants in unaffected individuals (i.e. controls, or general population) is 0.09% (1 in

1,1134 individuals) for *BRCA1* and 0.25% (1 in 400 individuals) for *BRCA2* [38]. Although the prevalence of founder mutations in some ethnic groups such as Ashkenazi Jews can be as high as 2.5% (1 in 40 individuals) [39].

Approximately 20% of HBOC cases can be ascribed to LoF variants in these two genes [40], but a small proportion of HBOC is explained by high-penetrance variants in tumour suppressor genes associated with other rare pleiotropic tumour syndromes. These suppressor genes include *TP53* (Li-Fraumeni syndrome), *PTEN* (Cowden syndrome), *STK11* (Peutz-Jeghers syndrome) and *CDH1* (Hereditary diffuse gastric cancer syndrome). The cumulative risk of developing BC at age 60 for women with Li-Fraumeni syndrome is 85% [41]. For individuals with *PTEN* pathogenic variants, the risk of developing BC throughout life is in the range of 67-85% [42–44]. In *STK11* pathogenic variant carriers, the cumulative risk of female breast and gynaecological cancers by the age of 60 years is 32% and 13%, respectively [45]. The cumulative risk for BC by 80 years is 42% for women with *CDH1* loss of function variant [46]. Pathogenic mutations in these genes are extremely rare (prevalence ~1 in 50-10,000 for each) [47], and represent less than 5% of HBOC cases [39,48] (**Figure 4**).

In the last decades, HBOC susceptibility gene list has been widely increased thanks to segregation analyses together with case-control studies investigating candidate genes which shared common cellular pathways with *BRCA1* and *BRCA2*. Nearly all known susceptibility genes for HBOC and other hereditary cancer syndromes encode for tumour suppressor proteins that participate in genome stability pathways such as HRR, DNA damage checkpoint control, DNA mismatch repair and stabilization of replication fork (**Figure 6 A-D**).

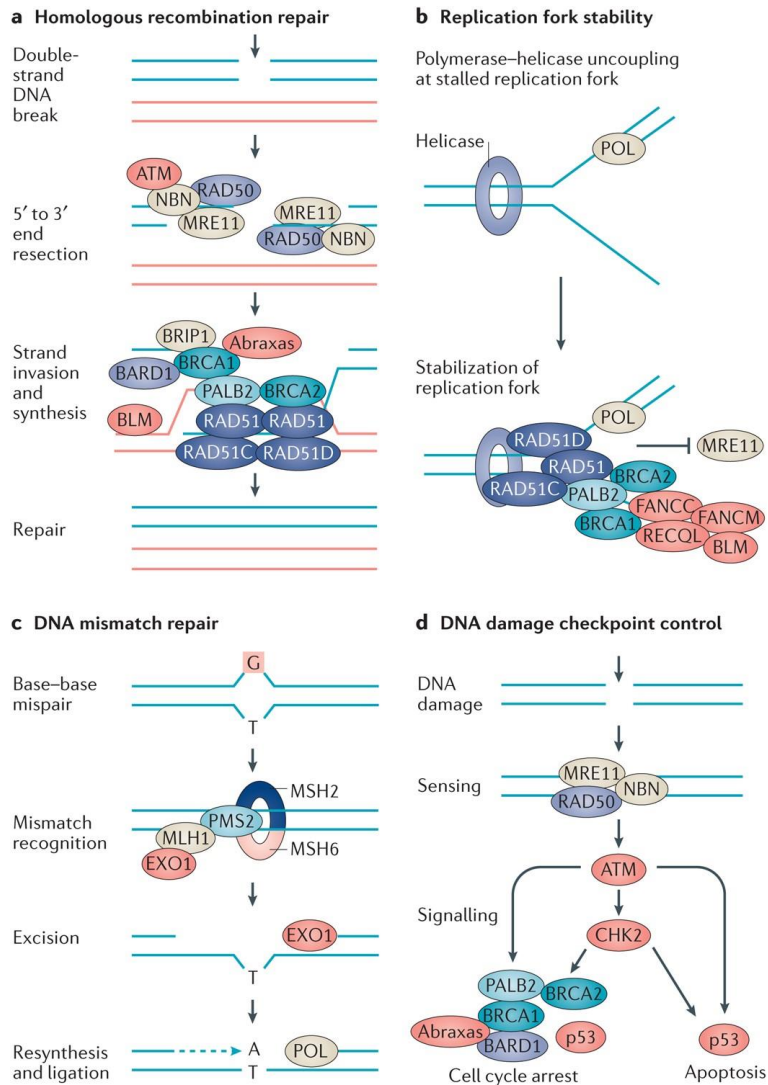


Figure 6. Role of HBOC and other hereditary cancer syndromes susceptibility genes in genome stability pathways.

Taken from Nielsen et al. (2016) [10].

PALB2 encodes for a key protein in HRR, as it acts as a bridge between BRCA1 and BRCA2 for the recruitment of RAD51, one of the final effectors of this pathway (**Figure 6 A**). It was originally identified as a moderate-penetrance BC susceptibility gene risk [49]. However, subsequent assessment of risk in families

confirmed high risk for BC, with an OR of 7.18 (95% CI 5.82–8.85; $p = 6.5 \times 10^{-76}$), and a lifetime risk of around 50% [50].

Case-control and segregation analyses also identified LoF variants in *ATM*, *CHEK2*, *BRIP1*, *RAD51C*, *RAD51D* and *BARD1* as moderate risk of developing BC and/or OC [51–58]. All these genes encode for BRCA1/2-interacting proteins, and/or are involved in DNA damage checkpoint control (**Figure 6 A, B and D**).

Loss of function variants in DNA mismatch-repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) (**Figure 6 C**) are primarily associated with Hereditary non-polyposis colorectal cancer (HNPCC, or Lynch syndrome), but contribute to moderate risk of OC as women with HNPCC have a 12-15% lifetime risk of developing OC [59].

Deleterious variants in *NF1* are associated with Neurofibromatosis type 1, another pleiotropic tumour syndrome. It has also been reported that *NF1* carriers have an increased risk of BC and OC despite this gene is not involved in genome stability maintenance [60,61].

Most of these genes have been included in HBOC susceptibility test panels, but the clinical validity classification under the Clinical Genome Resource (ClinGen) framework revealed a definitive clinical validity for *BRCA1*, *BRCA2*, *ATM*, *PALB2*, *CHEK2*, *BARD1*, *CDH1*, *TP53*, *PTEN*, *STK11* (**Figure 7 A**) for BC and *BRCA1*, *BRCA2*, *RAD51C*, *BRIP1*, *RAD51D*, *MLH1*, *MSH2*, *MSH6*, *EPCAM* and *PMS2* for OC (**Figure 7 B**) [62].

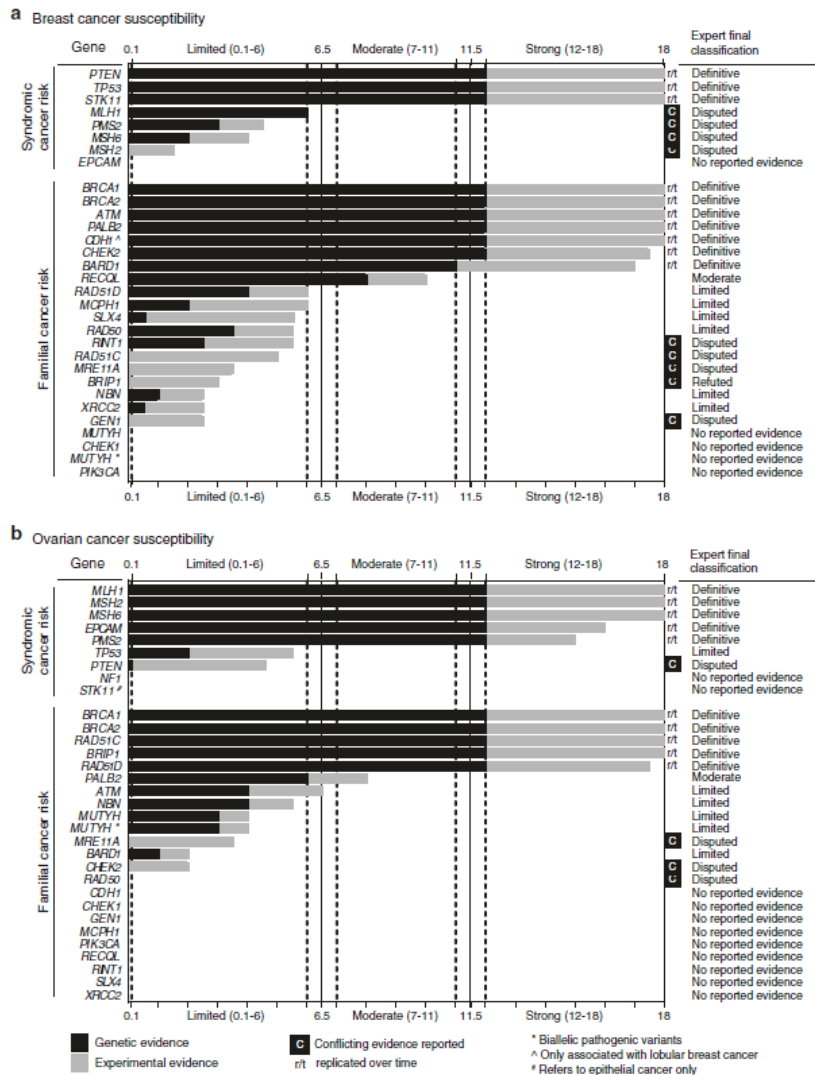


Figure 7. Clinical validity classifications of candidate susceptibility genes associated with syndromic or familial BC and OC.

A) Clinical validity classifications of 31 gene-disease pairs associated with BC susceptibility. **B)** Clinical validity classifications of 32 gene-disease pairs associated with OC susceptibility. Consensus genetic and experimental evidence scores are depicted for each gene-disease relationship wherein a limited preliminary classification scored 0.1–6 total points, a moderate preliminary classification had 7–11 total points, a strong preliminary classification scored 12–18 total points, and a definitive classification scored 12–18 total points and achieved replication over time (indicated by r/t). *Taken from Lee et al. (2019) [62].*

Additionally, two recent large case-control studies, BRIDGES [37] and CARRIERS [63], sequenced putative BC susceptibility genes in approximately 180,000

women including unselected BC cases and controls, with the final aim to define the genes that are most clinically useful for inclusion on testing panels for the prediction of BC risk and provide estimates of the risks associated protein-truncating variants to guide genetic counselling.

In the BRIDGES study protein-truncating variants in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, and *PALB2* were associated with a significant risk of breast cancer overall ($P < 0.0001$), while there was more modest evidence of an association for protein-truncating variants in *BARD1*, *RAD51C*, *RAD51D*, and *TP53* ($P < 0.05$) (**Table 2**) [37]. Moreover, truncating variants in most of these genes showed subtype-specific associations. *BARD1*, *BRCA1*, *BRCA2*, *PALB2*, *RAD51C*, and *RAD51D* truncating variants were more strongly associated with ER-negative tumours, whereas *ATM* and *CHEK2* truncating variants were more strongly associated with ER-positive disease [37].

Table 2. Risk of BC overall associated with protein-truncating variants in susceptibility genes from the population-based case-control BRIDGES study.

Gene	BRIDGES (Population-based studies: 48,2826 cases and 50,703 controls)	
	Odds Ratio (95%CI)	P value
ATM	2.10(1.71–2.57)	9.2×10 ^{−13}
BARD1	2.09(1.35–3.23)	0.00098
BRCA1	10.57(8.02–13.93)	1.1×10 ^{−62}
BRCA2	5.85(4.85–7.06)	2.2×10 ^{−75}
BRIP1	1.11(0.80–1.53)	0.54
CDH1	0.86(0.37–1.98)	0.72
CHEK2	2.54(2.21–2.91)	3.1×10 ^{−39}
<i>c.1100delC</i>	2.66(2.27–3.11)	1.1×10 ^{−33}
<i>Other</i>	2.13(1.60–2.84)	3.0×10 ^{−7}
FANCM	1.06 (0.90–1.26)	0.48
MLH1	0.58(0.19–1.77)	0.34
MSH2	1.06(0.47–2.36)	0.89
MSH6	1.96(1.15–3.33)	0.013
NBN	0.90(0.67–1.20)	0.48
NF1	1.76(0.96–3.21)	0.068
PALB2	5.02(3.73–6.76)	1.6×10 ^{−26}
PMS2	1.16(0.73–1.85)	0.53
PTEN	2.25(0.85–6.00)	0.10
RAD51C	1.93(1.20–3.11)	0.0070
RAD51D	1.80(1.11–2.93)	0.018
STK11	1.60(0.48–5.28)	0.44
TP53	3.06(0.63–14.91)	0.17

CI: confidence interval. Adapted from Dorling et al. (2021) [37]

The estimated absolute risks by 80 years of age place protein-truncating variants in *BRCA1*, *BRCA2*, and *PALB2* (>30%) in the high-risk category and *ATM*, *BARD1*, *CHEK2*, *RAD51C*, and *RAD51D* in the moderate-risk category (17-30%) (**Figure 8**)

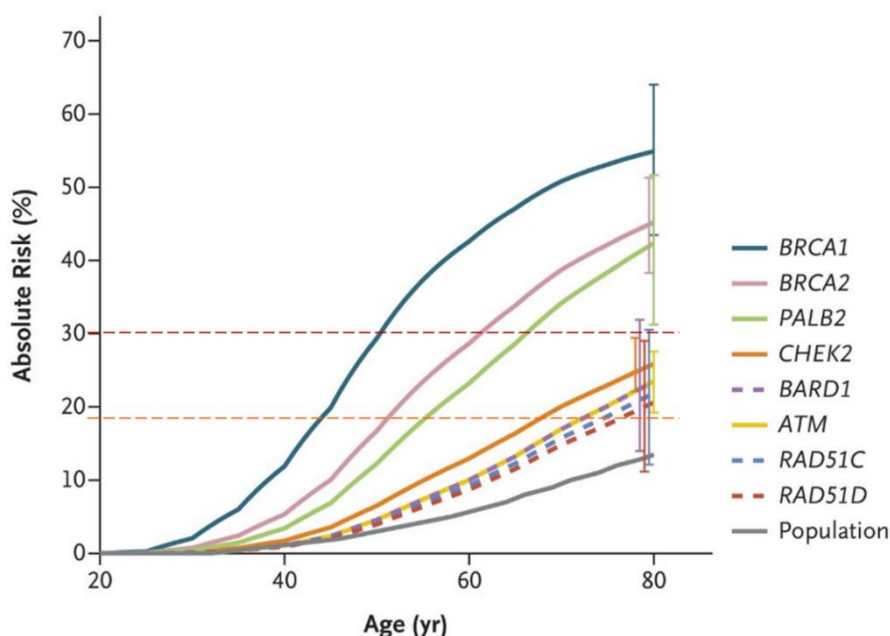


Figure 8. Estimated absolute risk of BC associated with protein-truncating variants in eight genes compared with the general population risk.

Dashed red and orange lines indicate the threshold for high (30%) and moderate (17-30%) risk according to National Institute for Health and Care Excellence guidelines[64]. Adapted from Dorling et al. (2021) [37].

Similar findings were reported in the CARRIERS study, as in population-based analyses, pathogenic variants in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, and *PALB2* were associated with increased risks of BC. Notably, pathogenic variants in *BARD1*, *RAD51C* and *RAD51D* were associated with increased risks of ER-negative and TN BC but no with ER-positive BC. In contrast, pathogenic *ATM* and *CHEK2* variants were associated exclusively with ER-positive BC. Furthermore, pathogenic variants in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, and *PALB2* were associated with a lifetime absolute risk of BC greater than 20% by age 85 years [63].

These large studies permitted risk estimates much more precise than those obtained previously, stratifying genes between high and moderate risk, and discarding other candidates that did not show association with the disease (e.g.,

NBN)[63]. Nevertheless, the evidence of association with BC for some genes remains equivocal despite the size of these studies (e.g., *FANCM*, *MSH6* and *NF1*)[37].

Lastly, genome-wide association studies (GWAS) have identified more than 300 single nucleotide polymorphisms (SNPs), that individually have a small effect on risk, but combined show substantial risk estimates, expressed as polygenic risk scores (PRS) [30]. It has been estimated that these SNPs currently explain 28% of the familial risk of BC. Furthermore, it has been shown that SNPs profile can affect the absolute risk of BC in carriers of pathogenic variants in high or moderate risk genes [65–67].

Overall, genetic landscape of HBOC is composed of pathogenic variants associated with high and moderate risk, as well as common SNPs. However, around half of HBOC cases remain genetically undiagnosed, it being very likely that a significant part of the missing heritability is due to the presence of undiscovered low-penetrance genetic modifiers [30].

1.5. Patients' Clinical Management

The detection of pathogenic variants in HBOC genes can benefit carriers by informing risk-reducing strategies and guiding personalized medical care. Furthermore, relatives may also benefit from this information, as they could opt for similar risk-reducing treatments if they are found to carry the same pathogenic variant.

1.5.1. Selection criteria for germline testing

Although several European guidelines concerning recommendations for *BRCA1/2* testing have been published, all clinical criteria for testing are based on risk factors such as age, hormone receptor status, ancestry with founder

mutations, and personal and familial history of cancer. The selection criteria from the Sociedad Española de Oncología Médica (SEOM) are detailed in **Table 3** [68].

Table 3. SEOM's selection criteria for germline HBOC testing.

Regardless of family history:
• Women with synchronous or metachronous breast and ovarian cancer • Breast cancer ≤ 40 years • Bilateral breast cancer (the first diagnosed ≤ 50 years) • Triple-negative breast cancer ≤ 60 years • High-grade epithelial non-mucinous ovarian cancer (or fallopian tube or primary peritoneal cancer) • Ancestry with founder mutations • <i>BRCA1/2</i> somatic mutation detected in any tumour type with an allele frequency $> 30\%$ (if it is known) • Metastatic HER2-negative breast cancer patients eligible to consider PARP inhibitor therapy
Two or more first-degree relatives with any combination of the following high-risk features:
• Bilateral breast cancer + another breast cancer < 60 years • Breast cancer < 50 years and prostate or pancreatic cancer < 60 years • Male breast cancer • Breast and ovarian cancer • Two cases of breast cancer diagnosed before age 50 years
Three or more direct relatives with breast cancer (at least one premenopausal) and/or ovarian cancer and/or, pancreatic cancer or high Gleason (≥ 7) prostate cancer

SEOM: Sociedad Española de Oncología Médica, HBOC: Hereditary breast and ovarian cancer, PARP: Poly (ADP-ribose) polymerase. Adapted from González-Santiago et al. (2020) [68].

The application of these criteria is associated with a minimum pathogenic variant detection rate of 10%. However, the sensitivity to identify pathogenic mutations in other high or moderate-risk genes beyond *BRCA1/2* genes is limited, as guidelines are based predominantly on the probability of carrying pathogenic mutations in these two genes. Recent studies have shown that the traditional approach based on clinical criteria may miss up to 50% of pathogenic variant carriers, supporting *BRCA1/2* testing in a broader range of individuals, if not in every breast cancer patient [69].

The SEOM recommends genetic counselling before and after genetic testing to guarantee the discussion of benefits and limitations of testing, provide risk estimates of developing cancer, recommendations for early detection and prophylactic interventions, advice regarding reproductive options and offer psychological support [68].

1.5.2. Genetic testing in HBOC

Germline genetic testing is based on massively parallel sequencing of panels which include high and moderate HBOC risk genes. Although there are recommendations at national and international levels as to which genes should be analysed in the context of HBOC diagnosis, the individual composition of gene panels varies greatly and there are still major differences in the diagnostic approach [37,62] and consequently in the diagnostic yield [70]. Despite these differences in panel strategies, studies have shown that the usage of multi-gene panels increases the likelihood of detecting cancer-predisposing variants and offers advantages in terms of time and cost compared with single-gene testing [71,72].

Clinical validity for *BRCA1/2* and *PALB2* (in BC/OC), and *BRIP1*, *RAD51C*, *RAD51D*, *MLH1*, *MSH2*, and *MSH6* (in OC) has been established with subsequent surveillance and preventive clinical options. Specifically in Catalonia, the clinical recommendations from the Servei Català de la Salut for multi-gene panel testing in HBOC include all above-mentioned genes, and additionally *BARD1*, as well as *TP53* and *CDH1* in case of suggestive phenotype [73].

Genetic testing should include sequence and fragment analyses since reported pathogenic variants mainly consist of small deletions/insertions, nonsense

variants, splice variants and, less frequent, large rearrangement alterations (>10%) [68].

Clinical guidelines are being developed to provide recommendations to prompt germline testing after identifying pathogenic variants in tumour sequencing [74,75], a strategy recommended for poly (ADP-ribose) polymerase (PARP) inhibitor (PARPi) treatment decision-making in OC. Specifically, germline analysis for *BRCA1*, *BRCA2*, *PALB2*, *BRIP1*, *RAD51C*, and *RAD51D* is recommended following the identification of a pathogenic single nucleotide variant (SNV) with > 30% variant allele frequency (VAF) or >20% for pathogenic small insertions/deletions, in any of these genes in any tumour at any age [68].

1.5.3. Early detection, risk-reducing and therapeutic strategies

Patients carrying pathogenic variants in *BRCA1* and *BRCA2* have well-defined risk profiles and guidelines for clinical management, as screening options for early detection including mammography, magnetic resonance imaging and ultrasonography; options for risk reduction, including chemoprevention, risk-reducing oophorectomy and/or mastectomy) [76] as well as therapeutic options [77]. Specifically, clinical management of carriers of pathogenic variants in HBOC risk genes include:

- **Strategies for early detection of cancer:** in the presence of a *BRCA1/2* or *PALB2* pathogenic variant, intensified screening by breast magnetic resonance imaging (MRI) should start at age 30 or five years younger than the youngest family member with BC [28]. Annual screening intervals are recommended, excepts for *BRCA1* where 6-monthly screening should be considered [28]. However, in most cases

biannual MRI is not available; and annual MRI may be supplemented by ultrasound or mammography depending on age [28]. Regarding OC screening in *BRCA1/2* carriers, transvaginal ultrasound and serum cancer antigen 125 determination may be considered 6-monthly until the age of risk reduction salpingo-oophorectomy (RRSO) or for those who have not elected RSSO [68]. Deleterious variants in HBOC genes have also been associated with an increased risk of prostate cancer. Evidence for a significantly elevated risk is more robust for *BRCA2* [78,79], with a moderately increased risk observed for *ATM* [80]. Hence, annual blood Prostate-Specific Antigen (PSA) screening should be offered to male *BRCA2* carriers from age of 40 years and may also be considered for male *ATM* carriers from age 40 years [28].

- **Risk-reducing surgeries:** bilateral risk reduction mastectomy (BRRM) is recommended for healthy *BRCA1/2* pathogenic variant carriers since it decreases the occurrence of BC by 90% without a decrease in all-cause mortality [81]. *BRCA1/2* carriers with a history of unilateral breast cancer have a higher risk of contralateral BC than non-carriers. Therefore, contralateral risk-reducing mastectomy (CRRM) should be considered as it is associated with improved overall survival [82]. Risk reduction salpingo-oophorectomy is recommended for women who carry a *BRCA1/2* pathogenic variant aged 35-45, as several studies demonstrated that it reduces the risk of OC, fallopian tube cancer, and primary peritoneal cancer by approximately 80% [83]. For *BRIP1* carriers, RRSO is recommended from 45 to 50 years. The same recommendations apply to *RAD51C/D* carriers unless family history suggests an earlier risk of developing

OC. Lastly, RRSO is also an option for Lynch syndrome patients given the cumulative lifetime risks of endometrial and ovarian cancer are 60% and 24%, respectively [68].

- **Pharmacological prevention (chemoprevention):** tamoxifen can be used as a preventive treatment for female *BRCA1/2*-mutated carriers who do not want to undergo BRRM [68,84] or with a previous diagnosis of BC, since adjuvant tamoxifen demonstrated to reduce the risk of contralateral BC in *BRCA1/2* carriers [85]. Similarly, oral contraceptives can reduce the risk of OC by ~50%, although there is a possibility of an increased risk of BC and their use in this population is controversial [86].
- **Therapeutic strategies:** As previously mentioned, *BRCA1* and *BRCA2* are primarily involved in HRR (**Figure 6 A**), a crucial pathway to repair DNA double-strand breaks (DSBs) to maintain genome integrity and stability [87]. Loss of function variants on both *BRCA1/2* lead to HRR deficiency (HRD) [35] which is a perfect target for therapies with DNA crosslinkers, such as platinum-based drugs, or other agents, such as PARPi [88]. Platinum salts, such as carboplatin, might be considered in neoadjuvant and metastatic TN BC setting among *BRCA1/2*-mutated patients, since it has been proved that platinum-based neoadjuvant chemotherapy increased pathological complete response [89], and carboplatin has shown a clinical benefit compared to docetaxel among *BRCA1/2* pathogenic variant carriers in advanced TN BC [90]. The sensitivity of HRD tumours to PARPi is explained through the concept of synthetic lethality: PARP inhibition causes an increase in DNA single-strand breaks, which, during replication, are

converted to irreparable DSBs in *BRCA1/2* defective cells, leading to cell death due to increased genomic instability (**Figure 9**) [91,92]. Currently, PARPi treatment is indicated for *BRCA*-mutated patients with OC (germline/somatic pathogenic variant and/or homologous recombination deficiency), HER2- BC (germline), metastatic prostate (germline/somatic) and metastatic pancreatic cancers (germline) [93–97].

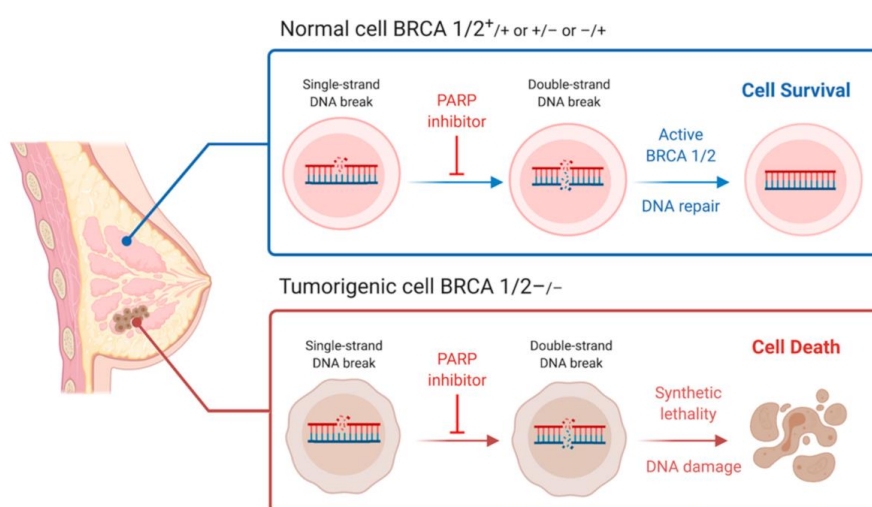


Figure 9. PARP inhibitor's function in *BRCA1/2* deficient cells is based on synthetic lethality.

In the normal cell with physiologic *BRCA1/2* (above, blue), PARP inhibition will cause SSBs to progress to DSBs. However, functional *BRCA1/2* will have the capability to repair resulting DSB damage and maintain cell viability. In the tumorigenic cells (below, red), PARP inhibition will cause SSBs to progress to DSBs, but deficient *BRCA1/2* allow the accumulation of DSBs ultimately leading to cell death. Taken from Neiger et al. (2021) [92].

2. Variant Classification

Genetic testing can yield three possible outcomes: i) the detection of a well-known LoF variant associated with increased cancer risk; ii) the detection of only neutral variants, which are not reported by laboratories; and iii) an inconclusive result, due to the identification of a variant of unknown significance (VUS). These

variants have an unknown effect on gene function and unknown associated cancer risk.

While patients with an identified pathogenic variant benefit from options available for early cancer detection, preventive measures and targeted therapies (detailed in section **1.5.3**), VUSs carriers are generally managed without regard to the result of genetic testing, due to its uncertainty, thus limiting the options for these patients.

2.1. Variants of Unknown Significance

The frequency of VUS detection can be significant. A study of multigene panels, exomes, and genomes conducted in 19 laboratories revealed that 32% of individuals received an inconclusive report due to the detection of one or more VUSs in the absence of any clear explanation for disease [98]. In another study, the analysis of a routine panel for HBOC showed a VUS detection rate of ~40% [70].

Concerning *BRCA1* and *BRCA2* genes, about 44% of the total variants reported on ClinVar were classified as VUS or showed conflicting interpretations of pathogenicity (**Figure 10**).

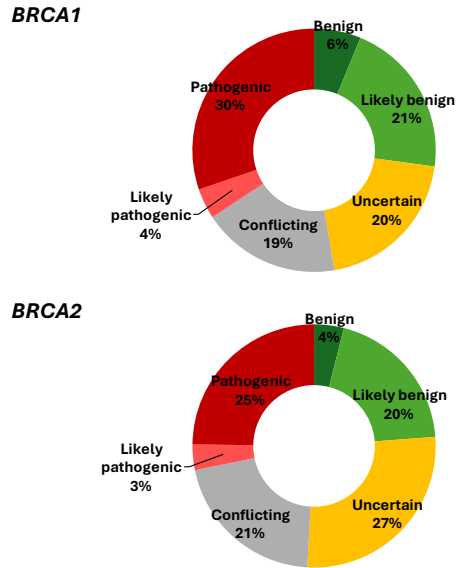


Figure 10. Proportion of *BRCA1* and *BRCA2* variants reported in ClinVar database according to their clinical significance.

Data was obtained from the total number of unique variants with a classification reported in ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>, Accessed 12th, August 2024) (*BRCA1* n=10,707; *BRCA2* n=18,777). Standard terms for clinical significance (Pathogenic, Likely pathogenic, Uncertain, Likely benign and Benign) are taken from the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines [99]. “Conflicting interpretations” refer to those reported in more than one class.

The inability to classify VUS primarily stems from the nature of the variants themselves, as most are in-frame deletion/insertions, missense or synonymous mutations, or alterations in intronic and regulatory regions. Thus, the impact of these alterations generally cannot be inferred from previous knowledge and must be verified experimentally [100]. VUS classification is also hampered by a lack of sufficient population-based statistical evidence, as many variants are rare, requiring massive and widespread data-sharing across large populations [101]. Additionally, divergent interpretations among different laboratories and clinicians, due to varying standards and methodologies for variant classification, further complicate VUS interpretation [102].

Given that uncertainty represents a major challenge; several efforts have been made to develop classification guides to unequivocally classify variants as neutral or pathogenic for appropriate patients' clinical management.

2.2. Variant Classification Guidelines

To reduce conflicting interpretations in describing the clinical significance of a VUS, it's essential that variant classification is based on a common set of rules among all clinical laboratories.

In 1998, Cotton et al. set the first tempt of classification guide delineating several types of evidence that might aid in understanding a missense variant's role in disease causation, including segregation analyses, nature of the amino acid substitution, and functional assays [103]. Later, the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) published standard recommendations for sequence variant interpretation based on the type of variant identified and whether the variant had been previously reported as a "recognized" cause of disease and highlighted the utility of follow-up family testing and functional studies to identify additional helpful evidence toward classification [104,105]. However, it was not defined what degree of certainty required to classify a variant as clinically significant, how much weight should be assigned to different types of evidence, or how to combine different pieces of evidence to get a classification. It was not until 2004 when Goldgar et al. proposed the first quantitative model for combining evidence, which incorporated several clinical data points together with conservation and functional data to arrive at likelihood ratios (LRs) that authors deemed high enough to support (1000:1) or refute (100:1) disease-causality for variants in *BRCA1/2* (**Figure 11**) [106].

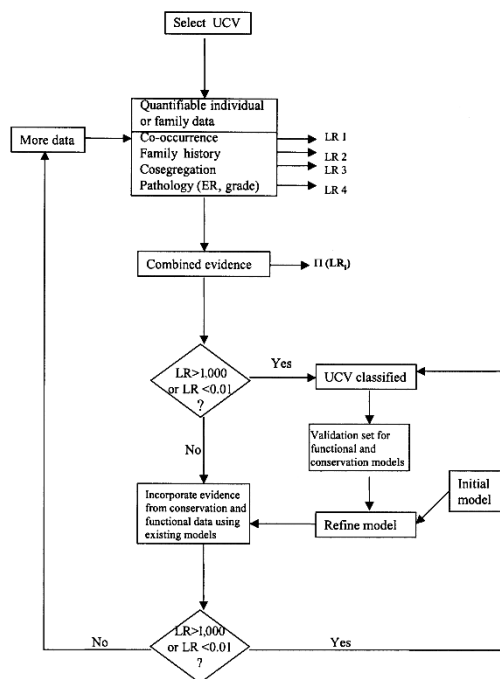


Figure 11. Flowchart of the procedure for classification of sequence variants of unknown clinical significance described by Goldgar et al. (2004) [106].

UCV: uncertain variant, ER: oestrogen receptor status, LR: likelihood ratio.

The model was revised to incorporate likelihoods from additional data sources and refine information originally included [107–114]. Further, LR estimates derived from a large dataset [108] were used to estimate the prior probability of pathogenicity of a variant based on *in silico* properties [115]. This updated multifactorial model allowed for a given variant to calculate the posterior probability of being pathogenic by combining its prior probability with all other independent lines of evidence. Based upon the numerical value of the posterior probability calculations, the International Agency for Research on Cancer (IARC) Working Group on VUS introduced a clinical translation step, creating a five-tier classification system [116]. It comprised class-1 variants (not pathogenic), class

2 (likely not-pathogenic), class 3 (uncertain), class 4 (likely pathogenic) and class 5 (pathogenic), with proposed clinical recommendations for each class (**Table 4**) [116].

Table 4. IARC variant classification system and corresponding clinical recommendations.

Class	Definition	Posterior	Clinical Testing	Surveillance recommendations
		Probability		
5	Pathogenic	>0.99	Test at-risk relatives for the variant	Full high-risk surveillance
4	Likely pathogenic	0.95-0.99	Test at-risk relatives for the variant	Full high-risk surveillance
3	Uncertain	0.05-0.949	Do not use as predictive testing in at-risk relatives	Counsel based on family history and other risk factors
2	Likely not pathogenic	0.001-0.049	Do not use as predictive testing in at-risk relatives	Counsel as if no mutation detected
1	Not Pathogenic	<0.001	Do not use as predictive testing in at-risk relatives	Counsel as if no mutation detected

Adapted from Plon et al. (2008) [116].

Multifactorial models need enough occurrences for a given variant in unrelated families, but VUS are individually rare detected in national or regional laboratories and therefore family-based clinical analyses lack statistical power. To address this limitation, the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA, <https://enigmaconsortium.org/>) consortium was created in 2009. The overall aim of this consortium is to facilitate variant classification by pooling genetic, clinical, histopathological, functional and splicing data from a network of laboratories (further information in sections **2.3** and **2.4**)[117].

In parallel to these efforts, there was increasing international uptake of variant classification guidelines by the ACMG/AMP, with applicability to any Mendelian disease. These guidelines, released in 2015, consist of a system that weights 28 independent lines of evidence (e.g. population data, computational data, functional data, segregation data, etc) for and against pathogenicity as “stand-alone (A)”, “very strong (VS)”, “strong (S)”, “moderate (M)” and “supporting (P)” (Figure 12) [99].

		Benign		Pathogenic			
		Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population Data	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>				Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>	
Functional Data	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>		
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →			
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>		
Allelic Data		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>			
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>				
Other Data		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>				

Figure 12. The ACMG/AMP evidence framework.

This chart organizes each of the criteria by the type of evidence as well as the strength of the criteria for a benign (left side) or pathogenic (right side) assertion. *BS*: benign strong, *BP*: benign supporting, *FH*: family history, *LOF*: loss-of-function, *MAF*: minor allele frequency, *PM*: pathogenic moderate, *PP*: pathogenic supporting, *PS*: pathogenic strong, *PVS*: pathogenic very strong. Taken from Richards et al. (2015) [99].

For a given variant, all criteria applicable based on all available evidence must be combined according to a qualitative scoring rules system to choose a classification from the five-tier system (likely pathogenic, pathogenic, uncertain, likely benign and benign), as detailed in **Table 5**.

Table 5. ACMG/AMP rules for combining criteria to classify sequence variants.

Classification	Combining rules
Pathogenic	(i) 1 PVS AND (a) ≥ 1 PS OR (b) ≥ 2 PM OR (c) 1 PM and 1 PP OR (d) ≥ 2 PP (ii) ≥ 2 PS OR (iii) 1 PS AND (a) ≥ 3 PM OR (b) 2 PM AND ≥ 2 PP OR (c) 1 PM AND ≥ 4 PP
Likely pathogenic	(i) 1 PVS AND 1 PM OR (ii) 1 PS AND 1–2 PM OR (iii) 1 PS AND ≥ 2 PP OR (iv) ≥ 3 PM OR (v) 2 PM AND ≥ 2 PP OR (vi) 1 PM AND ≥ 4 PP
Benign	(i) 1 BA OR (ii) ≥ 2 BS
Likely benign	(i) 1 BS and 1 BP OR (ii) ≥ 2 BP
Uncertain	(i) Other criteria shown above are not met OR (ii) The criteria for benign and pathogenic are contradictory

PVS: pathogenic very strong, PS: pathogenic strong, PM: pathogenic moderate, PP: pathogenic supporting, BA: benign stand-alone, BS: benign strong, BP: benign supporting.

Later, the Sequence Variant Interpretation (SVI) Working Group modelled this qualitative classification system into a Bayesian framework whereby the relative odds of pathogenicity for supporting, moderate, strong, and very strong pathogenic evidence were estimated to be 2.08:1, 4.33:1, 18.7:1, and 350:1, respectively, enabling the adjustment of evidence weight by a data-driven calibration [118]. To simplify the implementation of this Bayesian framework, the

strengths of evidence categories were abstracted into a point scale, where supporting scores 1, moderate 2, strong 4, and very strong 8 (negative values used for evidence against pathogenicity). These points are proportional to Log(odds), are additive, and produce a system that recapitulates the Bayesian formulation of the ACMG/AMP guidelines (**Figure 13**)[119].

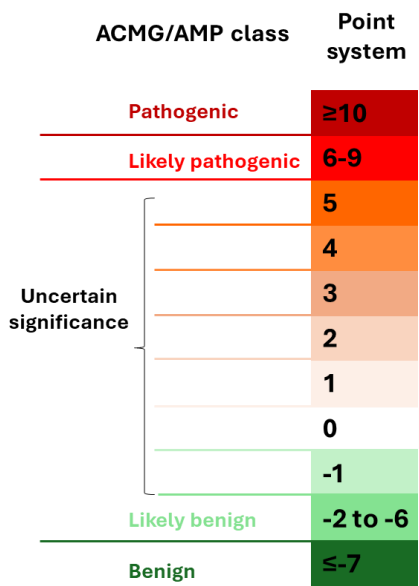


Figure 13. Point system ranges to assign variant classes according to Tavtigian et al. (2020)[119].

This method simplifies the process for code combinations and shows higher plasticity in combining different ACMG/AMP criteria and strengths, allowing new combinations not considered in the 2015 framework, including combinations where evidence is in the opposing directions (pathogenic and benign).

2.3. ENIGMA Consortium

ENIGMA is an international research-based consortium focused on determining the clinical significance of sequence variants in *BRCA1*, *BRCA2* and other known or suspected breast and/or ovarian predisposition genes, to provide this

expert opinion to global database and classification initiatives, and to explore optimal avenues of communication of such information at the provider and patient level [117].

Membership is both international and interdisciplinary and currently there are about 270 members from 40 countries (June 2024). An ENIGMA member is defined as a researcher or research group who is willing to work collaboratively towards the classification of variants and contribute data from families with VUS (or reference control variants), as required to aid in the variant classification projects of ENIGMA and/or conduct statistical analysis, bioinformatic studies and/or laboratory-based assays aimed at classification of variants within a working group framework.

This consortium is organized into five working groups with independent purposes but interacting with one another to access complementary expertise and information (**Figure 14**).

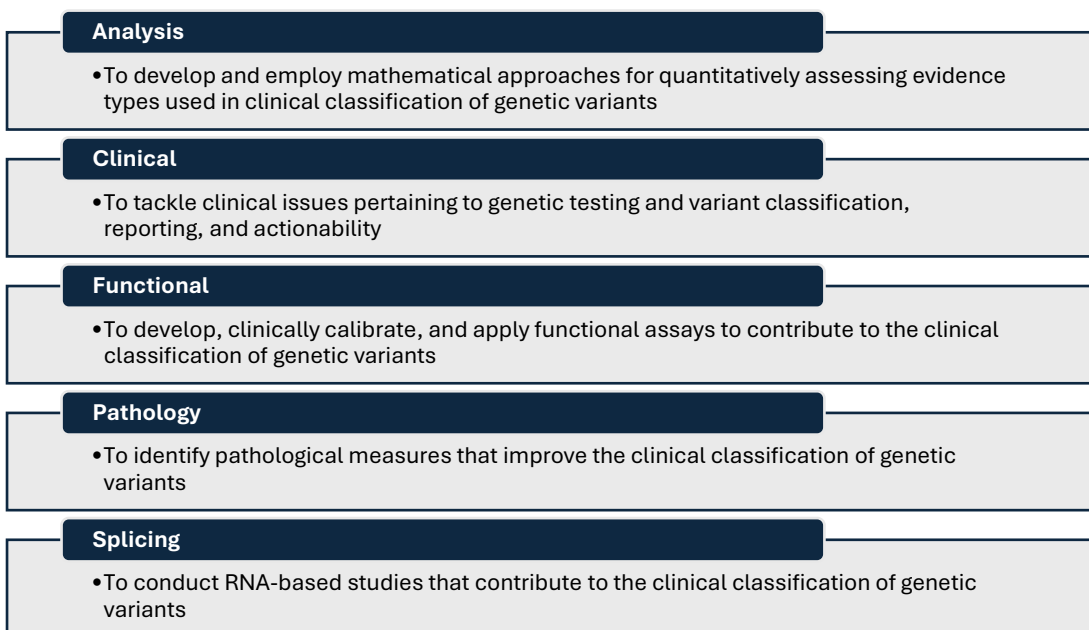


Figure 14. ENIGMA working groups.

Specific purpose/scope are detailed for each group.

Data source: <https://enigmaconsortium.org/enigma-consortium/working-groups/>

Any member can submit proposals for collaborative research projects, which are then subject to approval by the ENIGMA steering committee. The Hereditary Cancer Genetics group (formerly Oncogenetics group) from Vall d’Hebron Institute of Oncology (VHIO) has been an active member of ENIGMA since 2012 and has led two projects within two different working groups. These projects are included in **Part 2** and **3** of this thesis. Specifically, **Part 2** of this thesis is the “*Analysis of potentially spliceogenic variants affecting BRCA1 exon 18*” collaborative project, which was approved in 2015 within the splicing working group, to perform a comprehensive study of potentially splicing altering variants affecting the exon 18 of *BRCA1*. **Part 3** of this thesis prompted to propose a collaborative project within the pathology working group, which was approved in 2021. This project, titled “*The use of a RAD51-based functional test in FFPE tumour tissue as a predictor of BRCA1, BRCA2, and PALB2 variant*

pathogenicity”, was proposed to obtain enough samples to assess the usage of tumour HRR status for variant classification.

2.4. ENGIMA as Variant Curation Expert Panel (VCEP)

As above mentioned, ACMG/AMP guidelines were intended to be generic, and thus some criterion may not be relevant for variant curation for a specific gene. The Clinical Genome Resource (ClinGen) consortium prompted the creation of variant curation expert panels (VCEPs) to adapt these guidelines to specify which rule codes and strengths are appropriate for a specific gene-disease relationship, interpreting variants according to their specifications and publish the interpretations through NCBI’s publicly available ClinVar database [120,121].

At this point, ENIGMA was a pre-existing consortium focused on determining the clinical significance of variants in HBOC genes, thus ClinGen requested to form an external expert panel for curation of *BRCA1* and *BRCA2* variants. The classification criteria documented for this purpose captured qualitative criteria and quantitative multifactorial likelihood analysis methods. These guidelines were used to assign 7,456 expert curations for *BRCA1/2* in ClinVar. However, these specifications are no longer in use since in 2020, ENIGMA sought to become an internal ClinGen VCEP to align with Food and Drug Administration (FDA) recognized processes for expert panel contributions to ClinVar (**Figure 15**) [122]. As a VCEP, all ACMG/AMP criteria have been reviewed for relevance to interpreting *BRCA1* and *BRCA2* variants, and strengths have been statistically calibrated. Consequently, current ClinGen ENIGMA specifications to the ACMG/AMP variant interpretation guidelines considered 13 original codes as

non-applicable or overlapping with other criteria, 15 codes were calibrated and 4 were repurposed [122].

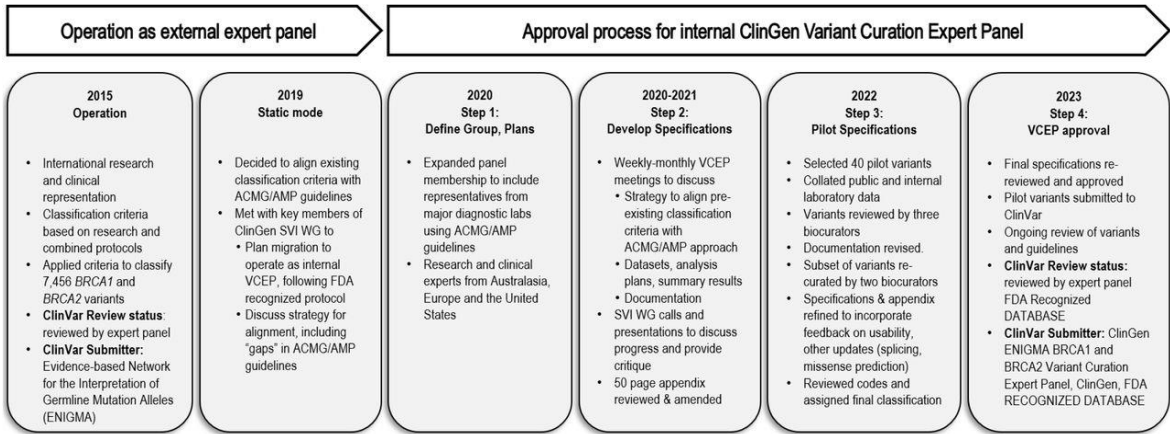


Figure 15. Timeline for migration from “external” expert panel to operation as a ClinGen Variant Curation Expert Panel following the ClinGen FDA-recognized protocol.
Details of ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel membership and biocurator workforce are available at <https://clinicalgenome.org/affiliation/50087/>, and the final specifications first approved for internal ClinGen BRCA1/2 VCEP use are available via <https://cspec.genome.network/cspec/ui/svi/affiliation/50087>. Taken from Parsons et al. (2024) [122].

3. mRNA Splicing in HBOC

A possible pathogenic mechanism for a subset of VUS is the disruption of normal mRNA splicing. These variants, known as ‘spliceogenic’, have been reported to be a common cause of genetic disease, and hereditary cancer risk genes (including HBOC and Lynch syndrome genes) are enriched for this type of variants [123]. Of note, spliceogenic variants account for the third most common pathogenic variant type in *BRCA1* and *BRCA2*, representing 10.1% and 7.6% of total variants on both genes, respectively [124]. Considering the relevance of splicing-altering variants in HBOC syndrome among other genetic diseases, several efforts have been made to enhance their identification and clinical classification.

3.1. Splicing Elements

RNA splicing consists in the transformation of a newly made precursor messenger RNA (pre-mRNA) transcript into a mature messenger RNA (mRNA). During splicing, introns are removed, and exons are joined together. This process is held in the nucleus, and it is catalysed by a large RNA-protein complex (the spliceosome), reliant on numerous *cis* and *trans* acting factors (**Figure 16**) [125,126]. The *cis* DNA elements are the DNA sequences that define exons, introns and other regulatory elements necessary for proper splicing. They include:

- **Splicing sites** include the splice **donor** (+1,2 dinucleotide flanking the 5' end of the intron) and the splice **acceptor** sites (-1,2 dinucleotide flanking the 3' end of the intron). These sequences are recognized during spliceosome assembly and introns are removed from primary transcripts by cleavage at these sites. Most introns (98.7%) are recognized by highly conserved dinucleotides at the 5' boundary (GT) and 3' boundary (AG)

(**Figure 16**)[127,128]. Noncanonical GC-AG and AT-AC sequences occur in 0.56 and 0.09% of the splice site pairs, respectively. The noncanonical GC-AG splice site, like the canonical GT-AG, is processed by the major spliceosome complex. But the AT-AC splicing motifs, that are present in U-12 type introns, are recognized by U12-dependent minor spliceosome [129]. Intronic and exonic nucleotides adjacent to these nucleotides are also highly conserved and have been found to be critical for splice site selection: CAG/**GUA**AGU in donor sites and NY**AG**/G in acceptor sites [127,130].

- **Branch point** and **polypyrimidine tract** are intronic sequence motifs essential for the spliceosome complex formation (**Figure 16**). The branch point is usually found at -18 to -44 nucleotides upstream of the 3' splice site and is essential for the formation of the lariat structure during the removal of introns [131]. The polypyrimidine tract consists of a sequence enriched in polypyrimidine nucleotides located at -5 to -40 bases far from the 3' splicing site. It serves as a binding site for the U2AF splicing factor, helping to 3' recognition and stabilizing the spliceosome assembly [130].
- **Splicing Regulatory Elements** (SREs) are intronic and exonic sequences that act as splicing intronic/exonic enhancers (ISE/ESE) or intronic/exonic silencers (ISS/ESS), binding serine and arginine-rich proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs), respectively (**Figure 16**)[130,132]. The spliceosome must correctly recognize and cut out the intronic sequences from the pre-mRNA for an accurate splicing process. Exon definition is the initial step in pre-mRNA splicing, and it has been suggested that accurate splice site recognition resides in an SRE balance, i.e., exons enriched with Exonic Splicing

Enhancers (ESEs) and introns with Intronic Splicing Silencers (ISSs) [133–135]. The proper identification of the splice sites is challenging since the consensus sequences are very short and there are many similar sequences across the whole genome, known as cryptic splice sites. Sometimes a cryptic site matches the consensus sequence of a natural splice site but is not used in the splicing process because of the lack of *cis* acting regulatory elements (enhancers and silencers) necessary for the proper exon identification [132].

Trans acting elements include the spliceosome, a protein-RNA complex containing five small nuclear ribonucleoproteins (snRNPs): U1, U2, U4, U5 and U6, and over 300 different proteins. The snRNPs have small nuclear RNAs which allow to form complementary RNA-RNA complexes and thus identify the specific sequences of the splicing sites by the spliceosome [130].

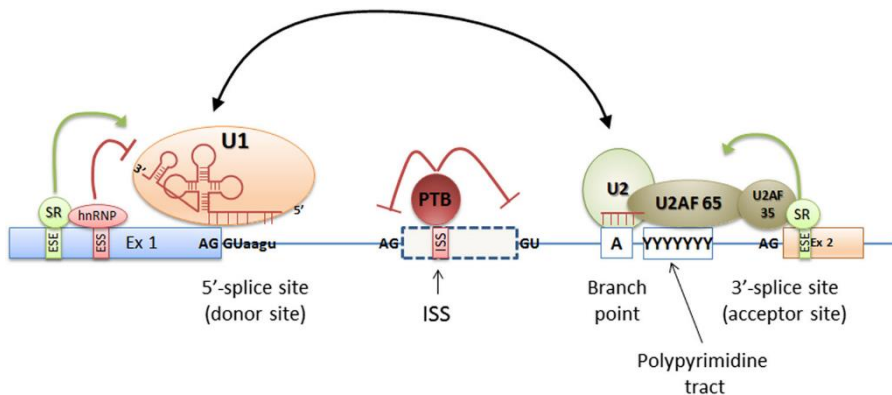


Figure 16. Schematic representation of *cis* and *trans* splicing elements.

Cis elements include DNA sequences defining canonical splicing sites (acceptor and donor), branch point, polypyrimidine tract and splicing regulatory elements (ESE: exonic splicing enhancer, ESS: exonic splicing silencer, ISE: intronic splicing enhancer, ISS: intronic splicing silencer). *Trans* acting elements include spliceosome proteins. Taken from Abramowicz and Gos. (2018) [130].

3.2. Alternative Splicing

The complex regulation of the splicing process and the dynamic composition of the spliceosome confer great flexibility and versatility to this mechanism, allowing to process pre-mRNAs in alternative manners, generating different mRNA transcripts or isoforms [136,137]. This phenomenon contributes to cell differentiation and lineage determination, tissue-identity acquisition and maintenance, and organ development, and allows cells to respond to their environment, conferring phenotypic plasticity [138].

Alternative splicing can generate mRNAs that differ in their untranslated regions (UTRs) or coding sequence through several events that include exon skipping, a choice between mutually exclusive exons, the use of alternative splice sites and intron retention (**Figure 17**). These differences might affect mRNA stability, localization or translation. Furthermore, some splicing mRNA isoforms could change the reading frame, resulting in the generation of different protein isoforms with diverse functions and/or localizations [138].

The precision and diversity of alternative splicing events are influenced by several factors such as exon and intron length, the order in which introns are removed, splice site strength, the presence and location of cryptic sites, the abundance of SREs, the expression levels of *trans*-acting factors (which indeed is tissue-specific), RNA polymerase II (RNAPII) elongation rates and chromatin structure [138,139].

Genome-wide studies have revealed that 90-95% of human genes undergo some level of alternative splicing and it has been reported that more than 50% of genes express different alternative spliced isoforms among tissues [139–141].

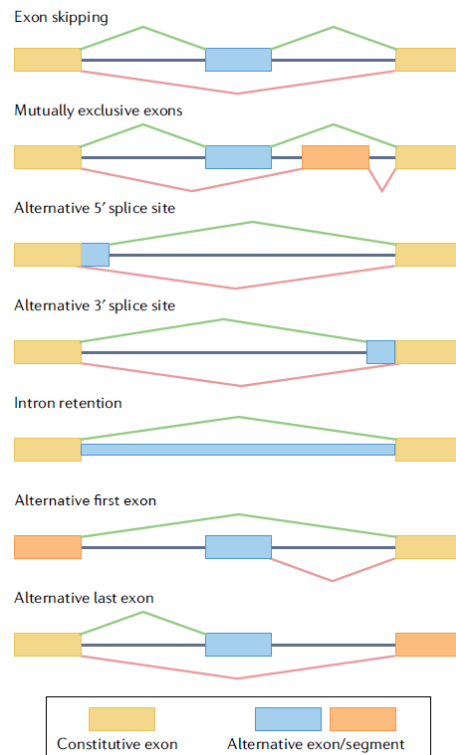


Figure 17. Alternative splicing events.

Green and red lines represent different splicing options. Taken from Wright et al. (2022) [142].

3.3. Aberrant Splicing and Disease

Although the splicing process is a regulated step in gene expression, it is frequently perturbed in human disease. Variants that alter gene splicing are estimated to comprise up to a third of all disease-causing variants [143]. Any substitution or insertion/deletion (indel) sequence variant could alter splicing regardless of its coding effect. Most of the spliceogenic variants are believed to affect *cis*-acting splicing motifs, impairing or strengthening their functionality. Others create *de novo* functional splicing motifs, since these motifs are relatively short and degenerated, the probability of a genetic variant creating a new cryptic site is not negligible [144]. There are many effects of splicing disruption on mRNA, detailed in **Figure 18**:

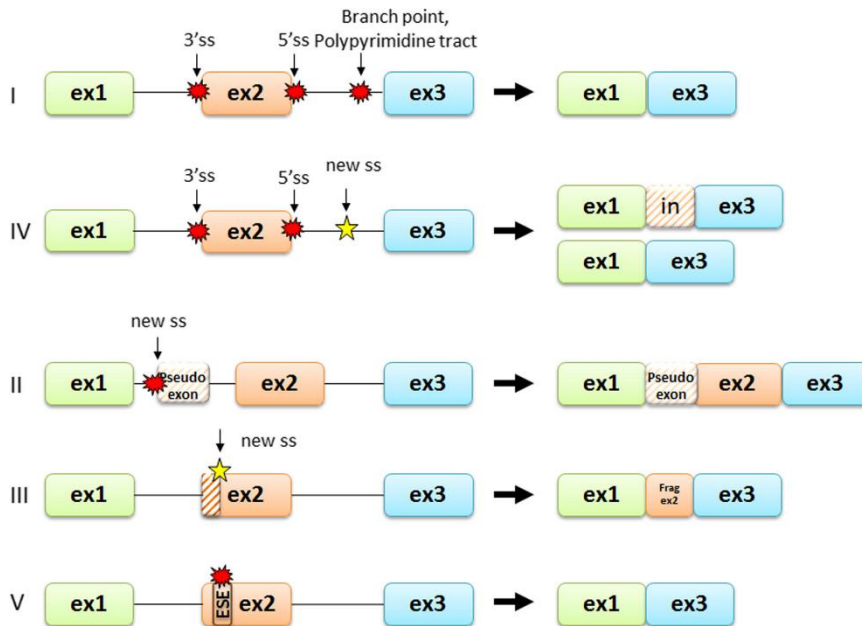


Figure 18. Schematic representation of effects of splicing disruption on mRNA.

(I) Splicing mutations within the canonical splice sites lead to whole exon skipping, (II) deep intronic variants creating new splice sites resulting in the inclusion of cryptic exons, (III) exonic single nucleotide variants creating new splice sites that result in the loss of an exon fragment, (IV) variants in the canonical splice sites resulting in the usage of the cryptic exonic or intronic splice site that leads to the inclusion of an intron fragment or exon fragment skipping, and (V) mutations within the exon, usually leading to ESE disruption, resulting in the entire exon skipping. *Taken from Abramowicz and Gos. (2018) [130].*

As a result of the splicing defects, the transcript's open reading frame can be altered, introducing a premature termination codon (PTC). When this occurs, transcripts are generally targeted for nonsense-mediated decay (NMD), a mechanism to prevent the production of truncated proteins [145]. In such instances, the transcripts will be undergoing active degradation. Otherwise, variants could lead to splicing aberrant transcripts with sequence insertions or deletions that preserve the reading frame, thus encoding for shorter proteins with unknown effects. Of note, also exists the possibility that variants affecting *cis*-acting elements disrupt or enhance the usage of cryptic sites shifting the ratio of naturally occurring transcripts [146].

Recent estimations have shown that an unexpectedly large fraction of genetic diseases (ranging from 15% to >60%) are caused by substitution or small indel variants that disrupt the splicing process [146–148]. According to the Human Gene Mutation Database (HGMD, <https://www.hgmd.cf.ac.uk/ac/index.php>, accessed 25th August 2023), nearly 9% of all variants linked to human genetic diseases affect pre-mRNA splicing. On average, 13.4% of all hereditary disease alleles are classified as splicing mutations mapping to canonical splicing sites. However, the rate of mis-RNA splicing variants causing disease is thought to be much higher as synonymous, missense, and deep intronic variants outside of the donor and acceptor splice sites are rarely considered and evaluated as splicing disruptors [149,150].

In general, mutations in the canonical acceptor and donor sites affect strongly conserved sequences that define exon-intron boundaries. Therefore, any variant in these canonical sequences might alter the interaction between pre-mRNA and proteins involved in the intron removal. The most classical mutations affect $\pm 1/2$ residues at the 5' donor and 3' acceptor splice sites, and usually lead to single exon skipping [130]. However, variants beyond these positions can also alter splicing by affecting splicing regulatory motifs (**Figure 18**). In studies of neurofibromatosis type 1 (*NF1*) and ataxia telangiectasia (*ATM*) where analyses were performed both at the DNA and RNA level, approximately 50% of the patients were found to have disease due to mutations that resulted in aberrant splicing [126]. In both studies most splicing-disrupting variants did not involve the essential dinucleotides, suggesting over half of the splice variants may be missed by the current focus on essential splice site variants [126]. Likewise, a previous study from our group collected data from over 300 germline variants in HBOC and Lynch Syndrome patients, and 51% of splicing altering variants within

the 30 nucleotides at exon-intron boundaries (10 exonic, 20 intronic) were outside dinucleotide splice sites (**Figure 19**) [151].

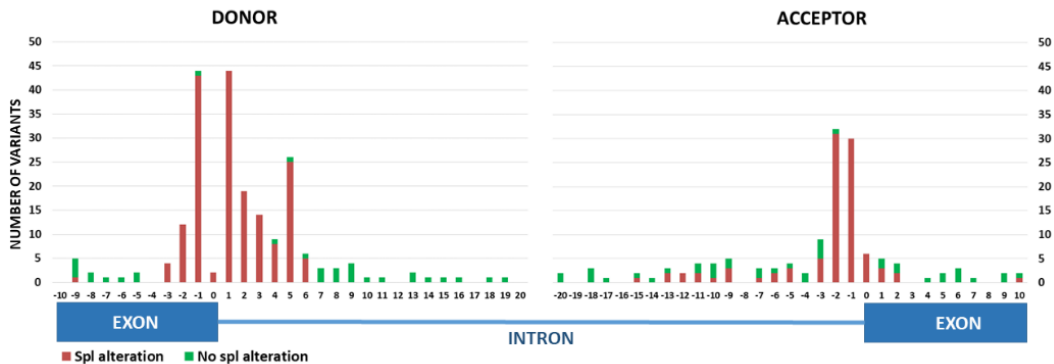


Figure 19. Graphical representation of *in vitro* splicing effect of 346 variants in HBOC and Lynch Syndrome risk genes.

In red are indicated the number of variants altering splicing (exon skipping, usage of cryptic sites and alteration of alternative splicing profile), and non-spliceogenic variants are depicted in green. Taken from Moles-Fernandez et al. (2018) [151].

There is a limited knowledge about the frequency of deleterious variants occurring in deep intronic regions (± 20 bp from canonical splicing sites), which is largely due to conventional genetic diagnosis usually restricted to coding exons and flanking intronic regions. However, nucleotide changes in these regions could generate aberrant transcripts by introducing intronic sequences in mRNA (**Figure 18**) [152–154]. Pathogenic deep intronic variants have been described in more than 75 disease-associated genes including monogenic disorders such as hereditary cancer syndromes [155]. These variants can alter splicing by two different mechanisms [156][17]: the creation/enhancement of cryptic splice sites and the alteration of an intronic SRE by the disruption of an ISS or the creation/strengthening of an ISE (**Figure 20**).

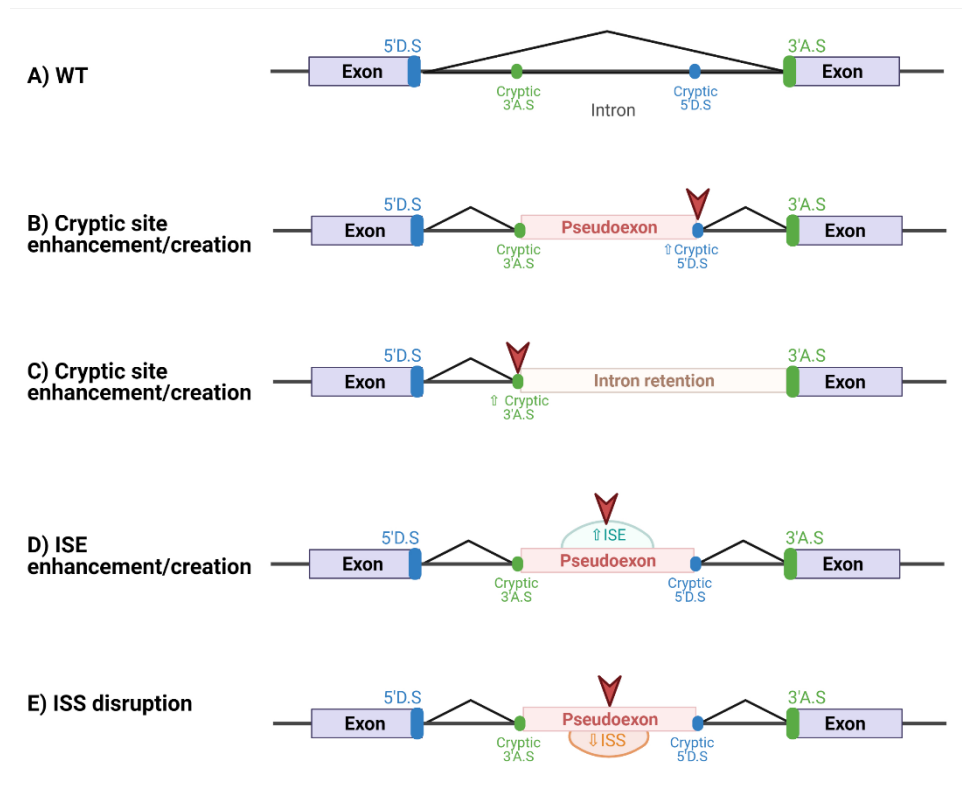


Figure 20. Splicing effects caused by deep intronic variants.

A) Normal splicing using natural splicing sites. **B)** Deep intronic variant creating/enhancing a cryptic splice site, resulting in the inclusion of a pseudoexon by using a complementary cryptic site. **C)** Intronic retention caused by a deep intronic variant that creates/enhances a cryptic site, which is used instead of the canonical splice site. **D)** Deep intronic variant creating/enhancing an ISE, resulting in the inclusion of a cryptic exon using two cryptic splice sites. **E)** Deep intronic variant disrupting an ISS, resulting in the inclusion of a cryptic exon using two cryptic splice sites. *WT: wild type, D.S: donor site, A.S: acceptor site, ISE: intronic splicing enhancer, ISS: intronic splicing silencer.*

A few examples of diseases driven by the inclusion of a pseudoexon (PSE) are the HBOC caused by the c.4185+4105C>T variant in *BRCA1*, the first reported deep intronic variant in this gene that activates a pre-existing cryptic donor site [157]; and the Ataxia-telangiectasia disease due to the c.2839-581_2839-578del variant in the *ATM* gene, which creates an ISE [158]. These examples highlight the relevance of screening deep intronic regions to identify germline pathogenic variants leading to aberrant mRNA processing.

Exonic variants outside splice sites can also alter splicing by creating/enhancing cryptic sites or disrupting SREs (**Figure 18**). One of the first documented examples of an exonic variant disrupting an SRE in *BRCA1* was the c.5080G>T variant, which results in exon 18 skipping due to the creation of an ISE [159,160]. Similarly, several variants within the exons 17 and 18 of *BRCA2* have been shown to induce aberrant splicing by affecting ESEs necessary for exon recognition [161]. However, many of these variants can be easily misclassified as synonymous, missense, or nonsense mutations, as their molecular effect is generally assumed solely from their coding impact [130].

3.4. *In silico* Splicing Assessment

Since experimental testing of all potentially spliceogenic variants is not currently feasible in a straightforward manner within a clinical setting, multiple computational prediction tools have been developed to infer their effects and prioritize variants for experimental evaluation [144].

Historically, *in silico* tools have evolved from approaches based on position weight matrix (PWM, assessment of the frequency of particular nucleotides within sequences predicted to bind splicing factors), to machine learning methods (computer algorithms that improve automatically through experience) [162–165].

Bioinformatic tools predicting the alteration of donor and acceptor sites are more reliable than those applied to more loosely conserved elements like SREs [149,151,166]. Notably, high-performance levels have been shown by the SPiCE (Splicing Prediction in Consensus Elements) web-based tool, which combines predictions from SpliceSiteFinder-like (SSF, built on PWM) and MaxEntScan (MES, maximum entropy approaches) [167]. Given that the performance of

specific predictors can differ between donor and acceptor canonical splicing sites, previous work from our group proposed a differentiated analysis for each site [151].

All mentioned tools can be used for identifying potential splicing altering variants beyond exon/intron boundaries, although with lower accuracy for deep intronic regions [149,157]. The deep neural network learning-based approach, SpliceAI, outperformed other tools both in the assessment of natural splice site disruption as well as in the prediction of new/cryptic usage elements caused by either exonic or intronic variants [150,168–170].

For identifying potential spliceogenic variants in SREs, QUEPASA (ESRseq, based on experimental data) [134] and HAL (combining machine learning and synthetic biology) [171] tools showed the best accuracies either as standalone or in combination [172].

Variants affecting the polypyrimidine tract can be optimally identified using MES [167], and the Branch Point Prediction (BPP) [173] showed the best performance identifying splicing variants affecting these *cis*-acting elements [174].

Apart from prioritising potential splicing variants for experimental validation, *in silico* predictions may be utilized as supporting evidence in variant classification according to ACMG/AMP guidelines (**Figure 12**)[99]. Typically, *in silico* tools provide a score or probability reflecting the likelihood that a given variant affects splicing. However, the choice of what threshold to use as a cutoff to classify a variant as predicted splice affecting is often relatively arbitrary, and/or based on assessment against a relatively small number of variants with known outcome [126].

Recently, the ClinGen SVI Splicing Subgroup has updated the recommendations to capture evidence related to both predicted and observed splicing defects [170]. After comparing the sensitivity and specificity of 11 different tools, SpliceAI, which outperformed all other predictors, was selected to demonstrate the calibration of a well-performing tool for predicting splice impact [170]. Although this calibration is based on a limited number of genes, the obtained thresholds are proposed for generic use for PP3 and BP4 evidence code application, given that the mechanism of RNA splicing is not gene-specific [170]. Consequently, several VCEPs, such as ENIGMA, have restricted the splicing *in silico* evidence to SpliceAI predictions.

Nevertheless, ACMG/AMP recommendations indicate that bioinformatics predictions should not be used as the sole source of evidence for asserting the clinical relevance of genetic variants. This is due to the necessity of an accurate definition of the nature of associated splicing events, a task that *in silico* analysis is unable to achieve [99,144].

3.5. *In vitro* Splicing Analyses

The most straightforward and effective method to determine whether a specific variant affects splicing is to analyse the RNA extracted from a relevant patient tissue or patient-derived cell line [130]. RNA is typically obtained from clinically accessible tissues, such as blood; however, this tissue may not necessarily be the affected tissue in a specific genetic disorder. As splicing is known to be highly tissue-specific [175], the clinically accessible tissue must reflect the natural splicing profile of the relevant disease tissue for accurate variant interpretation [170]. Specifically, for *BRCA1* and *BRCA2*, blood has been demonstrated to be a

good surrogate of breast tissue since splicing-naturally occurring events are similar in both tissues [176,177].

Additionally, the gene-specific catalogue of naturally occurring splicing events in non-carrier controls is also relevant for assay design and correct interpretation of splicing results yielded from genetic variants [176–179].

The conventional molecular procedure to ascertain the spliceogenic effect of a given variant involves reverse transcription followed by PCR amplification (RT-PCR)[180,181]. The resulting transcript profile must then be compared to those observed in tissue-matched non-carrier controls through electrophoresis and/or sequencing [179].

The accurate interpretation of splicing defects requires the identification of all splicing events associated with the variant allele, including full-length (FL), naturally occurring isoforms and aberrant transcripts. Qualitative analyses based on gel-electrophoresis and Sanger do not provide relative levels of transcript isoforms which is crucial for the comprehensive characterization of the genetic variant, especially in the presence of naturally occurring alternative transcripts [178,179]. Semi-quantitative analysis of PCR products by high-resolution capillary electrophoresis, and quantitative methods based on real-time PCRs (qPCR or digital PCRs) have demonstrated a greater resolution than conventional qualitative methods [178]. This was exemplified by the accurate interpretation of the *BRCA2* c.7976+5G>T and *PALB2* c.3201+5G>T variants, which altered the abundance of several alternative transcripts also present in non-carrier controls, achieved through semi-quantitative splicing profiling combined with the quantification of target isoforms by qPCR [182,183].

Remarkably, when using patients' samples, allele-specific assays targeting the expression of the normal transcript are highly recommended for uncovering possible incomplete splicing effect of a variant (leaky splicing effect) [184]. Additionally, assays to test allelic imbalance can be done to verify a normal splicing profile and exclude the possibility of aberrant profile not detected due to the active degradation of PTC-containing transcripts in the absence of NMD inhibitors [184].

Reporter minigenes constructs are useful alternative tools to study *ex vivo* the impact of a variant on splicing. They are circular plasmids containing everything needed for expression in eukaryotic cells, featuring exons flanking an intron, into which can be inserted a region of interest. For splicing assessment, the region of interest (typically an exon and short flanking intronic sequences) is inserted into the plasmid, both in reference and variant form, such that after expression in a cell line, the splicing of the wild-type (WT) and alternative forms can be compared to assess the effect of the variant on splicing (**Figure 21**). Minigenes have the benefit that are hemizygous constructs allowing allele-specific quantitative measurements of splicing defects, they can be used in almost any genomic context, and do not rely on the availability of patient samples, since variants can be generated artificially, e.g., with site-directed mutagenesis. However, the artificiality of the construct removes much of the larger context in which the variant occurs, and results depend on both cell type and promoter used, so they may not reflect the true picture of splicing within the patient [126,170].

Although both RT-PCR and minigenes approaches have been used extensively in the characterization of splicing variants over the years [150,185–188], their widespread in clinical diagnostics is hindered by their relatively low throughput

nature since in both cases assays must generally be tailored to each variant to be tested [126]. Transcriptome sequencing (RNA-seq) allows the total RNA content of a given sample to be surveyed in a single experiment, allowing the detection of both known and novel splice variants. In the last few years, it has been demonstrated the utility of this approach for detecting splicing altering variants. However, many different tools exist for detecting aberrant splicing events from RNA-seq, and there is not a clear consensus on which tool, or combination of tools is the most appropriate [126,189–191].

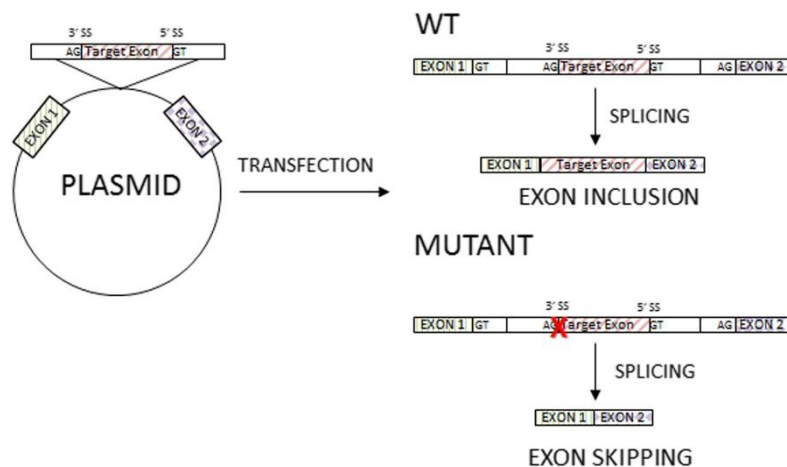


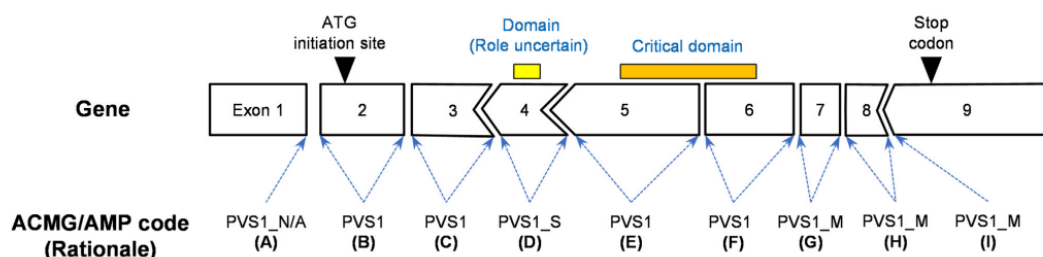
Figure 21. The principles of minigene assay.

The region of interest of a specific gene (e.g. target exon with surrounding intronic sequences) is cloned into an expression plasmid between two construct exons. These constructs enable the assessment of the splicing process for wild-type and variant constructs. *Taken from Abramowicz and Gos. (2018) [130].*

3.6. Splicing Observed Impact for Variant Classification

According to ACMG/AMP guides, the only criterion designated with Very Strong strength level for pathogenicity was PVS1 which was defined as a “null variant in a gene where LoF is a known mechanism of disease [99]. Mutations in the canonical $\pm 1/2$ splice sites are often presumed to be LoF due to their high likelihood of altering splicing. However, not all spliceogenic events are necessarily pathogenic. This distinction was recognized by the ClinGen Sequence Variant Interpretation (SVI) working group, which specified that individual $\pm 1/2$ variants, despite having a high prior probability of being spliceogenic, deserve different prior probabilities of pathogenicity (i.e. different strength levels of PVS1 according to ACMG/AMP notation) depending on the predicted splicing outcome (e.g., skipped exon, use of a cryptic splice site, etc, and the impact on the resulting encoded proteins) [192]. Of note, these specifications have been recently updated by the SVI splicing subgroup as detailed in **Figure 22** [170].

Among the recent ClinGen SVI Splicing Subgroup updates, the PS3/BS3 functional data codes are no longer recommended for capturing splicing-confirmed defects, as outlined in Richards et al. (2015)[99]. Instead, PVS1 has been repurposed to capture splicing assay data providing experimental evidence for variants generating LoF transcripts, regardless of their position or assumed coding effect. Similarly, BP7 (computational and predictive data) has also been repurposed to capture RNA results demonstrating no splicing impact for intronic and synonymous variants [170].



- (A) 5' UTR region - No splicing alteration predicted or use of a cryptic splice motif does not affect the coding sequence.
- (B) Exon skipping or use of a cryptic splice motif eliminates the initiation codon and there are no alternative start codons.
- (C) Exon skipping or use of a cryptic splice motif disrupts reading frame and is predicted to undergo NMD
- (D) Exon skipping or use of a cryptic splice motif preserves reading frame, and removes a region (>10% of the protein) which has not been established as critical to protein function.
- (E) Exon skipping or use of a cryptic splice motif disrupts reading frame and is predicted to undergo NMD
- (F) Exon skipping or use of a cryptic splice motif preserves reading frame, and removes a region which has been established as critical to protein function
- (G) Exon skipping or use of a cryptic splice motif preserves reading frame, and removes a region (<10% of the protein) which has not been established as critical to protein function.
- (H) Exon skipping or use of a cryptic splice motif disrupts reading frame and is not predicted to undergo NMD, and removes a region (<10% of the protein) which has not been established as critical to protein function.
- (I) Exon skipping or use of a cryptic splice motif disrupts reading frame and is not predicted to undergo NMD, and removes a region (<10% of the protein) which has not been established as critical to protein function.

Figure 22. PVS1 evidence assignment of gene-specific codes to splice donor/acceptor $\pm 1,2$ dinucleotide variants according to ClinGen SVI Splicing Subgroup.

N/A: non-applicable, S: strong, M: moderate. Taken from Walker et al. (2023) [170].

PVS1 evidence must be considered for variants that lead to aberrant transcripts with a PTC or that preserve the reading frame but resulting in a total or partial in-frame loss of sequences encoding protein critical domains (**Figure 22**). In contrast, the application of PVS1 is less clear for variants that produce in-frame transcripts coding for potentially fully or partially functional proteins, due to their potential rescue effect [170].

Variants inducing incomplete (“leaky”) splice defects are also challenging to classify because in addition to giving rise to aberrant transcripts, and/or quantitatively altering transcript patterns, they maintain the ability to produce a certain amount of functionally proficient mRNA. However, the level of splice perturbation required to impair gene function (i.e. the minimum level of functional transcripts to keep normal gene function) is an unanswered question and is likely to vary gene by gene, or even exon by exon.

Regarding HBOC genes, the c.68-7T>A in *BRCA2* remained controversial for a long time until a large collaborative study demonstrated that this variant induced a 4.5-fold higher expression of exon 3 skipping (20%), but not to completely abolish the expression of a normal RNA (full-length, FL) and was not associated with a high risk of BC [193]. Additionally, results from another leaky variant on the same exon (c.231T>G) helped establish that up to 60% of exon 3 skipping does not cause any increase in cancer risk [188]. Functional data from further exon 3 variants also suggested that variants causing a decline in FL *BRCA2* up to approximately 30% are not pathogenic [188].

Similarly, for *BRCA1*, variants resulting in 20%–30% expression of a functional transcript should not be considered pathogenic. These levels are based on results observed in carriers of *BRCA1* c.[594-2A>C; 641A>G]. This haplotype is not associated with high cancer risk, despite generating ≈70-80% truncating transcripts, and ≈20-30% of naturally occurring in-frame Δ9,10 transcript predicted to encode a protein with tumour suppression function [194]. Based on these findings, ENIGMA adopted the 20-30 % threshold for applying splicing observed evidence in germline variant classification (**Figure 23**).

PVS1 (RNA) or BP7 (RNA) code weight to apply, according to mRNA source, assay design/quantitation AND relative level/s of functional and non-functional transcript/s [§]						
Assay results with allele-specific quantitation [#]						
PVS1 Code Weight as per PVS1 flowcharts [*]	Patient with quantification [#]	Minigene only	Patient with quantification [#]	Minigene only	Patient with quantification [#] OR minigene only	
Proportion WT or (assumed) functional transcript [§]	≤10%	≤10%	>10 & <20%	>10 & <20%	20-30%	>30%
Proportion (assumed) non- functional transcript	≥ 90%	≥ 90%	< 90% & > 80%	< 90% & > 80%	70-80%	<70%
PVS1	PVS1 (RNA)	PVS1_Strong (RNA)	PVS1_Strong (RNA)	PVS1_Moderate (RNA)	PVS1_N/A (RNA)	BP7_Strong (RNA)
PVS1_Strong	PVS1_Strong (RNA)	PVS1_Moderate (RNA)	PVS1_Moderate (RNA)	PVS1_Supporting (RNA)		
PVS1_Moderate	PVS1_Moderate (RNA)	PVS1_Supporting (RNA)	PVS1_Supporting (RNA)	PVS1_Supporting (RNA)		
PVS1_Supporting	PVS1_Supporting (RNA)	PVS1_Supporting (RNA)	PVS1_Supporting (RNA)	PVS1_Supporting (RNA)		
Assay results using patient mRNA without allele-specific quantitation ^{##}						
PVS1 Code Weight as per PVS1 flowcharts [*]	Apparent (near) complete splicing		Apparent incomplete impact		Apparent no impact (relative to controls)	
PVS1	PVS1_Strong (RNA)		PVS1_N/A (RNA)		BP7_Strong (RNA)	
PVS1_Strong	PVS1_Moderate (RNA)					
PVS1_Moderate	PVS1_Moderate (RNA)					
PVS1_Supporting	PVS1_Moderate (RNA)					
§ Application of weights must also consider other factors relating to assay design. These may include the following: primer and/or construct design that enables detection of naturally occurring isoforms, or detection of aberrations that may be expected based on cryptic site prediction; sensitivity of the methodology to detect transcripts of different size ranges; approaches to measure allele-specific quantitation (sanger sequencing, snapshot etc.); possible amplification bias of shorter products; use of nonsense-mediated decay (NMD) inhibitors, and quantitation of PTC transcripts predicted to lead to NMD. See Appendix for justification for use of these operational thresholds.						
* Baseline PVS1 code weight is determined by interpretation of the transcript profile from available splicing assay(s) against PVS1 decision trees (See Table 4 searchable excel, and Appendix D).						
^ Assumed functional transcript: in-frame and not affecting a critical protein domain, initiation or termination codon, expressed as % per-allele expression. If multiple transcripts with different predicted effect observed – apply weight considering total % per-allele expression of (potentially) functional transcripts. When assessing mRNA from patient material for variants that lead to expression of one or more naturally occurring isoforms, an inferred per allele exclusion level is calculated as follows: Two times level of aberration/s observed in the variant carrier minus the average level of aberration/s in multiple samples of the same type, from control individuals.						
# Patient allele-specific OR patient not allele-specific plus minigene to quantify transcripts.						
## Interpretation of mRNA assay data without quantitation is based on consensus curator judgement						

Figure 23. ENIGMA VCEP recommendations for applying PVS1/BP7 (RNA) evidence.

WT: wild type, N/A non-applicable. Taken from Parsons et al. (2024) [122].

4. High-risk HBOC Genes: Carcinogenesis, Structure and Functions

BRCA1, *BRCA2* and *PALB2* encode for key proteins involved in HRR. All three are tumour suppressor genes for which LoF is a known disease mechanism. When one copy of any of these genes is altered in the germ line, the result is the HBOC syndrome, which is inherited in an autosomal dominant pattern. In these cases, and according to Alfred G. Knudson's "2-hit hypothesis", tumourigenesis requires a somatic alteration that inactivates the wild-type (WT) allele [195]. In contrast to heterozygous carriers, homozygous or compound heterozygous pathogenic variant germline carriers develop Fanconi Anaemia (FA), a different phenotype. Biallelic inheritance of *BRCA2* LoF alleles causes a recessive condition known as Fanconi Anaemia type D1 (FA-D1), a very early onset, severe disorder with specific clinical features (OMIM # 605724). Fanconi Anaemia Type

S (FA-S) is caused by biallelic inheritance of *BRCA1* LoF alleles (OMIM # 617883). Likewise, biallelic LoF in *PALB2* cause FA subtype N (FA-N) (OMIM # 610832).

Focusing on HBOC syndrome, it has been described that specific regions of *BRCA1* and *BRCA2* are differently associated with BC and OC. Specifically, in *BRCA1* there are three BC cluster regions (BCCR) (coding positions c.179-c.505, c.4328-c.4945 and c.5261-c.5563), and one OC cluster region (OCCR)(c.1380-c.4062). *BRCA2* also presents three BCCR (c.1-c.596, c.772-c.1806 and c.7394-c.8904) and two OCCR (c.3249-c.5681, c.6645-c.7471)[196]. Furthermore, not all high-risk genes are associated with similar BC subtypes. *BRCA1*-associated BC, which are mostly of TN (70%–85%), differ from *BRCA2*-associated and *PALB2*-associated BCs in their distribution into ER and HER2 clinical subgroups, which is similar to that of sporadic cancers [197]. Indeed, histopathological BC features had been used for variant classification during the last decade [198]. Recently, it has also been proposed that ovarian histotypes might be also informative of variants' pathogenicity/benignity, although germline pathogenic variants in both *BRCA1* and *BRCA2* occur predominantly in patients diagnosed with HGSC [199].

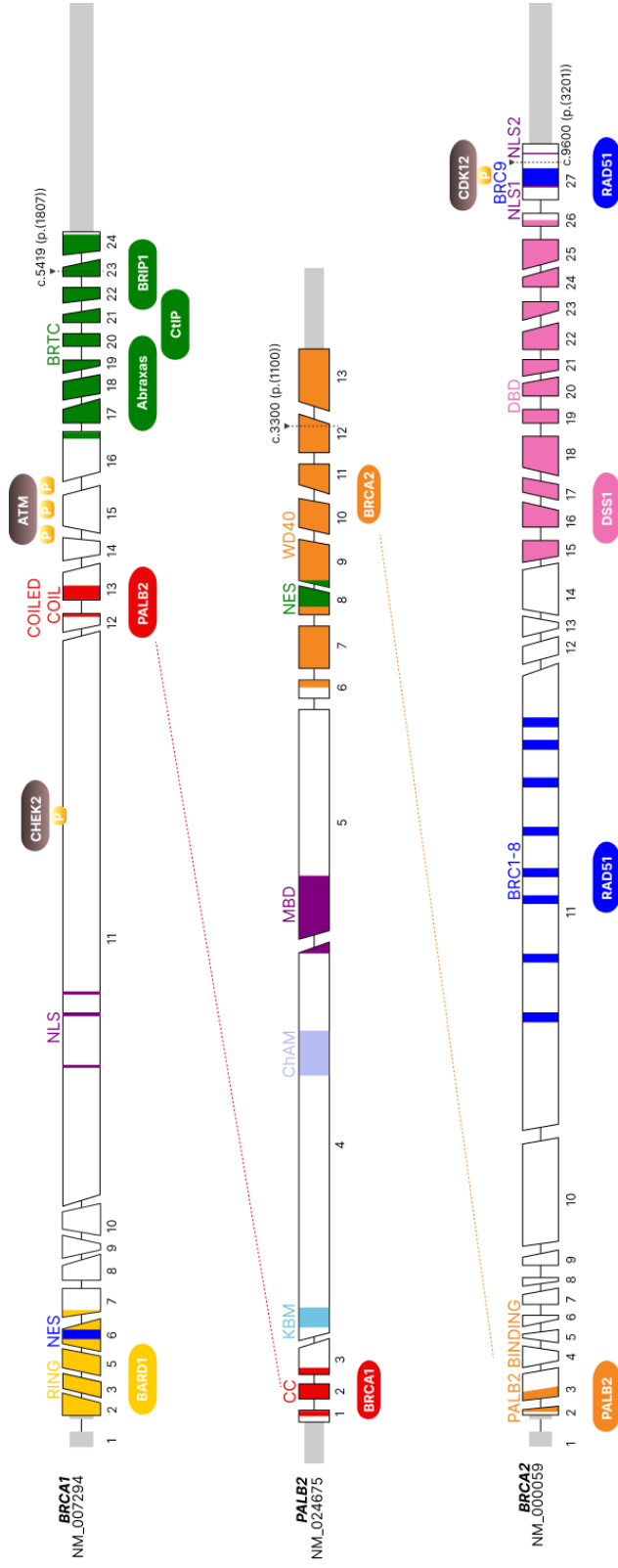


Figure 24. *BRCA1*, *BRCA2*, and *PALB2* exon and domain maps.

MANE reference transcript is indicated for each gene. Boxes represent exons and the number below indicates the exon number. Exon legacy numbers were used for *BRCA1*. The reading frame is indicated by exons' shape: the overhang on top indicates that the exon contains the two bases of a codon and the overhang on the bottom indicates that the exon contains one nucleotide of a codon. Parallel lines represent in-frame events (e.g. exon 18 of *BRCA1*). Dashed arrows indicate non-mediated decay (NMD) threshold positions. Transcripts with a premature termination codon (PTC) beyond the indicated coordinates are not expected to be degraded by NMD. Protein domains for *BRCA1*: RING (Really Interesting New Gene, AA 2-101), NES (Nuclear Export Signal, AA 81-99), NLS1-3 (Nuclear Localization -Sequence, AA 503-508, 607-614 and 651-656), Coiled Coil (AA 1391-1424) and BRTC (BRCA C Terminus, AA 1650-1863). For *PALB2*: CC (Coiled Coil, AA 9-44), KBM (KEAP1-binding motif, AA 76-100), ChAM (Chromatin Association Motif, AA 395-450), MBD (MRG15-binding domain, AA 562-629), NES (AA 928-945), WD40 repeats (AA 854-1186). For *BRCA2*: PALB2 binding domain (AA 10-40), BRC-1 (AA 1002-1036), BRC-2 (AA 1212-1246), BRC-3 (AA 1422-1453), BRC-4 (AA 1518-1549), BRC-5 (AA 1665-1696), BRC-6 (AA 1837-1871), BRC-7 (AA 1971-2005), BRC-8 (AA 2051-2085), DBD; (DNA binding domain consists of an α -helical-DSS1 binding domain (H) and three oligonucleotide binding (OB) folds, AA 2481-3186), NLS1 (AA 3263-3269), BRC-9 (or TR2, AA 3265-3330), and NLS2 (AA 3381-3385). Domains' information is also summarized in the current versions of VCEP guidelines (v.1.1 *BRCA1/2* and v1.1 *PALB2*). The *BRCA1* and *BRCA2* and phosphorylation sites had been adapted from Roy et al. (2011)[35].

4.1. **BRCA1**

The *BRCA1* gene is located at locus 17q21.31, encompassing genomic positions of 43,044,295 to 43,170,327 (*Homo sapiens* GRCh38.p14, <https://www.ncbi.nlm.nih.gov/gene/672>). It is structured in 23 exons that encode for a protein of 1863 residues. Exons are sequentially numbered to match the exon descriptions of the MANE Select transcript (GenBank: NM_007294.4). Exon numbering of *BRCA1* has historically been according to GenBank U14680.1 with exon 4 missing due to a correction made after the initial description of the gene, termed as legacy exon numbering (**Figure 24**).

BRCA1 is a pleiotropic DNA damage response (DDR) protein that functions in both checkpoint activation and DNA repair [35]. It has a Really Interesting New Gene (RING) domain at its N-terminus that binds BARD1, which is required for the E3 ubiquitin ligase activity [200]. The BRCA1/BARD1 heterodimer has been implicated in maintaining genome stability and tumour suppression through its involvement in DNA damage signalling, DNA repair and transcriptional regulation [201]. Additionally, heterodimerisation is also important for their stability *in vivo* and their nuclear localization [200,202,203]. The central region of BRCA1 is mainly encoded by exon 11. It contains a CHEK2 phosphorylation site and the nuclear localisation sequences (NLS). At C-terminus, there is a coiled-coil (CC) motif that binds PALB2 to form the BRCA1-PALB2-BRCA2-RAD51 complex, involved in the HRR pathway [204–206]; an SQ/TQ cluster domain (SCD) that contains ATM phosphorylation sites, involved in BRCA1-mediated G2/M and S-phase checkpoint activation [207,208]; and two BRCT (BRCA1 C terminus) repeats, which mediate binding to Abraxas, BRIP1 and CtIP, which aid

in its recruitment to sites of DBSs, DNA repair during replication and end resection mediation respectively [35,209,210].

The function of BRCA1 in HRR and cellular resistance to DNA damage requires RING, CC and BRCT domains. Thus, nearly any truncation or variant affecting BRCA1 protein stability inhibits HRR [211]. Several missense variants on these domains have been shown to affect BRCA1 function and have an association with high cancer risk [212–214]. Indeed, the presence of these variants identifies these three domains as clinically relevant, and it has been demonstrated that variants outside these regions are very unlikely to be pathogenic (**Figure 25**) [215].

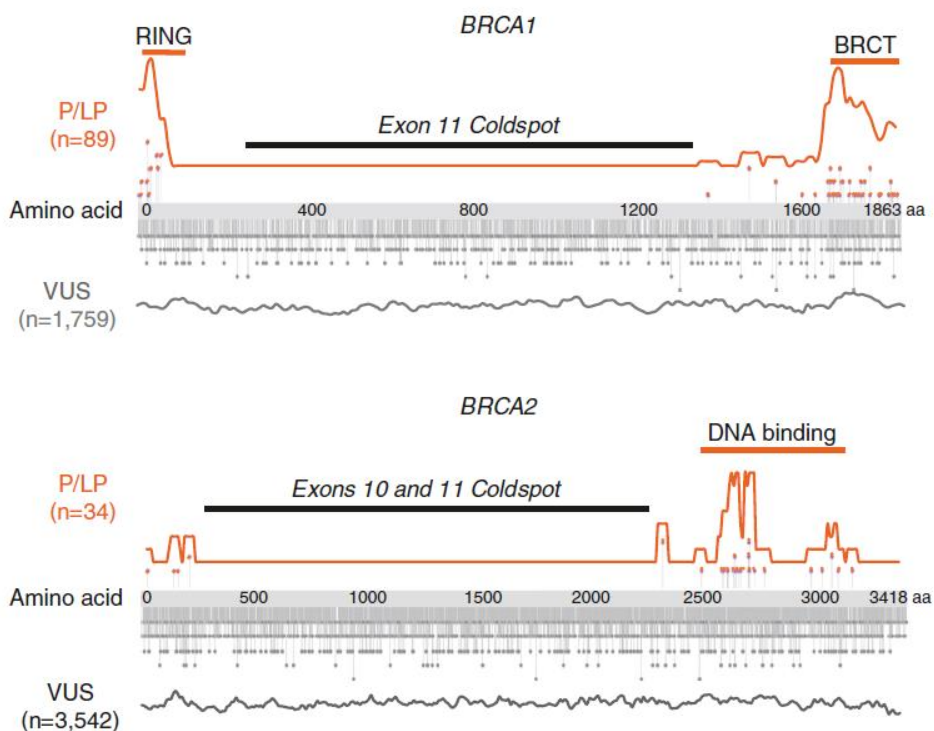


Figure 25. Distributions of missense variants in *BRCA1* and *BRCA2*.

Missense variants are indicated by lollipops. For each gene, distributions of (likely) pathogenic missense variants are shown in orange above the gene and of VUSs in grey below the gene. *BRCA1* exon 11 and *BRCA2* exons 10 and 11 harbour no (likely) pathogenic missense variants and are defined as cold spots. *P/LP*: pathogenic/likely pathogenic, *aa*: amino acid. Taken from Dines et al. (2020)[215]

4.2. **BRCA2**

BRCA2 is on chromosome 13 q13.1, encompassing genomic positions from 32,315,077 to 32,400,268 (*Homo sapiens*, GRCh38.p14, <https://www.ncbi.nlm.nih.gov/gene/675>). It contains 27 exons, encoding a peptide sequence spanning 3,418 amino acids (**Figure 24**).

In contrast to the multifunctional activities of *BRCA1*, the primary function of *BRCA2* is in HR [35]. It has an N-terminal domain that binds to *PALB2* [216]. The central region encoded by exon 11 contains eight interspersed BRC motifs that bind *RAD51* [217,218]. At the C-terminus, there is the DNA binding domain, which includes a helical domain (H), three oligonucleotide binding (OB) folds and a tower domain (T), which may facilitate *BRCA2* binding to both single-stranded DNA and double-stranded DNA [219]. Additionally, the H and OB1,2 domains associated with *DSS1*, has been linked to *BRCA2* protein stabilization [220,221]. The C-terminus also contain a two nuclear localization sequence (NLS), and another BRC domain that binds *RAD51* [222].

Specifically, *BRCA2* regulates the recruitment of *RAD51* to sites of DNA damage and the oligomerisation of *RAD51* with single-strand DNA that is necessary to form recombination intermediates [223–225]. Several missense/in-frame variants in the *PALB2* binding domain and DBD, disrupt *BRCA2* function and are associated with high cancer risk [184,226]. Consequently, both have been considered clinically relevant domains (**Figure 25**)[122].

4.3. **PALB2**

PALB2 gene is in 16p12.2, encompassing positions from 23,603,165 to 23,641,310 (*Homo sapiens*, GRCh38.p14,

<https://www.ncbi.nlm.nih.gov/gene/79728>). It contains 13 exons encoding a sequence of 1186 residues. Its N-terminal coiled-coil domain interacts with the coiled-coil domain of BRCA1, and the WD40 domain in the C terminus interacts with the BRCA2, hence it provides a physical link between the BRCA proteins (**Figure 24**) [227]. The BRCA1–PALB2 interaction is a prerequisite for the recruitment of BRCA2 and RAD51 to the site of DNA damage and for HR, and the PALB2-BRCA2 was shown to be essential for loading RAD51 onto RPA-bound single-strand DNA (ssDNA) [206,216].

It has been described that variants in both coiled-coil and WD40 impair protein function [228], and VCEP for *PALB2* defined both as clinically relevant domains (<https://clinicalgenome.org/affiliation/50039>).

4.4. Homologous Recombination Repair (HRR)

Cells are continually exposed to a variety of DNA-damaging substances, which can be from endogenous (i.e., reactive oxygen species and free radicals from normal metabolism, intrinsic replication errors) or exogenous sources (i.e., ultraviolet radiation, chemical exposures). To combat DNA damage, cells have evolved mechanisms—collectively termed the DNA-damage response (DDR)—to detect DNA lesions, signal their presence and promote their repair [229]. The wide diversity of DNA-lesion types necessitates multiple, largely distinct DNA-repair mechanisms including nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), nonhomologous end joining (NHEJ) and homologous recombination repair (HRR), in which nearly all known HBOC susceptibility genes are involved (**Figure 6**).

The two-ended DNA DSBs, the most cytotoxic form of DNA damage, pose an immediate threat to the stability of the genome by provoking chromosome

rearrangements and disrupting gene structure and function [87]. The HRR and NHEJ are the major pathways predominantly involved in the repair of two-ended DSBs. The HRR uses sister chromatids as templates for the accurate repair of replication-associated DSBs during late S and G2 phases of the cell cycle. While NHEJ, are error-prone and result in translocations and deletions, the HR pathway is error-free and contributes to the maintenance of genome stability [230].

The protection of the genome by HRR involves sensors that detect the DNA damage, signalling mediators that recruit and activate the effectors, that execute the repair and activate cell cycle checkpoints. After the initial sensing of the DSB by MRN and the activation of ATM, H2AX is phosphorylated, forming γ -H2AX[231]. It acts as a mediator of MDC1 and RNF8 that allows the BRCA1-Abraxas-RAD80 complex to associate with ubiquitylated histones near the sites of DNA damage. Next, BRCA1 is involved in processing DSBs through its interaction with CtIP and the MRN complex (MRE11-RAD51-NBS1), which process the DNA ends and generates ssDNA with 3' overhangs. The ssDNA is coated by RPA, which is subsequently displaced by recombination mediators to enable the formation of a RAD51 nucleoprotein filament. BRCA1 is also required for RAD51 recruitment to DSBs through its interaction with PALB2 linked to BRCA2. This last loads RAD51 onto the RPA-coated ssDNA. RAD51 filament assembly is stimulated by RAD51 paralogs, RAD51B, RAD51C, RAD51D, XRCC2, XRCC3 and SWSAP1. This RAD51-ssDNA nucleoprotein filament mediates the homology search that defines HRR by invading duplex DNA molecules, forming a D-loop. The invading strand is then extended by DNA polymerase using the invaded DNA molecule as a template. Finally, DNA ligases and endonucleases solve all DNA structures to complete the HR (**Figure 26**)[35,87,231].

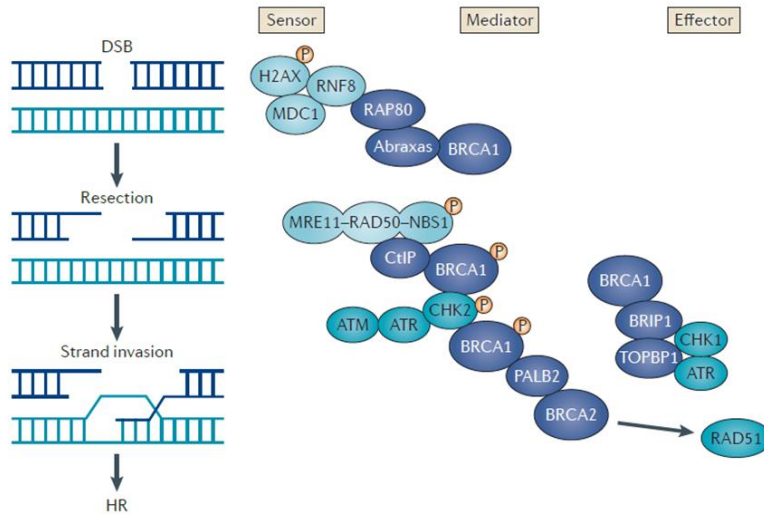


Figure 26. DNA double-strand break (DSBs) repair by homologous recombination (HR).

In response to DSBs, sensors (light blue) detect the damage, and signalling mediators recruit or activate effectors that repair the damage and activate cell cycle checkpoints. BRCA1-containing macro-complexes (dark blue) are crucial mediators of the DNA damage response. *Adapted from Roy et al. (2011)[35].*

4.5. Alterations in HRR Genes and Homologous Recombination Repair Deficiency (HRD)

The biallelic inactivation of any of the key genes involved in the HRR pathway, including *BRCA1*, *BRCA2* and *PALB2*, results in a status of homologous recombination deficiency (HRD) [232–234]. Tumours exhibiting HRD are dependent on alternative, error-prone pathways, such as NHEJ or microhomology-mediated end joining (MMEJ)[235] for the repair of DSBs. This results in a characteristic mutational footprint across the genome. Analyses of tumour-derived genome sequences have demonstrated that the loss of *BRCA1* or *BRCA2* results in a distinctive pattern of base-substitution mutations, designated as signature 3 [236]. Moreover, nonsense and frameshift germline *PALB2* variants, in addition to epigenetic silencing of *RAD51C* and *BRCA1*, are

also strongly associated with this mutational signature due to HRR impairment [234].

Although HRD may be advantageous for tumour development, it also makes cells selective sensitive to compounds that increase the demand on HRR [88]. PARPi are one example of such compounds, as they induce replication fork stalling and collapse, leading to increased DSBs. This, combined with the inability to repair DSBs through HRR, ultimately results in selective tumour cell death (**Figure 9**) [237]. Thus, several efforts have been made to identify the molecular features of tumours exhibiting a “BRCAness” phenotype -tumours that either are or mimic the *BRCA1/2* deficiency, i.e., HRR deficient tumours [238].

Testing for the consequences of an impaired HRR pathway can be performed by probing the genome for evidence of genomic abnormalities, such as genetic alterations in HRR genes and genomic scars or signatures (**Figure 27**)[239]. For instance, the HRDetect is a breast cancer-specific predictor of HRD based on six mutational signatures associated with *BRCA1/2* dysfunction [233]. Similarly, two commercial tests based on genomic scars have been approved by the FDA as predictive biomarkers of PARPi response in OC. The “myChoice HRD” kit by Myriad Genetics, which inquiries about the three genomic scars: i) the genomic patterns of loss of heterozygosity (LOH)[240]; ii) the number of telomeric allelic imbalances (TAI)[241]; and iii) large-scale transitions (LST) [242]. And the “FoundationFocus™ CDx BRCA LOH”, which is conceived to identify the occurrence of variants in *BRCA1/2* and the amount of the genome impacted by LOH (reviewed in [243]). However, genomic tests are static, meaning that they can detect if there has been an HRR defect in the history of the tumours but fail

to provide accurate HRR status information at the time of treatment. In addition, these assays are technically complex and carry high costs [244,245].

As an alternative, a functional test to assess HRD was developed by the Experimental Therapeutics group from VHIO (**Figure 27**). It consists of an immunofluorescent nuclear *foci* detection of HRR effector RAD51, together with γ H2AX (as a marker of DSBs) and BRCA1 in tumour S/G2-cells. The rationale of this test was supported by basic evidence that the formation of RAD51 nuclear *foci* of the downstream effector of the HRR pathway indicates its functionality (**Figure 26**)[246]. RAD51 test has been demonstrated to be a good predictive biomarker of response for PARPi treatment, and it also had shown high sensitivity in the identification of HRD tumours from germline carriers of *BRCA1/2* and *PALB2* variants [247,248].

Given that has been widely demonstrated that HRR status (genomic and functional) identify *BRCA1*, *BRCA2* and even *PALB2* deficient tumours, some works proposed its usage for classifying VUSs [183,234,249]. As exemplified by Duran-Lozano et al. (2019)[183] which reported the use of RAD51 tumour assay as an additional tool in the classification of a VUS in *PALB2* [183]. However, clear guidance on how to use this HRR tumour status in variant classification is still lacking and only very few tumour characteristics are robust enough to be included in current ACMG/AMP rules for germline variant classification in cancer predisposition genes [250]. Indeed, for *BRCA1/2* only histopathological tumour features have been calibrated to be used in variant classification[198,199], and no tumour evidence is currently available for *PALB2*.

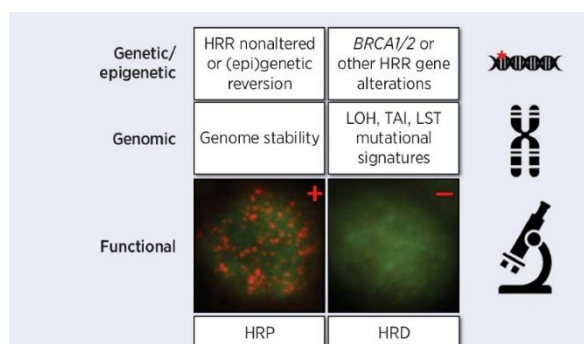


Figure 27. HRD tumour detection approaches.

Functional images show the presence (+) or absence (-) of nuclear RAD51 foci in geminin-positive tumour cells. *HRR*: Homologous recombination repair, *LOH*: loss of heterozygosity, *TAI*: telomeric allelic imbalance, *LST*: large-scale transitions, *HRP*: Homologous recombination proficiency, *HRD*: Homologous recombination deficiency. Adapted from Pellegrino et al. (2022) [245].

5. Rationale for the Thesis

The missing heritability of HBOC may be attributed to many potentially pathogenic variants that are not detected by current panel testing, such as deep intronic variants. Additionally, when detected, many of these variants are typically overlooked during standard diagnostic filtering pipelines due to the incomplete understanding of elements governing splicing on these regions. Thus, leading to a limited ability to identify and interpret potential pathogenic variants in there.

Although different bioinformatic tools accurately identify splicing exonic variants affecting canonical splice sites (MES, SSF, HSF)[151,165,251,252] or altering SREs (ESRseq, HZEI, HAL) [134,171,253], there has been limited success in identifying deep intronic variants [149,157,172]. Deep learning tools such as SpliceAI [168] could outperform classical prediction approaches [163], but little is known about its performance for identifying deep intronic variants affecting splicing either by creating de novo/enhancing splice sites or SREs [149,172].

Moreover, the arrangement of *cis* splicing elements in deep intronic regions, especially those corresponding to SREs, would configure sequences with structures like those of canonical exons [254,255]. This landscape of regulatory zones would favour the generation of pseudoexons if a new deep intronic variant helps to define these structures, but there are no algorithms dedicated to identifying potentially “exonizable” regions across intronic sequences. **Part 1** of this thesis assesses SpliceAI’s performance combined with ESRseq scores to identify spliceogenic deep intronic variants using literature and experimental datasets. It provides and tests the first *silico* framework to prioritize deep intronic variants for their experimental RNA analysis, considering the landscape of splicing elements and highlighting the importance of SRE balance in the inclusion of sequences in mature RNA.

Despite that *in vitro* splicing assessment provides one of the strongest pieces of evidence for proper variant interpretation, most laboratories limit experimental analyses to variants at canonical sites. Consequently, many variants outside these regions are not consistently assessed across laboratories, leading to conflicting or uncertain interpretations. This inconsistency complicates variant interpretation and hinders patients’ clinical management. Additionally, variants leading to in-frame isoforms are challenging to classify due to limited knowledge about in-frame encoded proteins. One clear example of both issues are the variants located in the in-frame exon 18 of *BRCA1*. This exon encodes for 26 amino acids from the BRCA1 C-terminal (BRCT) domain (from Asp1692 to Trp1718), which represent 1.4% of the BRCA1 protein. Additionally, several missense pathogenic variants in exon 18 have been associated with high cancer risk, such as c.5089T>C p.(Cys1697Arg) [256], c.5095C>T p.(Arg1699Trp) [108]

and c.5117G>A p.(Gly1706Glu) [257], thus highlighting the clinical relevance of this exon.

Exon 18 skipping ($\Delta 18$) is a minor alternative in-frame splicing event in healthy controls [176], and variants at canonical splicing dinucleotide positions have been described to produce exclusively this transcript, such as the c.5075-2A>G and the c.5152+1C>G [258]. However, variants outside canonical splicing sites, such as c.5075-18G>A, c.5075-6C>A, c.5075-3C>T, c.5152+3A>C, c.5152+6T>C, and c.5152+6T>G, show conflicting or uncertain interpretations in ClinVar (accessed in April 2024) depending on whether mRNA splicing assays were or not performed. **Part 2** of this thesis aims to conduct a comprehensive study of variants located within exon 18 and adjacent intronic regions with a focus on interpreting these variants through mRNA experiments. With this purpose, the ENIGMA consortium approved a collaborative project inviting all members to contribute mRNA samples from carriers of exon 18 variants, along with splicing data (e.g., minigene assays) and functional or clinical information related to variants in this exon.

While variant-specific features such as splicing evidence, population data, computational data, functional data, and segregation data can be used for classifying *BRCA1/2* and *PALB2* variants, the usage of tumour characteristics is poorly explored. Currently, only histopathological tumour features are considered in multifactorial clinical evidence for variant classification for *BRCA1/2* [198,199], but any tumour characteristic has not been yet considered for *PALB2*. In **Part 3** of this thesis, we propose that tumour HRR status by RAD51 *foci* detection might be a good predictor of variant pathogenicity or benignity and could be integrated as tumour phenotype criteria in variant classification guidelines for HBOC high-risk genes.

HYPOTHESIS AND OBJECTIVES

Hypothesis

VUS can affect the sequence or abundance of mRNA, and it could be the genetic cause of HBOC susceptibility. We hypothesize that optimizing the usage of *in silico* tools dedicated to identifying splicing defects will prioritize with high accuracy variants for RNA analysis, helping to characterize their effect and, in consequence, to classify them. Secondly, the comprehensive mRNA, functional, and clinical information analyses of variants, especially those leading to in-frame transcripts that encode proteins with unknown functionality, will decrease the uncertainty of these variants and provide specifications to improve their classification. Lastly, we hypothesize that tumour HRR status by RAD51 *foci* detection can be used as evidence in favour or against the pathogenicity of *BRCA1/2* and *PALB2* variants.

Objectives

The main objective of this work is to increase the number of patients receiving a genetic diagnosis of HBOC by improving the accuracy of classifying germline genetic variants, through optimising the use of splicing evidence and introducing new tumour evidence. This will lead to more accurate risk predictions and improved personalized treatment strategies for individuals with these genetic predispositions.

According to this, the specific aims of this project are:

1. To evaluate the splicing effect of deep intronic variants of uncertain significance of HBOC risk genes by several bioinformatic tools and confirm experimentally the predicted effects.
2. To standardize the classification of variants disrupting splicing of in-frame exons using *BRCA1* exon 18 as an example (a collaborative project of the ENIGMA consortium).
3. To investigate the feasibility of determining the tumour HRR status, by RAD51 *foci* detection, as evidence for informing pathogenicity or benignity of *BRCA1/2* and *PALB2* germline variants.

MATERIALS AND METHODS

Part 1

Datasets

Four datasets were used in this study: (i) 233 blood-detected germline deep intronic variants (located >20 nucleotides (nt) from known exon-intron boundaries) collected from literature using the keywords “pseudoexon” and “deep intronic” (**Supplementary Table 1**). Their splicing effect had been assessed experimentally by gel electrophoresis or Sanger sequencing using blood, lymphoblastoid cell lines, midigenes, or minigenes. We defined as deep intronic variants those that were more than 20 nt away from the nearest exon-intron junction because these nucleotides are outside of the known splice site consensus sequences [130,132]; (ii) 1161 exonic variants compiled by Tubeuf et al. (2020) [172] to compare the accuracy of SREs-dedicated algorithms for predicting splicing alteration by affecting SREs (**Supplementary Table 2**); (iii) an additional supportive dataset of 207 somatic deep intronic splice-altering variants detected and characterised in tumour by both whole-genome sequencing and RNA-sequencing, retrieved from Jung et al. (2021) [259] (**Supplementary Table 3**); and (iv) an experimental dataset from patients of Hospital Universitari Vall d’Hebron, comprised of 33 unique intronic germline variants and selected according to their blood RNA availability (**Supplementary Table 4**). The splicing impact of all 33 variants was assessed *in silico* with SpliceAI and ESRseq and experimentally characterised using whole blood RNA.

Additionally, we retrieved the sequences of all exons and adjacent 100 intronic nucleotides of HBOC and Lynch genes (*BRCA1*, *BRCA2*, *PALB2*, *BRIP1*, *RAD51C*, *RAD51D*, *PTEN*, *TP53*, *CDH1*, *CHEK2*, *BARD1*, *STK11*, *MSH2*, *MSH6* and *MLH1*) according to their NCBI reference transcripts (**Supplementary Table 5**), to compare the splicing regulatory elements balance between canonical exons

and pseudoexons from the literature dataset. Taking into account that most exons contain less than 300 nt, and that the mean length is 147 nt [260], exceptionally long exons (e.g., exons 11 of *BRCA1* and 10 and 11 of *BRCA2*) were not included.

***In Silico* Variant Annotation and Analysis for All Datasets**

SpliceAI (v1.3; <https://github.com/Illumina/SpliceAI> accessed January 2021) was run to obtain the delta (Δ) score of a variant (DS), defined as the maximum of DS for acceptor gain (DS_AG) and DS for donor gain (DS_DG) for deep intronic variants and the maximum of DS for acceptor loss (DS_AL) and DS for donor loss (DS_DL) for exonic variants [168]. The DS value ranges from 0 to 1 and can be interpreted as the probability of a variant being splice-altering. We considered the SpliceAI estimations of a gained/lost splice site in the 4999 nucleotides located on each side of the variant.

For the prediction of variant-induced SREs alterations, the Δ ESRseq value was calculated according to Ke et al. (2011) [134] as the difference between the ESRseq scores of a variant sequence of 11 nucleotides (5 nucleotides at each side of the variant) and wild type sequence scores (ESRseq VAR- ESRseq WT).

To get the SRE distribution along any genomic region, we focused exclusively on the WT ESRseq scores, which were also calculated according to Ke et al. (2011)[134], obtaining individual nucleotide scores of each one of the positions of a sequence. The sum of the scores for each nucleotide in a given sequence was defined as “area”. To account for the differences in size between exons, the area was divided by the number of nucleotides of each exon obtaining a normalized SRE area value (Normalized SRE area = $\sum$ ESRseq scores of all nt of region of interest/size of region of interest). This value was used to compare the

SREs balance of all constitutive exonic sequences from HBOC and Lynch genes, with all pseudoexons collected in the literature dataset.

Alamut Visual software v.2.10 (Interactive Biosoftware) was used for annotation of variants included in the patients' experimental validation dataset, providing data of allele frequencies in general population from the Genome Aggregation Database (gnomAD v2.1) and variant classification reported in ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/> accessed March 2021), considering the number of variant submitters and reviews by expert panel.

Statistical Analysis

Performance values of sensitivity, specificity, overall accuracy, Positive and Negative Predictive Values, False and Positive Discovery Rates, and Matthew's correlation coefficient (MCC) were calculated with different *in silico* tool thresholds individually and in sequential combinations. The statistical measures used for evaluation of the performance are depicted in **Table 6**.

Table 6. Formulas for performance evaluation of *in silico* tools.

Measure	Formula
Accuracy	$TP+TN/(TP+TN+FP+FN)$
PPV (Positive Predictive Value)	$TP/(TP+FP)$
NPV (Negative Predictive Value)	$TN/(FN+TN)$
Sensitivity	$TP/(TP+FN)$
Specificity	$TN/(TN+FP)$
MCC (Matthew's correlation coefficient)	$((TP*TN)-(FP*FN))/SQRT((TP+FP)(TP+FN)(TN+FP)(TN+FN))$

TP: true positives, TN: true negatives, FP: false positives, FN: False negatives

The SpliceAI (4999 bp) optimized threshold was calculated based on the highest MCC using the literature dataset. In the case of ESRseq, the cut-off optimization

was estimated by maximizing the sum of specificity and sensitivity since in the optimization based on MCC, the sensitivity/specificity was too unbalanced.

Analysis of variance (ANOVA) and T-test were used to compare the means of SREs scores between canonical exons, adjacent intronic sequences and pseudoexons with their adjacent sequences. T-test was used to compare the absolute SRE differences between the group of variants causing pseudoexons and variants without any effect. All tests were performed using Graphad Prism 6.

Experimental RNA Analysis in Patient's Data Set

The patient dataset included unrelated cases from HBOC families ascertained through the Familiar Cancer Unit of Hospital Universitari Vall d'Hebron, HUVH (Barcelona, Spain). A total of 33 unique intronic germline variants were selected based on blood RNA availability in *ATM*, *BARD1*, *BRCA2*, *FAM175A*, *MLH1*, *MSH2*, *MUTYH*, *NF1*, *PTEN*, *RAD51C*, *RBBP8*, and *TP53* (**Supplementary Table 4**). All variants were identified in DNA by massively parallel sequencing using Illumina technology with a diagnostic routine panel of coding exons and exon-intron boundaries or by a research panel specifically designed to sequence whole intronic regions and confirmed by Sanger sequencing [72,261]. Healthy individuals without familial cancer history were included as negative controls.

Reverse Transcription-PCR (RT-PCR) and Sanger Sequencing

Total RNA from variant carriers and controls was isolated from 10 mL of peripheral blood using Trizol reagent (Invitrogen, Waltham MS, US) following the manufacturer's protocol. RNA was cleaned-up using RNeasy Mini Kit (QIAGEN, Hilden, Germany) with an additional step of DNase digestion using RNase-Free DNase Set (QIAGEN) or Ambion™ DNase I RNase-free (ThermoFisher, Waltham MS, US) in samples with a low RNA concentration. A total of 100 ng of RNA were

retrotranscribed to yield cDNA using PrimeScript RT reagent kit (Takara Bio, Shiga, Japan), combining random and oligo-dT primers. PCR primers were designed to amplify a whole exon upstream and downstream from the intron containing the variant of interest. PCR assays were performed in 25 μ L reaction volume containing 50 ng of cDNA as template, using BioTaq DNA Polymerase (Meridian Bioscience, Cincinnati OH, US). Samples were denatured at 95° for 10 min, followed by 35 cycles consisting of 95° for 30 s, 56–62° for 30 s, and 72° for 1–7 min; and a final extension step at 72° for 7 min. All primers used in this study and amplification conditions are detailed in **Supplementary Table 6**.

Capillary electrophoresis using the 4200 TapeStation device (Agilent, Santa Clara CA, USA) with High Sensitivity D1000 ScreenTape reagents (Agilent) was used to assess the quality of PCR products. These products were enzymatically cleaned using ExoSAP-IT[®] PCR Product Cleanup (Affimetrix, ThermoFisher Scientific, Waltham MS, US) and bidirectionally sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Waltham MS, US). Sequencing products were run in an ABI3130xl Genetic Analyzer (Applied Biosystems) and were analysed using Sequencing Analysis v6.0 software (Applied Biosystems). The reference transcripts based on GRCh37 (hg19) genome and listed in the **Supplementary Table 5** were used for sequence alignment and transcript annotation.

Qualitative Analysis by Capillary Electrophoresis of Fluorescent Amplicons

RT-PCRs using primers labelled with 6-Carboxyfluorescein (6-FAM) at the 5' end were performed in triplicate (see labelled primers in **Supplementary Table 6**). These fluorescent products were assessed by high-resolution capillary electrophoresis to detect and annotate all amplified transcripts. Specifically, 0.5 μ L of the PCR products were run in an ABI3130xl Genetic Analyzer instrument

(Applied Biosystems) for fragment analysis. GeneScan 500 and 1000 ROX (Applied Biosystems) was used as internal size standard. Electrophoresis conditions were the same for all samples: 60 °C, 12 sec injection at 1.2 KV and 2000 s run at 12 KV. Data visualization and peak size-calling was performed using GeneMapper software v5.0 (Applied Biosystems).

The maximum fragment size that could be detected was 946 bp, using the internal size-standard 1000 ROX.

Ethical Considerations

This study was approved by the Clinical Research Ethics Committee (CEIC) of Hospital Universitari Vall d'Hebron, Barcelona, Spain. All individuals received genetic counselling and signed written informed consent for HBOC panel genetic testing and research studies.

Part 2

ENIGMA Proposal

In 2015, ENIGMA Steering Committee approved the collaborative project proposal detailed in **APPENDIX II**.

Thanks to ENIGMA Splicing Working Group collaborators a total of 166 variants located in *BRCA1* exon 18 and/or in adjacent intronic regions were evaluated (**Figure 35**). Each variant was studied with a minigene system, 32 variants were analysed in blood-derived RNA from patients, and 18 variants were also analysed in RNA from mouse embryonic stem cells (mESCs).

Annotation and *In Silico* Analysis

We used Alamut Visual Plus software v.1.7 (Sophia Genetics) to annotate variants using the reference transcript NM_007294.3 and the legacy exon numbering of *BRCA1*, with exon 4 missing due to a correction made after the initial description of the gene (GenBank U14680.1). Alamut software provided information about the coding effect, and variant classification reported in the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>) according to the number of variant submitters and review status. Of note, all variants *reviewed by expert panel* in ClinVar were those *BRCA1* variants classified by ENIGMA as external expert panel using former guidelines, except for the c.5090G>A variant which was included in the pilot dataset of variants classified under the new specifications of ENIGMA as VCEP outside this project [122].

SpliceAI (v1.3; <https://github.com/Illumina/SpliceAI> [168]) was used to obtain the delta score (DS) of 162 variants. For the deletion-insertion (delins) variants c.5077_5079delinsTTCATTCTGCA, c.5077_5080delinsTTCATTCTGC, c.5077_5080delinsTTGATTCTG and c.5092_5093delinsAG, the DS was obtained

through <https://spliceailookup.broadinstitute.org/>. We considered the SpliceAI raw estimations of a gained/lost splice site in a ± 4999 window. SPIP[262] was used to identify variants potentially altering the branch point through BPP (a sequence-based algorithm for branch point prediction)[173] and the polypyrimidine tract using MES (a maximum entropy distribution model)[252]. For detecting variants disrupting splicing regulatory elements (SREs), first we mapped all SREs along exon 18 and adjacent intronic regions, covering a region that included all 166 variants analysed in this study (from position c.5075-107 to c.5152+105). Specifically, we calculated ESRseq wild-type scores (ESRseq WT) according to Ke et al. (2011)[134], which is based on the quantitative evaluation of all possible RNA hexamers as potential exonic splicing regulatory sequences. All ESRseq positive values were considered as splicing enhancers (ESE: exonic splicing enhancer, ISE: intronic splicing enhancer) and negative values as silencers (ESS: exonic splicing silencer, ISS: intronic splicing silencer). Then, to identify variants potentially altering SREs we obtained Δ ESRseq scores (ESRseq from variant minus ESRseq WT) through HexosplICE (<http://bioinfo.univ-rouen.fr/HEXoSplice/index.php>)[172].

Regarding missense variants, we obtained BayesDel scores that did not incorporate population allele frequencies (noAF) [263] from BRCA Exchange website (<https://brcaexchange.org/>).

Variant annotations and *in silico* predictions are included in **Supplementary Table 9**.

Patients' mRNA Analysis

The splicing profile of exon 18 was assessed in RNA isolated from blood-derived cells of 51 carriers of 32 variants: 13 from Institut Català d'Oncologia

(L'Hospitalet del Llobregat, Spain), nine from Centre François Baclesse (Caen, France), eight from Hospital Universitari Vall d'Hebron (Barcelona, Spain), six from Masaryk Memorial Cancer Institute (Brno, Czech Republic), four from Hospices Civils de Lyon - The French Genetic and Cancer Group Unicancer consortium (French GGC network), three from Hospital Universitario la Princesa (Madrid, Spain), three from University Hospital Southampton (Southampton, United Kingdom), two from Manchester University NHS Foundation Trust (Manchester, United Kingdom), one from Ghent University Hospital (Ghent, Belgium), one from Hospital Clínico San Carlos (Madrid, Spain) and one from Hospital Universitari Sant Joan de Reus (Reus, Spain) (**Supplementary Table 10**). In addition, 46 blood RNA samples from different non-carrier controls were included in the study.

All variants included in this study were initially detected through DNA molecular testing in index cases of high-risk families with a history of hereditary breast and ovarian cancer (HBOC). All patients underwent genetic counselling and provided written informed consent for HBOC panel testing, as well as for their participation in research studies according to the respective national guidelines. RNA from patients and non-carrier controls was isolated and purified following different techniques depending on the centre (**Supplementary Table 10**). All samples were sent to Vall d'Hebron Institute of Oncology (VHIO, Barcelona, Spain), where RNA analyses were performed. Per each sample, at least two separate reverse transcriptions (RT) were performed using a total of 100-400ng of RNA to yield cDNA using PrimeScript RT reagent kit (Takara Bio, Shiga, Japan) and combining random with oligo-dT primers. RT-PCR assays were performed in 25 µL reaction volume containing 50-100ng of cDNA as template, using BIOtaq DNA Polymerase (Meridian Bioscience, Cincinnati OH, USA). PCR conditions

and primers used for each experiment are detailed in **Supplementary Table 11 A**.

RT-PCR assays were performed using primers labelled with 6-Carboxyfluorescein (6-FAM) at the 5' end, and fluorescent amplicons were analysed by high-sensitivity capillary electrophoresis. For splicing fraction determination, these RT-PCR products were run in an ABI3130xl Genetic Analyzer instrument (Applied Biosystems, Waltham, MS, USA) for fragment analysis. GeneScan 500 ROX (Applied Biosystems) was used as an internal size standard. Electrophoresis conditions were the same for all the samples: 60°C, 12sec injection at 1.2KVolts, and 2000sec run at 12KVolts. Data visualization and peak size-calling were performed using GeneMapper software v5.0 (Applied Biosystems). The relative abundance (splicing fraction) of each transcript was calculated as the ratio between the peak area of one transcript and the Σ of all peak areas (all transcripts) detected in the same sample. Splicing events were confirmed by Sanger sequencing or inferred based on isoforms already described in the literature and by fragment size (**Supplementary Table 12**).

Additionally, we calculated the inferred per-allele splicing fraction for full-length (FL) and exon 18 skipping ($\Delta 18$) as two times the level of a specific transcript observed in the variant carrier minus the average level of the same transcript in multiple non-carrier control samples of the same type using data from assays performed in tandem [194].

GraphPad Prism 8 (GraphPad Software) was used for data analysis and graph building.

RT-PCR RNA sequencing for an additional carrier of c.5152+7A>G was performed by Ambry Genetics, as detailed in Horton et al. (2022)[264], using primers hybridizing in exons 16 and 24 of *BRCA1*. Moreover, Invitae analysed the splicing profile of another carrier of this variant by targeted RNA sequencing as detailed in Kamps-Hughes et al. (2023)[265].

Full-length (FL) specific RT-PCR assays followed by Sanger sequencing were performed to determine the contribution of a variant allele to the expression of the normal transcript. This specific RT-PCR encompasses a region containing an informative heterozygous position (**Supplementary Table 11 A**). The qualitative analysis was done by comparing the abundance of both alleles in cDNA with their presence in DNA (depending on sample availability). For Sanger sequencing, the RT-PCR products were enzymatically cleaned using ExoSAP-IT® PCR Product Cleanup (Affimetrix, ThermoFisher Scientific, Waltham, MS, USA) and bidirectionally sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing products were run in an ABI3130xl Genetic Analyzer (Applied Biosystems) and were analysed using Sequencing Analysis v6.0 software (Applied Biosystems). Reference transcript *BRCA1* NM_007294.3 from the GRCh37 genome build was used for sequence alignment and transcript annotation.

For three variants analysed using RNA from lymphoblastoid cell lines (LCLs), allele-specific FL expression was measured by performing a quantitative primer extension assay (SNaPshot Multiplex Kit, Applied Biosystems), using specific primers targeting the variants of interest as previously described in Soukarieh et al. (2016)[266] (**Supplementary Table 11**).

Minigene Splicing Assay

To evaluate the impact on splicing of the 166 variants, the Inserm Unit U1245 (Rouen, France) performed a cell-based splicing assay based on the comparative analysis of the splicing pattern of wild-type (WT) and mutant reporter minigenes. These 166 variants were selected according to the following criteria: (i) all analysed with patient samples to compare both results and (ii) all variants reported in human variant databases (BIC, ClinVar, Cosmic, dbSNP, ESP, ExAC, LOVD, and UMD) in 2017 within the *BRCA1* exon 18 locus including the exon and 20 nucleotide flanking intronic regions.

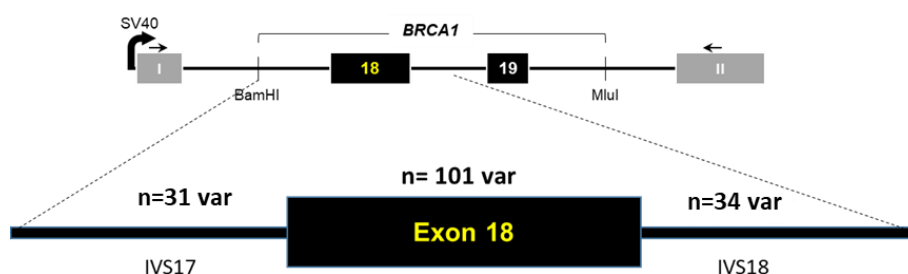


Figure 28. Schematic representation of the pSPL3mBK2-BRCA1-ex18_19 minigene.

Boxes indicate exons and lines in between represent introns. The minigenes were generated by inserting a genomic fragment encompassing *BRCA1* exons 18, 19 and their intronic flanking sequences (*BRCA1* NM_007294.3 sequence from c.5075-182 to 5193+157) into the intron of pSPL3mBK2-by using *Bam*HI and *Mlu*I restriction sites.

The variants of interest were introduced by site-directed mutagenesis using the two-stage overlap extension PCR method [267] and a combination of specific primers (**Supplementary Table 11 B**) and the WT construct as template consisting of *BRCA1* exons 18 and 19 and part of their intronic flanking sequences (*BRCA1* c.5075-182_5193+157; sequence depicted in **Figure 28**). WT and mutant segments were inserted into the intron of the pSPL3mBK2 minigene using *Bam*HI and *Mlu*I restriction sites, yielding pSPL3mBK2-BRCA1-ex18_19

minigenes (**Figure 28**). The integrity of all constructs was confirmed by Sanger sequencing. Then, minigenes (400ng/well) were transfected in parallel into HeLa cells grown at ~50% confluence in 12-well plates using the FuGENE 6 Transfection Reagent (Roche Applied Science). HeLa cells obtained from American Type Culture Collection (ATCC) were tested negative for mycoplasma contamination and cultured in Dulbecco's modified Eagle medium (Life Technologies) supplemented with 10% fetal calf serum in a 5% CO₂ at 37°C. Total RNA was extracted 24 hours post-transfection using the NucleoSpin RNA Kit (Macherey-Nagel). The splicing patterns were analysed by semi-quantitative fluorescence RT-PCR (30 cycles) by using the OneStep RT-PCR Kit (Qiagen), in a 25 µL reaction volume, with 200 ng total RNA and specific minigene primers (**Supplementary Table 11 B**). Fluorescent amplicons were resolved under denaturing conditions by capillary electrophoresis on an automated 3500 Genetic Analyzer (Applied Biosystems) using 500 ROX™ Size Standard (Applied Biosystems). Electropherograms were then analysed by using the GeneMapper v5.0 Software (Applied Biosystems) and peak areas were used to quantify the relative levels of each transcript as previously described for patients' mRNA analysis. Results are shown as the average of 2 independent experiments, and GraphPad Prism 8 (GraphPad Software) was used for data analysis and graph building.

Mouse Embryonic Stem Cells (mESCs)-Based Functional Assay

To directly correlate the mRNA splicing profile of *BRCA1* exon 18 variants with effects on BRCA1 protein function, a total of 18 variants were selected according to their splicing profile (normal, partial, complete) for functional characterization using a mouse embryonic stem cell (mESC)-based assay. This

assay was done by the Department of Human Genetics from Leiden University Medical Center.

Functional Analysis of hBRCA1 Variants

Variants were generated into the full-length human *BRCA1* gene and introduced in mESC as described previously for the analysis of *BRCA2* variants [268].

In brief, a pUC19 plasmid containing *BRCA1* exon 18 and about 100 nucleotides of intron sequence upstream and downstream of the exon was generated to enable the introduction of variants (QuickChange site-directed mutagenesis from Stratagene, La Jolla CA, USA). Following sequence confirmation, homology arms were added to the gene fragment containing the variant by PCR. Subsequently, Red/ET recombineering was used in combination with a positive/negative selection procedure to introduce the variant into the full-length human *BRCA1* gene located on the bacterial artificial chromosome (BAC) RP11-812O5. The BAC containing the variant was transfected in a *hemizygous mouse Brca1* mESC line (*Brca1*^{-/loxP} *Pim1*^{DR-GFP/WT} 129/Ola E14 IB10) using 6 µg of BAC DNA to transfect 5 x 10⁶ cells per variant using Lipofectamine 2000 (Invitrogen, Waltham MS, USA). Transfected cells were seeded in 90mm cell culture dishes and cultured in the presence of G418 (200 µg/ml), starting 24 hours post-transfection. Thirteen days post-transfection cells were treated for 16 hours with 1.0 µM 4-Hydroxytamoxifen (4-OHT) (Sigma-Aldrich, St. Louis, MO, USA) to remove the *mBrca1* conditional allele. Throughout the experiment, cells were cultured on gelatine-coated dishes using buffalo rat liver cell (BRL)-conditioned mESC medium.

Complementation Assay

As BRCA1 activity supports mESC viability, the impact of a variant on protein function can be assessed by measuring the capacity to form viable clones after the removal of the conditional *mBrca1* allele. For this purpose, 6000 cells were seeded on 60-mm dishes and cultured for six days in BRL-conditioned mESC medium containing puromycin (1.8 µg/mL). Viable clones were visualized by methylene-blue staining seven days post 4-OHT treatment.

For each variant, the number of viable clones was compared to the complementation of WT BRCA1 expressing cells and categorized into one of the following complementation categories: poor (<20% viability, including no complementation), reduced (20%–50% viability), or good (>50% viability).

Homology-Directed Repair (HDR) Assay

In addition to the viability assay, cells were seeded for the direct repeat-green fluorescent protein (DR-GFP) reporter assay after the removal of the conditional *mBrca1* allele. In this assay, the capacity of *BRCA1* variant-expressing cells to repair an I-Sce1 induced DNA double-stranded break via HDR can be measured. For each variant, 80×10^3 cells were seeded in four to six wells on a 48-well plate. On the next day, cells were transfected with an I-Sce1-mCherry plasmid (0.75 µg/well) using Lipofectamine 2000 (0.85µl/well) for 16 hours. Two days after transfection, flow cytometry (Novocyte, Agilent, Santa Clara CA, USA) was used to determine the fraction of GFP-expressing cells in the mCherry-positive population as an indirect measure of HDR activity. For each variant, the relative HDR activity was calculated as the ratio between the percentage of GFP-positive cells observed in *BRCA1* variant-expressing cells and the percentage of GFP-positive cells in WT *BRCA1* expressing cells.

mRNA Splicing Profile and Protein Expression Assessment

To study the effect of a variant on splicing, RNA was isolated from *BRCA1* variant-expressing mESC using a Trizol-based protocol nine days post 4-OHT treatment and sent to the Vall d'Hebron Institute of Oncology to be analysed like patients' samples.

BRCA1 protein expression in BAC-transfected mESC was confirmed by western blot. Proteins were separated by 4–12% gradient gels (NuPAGE™ Bis-Tris Gels from ThermoFisher Scientific) and transferred to an Amersham Protran premium 0.45 µm nitrocellulose membrane (GE Healthcare Life Sciences, Chicago, IL, USA). Membranes were blocked with 8% skimmed milk (Santa Cruz Biotechnology, Dallas, TX, USA) and stained with primary and secondary antibodies in the same buffer used for blocking. BRCA1 protein was detected with the mouse monoclonal antibody (Merck OP92; 1:1000) which binds within the N-terminal 304 amino acids of BRCA1. As loading control Tubulin was stained with Mouse α TUBULIN (Sigma T6199; 1:5000). After secondary antibody staining, the membranes were treated with the WesternBright ECL HRP Substrate kit (Advansta, San Jose, CA, USA) and imaged on an Amersham Imager 680 (Bioké, Leiden, The Netherlands).

Variant Classification

After a compilation of splicing *in silico* and experimental data, functional information from multiple sources, population allele frequency, case-control data, and multifactorial clinical data, the 166 variants were classified applying the ACMG/AMP guidelines specified for *BRCA1/2* by the ClinGen ENIGMA-VCEP(<https://cspec.genome.network/cspec/ui/svi/doc/GN092>)[122].

Variants were classified by combining criteria according to the minor adaptation of the traditional ACMG/AMP classification system described in the above-mentioned guides. For variants that met contradictory evidence or only met one code, we used the point system proposed by Tavtigian et al. (2020) [119].

In Silico Related Evidence: PP3, BP4, BP7 and PS1

Although several tools were used to predict variants' effect on splicing or protein function, only SpliceAI [168] and BayesDel [263] have been validated for *BRCA1* and *BRCA2* for bioinformatic code application against or in favour of pathogenicity (<https://cspec.genome.network/cspec/ui/svi/doc/GN092>).

Any intronic/synonymous variant with a SpliceAI DS greater or equal to 0.2 (indicative of splicing impact) met PP3_splicing criteria, whereas DS equal or lower than 0.1 met BP4_splicing criteria. Additionally, in synonymous as well as deep intronic variants (beyond -21 and +7 positions) that met BP4 criteria, the BP7 code was also applied to capture the low prior probability of pathogenicity of these type of variants.

Splicing codes could be also applied for missense/in-frame variants, but protein effect predictions must be also considered. BayesDel(noAF) score higher or equal to 0.28 would predict a protein functional impact (PP3_protein) whereas a score lower or equal to 0.15 is indicative of no impact (BP4_protein) for *BRCA1*. Accordingly, bioinformatic codes were applied as follows:

- SpliceAI ≤ 0.1 (BP4_splicing) and BayesDel(noAF) ≤ 0.15 (BP4_protein): BP4
- SpliceAI ≤ 0.1 (BP4_splicing) or uninformative, but BayesDel(noAF) ≥ 0.28 (PP3_protein): PP3_protein

- SpliceAI ≥ 0.2 (PP3_splicing), but BayesDel(noAF) ≤ 0.15 (BP4_protein) or uninformative: PP3_splicing
- SpliceAI ≥ 0.2 (PP3_splicing) and BayesDel(noAF) ≥ 0.28 (PP3_protein): PP3
- SpliceAI ≤ 0.1 (BP4_splicing) and BayesDel(noAF) uninformative: no code applicable
- SpliceAI uninformative and BayesDel(noAF) ≤ 0.15 (BP4_protein): no code applicable
- SpliceAI uninformative and BayesDel(noAF) uninformative: no code applicable

Regarding splicing PS1, the predicted event of the variant under assessment must precisely match the predicted event of the known (likely) pathogenic variant. In the case of exon 18, the c.5152+1G>T pathogenic variant at the donor natural site was used as a reference because no other variants were classified by an expert panel (ENIGMA old criteria) at splicing regions before this study, and after our reclassification with current guidelines this variant kept pathogenic classification. Hence, PS1_splicing only was used for variants in the donor splicing motif (from c.5050 to c.5152+6, both included). Specifically, we applied PS1_supporting for variants at c.5152+1/2 positions that met PVS1 criteria, and PS1_moderate for variants at donor motif that met PP3 criteria (exon 18 skipping predicted).

PS1 code was also applied for missense variants that lead to a substitution previously classified by an expert panel as pathogenic, and both reference and variant under assessment have no impact on splicing. The c.5143A>C missense variant, classified as pathogenic by ENIGMA (old criteria) as well as after our reclassification, was used as a reference variant for PS1_missense code in our

cohort. Given that all variants had an experimental mRNA result, the *in silico* codes (PP3, BP4, BP7, and PS1) were modified as detailed in **Figure 29**.

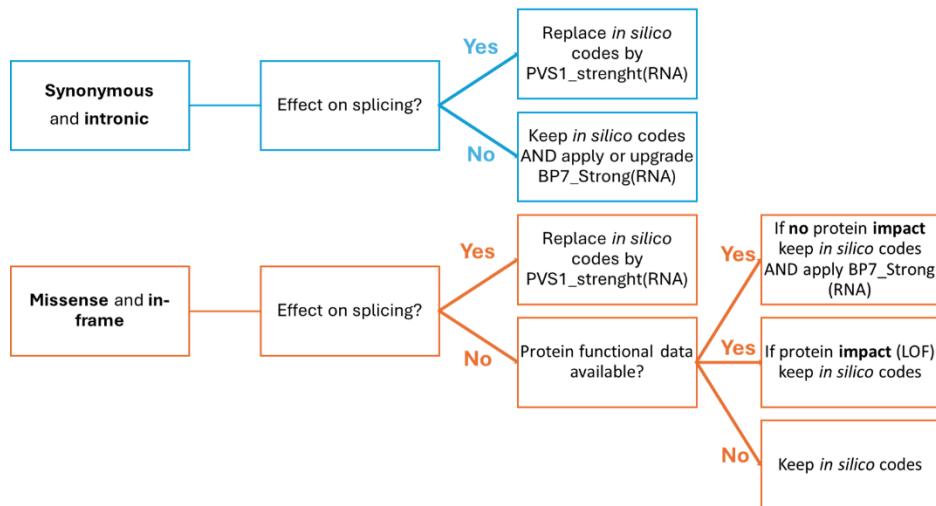


Figure 29. Modification of *in silico* codes by splicing observed impact in different variant types (synonymous/intronic and missense/in-frame).

LOF: loss of function

Variant Location and Coding Effect: PVS1 and PM5-PTC for Nonsense and Frameshift variants

BRCA1 exon 18 nonsense variants and those frameshift variants creating a premature stop codon (PTC) in exon 18, met PVS1 criteria because they affect an exon present in biologically relevant transcripts, and the PTCs on this exon will generate transcripts that will be degraded by the nonsense-mediated decay pathway (NMD). Frameshift variants creating a stop codon that would disrupt the reading frame beyond exon 18 also met PVS1 criteria regardless of NMD acting, because the altered region (BRCT domain) is relevant for protein function. Additionally, PTC variants in allow the usage of the PM5_PTC code, and its

strength varied depending on the exon in which the stop codon appears (Detailed in VCEP guidelines, Parsons et al. (2024)[122]).

For nonsense and frameshift variants altering splicing, coding (PVS1 and PM5_PTC), and splicing evidence (PVS1_strength (RNA)) was applied to each transcript, and we used conservative “overall” evidence for variant interpretation. We retained coding evidence unless a complete splicing defect was detected using patient samples, at which point PVS1(RNA) criteria was exclusively applied, replacing both coding specifications. If nonsense and frameshift variants caused leaky splicing defects or were only tested in the minigene system, we considered all splicing and coding evidence for variant classification to avoid underestimating pathogenicity.

mRNA Splicing Observed Evidence: PVS1(RNA) and BP7_strong (RNA)

Following PVS1 flowcharts from available guidelines, *BRCA1* exon 18 skipping met the PVS1 code. However, for applying PVS1(RNA) evidence, strength was applied depending on mRNA results and concordance between minigene and patients as detailed in **Table 7**.

We used the pre-established per-allele FL thresholds for RNA PVS1 and BP7 codes to define four groups: $FL \leq 10\%$; $10\% < FL < 20\%$; $20\% < FL < 30\%$ and $FL \geq 30\%$ [122,194]. We decided to focus exclusively on FL levels, since it was the only confirmed functional transcript detected in our assays. The results for patient mRNA and minigenes were considered concordant if both results for a specific variant fitted in the same group.

Concordant results met codes at full-weight as all variants assessed in patients were also studied through minigenes, regardless of whether the allele-specific

assessment had been performed or not in carriers since minigenes provided this allele-specific quantitative measurement.

For non-concordant results, we considered that the patient mRNA data was more relevant for clinical variant interpretation, and the RNA codes were applied depending on whether an allele-specific assay was or was not performed (**Table 7**). When only minigene results were available, strength was downgraded by one level (**Table 7**), except for variants affecting dinucleotide positions for which PVS1 was maintained as PVS1(RNA).

Table 7. PVS1(RNA) and BP7(RNA) specifications for variant classification.

<i>mRNA result</i>	FL ≤10%	10% < FL < 20%	20% < FL < 30%	FL >30%	No effect
<i>Patient Semi-quantitative result (per-allele) and with a concordant minigene result^a</i>	PVS1 (RNA)	PVS1_strong (RNA)	PVS1_NA	BP7_strong (RNA) for non-coding variants	BP7_strong (RNA) for non-coding variants
<i>Patient Semi-quantitative result (per-allele) with AE assay but, with a non-concordant minigene result^b</i>	PVS1 (RNA)	PVS1_strong (RNA)	PVS1_NA	BP7_strong (RNA) for non-coding variants	BP7_strong (RNA) for non-coding variants
<i>Patient Semi-quantitative result (per-allele) without AE assay and with a non-concordant minigene result^c</i>	PVS1_strong (RNA) for apparent (near) complete exon 18 skipping	PVS1_NA for incomplete exon 18 skipping		BP7_strong (RNA) for non-coding variants	
<i>Minigene only^d</i>	PVS1_strong (RNA)	PVS1_moderate (RNA)	PVS1_NA	BP7_strong (RNA) for non-coding variants	BP7_strong (RNA) for non-coding variants

^aConcordant results between patients and minigenes were considered "Patient quantitative results" and the maximum RNA weight was applied according to estimated full-length levels in patient (per-allele FL%) and minigene. ^b non-concordant results: patient's data was used for variant classification and if allele-specific data confirmed that FL expression mostly came from WT allele, the result was considered quantitative and PVS1 was applied according to FL % from deconvolution per-allele analysis. ^c If only semiquantitative and qualitative results from carriers were considered because the minigenes result does not reflect the patient's result, only PVS1 and BP7_strong were considered according to FL% (per-allele) observed in patients and no code was applicable for incomplete defects (FL% from 10% to 30%). ^dFor variants with only minigene results the strength was downgraded to strong or moderate depending on the percentage of FL. FL: full-length. AE: allele-specific full-length contribution by variant allele, measured by

sanger of RT-PCR using primers that exclusively hybridize on exon 18, or SNaPshot FL quantitative contribution measurement.

Functional evidence: PS3 and BS3

We used functional data from the ENIGMA-validated functional studies and BS3/PS3 codes were taken from Table 9 and supplementary Table 4 of ENIGMA specification's documents [269–273]. Additionally, PS3 and BS3 codes were applied for two delins variants as reported in Nepomuceno et al. (2022) [274].

For the functional data on HDR generated in our study, we followed the guidelines reported by Brnich et al. (2019)[275] to apply the respective derived functional evidence. Ten out of the 18 variants selected to be assessed using the mESC assay were already well-established benign (4) and pathogenic (6) variants. The clinical interpretation of these control variants reaches their classification independently of functional data, and their inclusion allowed us to define the ranges of HDR for benign and pathogenic variants. These ranges were defined using the maximum HDR value observed in pathogenic variants and the minimum value in benign variants. In addition, according to Brnich et al. (2019)[275] the inclusion of these well-known pathogenic and benign variants allowed the application of PS3 and BS3 criterion at supporting strength for individual tested variants.

When data from more than one functional assay was available and concordant, the highest strength of BS3/PS3 was considered for variant classification. When non-concordant results were observed across multiple studies, PS3/BS3 codes were not applied.

Population frequency evidence: PM2, BA1 and BS1

The highest non-cancer, non-founder population filter allele frequency in gnomAD v2.1 (only exomes) or gnomAD v3.1 were used to apply PM2_supporting, BA1, BS1 or BS1_supporting evidence for variant classification. PM2_supporting was applied for variants absent from gnomAD, but this criterion was not used for insertion, deletion, or delins variants. BA1 was applied for variants with a frequency above 0.001 (0.1%), BS1 above 0.0001 (0.01%), and BS1_Supporting greater than 0.00002 (0.002%) and lower or equal to 0.0001 (0.01%). Pathogenic code was not considered if the read depth was lower than 25. For benign codes, the read depth must be higher or equal to 20.

Case-control evidence: PS4

Variant-level case-control counts were taken from BRIDGES study [37](<https://bcac.ccge.medschl.cam.ac.uk/bcacdata/bridges-summary-results-breast-cancer-risk-genes-2021/>, accessed in February 2024). We calculated the odds ratio (OR), 95% confidence intervals, and p-value for each variant through MedCalc Software Ltd. Odds ratio calculator https://www.medcalc.org/calc/odds_ratio.php (Version 22.023; accessed February-March, 2024). Of note, Haldane-Anscombe correction was used to calculate an OR when the control or case frequency was 0. PS4 was considered according to guidelines, only when the p-value was equal or lower to 0.05, the OR was greater or equal to four and the 95% confidence interval did not include two.

Multifactorial clinical evidence: PP4 and BP5

Combined likelihood ratios (LRs) for pathogenicity estimated from clinical data, such as co-segregation with the disease, tumour (breast or ovarian) pathology, co-occurrence *in trans* with a pathogenic variant in the same gene and reported

family history were used to apply PP4 and BP5 codes. This data was obtained from the literature [106,108,256,257,276]. Additionally, ENIGMA collaborators unpublished data was included if provided. All LRs available for a given variant were combined by multiplication [109] and resulting combined LR was used to assign BP5/PP4 code as described in Tavtigian et al. (2018)[118]: pathogenic very strong >350, pathogenic strong >18.78 to 350, pathogenic moderate >4.3 to 18.7, pathogenic supporting >2.08 to 4.30; benign very strong <0.00285, benign strong <0.05 to 0.00285, benign moderate <0.23 to 0.05, and benign supporting <0.48 to 0.23. Combined LR between 0.48 to 2.08 were considered non-informative (PP4 and BP5 not met).

In addition, for the c.5152+7A>G additional data was compiled from several collaborators, and a multifactorial likelihood quantitative analysis was performed by the Department of Genetics from the Institute Curie (France. Specifically, LRs for breast and ovarian tumour pathology were applied according to Spurdle et al. (2014)[198] and O'Mahony et al. (2023) [199], respectively; LRs based on family history were calculated by using the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) risk-prediction scores obtained from CanRisk (<https://www.canrisk.org/>) and considering those obtained from *BRCA1/2* negative families in Spurdle et al. (2012)[277]; co-segregation LRs were assessed using the statistical model developed by Thompson et al.(2003)[278] and implemented in the COOL webserver (<http://bjfenglab.org/>) based on an improved penetrance model developed by Belman et al. (2020)[279] ; and co-occurrence LRs were calculated as detailed in Goldgar et al. (2004)[106].

Statistical Analysis

Kruskal-Wallis non-parametric test was used for comparing SRE distribution in exon 18 and its adjacent intronic regions since ESRseq WT scores did not show a normal distribution. Post-hoc Dunn's test was conducted to identify specific group differences following a significant Kruskal-Wallis test result.

Regarding mRNA splicing results, normality was assessed by the Shapiro-Wilk test for datasets with fewer than 50 samples, while the Kolmogorov-Smirnov test was used for datasets exceeding 50 samples.

Pearson's correlation coefficient analysis was utilized when normal distribution was observed; otherwise, Spearman's Rank correlation was applied for non-parametric data.

Linear regression analysis was conducted to explore the relationship between an independent variable (splicing event, denoted as 'x') and a dependent variable (functional outcome).

All statistical analyses were conducted using GraphPad Prism v8.

Part 3

Research Cohorts

This retrospective cohort study was part of the work-package 3 of the ERA PerMed 2019 call (RAD51predict ERA-NET Co-Fund ERAPERMED 2019-215), and of an ENIGMA project approved in 2021. Collaborators were invited to send treatment-naïve primary non-metastatic tumour samples. From each sample, germline *BRCA1/2* and *PALB2* status, tumour type (breast or ovarian), sample type (surgery or biopsy), age at diagnosis, and tumour pathology information such as hormonal receptor status (ER, PR and HER2) and tumour grade was asked or retrieved from medical records (**APPENDIX II**).

All participants provided informed consent before being included in this project and were codified by their respective centres. This study was reviewed and approved by the Vall d'Hebron Hospital Ethics Committee and by each institutional review board of the participating centres and conducted according to their ethical standards.

The RAD51 test has been performed in a total of 213 formalin-fixed paraffin-embedded (FFPE) samples of primary untreated tumours from individuals that underwent germline genetic testing for breast and/or ovarian cancer predisposition have been analysed. Samples were collected across five worldwide centres: Hospital Universitari de Vall d'Hebron (Spain)(n=146/213), Cancer Research Malaysia (Malaysia) (n=41/213), University Hospital of Parma (Italy) (22/213), Hospital Universitari Sant Joan Reus (Spain)(n=3/213) and Cancer Genetics Service National Cancer Centre Singapore (Singapore) (n=1/213).

All samples successfully evaluable by RAD51 test were clustered into three different cohorts based on patients' germline variant classification status:

- Pathogenic: This cohort initially included 85 tumours from germline carriers of *BRCA1* or *BRCA2* pathogenic variants reviewed by an expert panel in ClinVar. It also included carriers of loss of function variants in *PALB2* (nonsense, frameshift, splicing, and large in-frame deletions).
- Benign: This group initially comprised 23 tumours from carriers of exclusively benign and likely benign variants in *BRCA1* and *BRCA2* genes reviewed by an expert panel in ClinVar. Carriers of (likely) pathogenic variants or VUS in other clinically actionable hereditary cancer genes (*MLH1*, *MSH2*, *MSH6*, *TP53*, *BARD1*, *PALB2*, *CHEK2*, *ATM*, *BRIP1*, *RAD51C*, *RAD51D*, *PTEN*, and *CDH1*) were not included.
- Unreviewed: This cohort comprised 38 tumours from carriers of *BRCA1* or *BRCA2* variants that has not been reviewed by an expert panel in ClinVar, as well as missense variants in *PALB2*. In this cohort, additional genomic tests were performed to molecularly characterise these tumours. Furthermore, the germline variants of these patients were classified according to current VCEP guidelines (see details on “materials and methods, Part 3, variant classification” section) to enlarge the pathogenic and benign cohorts. Considering the achieved classification, 6 tumour samples were moved to pathogenic cohort (likely pathogenic and pathogenic variants) and 24 samples to benign cohort (likely benign and benign variants). The remaining eight samples were from carriers of variants that were classified as VUS and were not included into the benign nor the pathogenic definitive cohorts.

Functional HRD: RAD51 Test

Immunofluorescence (IF) staining was used to quantify RAD51 *foci* (as a functional surrogate of HR status), γ H2AX *foci* (as a biomarker of DNA DSBs) and BRCA1 *foci* (as a mediator of HRR) in 3 different slides from formalin-fixed paraffin-embedded (FFPE) samples (3 μ m tissue sections) of primary untreated breast and ovarian tumours. Each biomarker was dual-stained with geminin (as a marker of S/G2 cell cycle phase) and DAPI (4',6-diamidino-2-phenylindole) [247,248]. The primary antibodies used were: rabbit anti-RAD51 (Abcam, Cambridge, UK ab133534, 1:1000), mouse anti-BRCA1 C-terminus (residues 1843-1862 coded by exon 24) (Santa Cruz Biotechnology Inc, Dallas, TX sc-6954, 1 : 50), mouse anti- γ H2AX (Merk Millipore, Darmstadt, Germany #05-636, 1 : 200), mouse anti-geminin (Novocastra Laboratories Ltd, Newcastle upon Tyne, UK NCL-L, 1 : 60) and rabbit anti-geminin (Proteintech, Manchester, UK 10802-1-AP, 1 : 400). Goat anti-rabbit Alexa fluor 568, goat anti-mouse Alexa fluor 488, donkey anti-mouse Alexa fluor 568 and goat anti-rabbit Alexa fluor 488 (Invitrogen, Waltham, MA; 1: 500) were used as secondary antibodies [280]. The IF staining was performed as described in Castroviejo-Bermejo et al. (2018) [247]. The scoring was conducted blindly on live images using a Nikon Ti2-Eclipse microscope with a $\times$ 60 immersion oil objective. All IFs have been done by the Experimental therapeutics group from VHIO. Samples were jointly scored between the Experimental therapeutic (Alba Llop-Guevara, PhD) and Hereditary Cancer Genetics (Joanna Domènech-Vivó) groups at the VHIO. Samples with low γ -H2AX score (< 25% of positive cells with $\geq$ five *foci*) or with < 40 geminin-positive cells were non-evaluable (NE) due to insufficient endogenous DNA damage or insufficient tumour cells in the S/G2-phase of the cell cycle, respectively. RAD51% and BRCA1% scores were considered low or high based

on the predefined cutoff of 10% geminin-positive cells with 5 or more nuclear foci [245,247,281]. Functional HRD was defined by low RAD51% scores ($\leq 10\%$), and functional homologous recombination proficiency (HRP) by high RAD51% scores ($>10\%$).

Genetic and Genomic HRD: Myriad's MyChoice® HRD

For unreviewed cohort, further analyses were conducted to characterize molecularly these tumour samples, by determining genomic instability score (GIS) using the Myriad myChoice CDx assay as well as HRR genes mutation profile and gene-specific LOH (gsLOH)[282].

Tissue sections of 10 μm from the same FPPE primary untreated tumour samples were sent to the Philipps-Universität Marburg, where Myriad myChoice HRD Plus CDx assay was performed as previously described [283]. In brief, DNA was isolated and underwent targeted multiplex chain reaction amplification before library construction. Then DNA was hybridized to a custom capture panel (Probe Pool; Integrated DNA Technologies; catalogue number CAPP) with biotinylated probes for $>26,000$ single-nucleotide polymorphisms (SNPs) distributed across the human genome. This panel allows for whole detection of gene mutations and SNPs with a complete sequence coverage of *BRCA1/2*. The hybridized DNA mixture was amplified via standard PCR and library pools were sequenced on an Illumina NextSeq550 Dx (~ 30 million reads per sample). Post-sequencing data were uploaded and analysed by Myriad's sequencing analysis pipeline algorithm and software.

Sequencing results firstly were used to screen tumour mutations in the *BRCA1*, *BRCA2*, *PALB2* and other 12 HRR related genes (*ATM*, *BARD1*, *BRIP1*, *CDK12*, *CHEK1*, *CHEK2*, *FANCL*, *PPP2R2A*, *RAD51B*, *RAD51C*, *RAD51D*, and *RAD54L*).

All variants classified as deleterious or suspected deleterious under of the ACMG/AMP standards were reported. To estimate the gsLOH status of the HRR genes, the computationally most likely allele-specific copy number at each single-nucleotide variation location was analysed.

For testing genomic HRD, the genomic instability score (GIS) was calculated using a standardized automated bioinformatic analysis, based on the sum of loss of heterozygosity (LOH), telomeric allelic imbalance (TAI) and large-scale state transitions (LST) across the entire genome. The FDA cutoff for ovarian carcinoma was used to classify the HR status, which was defined as a GIS of ≥ 42 for HRD [282].

Variant Classification

The sample selection for the cohort of unreviewed variants started in 2019, but during this thesis, *BRCA1/2* and *PALB2* ClinGen VCEPs published ACMG/AMP adapted guidelines for variant classification. As previously mentioned, we classified all variants from the unreviewed group to increase the number of tumours for benign and pathogenic cohorts.

Variants were annotated with Alamut visual plus v1.7 (Sophia Genetics) using the reference transcripts NM_007294.3 and the legacy 24-exon numbering for *BRCA1*, with exon 4 missing due to a correction made after the initial description of the gene (GenBank U14680.1); NM_000059.3 for *BRCA2*; and NM_024675.3 for *PALB2*. ClinVar variants' review status was manually ascertained in June 2024.

The ClinGen ENIGMA *BRCA1* and *BRCA2* Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines for *BRCA1/2* Version 1.1.0 (released 21st, December 2023) and the ClinGen Hereditary Breast, Ovarian and

Pancreatic Cancer Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines for *PALB2* Version 1.1.0 (released 28th, November 2023), were used for variant classification.

SpliceAI (v1.3; <https://github.com/Illumina/SpliceAI>) [168] was used for *in silico* splicing assessment for PP3/BP4 evidence, based on VCEP recommendations for *BRCA1/2* and *PALB2*. However, *in silico* protein predicted effect can only be used for *BRCA1/2* classification. Specifically, BayesDel not considering allele frequency (noAF) scores were obtained from <https://brcaexchange.org/> (accessed May-June 2024) for PP3/BP4 specifications. For *PALB2*, published protein effect predictors have yet to be calibrated due to the lack of functional outcomes for missense variants.

Likewise, functional evidence (PS3/BS3) could only be considered for *BRCA1/2*. BS3/PS3 codes were taken from the table 9 and supplementary table 4 of ENIGMA specifications' files, available here: <https://cspec.genome.network/cspec/ui/svi/doc/GN092> [269,271–273].

For observed splicing impact evidence for *BRCA1/2*, we checked the datasets from Houdayer et al. (2012) [184], Walker et al. (2013) [180], Vallée et al. (2016)[284], and Walker et al. (2023) [170]. Moreover, internal results were used for two variants: c.5152+3A>C (see details of *BRCA1* exon 18 variants in **Part 2, Supplementary Table 9, Supplementary Table 17** and **Supplementary Table 22**) and c.791_794del (data available in a poster communication P13.079.A [285]). For *PALB2* observed splicing evidence, we search for splicing results by interrogating each single variant in Litvar2 (<https://www.ncbi.nlm.nih.gov/research/litvar2/>, accessed July 2024), but no results were found for variants in this gene.

For *BRCA1/2* multifactorial likelihood clinical data, we obtained combined LRs from published results. Moreover, we calculated tumour pathology LRs according to Spurdle et al. (2014) [198] for BC and to O'Mahony et al. (2023) [199] for OC.

Variant-level case-control counts were taken from BRIDGES study [37] (<https://bcac.ccge.medschl.cam.ac.uk/bcacdata/bridges-summary-results-breast-cancer-risk-genes-2021/>, accessed in February 2024). Odds ratio (OR), 95% confidence intervals and p-value for each variant were calculated through MedCalc Software Ltd. (Odds ratio calculator https://www.medcalc.org/calc/odds_ratio.php (Version 22.023; accessed July 2024)). Of note, Haldane-Anscombe correction was used to calculate an OR when the control or case frequency was 0.

As recommended for *BRCA1/2*, when contradictory evidence was met for a given variant, we applied the point system described in Tavtigian et al. (2020) [119]. This approach is not stated for *PALB2*, but we decided to also use it given that it has been previously applied in Valenzuela-Palomo et al. (2022) [286]. Point-based variant classification categories are defined as follows: pathogenic $\geq +10$; likely pathogenic +6 to +9; VUS -1 to +5; likely benign -2 to -6; and benign ≤ -7 .

Statistical Analysis

Study cohorts were descriptively analysed, continuous values were expressed as mean (maximum-minimum) values, and categorical variables as absolute values and percentages. T-test was used to assess differences between pathogenic and benign cohorts regarding continuous variables and chi-square test for analysing differences between frequencies in both groups (categorical variables).

Normal distribution of biomarker % in all groups was assessed by the Shapiro-Wilk test for datasets with fewer than 50 samples, while the Kolmogorov-Smirnov test was used for datasets exceeding 50 samples. Kruskal-Wallis non-parametric test and post-hoc Dunn's test were used in case of no-normality and ordinary one-way ANOVA was used to identify group differences when normal distribution was met.

The Cohen κ coefficient was used to analyse the concordance between categorical HRD tests.

All confidence intervals (CIs) reported were 95%, and all tests were two-sided with a value of P lower to 0.05 considered statistically significant.

Of note, RAD51% accuracy, specificity, sensitivity, positive predictive values and negative predictive values in identifying pathogenic and benign variant status, were calculated as detailed in **Table 6**. Specifically, RAD51% result was considered positive for HRD ($\leq 10\%$) and negative if HRP ($> 10\%$) detected, hence HRD in tumours from carriers of pathogenic variants were true positives, and HRP in tumours from carriers of benign variants were considered true negatives.

Receiver operating characteristic (ROC) analysis was performed to assess the performance of RAD51% score. Youden index ($J = \text{specificity} + \text{sensitivity} - 1$) was also calculated to find the exact thresholds of RAD51% score at which sensitivity and specificity were most balanced

All analyses above mentioned were conducted using GraphPad Prism v8, except Cohen κ coefficients and chi-square, that were calculated through MedCalc Software Ltd. (Inter-rater agreement. <https://www.medcalc.org/calc/kappa.php> Version 22.032; accessed August 2024 and Comparison of proportions

calculator. https://www.medcalc.org/calc/comparison_of_proportions.php Version 22.032; accessed August 2024).

RAD51% score was assessed as a predictor of germline *BRCA1/2* or *PALB2* pathogenic variant status by defining LR estimates. LR were calculated for each RAD51% group (HRP or HRD) by comparing their frequency between pathogenic and benign carriers:

$$HRD_{LR} = \frac{\rho_i \text{ with low RAD51(HRD)}}{\rho_o \text{ with low RAD51 (HRD)}} \text{ and } HRP_{LR} = \frac{\rho_i \text{ with high RAD51 (HRP)}}{\rho_o \text{ with high RAD51(HRP)}}$$

$$\text{Where, } \rho_i = \frac{\text{Pathogenic}}{\sum \text{Pathogenic}} \text{ and } \rho_o = \frac{\text{Benign}}{\sum \text{Benign}} .$$

Pathogenic denotes the number of tumours from germline pathogenic variants in *BRCA1/2* or *PALB2* for a given group, and benign denotes the number of tumours from germline benign *BRCA1/2* carriers for the same group. Variance and 95% CI were calculated as detailed in Walker et al. (2023)[170].

These LR estimations were then refined considering tumours and patients' characteristics (germline mutated gene, tumour grade, molecular subtype and age at diagnosis). Only BC samples were included in these grouped analyses due to the limited number of OC samples.

ACMG/AMP LRs Evidence Strength Alignment

Likelihood ratios were aligned to the code strengths of the Bayesian framework of the ACMG/AMP system [118] (**Table 8**), and we applied this tumour evidence under the PP4 pathogenic and BP5 benign codes.

Table 8. Likelihood ratios alignment with ACMG/AMP qualitative rules strengths.

Code strength	LRs range for PP4	LRs range for BP5
Very strong	$LR \geq 350$	$LR < 0.00285$
Strong	$18.7 \leq LR < 350$	$0.00285 \leq LR < 0.05$
Moderate	$4.30 \leq LR < 18.70$	$0.05 \leq LR < 0.23$
Supporting	$2.08 \leq LR < 4.30$	$0.23 \leq LR < 0.48$
Uninformative	$0.48 \leq LR < 2.08$	

LR: likelihood ratio

We defined as informative any LR with statistically significant 95% CI (not including 1) and with a value in the informative range, i.e., values greater to 2.08 were aligned to strengths in favour of pathogenicity, and values lower to 0.48 were aligned to strengths against pathogenicity.

RESULTS

Part 1

“Role of Splicing Regulatory Elements and In Silico Tools Usage in the Identification of Deep Intronic Splicing Variants in Hereditary Breast/Ovarian Cancer Genes”

This study was published in Cancers in 2021:

Moles-Fernández, A*.; **Domènech-Vivó, J.***; Tenés, A.; Balmaña, J.; Diez, O.; Gutiérrez-Enríquez, S. Role of Splicing Regulatory Elements and In Silico Tools Usage in the Identification of Deep Intronic Splicing Variants in Hereditary Breast/Ovarian Cancer Genes. *Cancers (Basel)* **2021**, *13*, doi:10.3390/cancers1313334 [287].

* Co-first authors

JD-V contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, writing—original draft preparation, review and editing, and visualization.

SpliceAI Optimally Predicts Deep Intronic Splice-Altering Variants but with Less Sensitivity Those Affecting Splicing by Altering Regulatory Elements

To establish the performance of SpliceAI in predicting deep intronic pseudoexon-generating variants, we interrogated a set of variants collected from the literature, after searching for variants located beyond 20 nucleotides from exon-intron boundaries and for which RNA data was available (**Supplementary Table 1**). This collection contains 233 deep intronic variants from 80 HBOC and

other rare mendelian disease genes, including 133 variants that promote the creation of pseudoexons or intron retention events and 100 variants that do not alter splicing. Once the delta (Δ) score for each variant was obtained by running SpliceAI (v1.3), it was compared with the experimental results in RNA to estimate sensitivity, specificity, and accuracy. Then, we estimated the threshold at which the performance of the tool was more optimal, obtaining at Δ score of 0.05 the highest MCC of 0.86 (**Table 9** and **Figure 30**).

Table 9. Performance of Δ SpliceAI for literature dataset with 233 deep intronic variants using the optimized threshold of 0.05.

Dataset	Splicing Altering Variants	No Splicing Altering Variants	Sensitivity	Specificity	Accuracy	MCC
All variants	133	100	93.99	92.00	93.13	0.86
Cryptic splice	117	100	95.73	92.00	94.01	0.88
SREs altering	16	100	81.25	92.00	90.52	0.66

SRE: splicing regulatory element, MCC: Matthews's correlation coefficient.

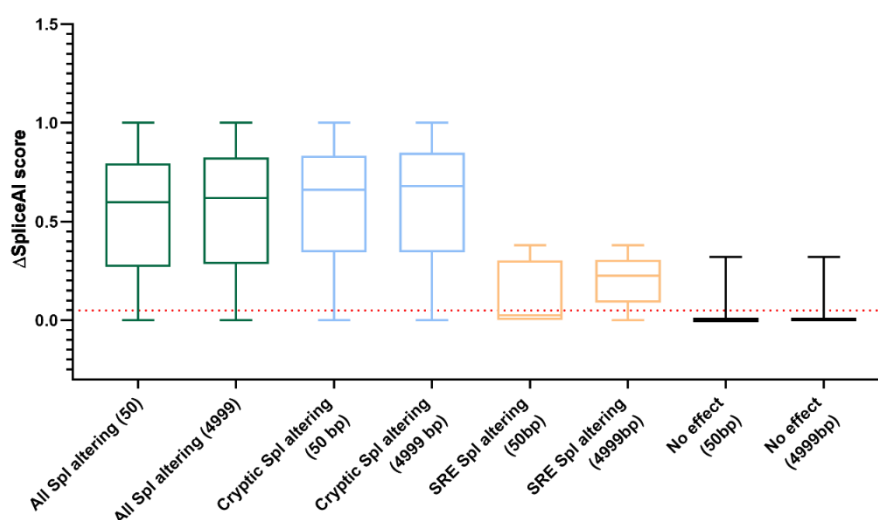


Figure 30. Δ SpliceAI scores of spliceogenic and not spliceogenic variants from the literature dataset, considering predictions of cryptic sites in a window of 50bp or 4999bp.

In the analysis with the 16 splicing altering variants by affecting SREs, the Δ SpliceAI evaluation of the 4999 bp located on each side of the variant reached higher predictive power in comparison with a 50 bp window analysis. Dashed red line indicates a 0.05 score cutoff and mean and $\pm$ standard deviation is represented. *Spl*: splicing, *SRE*: splicing regulatory element.

Since SpliceAI was specifically developed to detect altering splicing variants by activating or creating cryptic splice sites [168], we reassessed its predictions separately considering two groups of intronic-splice altering variants according to the *cis* element affected. We obtained better results with the group of cryptic splice variants than with the SRE disruptive variants (0.88 vs. 0.66 of MCC, respectively; **Table 9** and **Figure 30**).

The lower sensitivity showed by SpliceAI in predicting the impact of only 16 deep intronic variants on SREs prompted us to assess its performance with a large previous published dataset, composed of 360 exonic variants that affect splicing by altering SREs and 801 exonic variants without effect on splicing (**Supplementary Table 2**) [172]. The performance with our pre-established 0.05

cutoff was 69.17% sensitivity, 84.27% specificity and 0.53 MCC, while with the cutoff of 0.06 showing the highest MCC the prediction improves, reaching 0.55 MCC (**Table 10**).

Table 10. Performance of Δ SpliceAI (4999bp) optimized threshold for SREs disrupting exonic variants.

Δ SpliceAI Thresholds	TP	FN	FP	TN	Sensitivity	Specificity	Accuracy	MCC
0.05	249	111	126	675	69.17	84.27	79.59	0.53
0.06	242	118	104	697	67.22	87.02	80.88	0.55

*This is an adapted version of Table S2 B from Moles-Fernández et al. (2021)[287]. Δ SpliceAI(4999nt window) optimization was based on predictions regarding acceptor and donor loss, using the variants included in **Supplementary Table 2**. TP: true positives, FN: false negatives, FP: false positives, TN: true negatives.*

To supplement the performance of SpliceAI in identifying deep intronic variants disrupting SREs, we consecutively added the ESRseq evaluation. ESRseq is a computational algorithm specifically developed to predict SREs disruption that showed the best performance in predicting both variant-induced exon skipping and exon inclusion in a recent benchmarking study [172]. To do this, the Δ ESRseq values (WT ESRseq score - variant ESRseq score) were calculated as described in Ke et al. (2011)[134] for intronic variants negatively predicted by SpliceAI (<0.05). The threshold optimization by maximizing the sum of specificity and sensitivity, indicated that variants with a score change equal or higher than 0.63 were predicted to promote pseudoexon inclusion by altering SREs with higher sensitivity and specificity (**Supplementary Table 7**). The performance of the sequential pipeline applying to those variants with a SpliceAI Δ score of <0.05 and the Δ ESRseq threshold of ≥ 0.63 , showed higher sensitivity (96.24%) than those obtained by SpliceAI alone, but lower specificity (69%) and MCC (0.69) (**Table 11**).

Table 11. Performance of Δ SpliceAI alone and sequentially combined with ESRseq tool with 233 intronic variants from the literature dataset (133 altering and 100 non altering splicing).

Pipeline	Sensitivity	Specificity	Accuracy	MCC
(1) Δ SpliceAI ≥ 0.05	93.98	92.00	93.13	0.86
(2) Δ SpliceAI ≥ 0.05 + Δ SpliceAI < 0.05 and Δ ESRseq ≥ 0.63	96.24	69.00	84.55	0.69
(3) Δ SpliceAI ≥ 0.05 + Δ SpliceAI < 0.05 and Δ ESRseq ≥ 0.63 and Abs. Dif. 0.51	95.49	86.00	91.42	0.83

Abs. Dif.: absolute difference, MCC: Matthews's correlation coefficient

We next sought to know if the optimized SpliceAI cut-off of 0.05 with the whole set of variants, regardless of the *cis*-element affected, would be higher if they were evaluated using ESRseq primarily, i.e., SpliceAI would exclusively compute for variants affecting cryptic sites. However, this analysis of assessing the 233 intronic variants firstly with ESRseq and then with SpliceAI, led to the same previous optimized SpliceAI threshold of 0.05.

Splicing Regulatory Elements Balance Is Similar between Pseudoexons and Canonical Exons

To further investigate the role of splicing regulatory elements (SREs) in RNA included sequences, we compared the SREs landscape between constitutive exons and pseudoexons.

First, we extracted from canonical transcripts of HBOC and Lynch genes (**Supplementary Table 5**), the respective constitutional exon sequences and the 100 adjacent intron nucleotides for each gene. Next, an ESRseq value was assigned to each nucleotide according to Ke et al. (2011)[134], thus obtaining a map of the distribution of regulatory elements along the different exons. In addition, we calculated the sum of the total values (area), and the normalized

SRE area score (sum of total scores/number of region length nucleotides), of each exonic and adjacent intronic region for each gene. Comparing the values obtained from canonical exons and adjacent upstream and downstream introns, significant differences were observed. Exons were enriched in positive values, corresponding to exonic splicing enhancers (ESE), while in intronic regions predominated the negative values, indicating an abundance of intronic splicing silencers (ISS) (**Figure 31 A**).

We then determined the presence and proportion of SREs in the pseudoexons in our literature dataset. The sequences of the pseudoexonized regions or intron retentions and the adjacent up and downstream 100 nucleotides were analysed, following the cDNA position indicated in the corresponding publication (**Supplementary Table 1**). These sequences were mapped with ESRseq scores obtained for each nucleotide, thus obtaining the area and the normalized SRE area values. Next, we compared the ESRseq values obtained between the region included in the mRNA as a pseudoexon and in upstream and downstream intronic regions. As with the canonical exons, significant differences were also observed. The regions included in the mRNA as pseudoexons presented a higher percentage of positive values and for those that remained as introns, a greater proportion of negative values (**Figure 31 B**, data in **Supplementary Table 1**).

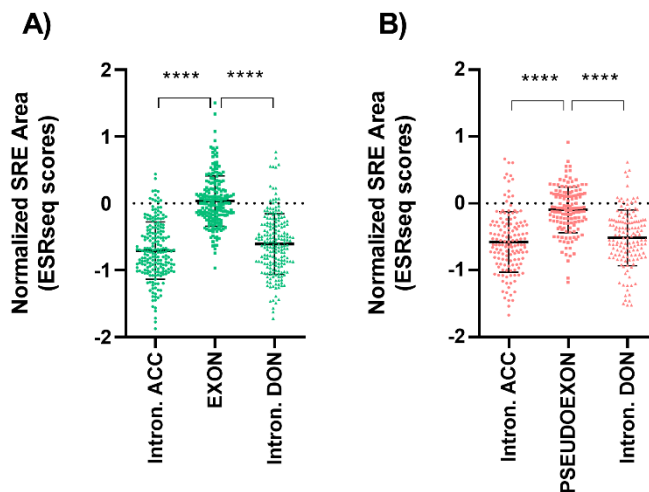


Figure 31. SREs abundance in different genomic regions, using the normalized SRE area calculated with ESRseq scores.

A) Normalized SRE area using ESRseq scores of exonic and adjacent 100 intronic nucleotides located upstream and downstream of canonical exons of HBOC and Lynch genes. Significant differences were identified between exons and intronic regions (Pair-wise significance levels calculated by Tukey test, **** p-value < 0.0001). **B)** Normalized SRE Area using ESRseq scores of pseudoexons and adjacent 100 intronic nucleotides located upstream and downstream of pseudoexons listed in the literature dataset. Significant differences were identified between pseudoexons and intronic regions (Pair-wise significance levels calculated by Tukey test, **** p-value < 0.0001). *Intron. ACC: intronic sequence adjacent to acceptor site; Intron. DON: intronic sequence adjacent to donor site.*

However, these differences were more pronounced between canonical exons and their surrounding intronic regions (**Figure 32 A**), since the presence of enhancer or silencer regulatory elements is more abundant in exonic and intronic canonical regions respectively (**Figure 32 B,C**).

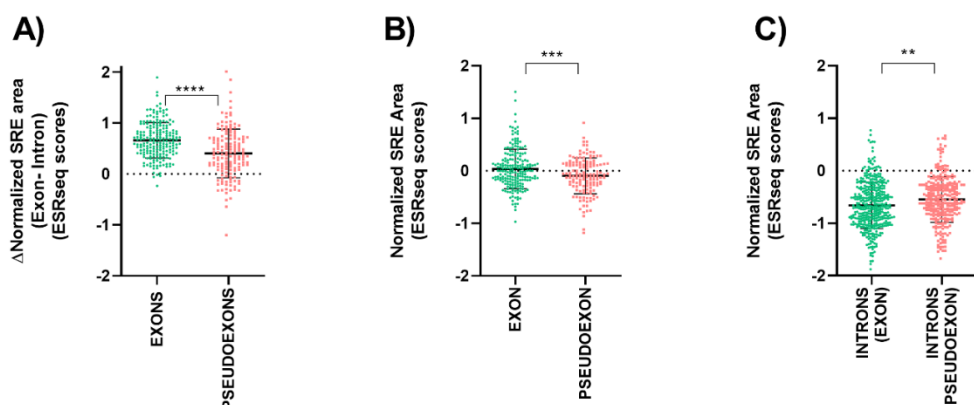


Figure 32. Comparison of SREs abundance in different genomic regions, using the Normalized SRE Area calculated using ESRseq scores.

A) Comparison of the exon-intron difference of normalized SRE areas between canonical exons and pseudoexons. First, the mean of normalized SRE area of adjacent donor and acceptor site intronic regions was calculated. Then, this mean was subtracted from the exon and pseudoexon normalized SRE area value. The difference between exonic and intronic regions in canonical exons was higher than in the case of pseudoexons, suggesting that they are more defined by a SRE balance. (t-test, **** p-value < 0.0001). Mean $\pm$ standard deviation is represented in each graph. Intron. ACC: intronic sequence adjacent to acceptor site; Intron. DON: intronic sequence adjacent to donor site. **B)** Comparison between the normalized SRE area from canonical exons and pseudoexons. A greater abundance of positive ESRseq values was observed in canonical exons (t-test, ***p-value < 0.001). **C)** Comparison between the normalized SRE area of introns adjacent to canonical exons and pseudoexons. For this comparison, the mean of both intronic acceptor and donor areas of each exon/pseudoexon was calculated. A greater abundance of negative ESRseq values was observed in introns adjacent to canonical exons (t-test, **p-value $\leq$ 0.01). Mean $\pm$ standard deviation is represented in each graph.

We also estimated for each variant of the literature dataset, regardless of its splicing effect, the ESRseq values for each position of the 100 intronic bases before and after the variant (without including the hexamers affected by the variant), and then the absolute difference of normalized SRE area between these two regions was calculated (**Supplementary Table 1**). This value was used to compare the two groups of deep intronic variants: 133 spliceogenic vs. 100 no effect variants. Our results demonstrated that variants with no effect presented a very low difference between adjacent regions, compared to variants with a spliceogenic effect (**Figure 33**).

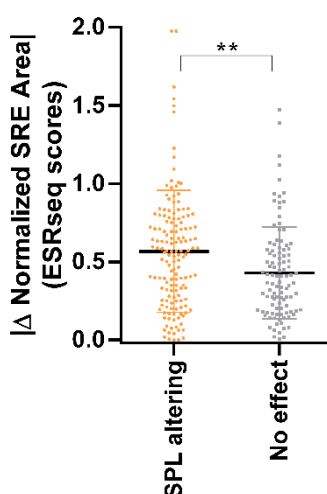


Figure 33. Comparison of the absolute difference of normalized SRE area from 100 nucleotides before and after each deep intronic variant compiled from the literature

Absolute values of normalized SRE area difference from 100 nucleotides upstream and downstream of each variant were used to compare those spliceogenic with those without any effect. Splicing variants (SPL altering) showed higher differences between previous and posterior sequences (t-test, ** p-value ≤ 0.01). Mean $\pm$ standard deviation is represented.

This is consistent with spliceogenic intronic variants being in regions with a significant difference in the balance of SRE between exons and introns (in line with the pseudoexon landscape showed in **Figure 31 B**), while variants that do not cause alteration are in intronic regions with no SREs that could make them susceptible to be pseudoexonized. Overall, these results suggest that: (i) intronic regions with a similar SRE balance to that of exons are more susceptible to be included in mature RNA; (ii) a SRE balance is relevant in the RNA misplacing caused by deep intronic variants and (iii) bearing in mind this balance will facilitate the *in silico* identification of intronic variants leading to pseudoexon inclusions.

Inclusion of SRE Landscape in the *In Silico* Detection of Deep Intronic Splice-Altering Variants

Although SRE *in silico* tools (such as ESRseq) can detect variant-induced SRE alterations, they are not able to identify whether the SRE landscape where the variant is located presents similarities to those of an exon and ignore the relevance of the SRE balance to identify variant-induced pseudoexon events. To address this limitation, we included in the pipeline of SpliceAI (Δ score cut-off of ≤ 0.05) and ESRseq (Δ score cut-off of $+0.63$), the estimation of the absolute difference between 100 nucleotides up and downstream for each variant (**Supplementary Table 1**). With the absolute difference threshold of 0.51 we obtained a MCC of 0.83, with 95.49% sensitivity, 86% specificity, and 91.42% accuracy (**Table 11** and **Supplementary Table 8**). Although this pipeline would theoretically detect variants that alter SREs, the sensitivity of 95.49% only slightly improves that observed using SpliceAI alone (93.98%) (**Table 11**).

The performance of the last pipeline (SpliceAI Δ score 0.05 $\rightarrow$ Δ ESRseq 0.63 $\rightarrow$ difference in absolute values 0.51), was tested with a set of 207 splicing-disrupting deep intronic variants identified in tumours by RNAseq [259] with 86% sensitivity which again slightly improves that observed using SpliceAI alone (85.5%) (**Supplementary Table 3**).

Experimental Analysis of Hereditary Cancer Gene Variants

Thirty-three unique variants were experimentally assessed. Thirteen variants passed the 0.05 SpliceAI threshold and six presented Δ ESRseq equal or greater than 0.63 and an absolute difference value greater than 0.51 (**Supplementary Table 4**). The remaining 14 variants were not predicted as spliceogenic (**Supplementary Table 4**). We characterised the variant effect by RT-PCR assays comparing their splicing profiles (by high-resolution electrophoresis) with those

in the healthy controls, and posterior Sanger sequencing. This analysis detected the inclusion of intronic regions in mature mRNA in only 5 of the 13 variants prioritized by SpliceAI alone (**Table 12** and **Figure 34**): *ATM* variants c.1899-123A>G, c.2466+1552G>C, c.8850+2029A>G, *FAM175A* variant c.476+158G>T and *MUTYH* variant c.998-27G>A. The six variants with a ΔESRseq equal or greater than 0.63 and absolute difference prediction in favour of altering a SRE region did not show an aberrant splicing. All 14 variants with a negative splicing alteration prediction presented a normal splicing pattern.

Table 12. Deep intronic variants with a spliceogenic effect detected in hereditary cancer genes.

Gene ^a	Variant ^b	Intron	ΔSpliceAI (Position of Predicted Splice site) ^c	ΔESRseq ^d	Abs. diff. ^e	Splicing Outcome ^f
						Three pseudoexons: ▼ 12A.1(r.1899_1900ins1899-174_1899-124), ▼ 12A.2 (r.1899_1900ins1899-177_1899-124), ▼ 12A.3 (r.1899_1900ins1899-213_1899-124)
ATM	c.1899-123A > G	12	AG 0.15 (-51 bp) and AG 0.74 (-90 bp)/ DG 0.71 (-1 bp)	0.63	1.22	Pseudoexon ▼ 16A (r.2466_2467ins2466 + 1555_2466 + 1650)
ATM	c.2466+1552G > C	16	AG 0.93 (3 bp)/ DG 0.69 (97 bp)	-0.144	1.82	Pseudoexon ▼ 61A (r.8850_8851ins8850 + 2030_8850 + 2131)
ATM	c.8850+2029A > G	61	AG 0.22 (1 bp)/ DG 0.16 (102 bp)	-0.187	0.77	Pseudoexon ▼ 5A (r.476_477ins476 + 156_476 + 252)
FAM175A	c.476+158G > T	5	AG 0.17 (2 bp)/ DG 0.22 (-94 bp)	0.56	0.59	Intron retention ▼ 12A (r.997_998ins998-23 + 998)
MUTYH	c.998-27G > A	11	AG 0.41 (-4 bp)/ DG 0.05 (-215 bp)	0.34	1.00	

^a Reference sequences for annotating variants: NM_000051.3 for ATM; NM_139076.2 for FAM175A; NM_001128425.1 for MUTYH. ^bHGVS nomenclature guidelines were used for variant annotation (<http://varnomen.hgvs.org/> accessed September 2020). ^c SpliceAI prediction: AG = Acceptor Gain/DG = Donor Gain. The ΔScore range from 0 to 1 and can be interpreted as the probability that the variant affects splicing at any position around it. For each variant, SpliceAI evaluates a nucleotide window (+/- 4999 in this case) to see how the variant affects the probabilities of different positions in the pre-mRNA being splice acceptors or donors. The values in brackets represent positions with the biggest probability of being used as splicing sites within the window. Negative values are upstream (5') of the variant, and positive are downstream (3') of the variant. In bold are the scores of ≥0.05. ^d ΔESRseq score was calculated as described by Ke et al. (2011)[134], and in bold are indicated values ≥0.63. ^e Absolute difference (Abs diff.): absolute value of the difference between the normalized SRE area of 100 bp before and after a variant. In bold are indicated values ≥0.51. ^f RNA splicing profiles were compared between carriers and controls by capillary electrophoresis and Sanger sequencing.

A) ATM(NM_000051.3) c.1899-123A>G

RT-PCR fragment (203 bp)

FL (203 bp)

▼12A.1 (256 bp)

▼12A.2 (253 bp)

▼12A.3 (291 bp)

Control

c.1899-123A>G carrier

Reverse sequence

Control

c.1899-123A>G carrier

B) ATM(NM_000051.3) c.2466+1552G>C

RT-PCR fragment (553 bp)

FL (553 bp)

Δ18 (353 bp)

▼16a (649 bp)

Control

c.2466+1552G>C carrier

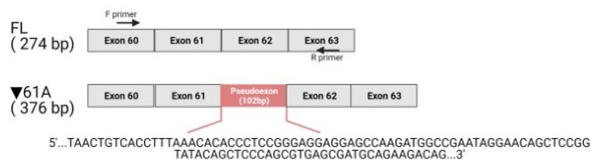
Reverse sequence

Control

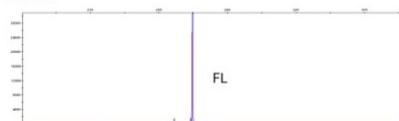
c.2466+1552G>C carrier

C) *ATM* (NM_000051.3) c.8850+2029A>G

RT-PCR fragment (274 bp)



Control

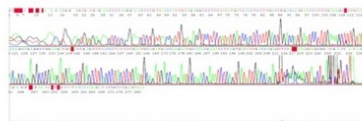


c.8850+2029A>G carrier

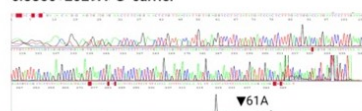


Forward sequence

Control

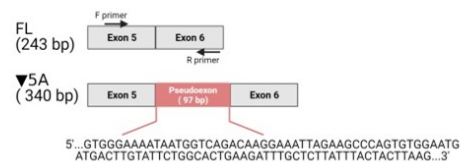


c.8850+2029A>G carrier

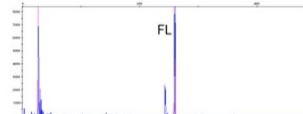


D) *FAM175A* (NM_139076.2) c.476+158G>T

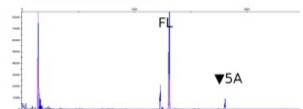
RT-PCR fragment (243 bp)



Control

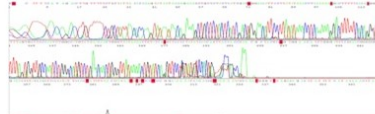


c.476+158G>T carrier

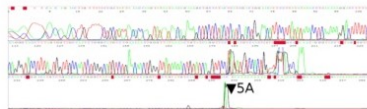


Forward sequence

Control



c.476+158G>T carrier



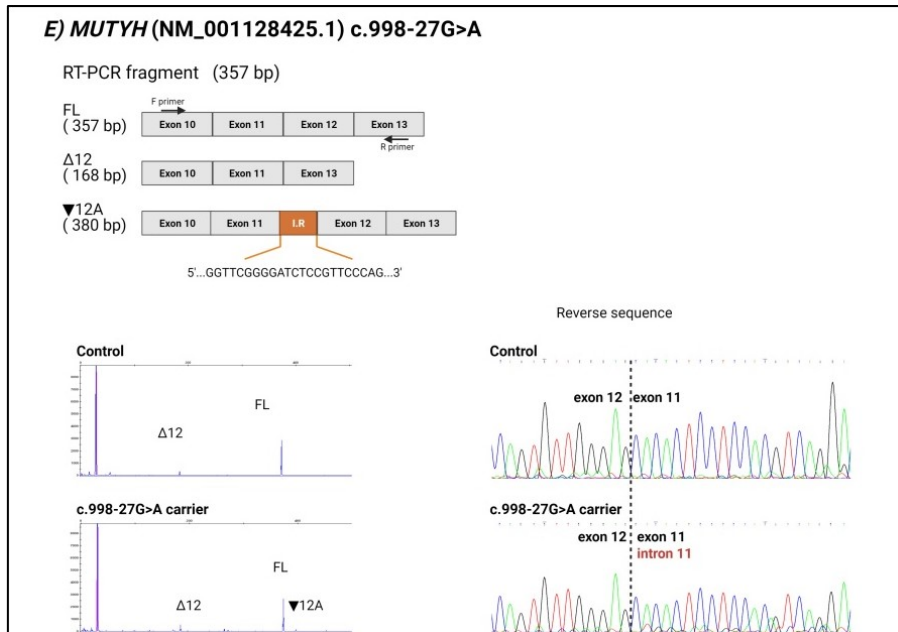


Figure 34. Spliceogenic variants characterization in patients' RNA.

For each variant, there is an RT-PCR assay graphical representation, the results of capillary electrophoresis of 6-FAM labelled amplicons and Sanger sequencing to confirm the expression of additional transcripts. **A)** The *ATM* c.1899-123A > G variant activates a cryptic donor site, which is used to yield three different pseudoexons: ▼12A.1, ▼12A.2, and ▼12A.3, each generated as result of the usage of different cryptic acceptor sites (c.1899-174, c.1899-177, and c.1899-213) and the cryptic donor site created by the variant. The ▼12A.1 and ▼12A.2 transcripts were equally expressed, and their abundance was greater than the ▼12A.3. **B)** The *ATM* c.2466 + 1552G > C variant generates the ▼16A additional transcript. This pseudoexon comprises nucleotides from the acceptor site created by the variant and the cryptic donor at c.2466 + 1650. **C)** The *ATM* c.8850 + 2029A > G variant presents an additional transcript (▼61A), from the cryptic acceptor site created by the variant to the cryptic donor at c.8850 + 2131. It was not possible to clearly read the sequence of the aberrant transcript because of its low expression levels, but in the Sanger sequence, we could detect the additional transcript with the insertion since it is marked by the FAM signal at the end of the fragment. **D)** The *FAM175A* c.476 + 156G > T variant leads to the inclusion of a pseudoexon (▼5A), which results in the usage of the cryptic acceptor site activated by the variant and the cryptic donor at c.476 + 252. Its abundance was very low, but the transcript with the insertion was also detected in the Sanger sequence since it is marked by the FAM signal at the end of the fragment. **E)** The *MUTYH* c.998-27G > A variant creates/enhances a cryptic acceptor site which is used instead of the natural acceptor site of exon 12, generating an intronic retention (IR) (▼12A transcript).

The sensitivity and specificity using SpliceAI alone for all variants with the cutoff of ≥ 0.05 was higher than the pipeline of applying ΔESRseq equal or greater than 0.63, and an absolute difference value greater than 0.51 for those variants with $\Delta\text{SpliceAI}$ scores of < 0.05 (**Table 13**).

Table 13. Performance of $\Delta\text{SpliceAI}$ alone and the proposed pipeline for the experimental dataset.

	Sensitivity	Specificity	Accuracy	MCC
$\Delta\text{SpliceAI} \geq 0.05$	100	71.42	75.75	0.62
$\Delta\text{SpliceAI} \geq 0.05$; $\Delta\text{ESRseq} \geq 0.63$; Abs. Dif. 0.51	100	50.00	57.57	0.43

$\Delta\text{SpliceAI}$ considering 4999bp window

had no assertion criteria provided, eight were reported but no classification was provided, and the remaining 24 were not reported in ClinVar (**Supplementary Table 9**).

A comprehensive summary of splicing and functional data generated in this study, *in silico* analysis, and ClinVar reports can be found in **Supplementary Table 9**.

***In Silico* Identification of Potentially Splicing-Altering Variants**

We used three *in silico* tools to identify all potential splicing-altering variants in our cohort and found that 64 variants were predicted to alter different *cis*-acting splicing elements (**Supplementary Table 9**). Among these tools, we used SpliceAI [168] to identify variants that disrupt natural splicing sites, activate cryptic ones or create *de novo* sites. The analysis indicated that 39 variants potentially altered the exon 18 natural splicing sites ($\Delta\text{SpliceAI AL/DL} \geq 0.2$ threshold according to the calibration done by Walker et al. (2023)[170]), theoretically leading to exon 18 skipping. Of these, 12 variants were in the splicing 3' acceptor site (-4 to +1: NYAG/G) while 16 were in the 5' donor site (-3 to +6: AGC/GUAAGU). The remaining 11 variants were located outside conserved regions, with two in intron 17, eight in exon 18, and one in intron 18 (**Supplementary Table 9**). Furthermore, SpliceAI also predicted that the intronic variant c.5075-8T>G creates a new acceptor site, and the exonic variant c.5141T>G creates a new donor site ($\Delta\text{SpliceAI AG/DG} \geq 0.2$).

MES and BPP *in silico* algorithms were used through SPiP [262] to identify variants that affect the polypyrimidine tract and branch point, respectively. One potential variant was predicted to disrupt the branch point: c.5075-18G>A (**Supplementary Table 9**).

SRE mapping using ESRseq scores estimated according to Ke et al. (2011) [134] revealed ESEs at the acceptor splicing region from c.5075-9 to c.5080 positions, another one between c.5089 to c.5103 nucleotides and one more between positions c.5120 to c.5131. We also identified potential ESSs from c.5083 to 5088, from c.5104 to c.5119, and the last one was identified at the exon-intron boundary from c.5132 to c.5152+21 cDNA positions (**Figure 36 A**). Regarding adjacent intronic regions, we detected a greater abundance of splicing silencers (ISS) in comparison to exon 18. Specifically, intronic donor positions show significantly lower ESRseq scores suggesting a higher presence of intronic silencers in this region (**Figure 36 B**).

To detect variants that could alter any of these SREs, we estimated the Δ ESRseq score for all single nucleotide variants (75 exonic, 57 intronic). According to Tubeuf et al. (2020)[172] exonic variants showing a score change lower than -0.5 is indicative of exon skipping due to the disruption of an ESE or by the creating/activating of an ESS, while a score variation greater than 0.36 is indicative of exon inclusion by creating/activating an ESE or by disrupting an ESS. Among the 75 single nucleotide exonic variants examined, the Δ ESRseq estimations pointed out 30 variants that potentially enhance the occurrence of exon 18 skipping (Δ ESRseq $\leq$ -0.5). For eight out of these 30 variants, SpliceAI also predicted the exon skipping effect. In addition, ESRseq identified 32 variants that might promote exon 18 inclusion (ESRseq score change >0.36) (**Supplementary Table 9**).

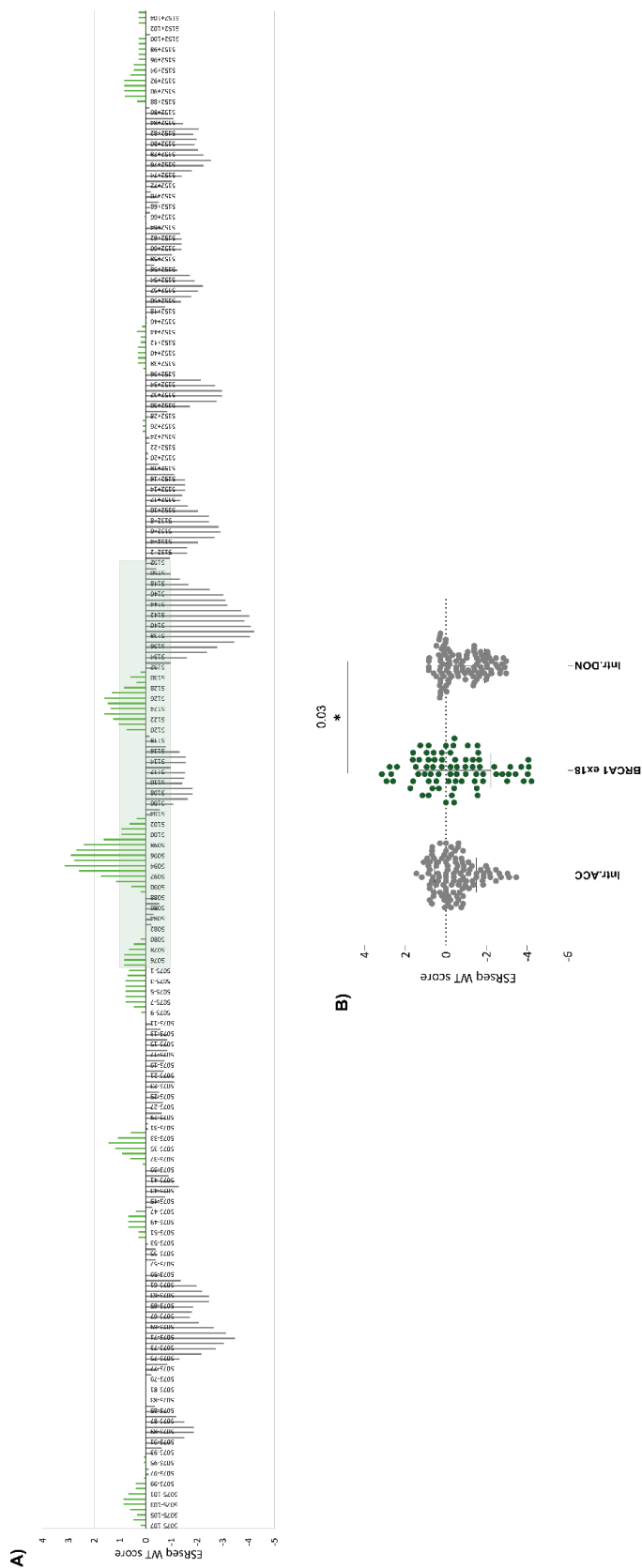


Figure 36. Splicing regulatory elements (SREs) distribution among exon 18 and intronic adjacent regions.

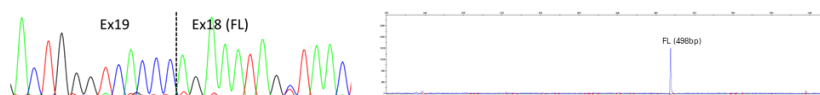
A) WT ESRseq scores were calculated according to Ke et al. (2011) [134] for all positions located between nucleotides c.5075-107 to c.5152+105 to map all SREs that might be involved in exon 18 splicing. In green are shown positions considered enhancers (exonic: ESEs; intronic: ISEs) and in grey the silencers (Exonic: ESS and intronic: ISS). The box englobes all exonic positions. **B)** ESRseq WT scores for all exon 18 and its adjacent intronic positions. A significant difference among groups was identified (Kruskal-Wallis, p -value=0.015), and a multiple comparison Dunn's test revealed that the mean of SRE is significantly different in intronic donor site in comparison to SREs in exon 18, being the donor region enriched with silencer elements (p -value=0.027). *Intr.ACC: intronic acceptor region (c.5075-107 to c.5075-1), BRCA1 ex 18 (ESRseq WT scores from c.5075 to c.5152), Intr.DONOR: donor acceptor region (c.5152+1 to c.5152+105).*

mRNA Splicing Profiling Characterization in Patient Samples

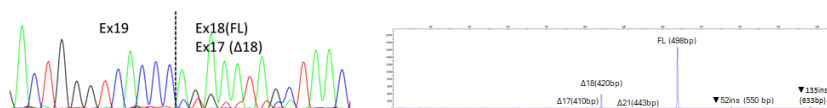
In this study, we analysed a total of 32 unique variants using RNA samples from 51 carriers, and results were compared with samples from 46 control non-carriers.

Patient cDNA study by capillary electrophoresis and Sanger sequencing revealed the presence of several transcripts: the reference full-length (FL, 498bp); which was present in all patients and control samples (**Figure 37 A**); exon 18 skipping ($\Delta 18$, 420 bp) (**Figure 37 B and C**), skipping of exons 17 and 18 ($\Delta 17-18$, 332 bp) (**Figure 37 D**), among other less abundant isoforms listed in **Supplementary Table 12**, which were inferred by amplicon size because they could not be confirmed by Sanger sequencing.

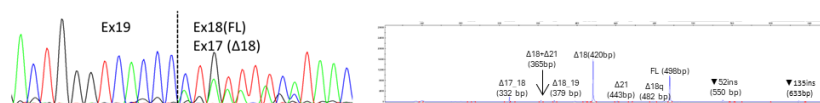
A) Normal splicing profile: **c.5144G>A** (FL=100%)



B) Intermediate skipping: **c.5123C>A** (Carrier 2, $\Delta 18=19.26\%$)



C) Complete skipping: **c.5152+3A>C** (Carrier 1, $\Delta 18=58.33\%$)



D) Complete skipping: **c.5152+2dup** (Carrier 2, $\Delta 18=43.43\%$)

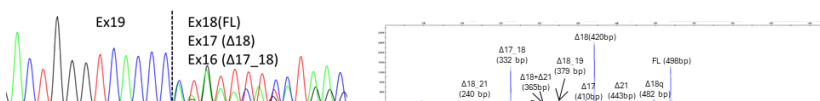


Figure 37. Representative Sanger and capillary-electrophoresis (C.E) results of *BRCA1* splicing events detected in patients' samples.

A) Example of a normal splicing from a c.5144G>A carrier (whole blood) detected by RT-PCR assay, using primers in exon 16 and 22 of *BRCA1*. Only the full-length (FL) (498bp) transcript was detected. **B)** Splicing profile of c.5123C>A variant carrier (whole blood), which showed intermediate levels of exon skipping, hence FL and $\Delta 18$ (420bp) were detected in C.E and confirmed by Sanger. **C)** Splicing profile of a c.5152+3A>C carrier (whole blood RNA). This variant induces (near) complete exon 18 skipping. In Sanger and C.E, FL and $\Delta 18$ (420bp) were detected, and in C.E less abundant isoforms involving the skipping of exon 18, such as $\Delta 17\_18$ (332bp) were inferred by size. **D)** Splicing profile of a carrier of c.5152+2dup variant, by using RNA from short-term cultured lymphocytes treated with a nonsense mediated decay inhibitor (Puromycin). In this sample, the three above-mentioned transcripts (FL, $\Delta 18$ and $\Delta 17\_18$) were detected in Sanger since $\Delta 17\_18$ (332bp) truncating transcript was not degraded. In all panels, transcripts indicated in C.E but not in Sanger were inferred by amplicon size. See Supplementary Table 12 for detailed information.

Splicing fraction (considering expression of both alleles) in controls revealed a normal range of exon 18 skipping between 0 and 3.8%, and a mean of 97.2% of FL expression (**Figure 38 A, Supplementary Table 13**). Given that some intronic and synonymous variants (c.5075-53C>T, c.5100A>G, c.5124G>A, c.5152+66G>A, and c.5152+85del) had been previously classified as benign by the ENIGMA (old criteria), we considered their splicing profile also as benign reference range. These benign variants showed a $\Delta 18$ splicing fraction between 0% to 12.6%, and a FL from 78.5% to 100% (**Supplementary Table 9**). Accordingly, a total of 13 variants showed $\Delta 18$ expression levels within this range: c.5075-107A>G, c.5075-49del, c.5075-6C>A, c.5075A>T, c.5091_5092del, c.5095C>T, c.5096G>A, c.5117G>A, c.5117G>C, c.5144G>A, c.5145G>C, c.5152+10A>G, and c.5152+20T>A (**Figure 38 B, Supplementary Table 13**).

In contrast, the splicing fraction (considering the expression of both alleles) profile observed in carriers of 12 variants was suggestive of complete exon skipping (Splicing fraction $\Delta 18$ = 43.4-66.8%, FL= 24.9- 43.5%). Most of these variants were located in splice site regions (acceptor 3': -4 to +1, donor 5' : -3 to +6) (**Figure 38 B, Supplementary Table 13**): c.5075-2A>G, c.5152+1G>C, c.5152+1G>T, c.5152+2dupT, c.5153+3A>C, c.5152+4A>T, 5152+5G>A,

c.5152+5G>C, and c.5152+6T>C, except c.5077_5080delinsTTCATTCTGC and c.5080G>T which were located in an ESE near to the canonical acceptor site (**Figure 36**) and c.5152+7A>G located beyond the donor site region. All 12 variants showed a FL splicing fraction lower than 50% (**Figure 38 B, Supplementary Table 13**). However, this semi-quantitative measure was obtained from RT-PCR experiments and there is a bias preferably amplifying smaller transcripts, such as $\Delta 18$ and $\Delta 17-18$, and thus could be masking a slight contribution of the variant allele to FL-expression. To address this concern, a FL allele-specific assessment was performed when possible. In total, we were able to confirm that the contribution of alleles carrying c.5075-2A>G, c.5152+1G>C, c.5152+2dup, c.5152+4A>T, c.5152+5G>A, c.5152+5G>C, c.5152+6T>C, c.5077_5080delinsTTCATTCTGC, and c.5080G>T, to FL expression was indeed negligible (**Figure 39 A-K**).

In addition, we detected exon 18 skipping in other samples with levels suggesting a partial or leaky splicing effect. The splicing fraction observed in two c.5078_5080del carriers indicated that the variant allele was expressing both r.5078_5080del and $\Delta 18$ (r.5078_5080del=10.2 % and 20.3 %; $\Delta 18$ = 56.2% and 32.4%) (**Figure 38 B and Supplementary Table 13**). In concordance, the respective FL-allele-specific analysis showed a substantial FL expression (**Figure 39 N and S**). Likewise, for two out of the three c.5123C>A carriers, the FL-allele specific results agreed with the skipping levels of exon 18, 19.3% (i.e FL=79.3%) and 16.7% (i.e FL= 80.1%), respectively (**Figure 39 P and R, and Supplementary Table 13**). However, the third c.5123C>A carrier whose RNA was obtained from a different source (LCL), exhibited a lower leakiness rate of exon 18 skipping ($\Delta 18$ =2.4%, FL=91.9 %) and a balanced bi-allelic expression (**Figure 39 T and Supplementary Table 13**).

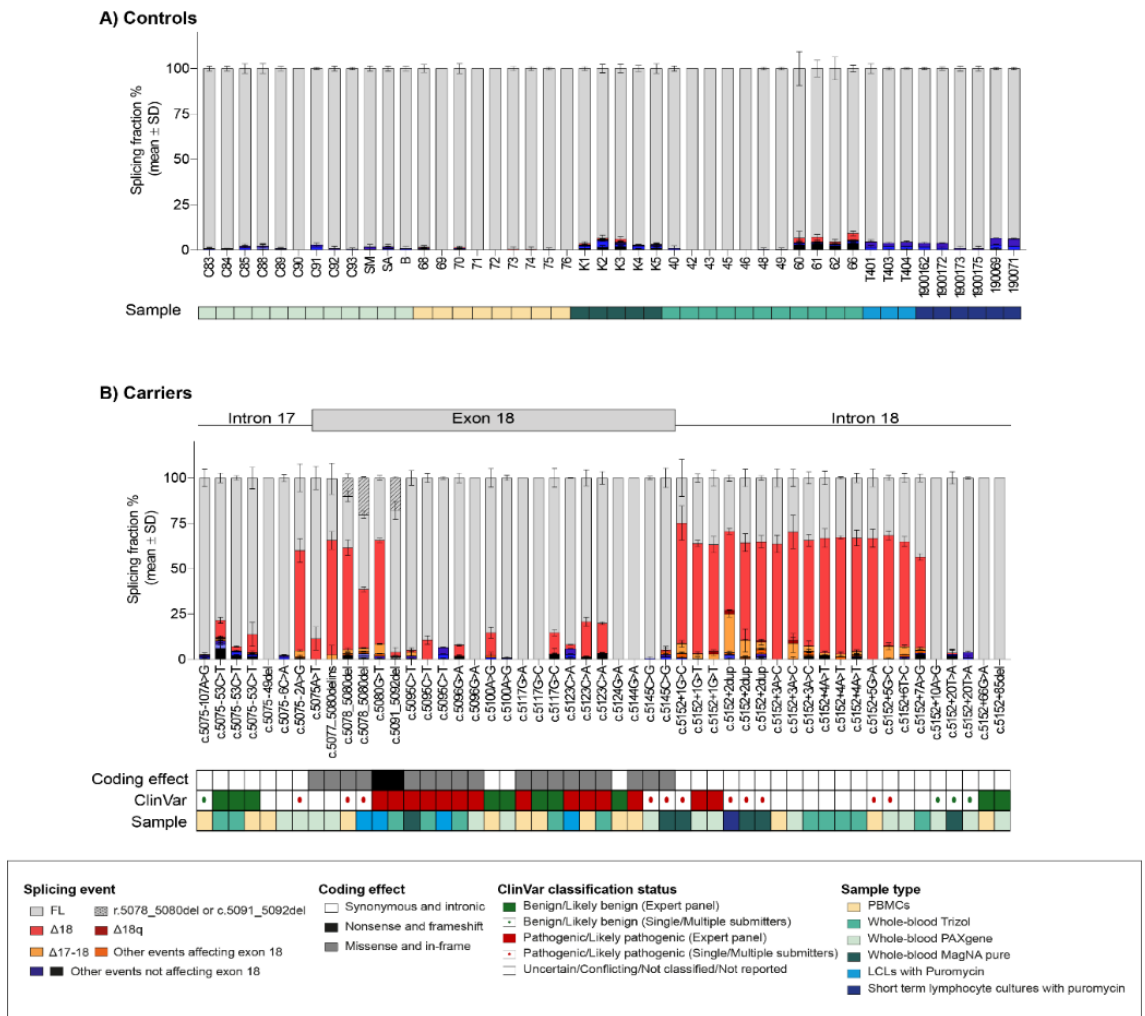


Figure 38. Semi-quantitative mRNA splicing profiling by capillary electrophoresis of fluorescent amplicons in controls (n=46), patients (n=51 samples, 32 unique variants).

Splicing fractions of RNA from non-carrier controls (A) and carriers (B) of *BRCA1* exon 18 variants. The mean and standard deviation (SD) of at least two RT-PCRs spanning exons 16 - 22 of *BRCA1* are shown. Isoform levels are calculated by dividing the peak area of each isoform by the total peak areas detected in the samples and then represented as %. Splicing events: FL (full-length), Δ18 (exon 18 skipping), Δ18q (cryptic donor usage), Δ17-18 (skipping of both exon 17 and 18), other isoforms including exon 18, other isoforms not affecting exon 18 (details in **Supplementary Table 12**). ClinVar classification and variant review status, sample type and coding effect is also indicated for each variant.

NEGLECTIBLE VARIANT ALLELE CONTRIBUTION TO FL SYNTHESIS		
ASSAY 1	DNA	Full-length transcript (cDNA)
A) c.5075-2A>G ($\Delta 18=55.2\%$)		
B) c.5152+1G>C ($\Delta 18=66.3\%$)		
C) c.5152+2dup (Carrier 2 $\Delta 18=53.7\%$)		
D) c.5152+2dup (Carrier 3, $\Delta 18=54.8\%$)		
E) c.5152+4A>T (Carrier 2 $\Delta 18=64.4\%$)		
F) c.5152+4A>T (Carrier 3 $\Delta 18=62.5\%$)		
G) c.5152+5G>A ($\Delta 18=66.8\%$)		
H) c.5152+5G>C ($\Delta 18=60.9\%$)		
I) c.5152+6T>C ($\Delta 18=58\%$)		
ASSAY 2	DNA	cDNA
J) c.5077_5080delinsT TCATTCTGC ($\Delta 18=63\%$)	DNA NOT AVAILABLE	
ASSAY 3 (SNaPShot)	DNA	cDNA
K) c.5080G>T: LCL-PURO (PURO+ $\Delta 18=57.1\%$)		

BOTH ALLELES CONTRIBUTE TO FL SYNTHESIS		
ASSAY 1	DNA 	Full-length transcript (cDNA)
L) Control (c85) (Δ18=0.1%)	Fw 	Fw
M) c.5075-53C>T (Carrier 3, Δ18=10.2%)	Fw 	Fw
N) c.5078_5080del (Carrier 1, r.5080_5078del=10. 2% Δ18=56.2%)	Fw 	Fw
ASSAY 1B	DNA 	cDNA
O) c.5123C>A (Carrier 3, Δ18=16.7%)	Rv 	Rv
ASSAY 2	DNA 	cDNA
P) c.5123C>A (Carrier 2, Δ18=19.3%)	Rv 	Rv
ASSAY 3 (SNaPShot)	DNA 	cDNA
Q) c.5078_5080del LCL+PURO (Carrier 2,LCL+PURO r.5078_5080del=20. 3% Δ18=32.4%)	DNA Rv 	cDNA Rv
R) c.5123C>A LCL- PURO (Carrier 1,LCL+PURO Δ18= 2.4%)	DNA Rv 	cDNA Rv

Figure 39. Allele-specific assays to assess the contribution of mutant allele to full-length (FL) expression.

PCR experiments were performed using samples from carriers of variants with an altered splicing profile ($\Delta 18$ splicing fraction $>12.6\%$), to know the contribution of mutant allele to full-length (FL) expression. Qualitative assays (Assay 1, 1B and 2) consist of RT-PCR experiments using primers designed to amplify a region containing exon 18. Allele-specific assay 1 (**A-I, L-N**) was based on the utilization of a forward primer at exon 15 and a reverse hybridizing with exon 18, due to it was the exon exclusively present in normal FL transcript and this region also contained the SNP c.4837A>G in exon 16, previously confirmed to be heterozygous in carriers. Assay 1B (**P**) was based on same rationale, but the primers were 18F and 20R and the heterozygous exonic position was in exon 18 variant (c.5123C>A). Two samples, one non-carrier control (**L**) and one carrier of the c.5075-53C>T (**M**) were included as controls. Assay 2 (**J, P**) consisted of the amplification of a region containing the exon 18, but no amplifying exclusively FL molecules, by using 16F and 22R primers and the heterozygous status the same exon 18 variant under assessment. DNA genotyping was included depending on sample availability and variants and splicing fractions are indicated for each sample. In assay 3 (**K, Q, R**), allele-specific expression was measured by SNaPshot analysis in LCL (without puromycin: PURO-). Genomic segment (gDNA) and cDNA containing exon 18 were amplified in parallel using specific primers and the SNaPshot quantitative primer extension assay was then performed with a variant-specific reverse primer (black thin arrow in exon 18) targeting the sequence immediately upstream the variant (star). Allele-specific expression levels representing the relative contribution of the variant allele to the expression of full-length (FL) *BRCA1* transcripts as compared to WT ("variant allele relative to WT"), were measured by normalization of the cDNA peak area ratio (variant/WT) to the corresponding gDNA peak area ratio (variant/WT). Results are representative of three independent experiments, and the mean value is indicated.

Then, we assessed the SpliceAI performance in detecting splicing events. Although we defined benign splicing profiles using the isoform expression levels from carriers of intronic and synonymous benign variants, we used the highest level of exon 18 skipping observed in non-carrier controls (3.8%, **Figure 38 A and Supplementary Table 13**) as a threshold to define a true positive SpliceAI spliceogenic prediction (≥ 0.2). This choice was made because benign variants might influence exon 18 skipping, albeit in a tolerable manner. Accordingly, SpliceAI showed a sensitivity and specificity of 80%, and 76.5%, respectively (**Supplementary Table 14**). SpliceAI predicted as spliceogenic a total of 15 variants (≥ 0.2). Of those, only two showed no impact on splicing (c.5075-6C>A and c.5145C>G). On the contrary, 12 variants were predicted ($DS \leq 0.1$) as non-spliceogenic, but we observed a splicing altered profile in four of these variants ($\Delta 18 > 3.8\%$). Out of these four, the c.5077_5080delinsTTCATTCTGC led to a

complete splicing defect. The remaining three variants showed a leaky splicing defect and not in all carriers (c.5075-53C>T, c.5100A>G, and c.5117G>C). The remaining variants analysed in patient RNA presented a Δ SpliceAI between 0.2 and 0.1 and the predicted effect was uncertain.

RNA Splicing Evaluation through Minigene Reporter Assay

A total of 166 variants were assessed using RNA from minigenes, including the 32 variants already analysed using RNA from patients (**Figure 40, Supplementary Table 15**). The minigene splicing assays showed 0% of Δ 18 and 100% FL in WT constructs. Similarly, benign synonymous and intronic variants previously classified by old ENIGMA criteria showed up to a 0.3% occurrence of Δ 18 (**Supplementary Table 9**). Therefore, we determined that 121 variants demonstrated a normal splicing profile with FL $\geq$ 99%, five out of them also associated with a slight activation of exonic cryptic sites (0.3-0.5%). In contrast, 39 variants induced exon 18 skipping with 1.9 to 100 % and 98-0% of FL, and 13 out of the 39 variants also presented the activation of an exonic cryptic 5' splice site (ss) leading to the loss of the last 16 exonic nucleotides of exon 18 (Δ 18q) with levels of 0.2-5.7%. Finally, six variants that exhibited normal levels of Δ 18 (i.e., $\leq$ 1%) but with FL<99%, showed the usage of cryptic or *de novo* ss with levels of 1.5-100% (**Figure 40, Supplementary Table 15**).

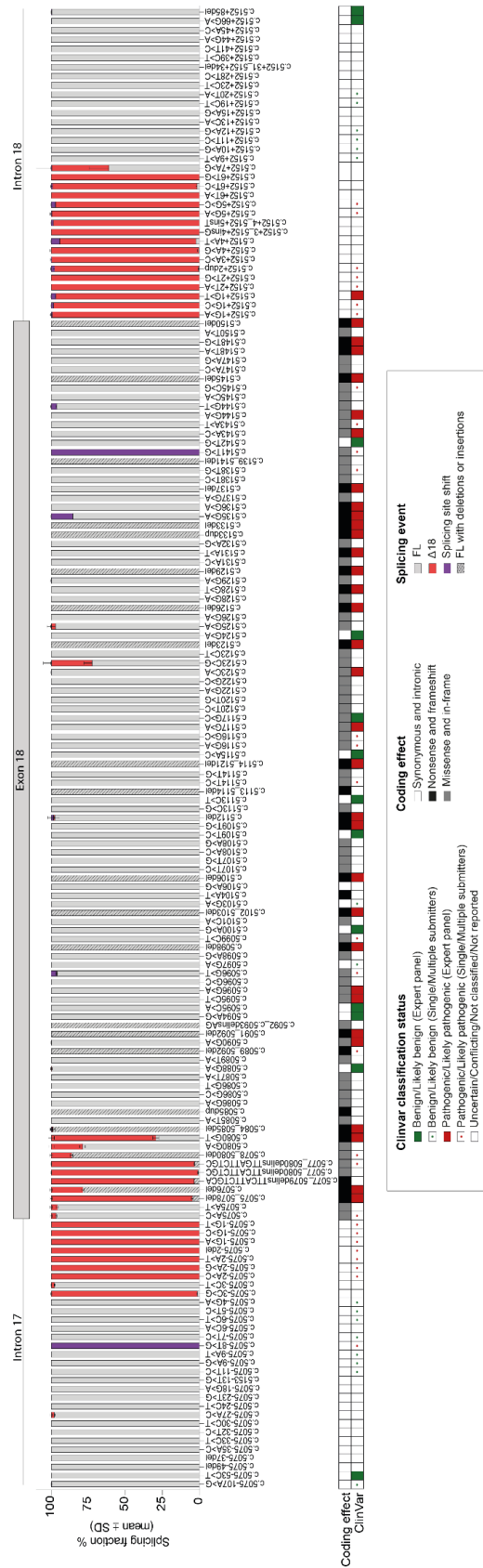


Figure 40. Semi-quantitative mRNA splicing profiling by capillary electrophoresis of fluorescent amplicons in minigenes (n=166).

The mean and standard deviation (SD) of at least 2 RT-PCRs is shown. ClinVar classification and variant review status, and coding effect are indicated for each variant. Detailed splicing fractions can be found in **Supplementary Table 15**.

Of the 39 variants showing exon 18 skipping: 12 were within the acceptor splicing region (c.5075-20 to c.5075, both included), with exon 18 skipping levels around 100% except c.5075-3C>T and the two variants at the c.5075 position, which showed a range of exon skipping of 2.2% to 4%; one intronic variant outside the acceptor site region expressed 1.9% of transcripts lacking exon 18 (c.5075-27A>C); 16 variants at the donor site region (c.5149 to c.5152+6, both included), lead to exon 18 skipping levels greater than 90% (91.9 to 100%); one intronic variant outside the donor site region led to a 38.4% of exon 18 skipping (c.5152+7A>G), and nine exonic variants outside splice site regions skipped exon 18 to a range of 2.6% to 99% (c.5076del, c.5077_5079delinsTTCATTCTGCA, c.5077_5080delinsTTCATTCTGC, c.5077_5080delinsTTGATTCTGC, c.5078_5080del, c.5080G>A, c.5080G>T, c.5123C>G, and c.5125G>A) (**Figure 40, Supplementary Table 15**).

SpliceAI showed an accuracy of 92.1% in predicting variant spliceogenic effect using minigene outcomes, with a sensitivity and specificity of 87.8% and 94.2%, respectively (**Supplementary Table 14**). Specifically, the computational tool identified 41 variants as potentially altering splicing ($\Delta\text{SpliceAI} \geq 0.2$), of which 36 impacted splicing; the five variants that did not affect splicing were c.5075-6C>A, c.5123C>A, c.5123del, c.5145C>G, and c.5148T>G. (**Figure 40 , Supplementary Table 15**). On the contrary, a total of 86 variants were predicted not to alter splicing ($\Delta\text{SpliceAI} \leq 0.1$). However, c.5077_5079delinsTTCATTCTGCA, c.5077_5080delinsTTCATTCTGC, c.5077_5080delinsTTGATTCTGC, showed exon skipping levels between 96.4-99.0%, and c.5112del and c.5125G>A variants presented splicing effects generating transcripts lacking exon 18 as well as transcripts generated from cryptic splice sites activation (**Figure 40, Supplementary Table 15**).

Spearman's rank correlation analysis indicated a strong positive association between patient per-allele and minigene FL transcript levels ($r=0.82$, p -value <0.0001) (**Figure 41**).

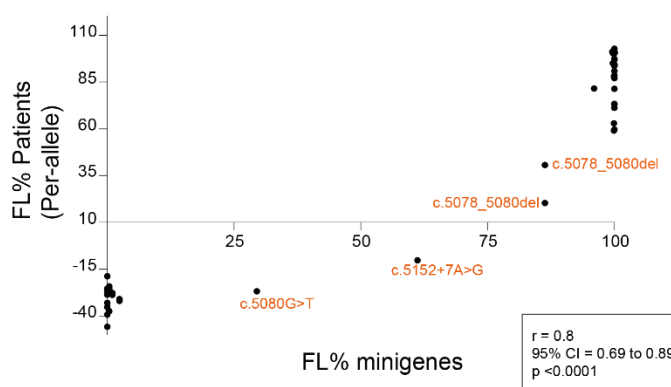


Figure 41. Correlation analysis of full-length (FL) expression in patients and minigenes.

Patient per-allele FL expression levels (Y-Axis) plotted against FL levels observed in minigenes (X-Axis). Per-allele FL expression in patients was needed for comparing splicing effects with the minigenes mono allelic model, and it was calculated as twice the FL level in carriers minus the mean FL expression in controls of the same sample type. Variants showing non-concordant results between patients and minigenes are indicated in orange (see methodology Part 2, variant classification, mRNA Splicing Observed Evidence: PVS1(RNA) and BP7_strong (RNA) section for detailed information).

Functional Tolerability Range of Exon 18 Skipping

As there was no functional evidence regarding the tolerability of exon 18 skipping concerning BRCA1 function, we aimed to define the $\Delta 18$ threshold for pathogenicity and the minimum expression level of FL transcript required to retain BRCA1's functionality in HDR. With this purpose, functional characterization was performed for 18 variants which displayed different expression levels of the $\Delta 18$ transcript in patients and/or minigene assays: complete, partial and normal skipping of exon 18 (**Table 14**). The variants were introduced into the FL human *BRCA1* gene and transfected into hemizygous mESC to measure variant impact on splicing and BRCA1 protein activity. The

mESC-based model system allows to directly link transcript profiles to protein functionality as measured by cell viability after removal of the conditional *mBrca1* allele and the capacity to repair a DSB via HDR in the DR-GFP reporter assay.

Table 14. Variants assessed in the mESC-based functional assay.

Variant	Coding effect	Expert panel classification ^A	Splicing selection criteria ^B	Δ18% splicing fraction in Carriers (Per-allele) ^C	Δ18% Minigenes	Δ18% mESC	HDR capacity (mean %)	Complementation ^D	Saturation genome editing assay ^E		
									FUNC score	FUNC class	RNA score
WT-BRCA1	-	-	Wild-type	0-3.8% (0-7.6%)	0.0%	0.0%	100%	Good	-	-	-
Null BRCA1	-	-	Null	-	-	-	4.4%	Poor	-	-	-
BRCA1-Exon 18 deletion	Deletion of exon 18	-	Deletion of exon 18	-	-	83.7%	6.7%	Poor	-	-	-
c.5077_5080delinsTTCATTC TGC	In-frame	-	Complete	C1: 63.01% (125.5%)	92.1%	92.1%	9.5%	Poor	-	-	-
c.5152+1G>C	Intronic	-	Complete	C1: 66.3% (131.9%)	98.5%	82.9%	8.0%	Poor	-1.80	LOF	-
c.5152+1G>T	Intronic	Pathogenic	Complete	C1: 60.7% (121.5%), C2: 58.7% (117.4%)	96.9%	81.9%	12.0%	Poor	-1.99	LOF	-
c.5152+2dup	Intronic	-	Complete	C1: 43.4% (86.9%), C2: 53.7% (106.7%), C3: 54.8% (108.9%)	97.4%	87.2%	13.2%	Poor	-	-	-
c.5152+3A>C	Intronic	-	Complete	C1: 63.6% (127%), C2: 59.7% (119.5%), C3: 58.3% (115.7%)	99.7%	82.5%	7.0%	Poor	-1.52	LOF	-
c.5152+4A>T	Intronic	-	Complete	C1: 63.4% (125.8%), C2: 64.4% (127.9%), C3: 62.5% (124.1%)	91.9%	88.0%	10.5%	Poor	-1.79	LOF	-
c.5152+5G>A	Intronic	-	Complete	C1: 66.8% (133.3%)	99.5%	85.5%	10.3%	Poor	-2.10	LOF	-
c.5078_5080del	In-frame	-	Partial	C1: 56.2% (112.1%), C2: 32.4% (64.8%)	13.7%	17.3%	17.0%	Poor	-	-	-
c.5123C>A	Missense	Pathogenic	Partial in two carriers, normal in one carrier and minigene	C1: 2.4% (4.7%), C2: 19.3% (38.3%), C3: 16.7% (32.4%)	0.0%	0.1%	12.5%	Poor	-2.01	LOF	-0.55
c.5152+7A>G	Intronic	-	Near complete in a carrier, partial in minigene	C1: 50.3% (99.6%)	38.4%	30.2%	75.8%	Good	-1.85	LOF	-
c.5075-53C>T	Intronic	Benign	Normal	C1: 9.4% (17.9%), C2: 2.6% (4.2%), C3: 10.2% (20.2%)	0.0%	0.0%	86.3%	Good	-	-	-
c.5095C>T	Missense	Pathogenic	Normal	C1: 1.2% (1.9%), C2: 10.6% (20.3), C3: 0.2% (0.4%)	0.0%	0.0%	16.3%	Poor	-	-	-
c.5096G>A	Missense	Pathogenic	Normal	C1: 6.01% (11.2%), C2: 0%	0.0%	0.0%	29.5%	Poor	-	-	-
c.5100A>G	Synonymous	Likely benign	Normal	C1: 12.6% (25%), C2: 0%	0.0%	0.0%	122.7%	Good	0.34	FUN	0.16
c.5117G>A	Missense	Pathogenic	Normal	C1: 0%	0.0%	0.4%	8.5%	Poor	-2.32	LOF	-0.51
c.5117G>C	Missense	Benign	Normal	C1: 0%, C2: 11.8% (22.6%)	0.0%	0.0%	74.0%	Good	-0.05	FUN	-0.20
c.5124G>A	Synonymous	Likely benign	Normal	C1: 0%	0.0%	0.0%	95.5%	Good	0.02	FUN	-0.15
c.5143A>C	Missense	Pathogenic	Normal	-	0.0%	0.4%	7.3%	Poor	-2.21	LOF	-

^A ENIGMA variant classification reported on ClinVar before this study (old criteria). ^B Splicing selection criteria: complete (patients: allele-specific (AE) assay confirming negligible contribution of the variant allele to FL transcript synthesis, minigenes: Δ18 ≥ 99%), near complete (patients: splicing fraction FL<50% but no AE assay available), partial (patients: splicing fraction Δ18 >12.6% and AE showing the contribution of variant allele to FL synthesis, minigene: 0%< Δ18 <99%) and normal (patients: splicing fraction Δ18 ≤12.6%, minigenes: Δ18<1%). ^C Splicing results from RT-PCRs experiments performed in VHIO. ^D Complementation categories: poor (<20% viability, including no complementation), reduced (20-50% viability), or good (>50% viability). ^E Results from Findlay et al. (2018)[271]. Null BRCA1: mBrca1 mESC cells after removal of the conditional mBrca1 allele, C: carriers, Wt: wild-type, FUNC: functional, LOF: loss of function.

The RNA splicing analysis from mESC identified the same isoforms observed in patients (**Figure 42, Supplementary Table 16**), and splicing profiles showed a positive significant correlation with both patients (Spearman $r=0.87$, $p<0.0001$) and minigenes (Spearman $r=0.85$, $p<0.0001$) (**Supplementary figure 1**).

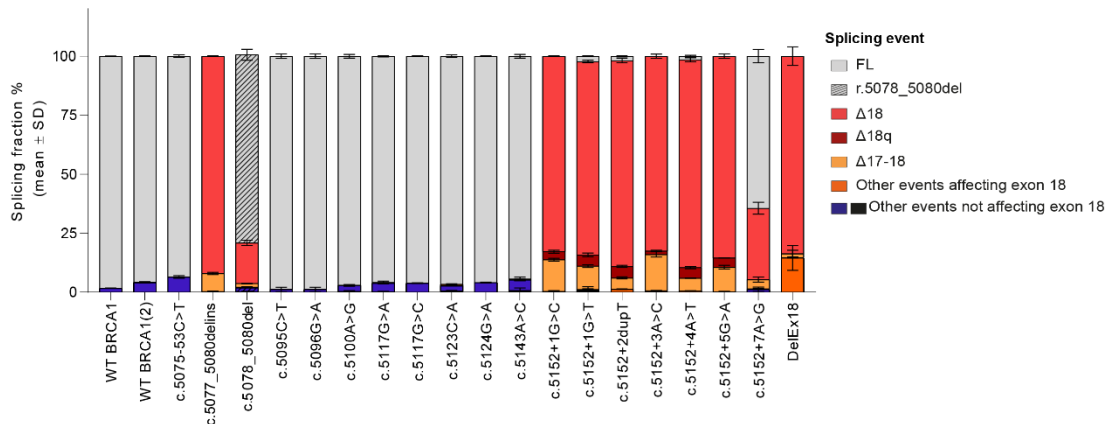


Figure 42. Mouse-embryonic stem cells (mESC) mRNA splicing profiling.

Semi-quantitative mRNA splicing profiling by capillary electrophoresis of fluorescent amplicons using RNA of variant- expressing mESC after *mBrca1* depletion. RT-PCR assays were performed amplifying the cDNA region between exons 16 and 22 of *BRCA1*. The mean and standard deviation of at least three RT-PCR experiments are shown. Detailed splicing fractions can be found in **Supplementary Table 16**. *WT*: wild-type, *FL*: full-length.

The seven variants with complete exon 18 skipping observed in patients and minigenes also showed a completely altered splicing profile in mESCs, showing Δ18 levels from 81.9 to 92.1% and FL up to 2.3% (**Table 14**). Regarding the variants with normal splicing profiles in patients and/or minigenes, all of them showed Δ18 levels under 1%, and FL close to 100% as WT-BRCA1 mESC (Δ18=0%, FL=95 to 98%) (**Table 14**).

The c.5123C>A variant with partial skipping in two out of three carriers, and no skipping in the minigene assay, did not show leaky splicing effects in mESCs (**Table 14**, **Figure 42**). In contrast, the c.5078_5080del showed partial skipping in mESCs as observed in patients and minigenes (**Table 14**, **Figure 42**). In mESCs carrying the c.5152+7A>G, both $\Delta 18$ and FL transcripts were detected at similar levels to those observed in minigenes ($\Delta 18=30.2\%$, FL=64.5%), nonetheless, exon skipping was higher in one carrier (Per allele $\Delta 18=99.7\%$) (**Table 14**).

Ten of the 18 variants were already classified by ENIGMA (old criteria) before this study, six as pathogenic and four as (likely) benign (**Table 14**). Their HDR activity was used to define two ranges. For the benign range the lower HDR threshold was set at 64% as that was the lowest HDR capacity measured for the benign variants (**Figure 43**). The threshold for the pathogenic range was set at 34% HDR capacity, which was the highest measurement observed for the pathogenic variants (**Figure 43**).

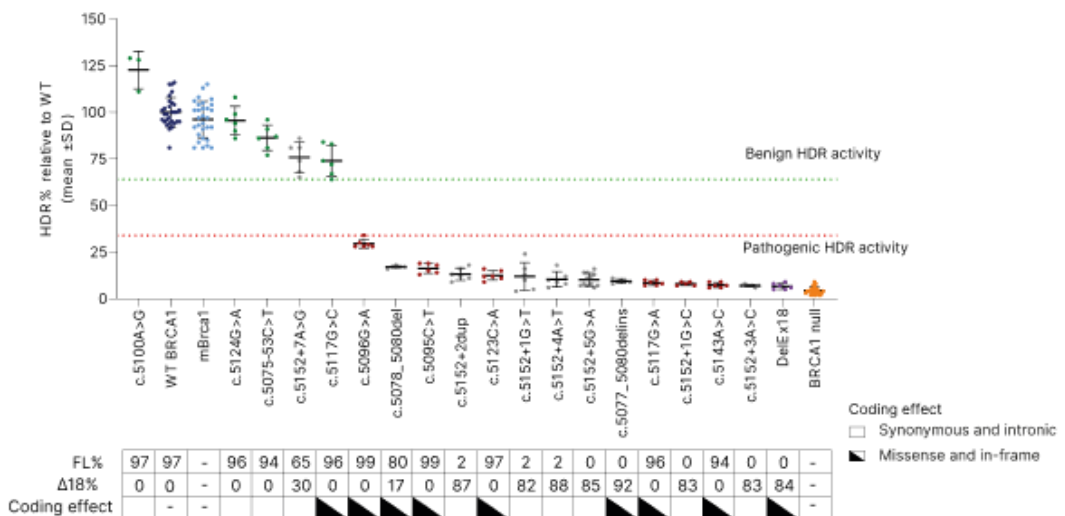


Figure 43. Homology direct repair (HDR) activity of BRCA1 exon 18 variants relative to wild-type BRCA1 capacity in mESC.

The fraction of GFP-expressing cells was determined by flow cytometry 2 days post-transfection with an I-Sce1 expression vector. Relative HDR activity is expressed as the ratio between the percentage of GFP-positive cells observed in BRCA1 variant-expressing cells and the percentage of GFP-positive cells in WT BRCA1-expressing cells (dark blue). The upper box represents the HDR range of benign variants. The lower box represents the HDR range of pathogenic ones. Tested variants are indicated in the X-axis: ENIGMA classified benign (green) and pathogenic variants (red), variants of unknown significance (grey), mBrca1 was the conditional Brca1 -/loxP Pim1DR-GFP/WT cell line expressing endogenous mouse Brca1 (light blue), BRCA1 null was the same cell line after removal of the conditional *mBrca1* allele (orange dots), the DelEx18 cell line was transfected with a *BRCA1* gene without the coding DNA sequence of exon 18 (purple). Error vertical bars indicate the standard deviation (SD) of at least three independent GFP measurements per variant. Below are indicated the levels of FL and $\Delta 18$ transcripts for each cell line, and the coding effect of each variant. *WT*: wild-type, *HDR*: homology direct repair capacity, $\Delta 18$: exon 18 skipping isoform, *FL*: full-length.

A severe functional impairment was measured for the seven variants showing complete exon skipping with HDR measurements in the pathogenic range (**Figure 43**). The HDR activity of these variants was close to the activity observed in BRCA1 deficient cells (HDR=4.4%) and cells expressing the BAC-variant deleted for exon 18 (HDR=6.7%) (**Figure 43**). In addition, all exhibited a poor complementation phenotype, as depicted in **Figure 44**. These two functional outcomes (HDR and complementation phenotype) suggest that the complete absence of exon 18, despite encoding an in-frame protein (as illustrated in **Supplementary Figure 8**), results in a non-functional protein. Contrarily, the c.5152+7A>G inducing 30% of $\Delta 18$, showed an HDR capacity in the functional range of benign *BRCA1* variants (HDR~76%) (**Table 14 and Figure 43**), suggesting that this splicing profile has no deleterious impact and allows for substantial BRCA1 protein activity.

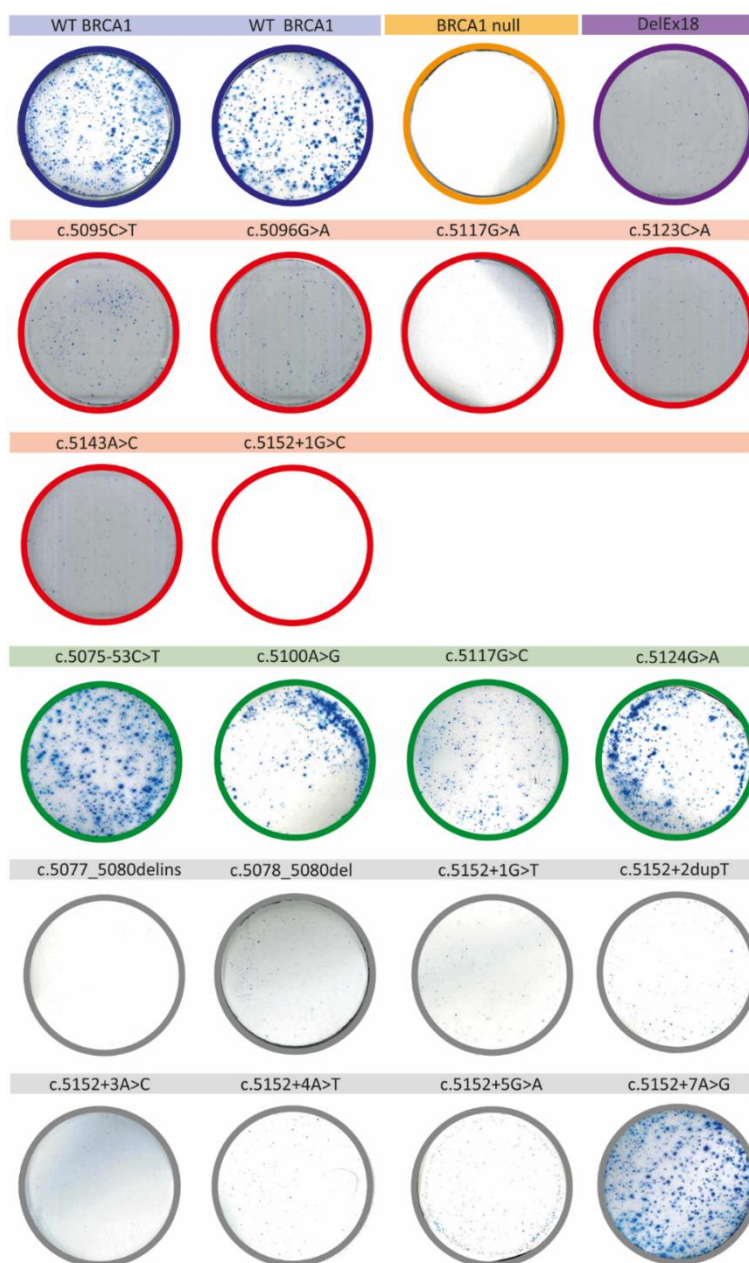


Figure 44. Representative images of complementation phenotypes of BRCA1 exon 18 variants in mESC.

As BRCA1 protein activity is essential for mESC viability, the capacity of BAC-transfected cells to form clones after removal of the conditional mBrca1 allele is indicative for variant impact on protein function. BRCA1 null represents the mESC line wherein endogenous expression of Brca1 is depleted by removal of the conditional allele. All other dishes show the complementation capacity after removal of the conditional allele of cells transfected with either a WT BRCA1 BAC or BRCA1 variants; The DelEx18 cell line is transfected with a BRCA1 gene wherein the DNA sequence encoding exon 18 is removed. The red circles indicate variants considered (likely) pathogenic whereas the green circles indicate variants considered (likely) benign according to ENIGMA old-criteria classification.

To estimate the maximum tolerated levels of $\Delta 18$ to retain BRCA1's tumour suppressor activity, we correlated the HDR capacity and exon 18 skipping levels from variants that can only be deleterious through mRNA alterations, including synonymous and intronic variants. Additionally, we included coding variants with a complete altered splicing profile since their functional impact is also completely imposed by their effect on splicing. Variants included were c.5100A>G (synonymous), c.5075-53C>T, c.5152+7A>G, c.5152+1G>T, c.5152+1C>T, c.5152+2dup, c.5152+3A>C, c.5152+4A>T, c.5152+5G>A and c.5077_5080delinsTTCATTCTGG (frameshift with complete splicing effect, FL=0%). Linear regression analyses in mESCs (hemizygous model) allowed us to estimate that a per-allele expression level of $\Delta 18 < 34\%$ (**Figure 45 A**) and FL $\geq 59\%$ (**Figure 45 B**) would lead to an HDR capacity in the benign range (HDR >64%). On the contrary, an expression profile with $\Delta 18 \geq 63\%$ and FL <27% would be deleterious and results in HDR capacity in the pathogenic range (HDR <34%) (**Figure 45 A, B** respectively).

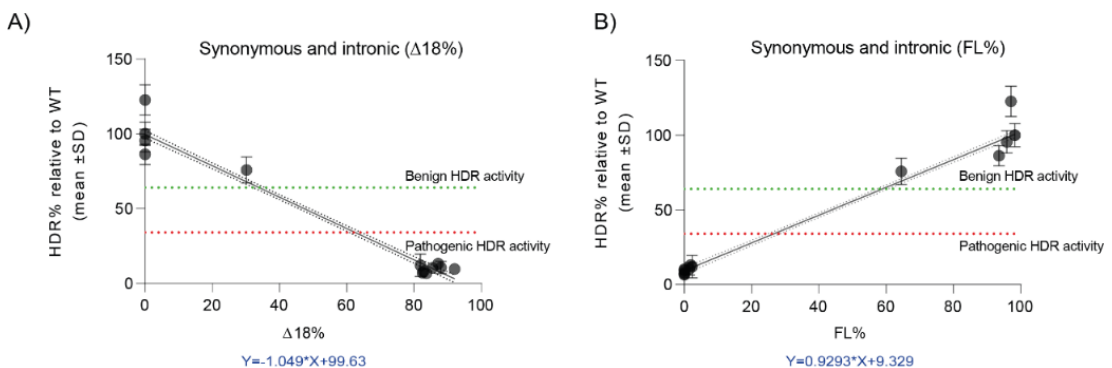


Figure 45. Linear regression analysis between *BRCA1* exon 18 skipping levels (A) or FL (B) (X-axis) and HDR activity (Y-axis).

Only results from of synonymous and intronic variants, plus variants exclusively expressing transcripts lacking exon 18, WT BRCA1 and the DelEx18 variant were considered. $\Delta 18$ and FL levels obtained from RT-PCR experiments were plotted against HDR capacity relative to WT cells. The y-green dotted line defines the HDR range associated with benign variants and red defines the HDR threshold for pathogenic variants. Linear regression equations allow to estimate the HDR activity for a specific exon 18 skipping or FL level. WT: wild-type, HDR: homology direct repair capacity, $\Delta 18$: exon 18 skipping isoform, FL: full-length.

We next aimed to corroborate these thresholds of splicing levels using functional data from a saturation genome editing assay [271], performed using a HAP1 cell line. This assay showed a positive, though non-significant, correlation (Spearman's $r = 0.5$ p -value = 0.08) with mESC HDR results (**Figure 46**), suggesting that both approaches can capture the functional impact of (spliceogenic) variants in this region of *BRCA1*. However, one discordant variant was found, the c.5152+7A>G which showed an HDR capacity in the benign range (HDR~76%, **Figure 43**) while it was reported as loss of function in the saturation assay (functional score= -1.85) [271].

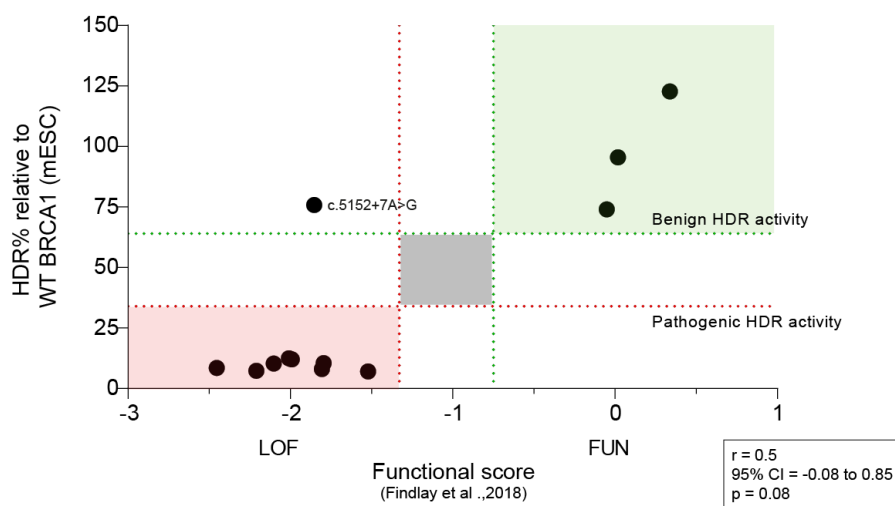


Figure 46. Correlation analysis between functional scores from saturation genome editing in HAP1 cells and HDR outcomes in mESCs.

Dashed lines indicate functional thresholds from both assays, and filled boxes include concordant results: the green box contains neutral variants in both assays (functional score > -0.74 and HDR >64), the grey includes intermediate impact variants, and the red one the deleterious variants in both assays (Functional score <-1.328 and HDR<34). Variants outside these areas are non-concordant. Saturation genome editing functional results were from Findlay et al. (2018)[271]. Spearman correlation analysis was performed (n=12). HDR: homology direct repair capacity, WT: wild type, mESC: mouse embryonic stem cell, LOF: loss of function, FUN: Functional.

Although HAP1 splicing profile was not directly interrogated in the saturation genome editing assay, the authors provided an “RNA score” that measured RNA

depletion for exonic synonymous and missense variants which may be associated with splicing alterations. Thus, we correlated FL levels observed in minigenes and patients with this RNA score to assess the consistency with our RNA findings, and a significant weak positive correlation with minigenes (Spearman $r=0.32$, $p\text{-value}=0.01$), but not with patients (Spearman $r=0.20$, $p\text{-value}=0.50$) was observed (**Figure 47 A, B**). Under the assumption that minigene splicing profiles are concordant with the RNA profiles of the HAP1 cells, we performed a linear regression analysis of minigene exon 18 skipping and FL with functional scores obtained in Findlay et al. (2018)[271]. This regression analysis estimated that per-allele $\Delta 18$ levels lower than 39% and FL% greater or equal to 61% would result in a neutral functional score (>-0.74) (**Figure 47 C, D**). Conversely, an $\Delta 18$ % higher or equal to 70% and an FL% under 30% would result in a functional score lower than -1.328, the threshold for being considered as loss-of function (**Figure 47 C, D**).

The c.5152+7A>G variant, which did not present a concordant HDR capacity in mESC with Findlay's functional score (**Figure 46**) led to 38.4% of exon skipping and 61.2% of FL in the minigene assay and presented a loss of function score of -1.85 (**Figure 47 C, D**). According to our linear regression analysis we would expect a neutral functional score of -0.7. Likewise, for its loss of function score of -1.85, we expected 98% of exon skipping (i.e. 1.94% of FL). All these results suggested that this variant is probably leading to higher levels of exon 18 skipping in HAP1 cells.

Overall, regression analyses supported that per-allele ~40% of $\Delta 18$ (range of up to 34-39%) together with an FL level ~60% (range of up to 59-61%) would be tolerable and maintain sufficient function.

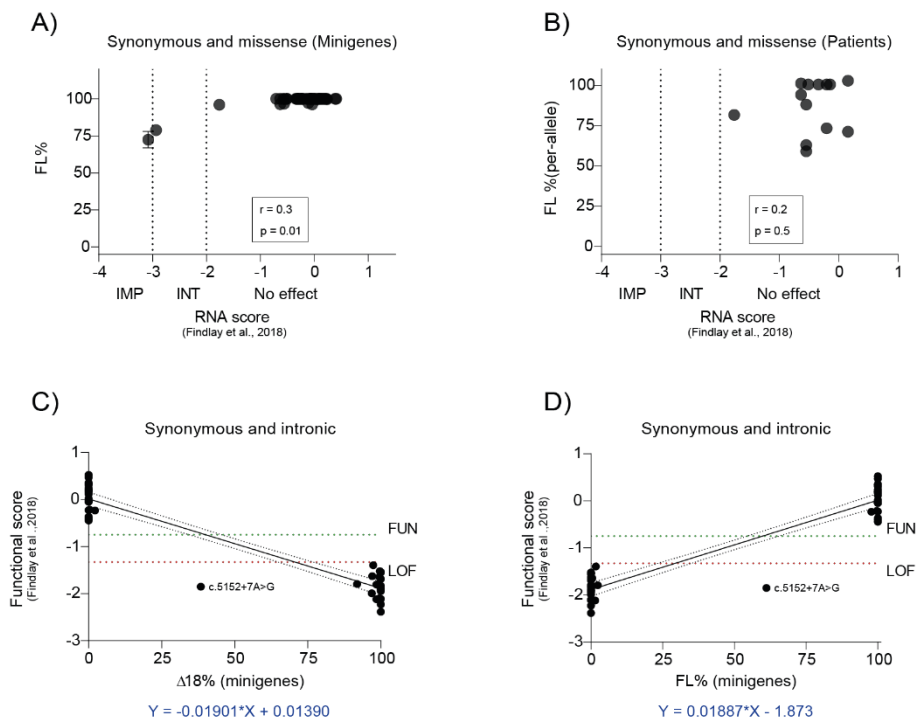


Figure 47. Splicing profile correlation with saturation genome editing functional assay.

A) Correlation between minigene FL levels and RNA scores from exonic missense and synonymous variants obtained in Findlay et al., (2018). Dashed lines indicate the RNA alteration thresholds: <-3 impaired (IMP), >-3 but <-2 intermediate (INT), or >-2 no effect. Spearman correlation analysis was performed ($n=61$). **B)** As in “A” except the Spearman correlation analysis was performed using patients’ per-allele FL levels ($n=13$). **C)** Linear regression analysis between functional scores and $\Delta 18$ levels in minigenes from synonymous and intronic variants ($n=41$; c.5075-8T>G variant excluded as its splicing defect is due to an acceptor shift and not $\Delta 18$). Dashed lines indicate functional score thresholds mentioned in **Figure 46**. **D)** As in “C”, except for FL levels observed in minigenes ($n=41$). Regression curves were fitted for both comparisons to estimate functional outcomes corresponding to each transcript level. IMP: impaired, INT: intermediate, LOF: loss of function, FUN: Functional.

Observed Splicing Impact Specially Benefits the Classification of Intronic/Synonymous Variants in the Absence of Additional Evidence

We aimed to evaluate the efficacy of mRNA data alone in classifying genetic variants, considering variant position, coding effect, *in silico* data, and experimental mRNA results from our study.

Of note, c.5152+1G>T and c.5143A>C, previously classified by ENIGMA as pathogenic (old criteria), were used as reference variants for PS1 codes.

Thus, considering coding effect, location, and *in silico* predictions (**Supplementary Table 17**), 18.3% (30/164) of variants were classified as pathogenic, 7.3% (12/164) as likely pathogenic, 54.3% (89/164) as uncertain and 20.1% (33/164) as likely benign (**Figure 48**). By adding RNA experimental results from patient and minigene assays, the percentage of pathogenic slightly decreased to 17.7% (29/164), the fraction of likely pathogenic increased up to 11.6% (19/164), the fraction of VUS decreased to 39.6% (65/164) and the percentage of likely benign variants raised to 31.1% (51/164) (**Figure 48**).

Regarding nonsense and frameshift variants, only one (c.5080G>T nonsense variant) out of 30 showed a complete aberrant RNA splicing profile in the carrier, thus replacing its PVS1 and PM5_PTC_strong codes with PVS1(RNA), which altered its classification from pathogenic to likely pathogenic, but not affecting patient's clinical management (**Figure 48 and Supplementary Table 17**). It is worth noting that despite the c.5075_5078del and c.5077_5079delinsTTCATTCTGCA frameshift variants leading to complete exon 18 skipping in minigenes, we decided to retain the PVS1 and PM5_PTC_strong codes for both variants. This decision was based on the inability to confirm that these variants solely generate $\Delta 18$ in carriers. Therefore, all possible transcripts from these alleles were considered ($\Delta 18$ and r.5075_5078del or r.5077_5079delins transcripts) maintaining the pathogenic classification for both variants (**Supplementary Table 17**).

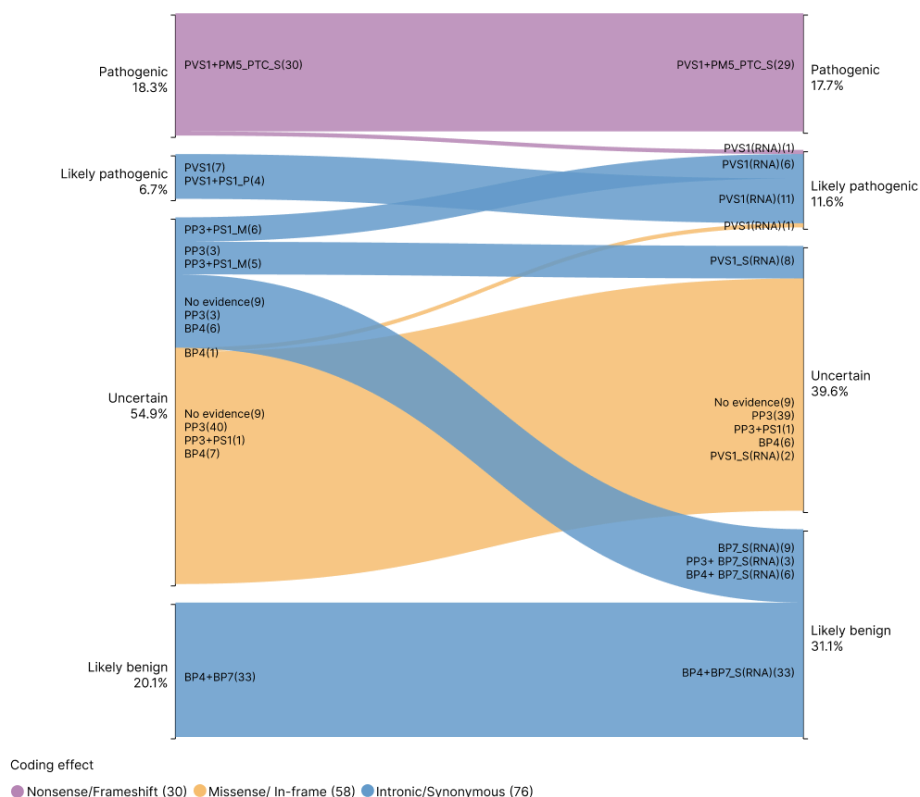
In silico prediction, coding effect and position*In silico* prediction, coding effect, position and splicing results

Figure 48. Sankey plot showing *in silico* and experimental splicing evidence usage for variant reclassification depending on coding effect for 164 variants.

The variants c.5152+1G>T and c.5143A>C, previously classified as pathogenic by ENIGMA expert panel (old criteria), were used as references for applying PS1 code after being reclassified separately as pathogenic.

S: strong, M: moderate, P: supporting

In the group of intronic/synonymous variants (n=76), five intronic variants outside canonical dinucleotides upgraded their classification from uncertain to likely pathogenic after considering their splicing result in patients and minigenes, being their PP3 and PS1_moderate codes replaced with PVS1(RNA) (**Figure 48**). The splicing results confirmed that all 12 intronic variants at canonical acceptor/donor dinucleotide positions (-1,-2/+1,+2) met PVS1(RNA) criteria and therefore retained their likely pathogenic classification (**Figure 48**).

For eight intronic variants, the initially assigned PP3 or PP3 + PS1_moderate criteria were replaced with the PVS1_strong(RNA) code, but it did not change the classification from uncertain (**Figure 48**). The use of PVS1_strong(RNA) was because seven out of the eight variants were tested only using minigenes. Additionally, one variant (c.5152+7A>G) showed conflicting results between patient data and minigene tests. However, the significant splicing alteration observed in one carrier (-10.3% per-allele FL without allele-specific assay) justified the consideration of PVS1_strong(RNA) for this variant (**Supplementary Table 17**).

Without considering RNA experimental findings, all 58 missense and in-frame variants were classified as uncertain (**Figure 48**). The mRNA splicing study indicated that only c.5077_5080delinsTTCATTCTGC variant showed a complete altered splicing. This led to its classification as likely pathogenic after applying PVS1(RNA) specification. In contrast, although the c.5078_5080del in-frame variant generated significant exon 18 skipping both in two carriers and in the minigene assay, the extent of this skipping was not consistent between the evaluated cell models. Therefore, the PVS1(RNA) was not considered for this variant and remained as uncertain (**Supplementary Table 17**).

In summary, applying experimental RNA data-related evidence to *in silico* and coding specifications resulted in a change in the classification status of ~16% of variants (26/164): one pathogenic nonsense changed to likely pathogenic, one in-frame and six intronic uncertain changed to likely pathogenic and five synonymous and 13 intronic switch to likely benign, suggesting that intronic and synonymous variants benefit the most from the application of experimental splicing impact evidence in the absence of additional evidence (**Figure 48**).

Experimental RNA Splicing Evidence Reduces Variant Uncertainty and Is Essential for a Correct Interpretation of Variants Under ENIGMA-Adapted ACMG/AMP Guidelines

According to the ClinVar review status asserted on 15th December 2023, from the 166 variants included in this project, 37.3% (62/166) of variants had been reported as (likely) pathogenic, 12.7% (21/166) as uncertain and 17.5% (29/166) as (likely) benign. Additionally, in 4.8% (8/166) of variants the criteria were not provided, 13.2% (22/166) of variants showed conflicting interpretations of pathogenicity, while 14.5% (24/166) of variants were not present in ClinVar (**Supplementary Table 9**).

Subsequently, we assessed how splicing experimental results could impact variant interpretation after all *in silico* (**Supplementary Table 17**), functional (**Supplementary Table 18**), population allele frequency (**Supplementary Table 19**), case-control (**Supplementary Table 20**), and multifactorial model (**Supplementary Table 21**) evidence consideration.

After a comprehensive classification without considering RNA splicing results, 28.6% (47/164) of variants were classified as pathogenic, 19.5% (32/164) as likely pathogenic, 18.9% (31/164) as uncertain, 29.3% (48/164) as likely benign and 3.7% (6/164) as benign (**Figure 49** and **Supplementary Table 22**). By applying mRNA evidence derived from our study, the number of variants classified as pathogenic or benign increased from 28.6% to 31.7% (52/164) and from 3.7% to 24.4% (40/164), respectively. Conversely, the number of variants classified as likely pathogenic, uncertain, and likely benign decreased from 19.5% to 17.7% (29/164), from 18.9% to 8.5% (14/164), and from 29.3% to 17.7% (29/164), respectively (**Figure 49**).

Specifically, our RNA splicing findings upgraded the classification of five likely pathogenic intronic variants located outside of donor/acceptor positions to pathogenic, as well as the likely benign classification of 10 synonymous, 11 missense, and 13 intronic variants to benign (**Supplementary Table 22**).

Apart from refining variant interpretations, uncertainty for ~10% (17/164) of variants was solved by considering experimental mRNA results. Specifically, 15 synonymous/intronic out of the 30 uncertain variants became likely benign and one in-frame and one intronic reached likely pathogenic class (**Supplementary Table 22**). Accordingly, a total of 14 variants remained uncertain after the mRNA result: 11 missense/in-frame variants and three intronic. Regarding the c.5086G>T, c.5096G>C, c.5114T>G, c.5122G>A, c.5123C>G, c.5128G>A, c.5138T>C, c.5139_5141del, and c.5150T>A missense/in-frame variants, all had been reported as uncertain, conflicting or not assertion provided in ClinVar (**Supplementary Table 22**). None of these variants showed a splicing alteration but BP7_strong(RNA) could not be applied because none of them met BS3 (functional assays showing no protein defect). Specifically, c.5086G>T, 5096G>C, c.5128G>A and c.5150T>A showed discordant results from multiple functional assays and the c.5114T>G showed intermediate effect at protein levels thus neither meeting PS3 nor BS3 codes (**Supplementary Table 18**). The variants c.5090G>A, c.5122G>A, c.5123C>G, and c.5138T>C demonstrated a deleterious protein impact, meeting the PS3 and PP3 criteria. The c.5090G>A variant additionally met the PM2_supporting and BP5_supporting criteria. Similarly, although the c.5077_5080delinsTTGATTCTGC in-frame variant and c.5152+3_5152+4insG and c.5152+4_5152+5insT intronic variants demonstrated complete exon 18 skipping in minigenes, which corresponded to the assignment of a PVS1_strong(RNA) specification, this was not sufficient to upgrade their uncertain classification as no other evidences were available for

these variants (**Supplementary Table 22**). The remaining uncertain variant c.5152+7A>G, despite having both functional and mRNA results, showed non-concordant results and no additional information was available at this point.

Overall, the experimental splicing results affected the interpretation of 34% (56/164) of the variants after considering all *in silico*, functional, population allele frequency, case-control, and multifactorial clinical evidence.

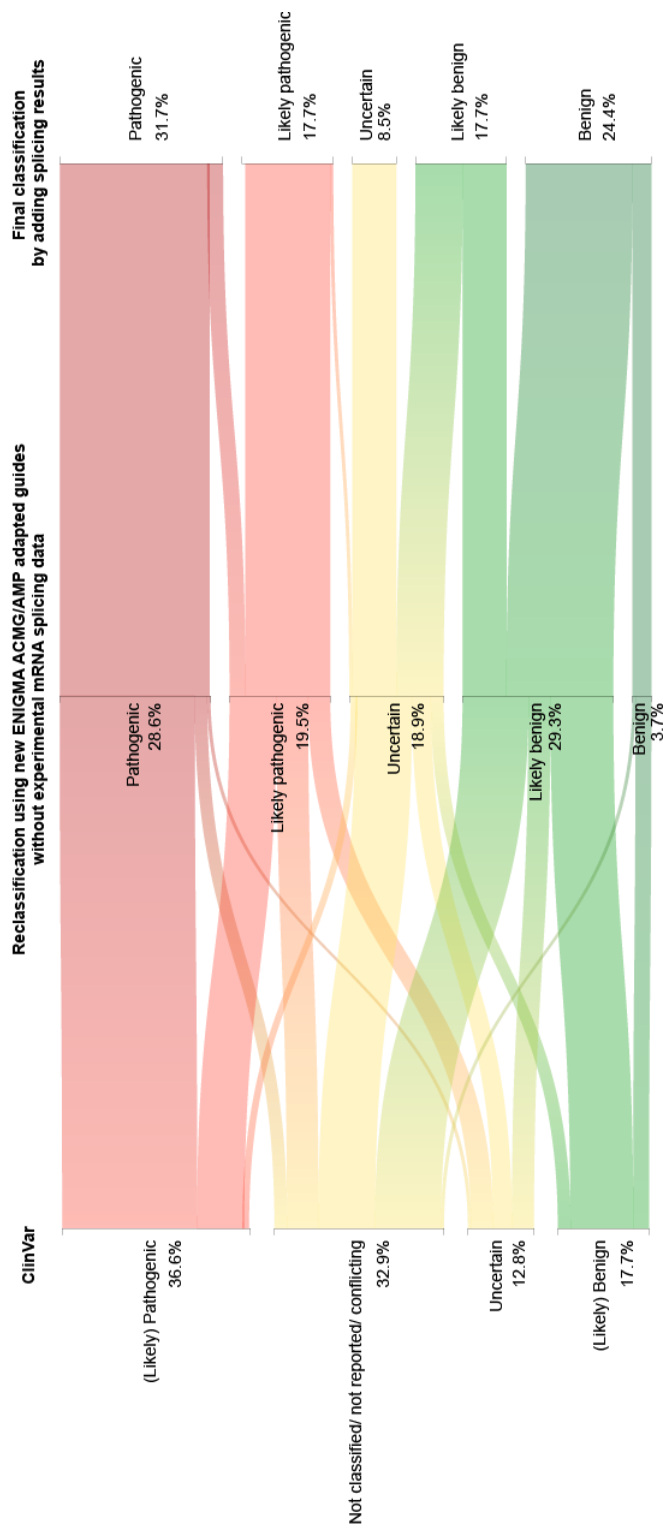


Figure 49. Overview of the classification for 164 out of 166 variants before and after this study by considering *in silico*, variant coding effect, position functional, allele frequency, multifactorial clinical data, and mRNA experimental evidence generated in this project.

The Sankey plot shows the transition in classification categories for variants, from initial ClinVar classification (December 2023) to final classification obtained in this study. This process first involved compiling all available evidence, including *in silico* predictions, population allele frequency data, case-control results, multifactorial clinical likelihood ratios, and functional assay data, followed by the inclusion of mRNA splicing results. The two remaining variants c.5152+1G>T and the c.5143A>C were reclassified separately considering all evidence available, since both were used as reference variants for PS1 codes after being reclassified as pathogenic.

Interestingly, upon implementing the latest ENIGMA recommendations, a small subset of the 45 variants previously classified by this expert panel underwent changes in their classification. Specifically, eight variants previously deemed likely benign were upgraded to benign, and one variant was reclassified from pathogenic to likely pathogenic. Interestingly, the c.5090G>A is the only variant out of the 166 of this study that has been classified by the new VCEP *BRCA1/2* ENIGMA guides in the pilot dataset from Parsons et al. (2024)[122] (**Supplementary Table 22, and Figure 50 A**). For this variant, the likely pathogenic classification in the pilot dataset was based on the variant meeting PS3 due to functional impact, PP3 for *in silico* predictions, and PM2_supporting [122]. Conversely, here we classified this variant as uncertain considering the same evidence together with BP5 code based on the multifactorial LR analysis from Li et al., (2020) [276] (0.475) (**Supplementary Table 21 and 22**), which have been omitted in the pilot dataset due to the LR value nearing the uninformative threshold for BP5 (0.48).

Importantly, we successfully assigned a classification to all 51 variants reported in ClinVar with conflicting interpretations, uncertain significance, or no provided classification, as well as to 24 variants not reported in ClinVar. This resulted in 17 variants being classified as benign, 23 as likely benign, 15 as likely pathogenic, eight as pathogenic, and 12 as uncertain (**Figure 50 B**).

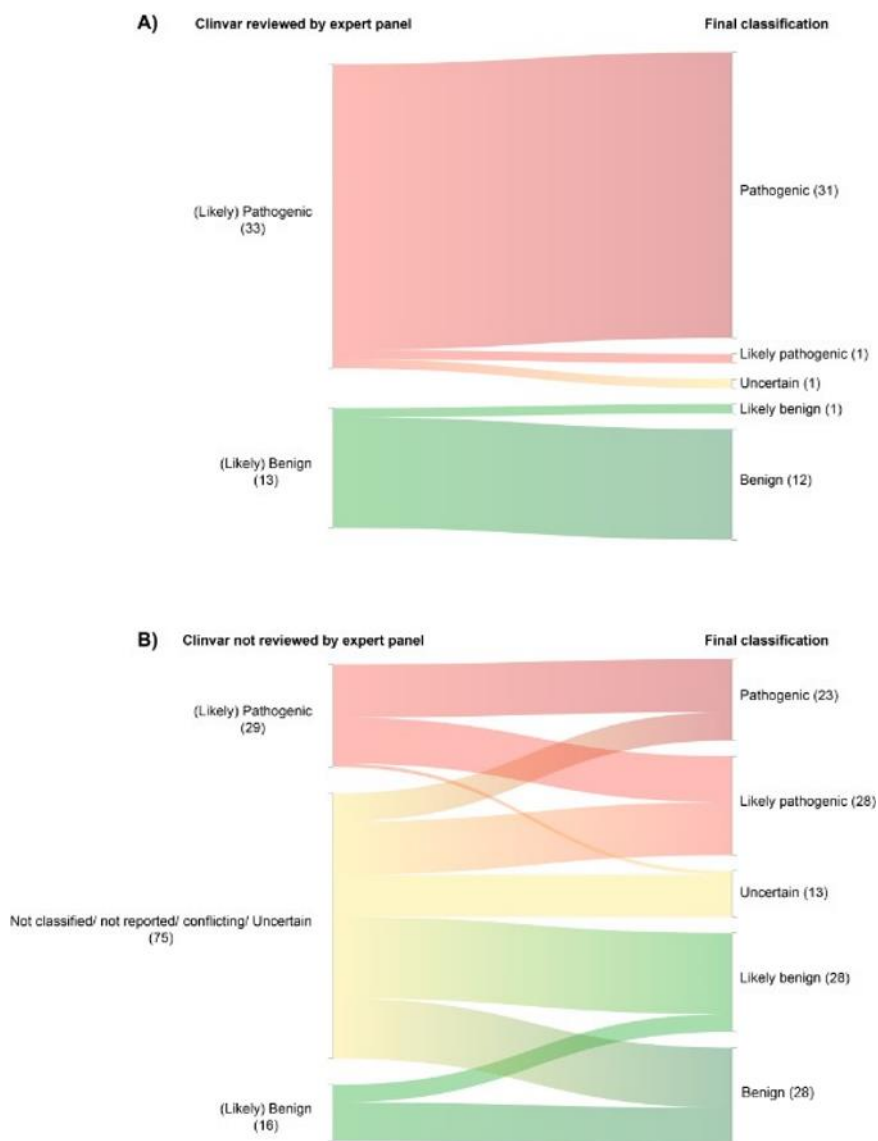


Figure 50. Overview of 166 variant classifications before and after this study by considering *in silico*, variant coding effect, position functional, allele frequency, multifactorial clinical data, and mRNA splicing experimental evidence generated in this project.

Sankey diagrams show the transition from the initial ClinVar classification to the final classification obtained in this study. **A)** Variants classified by ENIGMA outside this project: 45 were classified by old criteria, and the one changing from pathogenic to uncertain (c.5090G>A) had been classified in the pilot dataset from Parsons et al. (2024) [122] under the new VCEP rules. **B)** Variants not classified by an expert panel in ClinVar before this study. *VCEP: variant curation expert panel.*

The Splicing and Functional Ambiguity behind the c.5152+7A>G Variant

With all the evidence generated from this study, we observed that the c.5152+7A>G variant was the only one with a non-concordant result between different assay models, including mRNA splicing in patients and minigenes (**Figure 41**), as well as functional assays (**Figure 46**).

According to our criteria, PVS1_strong was applied because we observed a near complete splicing defect in one carrier (per-allele FL-10.3%), although minigene assay showed non-concordant results (FL 61.2%) and an allele-specific FL expression assessment was not possible (**Supplementary Table 17**). Since complete effect was inferred, but not confirmed in one carrier, we looked for more mRNA results in additional carriers. Thanks to the collaborative effort from Ambry Genetics and Invitae, we obtained splicing results from two additional patients (**Table 15**).

Table 15. Summary of splicing and functional results for the c.5152+7A>G variant.

Sample type	Methodology	Splicing profile	Splicing reference controls ^a	Functional outcome
Carrier 1-Whole blood (VHIO)	Splicing fraction (per-allele) from RT-PCR	FL: 43.5% (-10.3%) Δ18: 50.3% (99%)	c.5152+4G>C (Δ18: 63.5% (126%)) 46 non-carrier controls (Δ18 splicing fraction up to 3.8%)	-
Carrier 2- Whole blood (Ambry Genetics)	PSI (per-allele) from RT-PCRseq	FL: 70.7% (40.5%) Δ18: 26.3% (52.6%)	c.5152+1G>T (Δ18: 40.1% (81.2%)) Non-carrier control (Δ18:0%)	-
Carrier 3- Whole blood (Invitae)	PSI from targeted RNAseq	FL: 72% Δ18: 12%	Non-carrier controls (Δ18:0%)	-
Minigenes (Inserm U1245, Univ Rouen Normandie)	Monoallelic splicing fraction from RT-PCR	FL: 61.2% Δ18: 38.4%	c.5152+1G>T (Δ18:96.9%) WT (Δ18:0%)	-
mESC (Department of Human Genetics, Leiden University Medical Center)	Monoallelic splicing fraction from RT-PCR	FL: 64.5% Δ18: 30.2%	c.5152+1G>T (Δ18:81.9%) WT (Δ18:0%)	Benign HDR capacity
HAP1 cells (Findlay et al., 2018)	Saturation genome editing	NA	NA	Loss of function

a. Same sample type controls analysed in tandem with c.5152+7A>G in each assay.

mESC: mouse embryonic stem cell, PSI: percent spliced-in, FL: full-length, NA: not available result, HDR: homology direct repair capacity.

Specifically, whole blood splicing profile by RT-PCR RNA sequencing of one additional carrier revealed a $\Delta 18$ percent spliced-in (PSI) of 26.3% (5808/22093 reads) in contrast to the 40.1% (7650/19083 reads) and 0% for a c.5152+1G>T carrier and a non-carrier control, respectively (**Figure 51**). Whole blood targeted-RNA sequencing from a third c.5152+7A>G carrier revealed a $\Delta 18$ PSI of 12% and FL PSI of 72% versus the 98% FL PSI in controls (data not shown). Overall, these results highlighted that exon 18 skipping levels detected in two additional carriers were lower in comparison with RT-PCR semiquantitative analysis (**Table 15**), hence PVS1 code assignment was unclear and was not considered for variant interpretation.

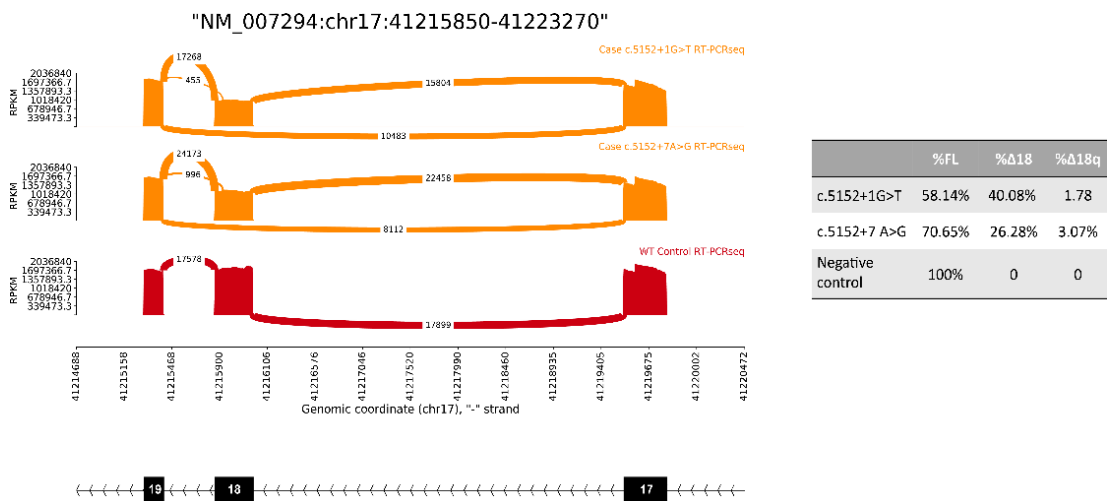


Figure 51. Sashimi plots from RT-PCRseq experiments showing *BRCA1* exon 18 splicing profiles. Experiments were performed using whole blood RNA and primers hybridizing in exons 16 and 24 of *BRCA1*. The relative expression of splicing events was measured by percent splicing index PSI (%), calculated as the number of reads supporting a specific splicing event divided by the number of all reads in the region covering the splicing event. Specifically, reads supporting full-length isoform were calculated by dividing by two the total of reads covering exon 17-18 junction plus exon 19-18 junction. Exon 18 skipping reads included the reads covering exons 17-19 junction, and cryptic donor usage included the reads covering the cryptic donor junction with the natural acceptor site of exon 19. Exon 18 skipping ($\Delta 18$) and cryptic donor usage ($\Delta 18q$) were detected in both carriers (c.5152+1G>T and c.5152+7 A>G), but not in negative control.

Additionally, this variant showed around 30% of exon skipping in mESC and an HDR capacity fell in the benign range (**Table 15**). However, in the saturation genome editing assay [271] this variant showed a loss-of-function score, but no splicing data was available (**Table 15**). Due to these inconsistencies, PS3/BS3 could not be considered.

Prior to this study, multifactorial analysis conducted by Parsons et al. (2019)[256] revealed an uninformative LR of 0.9. Consequently, we aimed to gather new clinical data related to tumour pathology, co-occurring pathogenic variants, and family cancer history. Thus, we obtained information about 12 carriers from ten families. Of note, Myriad Genetics provided relevant information from one carrier, which showed a co-occurrent well-known pathogenic variant, the c.5266dup, p.(Gln1756fs) in *trans*, but did not show a FA phenotype (Myriad Genetics internal data). When this co-occurrence was used in the multifactorial clinical interpretation (co-occurrence LR 0.0013287), the combined multifactorial LR was 0.005 which would allow to apply BP5_strong evidence against pathogenicity (**Supplementary Table 23**). Since this variant was absent in non-cancer controls from gnomAD (**Supplementary Table 19**), it accumulates 1 point supporting pathogenicity (PM2_supporting) and 4 against it (BP5_strong), resulting in likely benign interpretation (-3 points). However, the lack of recessive FA phenotype could also be considered separately, applying BS2_moderate since the patient's breast cancer occurred at age 40. When co-occurrence LR was not considered in the multifactorial evidence, the combined LR reached 4.4, meeting PP4_moderate. Accordingly, and following the point system, this variant would accumulate 3 points in favour of pathogenicity (PP4_moderate + PM2_supporting) and 2 against it (BS2_moderate), reaching an uncertain interpretation (1 point). Therefore, the only evidence that can be

considered unambiguously is PM2_supporting and the resulting class is uncertain (**Supplementary Table 22**).

Part 3

“*Tumour RAD51 Foci Detection: New Evidence for BRCA1, BRCA2 and PALB2 Germline Variant Classification*”

Manuscript in preparation.

Some of the RAD51 results from pathogenic carriers had been published by the Experimental therapeutics group from VHIO in manuscripts focused on testing RAD51% as PARPi response biomarker (detailed below). From these publications, only the RAD51 quantifications were incorporated into this thesis, alongside new RAD51 data generated for two additional patient cohorts. The combined analysis of these datasets produced novel results that are exclusive to this thesis.

- RAD51% in tumours from germline pathogenic *BRCA1/2* carriers: Cruz, C.; Castroviejo-Bermejo, M.; Gutiérrez-Enríquez, S.; Llop-Guevara, A.; Ibrahim, Y.H.; Gris-Oliver, A.; Bonache, S.; Morancho, B.; Bruna, A.; Rueda, O.M.; et al. RAD51 Foci as a Functional Biomarker of Homologous Recombination Repair and PARP Inhibitor Resistance in Germline BRCA-Mutated Breast Cancer. *Annals of Oncology* **2018**, 29, 1203–1210, doi:10.1093/ANNONC/MDY099. [248]
- RAD51% in tumours from germline pathogenic *PALB2* carriers: Castroviejo-Bermejo, M.; Cruz, C.; Llop-Guevara, A.; Gutiérrez-Enríquez, S.; Ducey, M.; Ibrahim, Y.H.; Gris-Oliver, A.; Pellegrino, B.; Bruna, A.; Guzmán, M.; et al. A RAD 51 Assay Feasible in Routine Tumour Samples Calls PARP Inhibitor Response beyond BRCA Mutation. *EMBO Mol Med* **2018**, 10, 23, doi:10.15252/emmm.201809172. [247]

- RAD51% in tumours from germline pathogenic *BRCA1/2* and *PALB2* carriers: Torres-Esquius, S.; Llop-Guevara, A.; Gutiérrez-Enríquez, S.; Romey, M.; Teulé, À.; Llor, G.; Herrero, A.; Sánchez-Henarejos, P.; Vallmajó, A.; González-Santiago, S.; et al. Prevalence of Homologous Recombination Deficiency Among Patients With Germline RAD51C/D Breast or Ovarian Cancer. *JAMA Netw Open* **2024**, 7, e247811–e247811, doi:10.1001/JAMANETWORKOPEN.2024.7811. [288]

Study Populations

A total of 213 tumours were analysed, and RAD51% was successfully scored in 148 samples (69.5%) (**Supplementary Table 25, Supplementary figure 3**). Four tumours were excluded due to insufficient sample availability for all stains, while 61 were NE (see detailed reasoning in **Supplementary Table 26**). Specifically, 59.0% (36/61) of samples were NE due to unclear staining, 26.2% (16/61) had fewer than 25% of cells with γ H2AX foci (DBSs marker) and 14.8% (9/61) had fewer than 40 geminin-positive (S/G2 marker) tumour cells (**Figure 52**). It is also important to note that 85.3% of the NE samples (52 out of 61) were derived from surgical specimens. On contrast, these sample types only represented the 29.7% of (44/148) of evaluable samples.

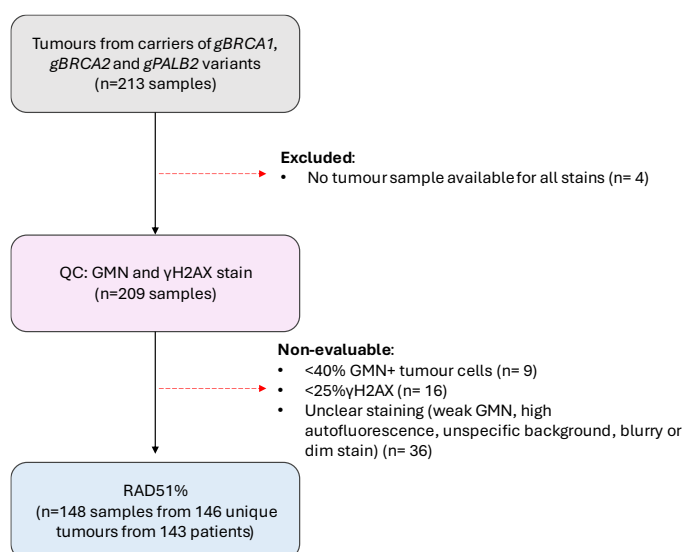


Figure 52. Flow diagram for RAD51 test.

A total of 213 primary untreated tumour samples from germline carriers of *BRCA1/2* and *PALB2* variants were selected. Four samples were excluded due to insufficient sample. Sixty-one samples were non-evaluable due to unclear staining (likely due to defective fixation/processing of samples) or due to no full-fill QC (quality check) of the RAD51% test. These QC are needed to ensure sufficient evaluable cells in S/G2 cell cycle phase (at least 40 geminin-positive cells) with dsDNA breaks ($\geq 25\%$ γ H2AX). A total of 148 samples were successfully scored, corresponding to 146 unique tumours (for two tumours, two different samples from each tumour were analysed, and a mean of the biomarker score obtained in the two respective samples was used for further analyses). *g*: germline, *n*: sample size, GMN: geminin.

The 148 successfully evaluated samples, were derived from 146 unique tumours from 143 patients (**Supplementary Table 25**).

Eighty-five tumours (83 BC, 2 OC) were from germline carriers of 56 unique pathogenic variants (*BRCA1* n=21, *BRCA2* n=22 and *PALB2* n=13), 38 tumours (34 BC, 4 OC) were from carriers of 40 unique variants not yet classified by an expert panel in ClinVar (“unreviewed”) (*BRCA1* n=10, *BRCA2* n=22 and *PALB2* n=8), and 23 tumours (22 BC, 1 OC) were from germline carriers of benign variants. Of note, the benign cohort included patients that carry more than one benign variant, although some centres only reported missense benign variants in *BRCA1/2* (**Supplementary Table 25**).

No differences were observed among DNA damage levels measured by γ H2AX% between pathogenic, benign and unreviewed groups (**Figure 53**).

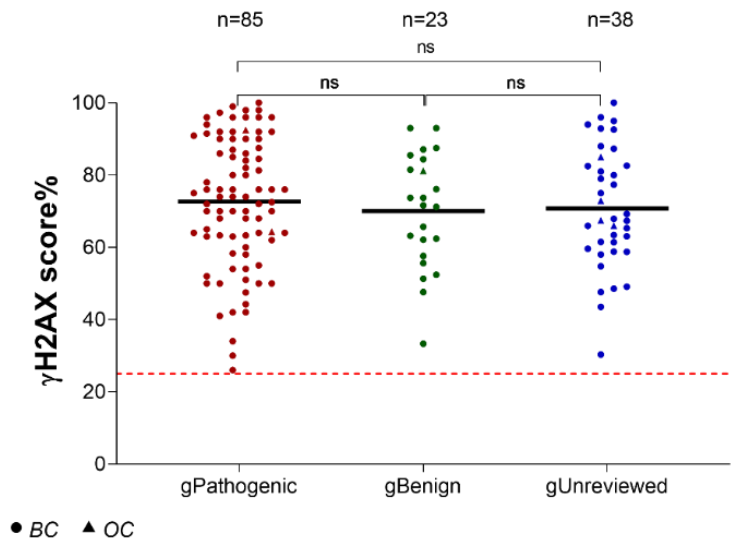


Figure 53. γ H2AX score distribution in breast and ovarian tumour samples from patients carrying germline pathogenic, benign and unreviewed variants.
The percentage of geminin-positive tumour cells that have γ H2AX foci is represented. Each dot represents a single tumour, and the mean of each group is shown. No statistically significant differences between groups were observed by one-way ANOVA (ns $p > 0.05$). Dashed red line indicates the test cutoff ($\geq 25\%$). n: number of data points, ns: non-significant, g: germline, BC: breast cancer, OC: ovarian cancer.

Prevalence of functional and genomic HRD in the Cohort of Unreviewed Variants

Forty-five percent (17/38) of tumours from carriers of unreviewed variants showed HRD by RAD51%, and no significant differences were observed between tumours from carriers of germline variants in different genes (**Figure 54 A**).

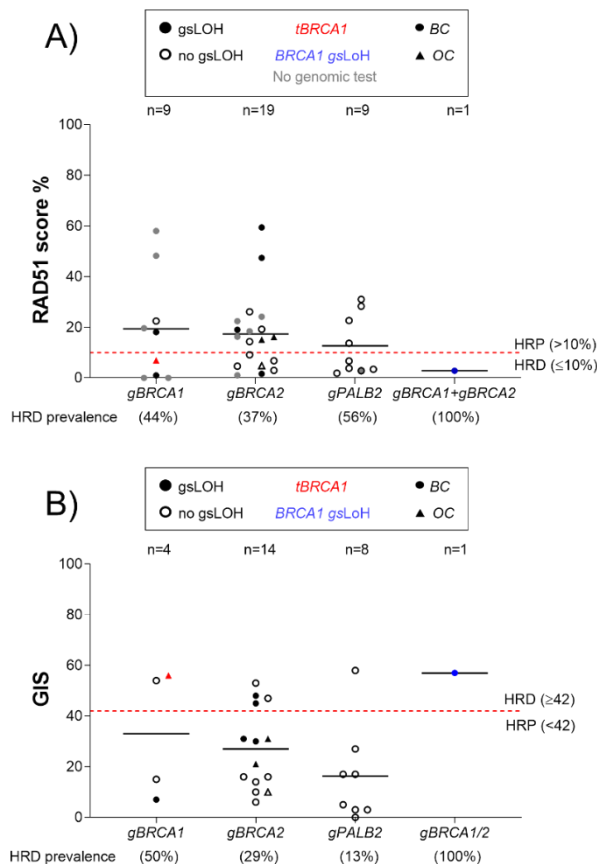


Figure 54. HRR status by RAD51% and GIS in 38 tumours from germline carriers of unreviewed variants.

A) Distribution of the RAD51 score (% of cells with at least 5 nuclear *foci*) according to the germline mutated gene. The prevalence of HRD tumours (RAD51 $\leq 10\%$), the mean of each group and the tumour type is indicated. Each dot represents a single tumour sample. Dashed lines indicate the test cutoff (HRD=RAD51 $\leq 10\%$). No statistically significant differences between groups including more than 1 sample were observed by Kruskal-Wallis non-parametric test ($p > 0.05$). **B)** Genomic Instability Score (GIS) distribution according to the germline mutated gene. The prevalence of HRD (GIS ≥ 42) tumours, the mean of each group, somatic mutations, gene-specific loss of heterozygosity (gsLOH) of germline mutated gene (*BRCA1*, *BRCA2* or *PALB2*) and tumour type is indicated. Each dot represents a single tumour. Dashed lines indicate the test

cutoff (HRD= GIS >42). No statistically significant differences between groups were observed by Kruskal-Wallis non-parametric test ($p > 0.05$). *n*: number of data points, *BC*: breast cancer, *OC*: ovarian cancer, *GIS*: genomic instability score, *gsLOH*: gene-specific loss of heterozygosity, *g*: germline; *t*: tumour (somatic mutations), *HRD*: homologous recombination deficiency; *HRP*: homologous recombination proficiency.

For this cohort (38 tumours), we decided to further characterise the tumours using the Myriad myChoice HRD test. Out of the 38 samples, one sample was not assessed due to insufficient DNA and five were not available for the test (**Table 16**). The Myriad myChoice HRD test was successful in 84% (27/32) of the samples. Although five samples were non-evaluable to obtain the GIS score due to too poor DNA quality, two out of them were evaluable for HRR gene mutation calling and *gsLOH* status (**Table 16**). HRD by GIS (≥ 42) was identified in 30% (8/27) of tumours, and no differences between *BRCA1*, *BRCA2* and *PALB2* carriers were observed (**Figure 54 B**).

From the total of 27 tumours evaluable by both HR tests, a concordance rate of 67% (18/27) (Cohen $k=0.32$, CI95%: -0.01 to 0.65) was observed (**Figure 55 A**). However, this concordance differed between sample type, since it was lower for surgical specimens (6/13, 46%, Cohen $k=0.10$, CI95%: -0.10 to 0.30) in comparison to biopsies (12/14, 86%, Cohen $k=0.71$, CI95%: 0.40-1.00 (**Figure 55 B, C**).

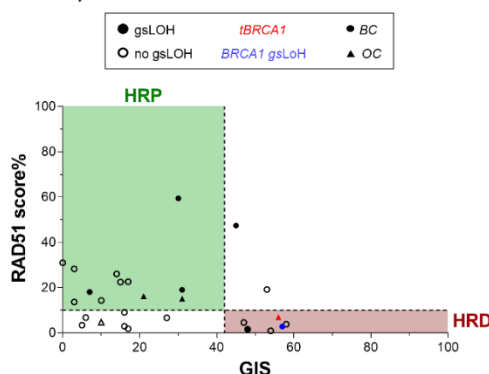
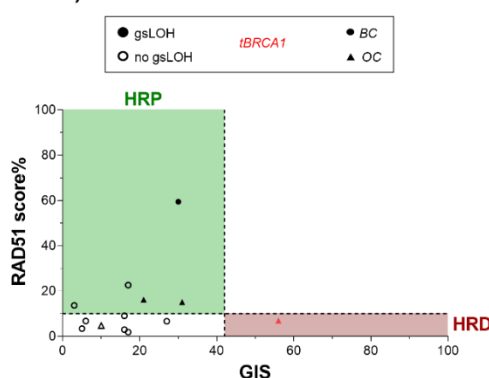
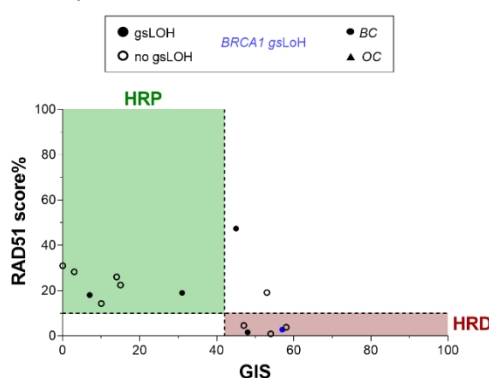
A) All (concordance 67%)**B) Sx (concordance 46%)****C) Bx (concordance 86%)**

Figure 55. Correlation between RAD51% and GIS tumour results in all samples (A), in surgical specimens (B) and in biopsies (C).

Concordance rate indicated for each analysis. Squares are delimited by tests' cutoff (10 for RAD51% and 42 for GIS), and filled ones contain concordant results (green: HRP by both tests, red: HRD by both tests). Each dot represents a single tumour (n=27). Tumour type, gsLOH status of germline mutated gene (*BRCA1*, *BRCA2* or *PALB2*) and somatic mutations are indicated. BC: breast cancer, OC: ovarian cancer, GIS: genomic instability score, gsLOH: gene-specific loss of heterozygosity, g: germline; t: tumour mutations (somatic mutations), HRD: homologous recombination deficiency; HRP: homologous recombination proficiency, Sx: surgical specimens, Bx: biopsies.

As VCEP ACMG/AMP adapted guides for *BRCA1/2* and *PALB2* were released in 2023, we decided to use them to classify the 40 unique germline variants present in the total of 38 tumours from 38 patients. Overall, 26 variants (24 tumours) were classified as (likely) benign, 6 as (likely) pathogenic (6 tumours) and 8 variants (8 tumours) as VUS (**Table 16, Supplementary Table 27, Supplementary Table 28, Supplementary Table 29, Supplementary Table 30, Supplementary Table 31 and Supplementary Table 32**).

Tumours from carriers of germline (likely) pathogenic variants are expected to exhibit HRD. In contrast, tumours from carriers of germline (likely) benign variants are expected to show a much lower prevalence of HRD [232,289]. Consistent with this, within this cohort, five out of the six tumours from variants reclassified as (likely) pathogenic displayed HRD by RAD51%, while 15 out of the 24 tumours from carriers of variants reclassified as (likely) benign showed HRP (**Table 16**). Tumour panel sequencing of HRR-related genes, combined with *BRCA1 foci* detection (*BRCA1%*) clarified 4 out of the 10 cases showing discrepancies between germline variant classification and the HRR status by RAD51%.

Notably, sample T116 showed HRD by both RAD51% and GIS, despite the patient's germline variant (*BRCA1* c.398G>A), being classified as benign (**Table 16**). Tumour sequencing identified one potential pathogenic *BRCA1* variant (c.213-2A>C) in this sample. This, along with the observed gsLOH for *BRCA1* and low *BRCA1%* suggested that the HRD was likely due to the somatic mutation (**Table 16**). Similarly, samples T37, T123 and T127, from carriers of germline variants in *BRCA2* and *PALB2*, showed HRD by RAD51% despite all germline variants being classified as (likely) benign (**Table 16**). However, these cases also exhibited low *BRCA1%*. These results indicate that the origin of HRD was

possibly not related to the germline variants but might be explained by epigenetic silencing of *BRCA1* [247].

HRD by both GIS and RAD51% was observed in the T130 tumour despite the germline *BRCA1* c.2123C>A and c.6560C>T (*BRCA2*) variants achieved a benign classification (**Table 16**). Nevertheless, the gsLOH detected for *BRCA1* with low *BRCA1*% *foci* detection, suggested that the germline *BRCA1* variant could be pathogenic. Likewise, for the T122, T125, T128 and T129 their RAD51% HRD did not agree with the (likely) benign classification of germline *BRCA2* and *PALB2* variants (**Table 16**). All these tumours displayed HRP by GIS, showed high-*BRCA1*%, and no potentially pathogenic variants were identified in other HRR related genes that would act as alternative causes of HRD (**Table 16**). However, we cannot rule out the possibility of somatic alterations beyond targeted sequenced regions, promoter hypermethylation of other HRR genes or a false HRD by RAD51% due to unforeseen technical reasons [290].

The T87 showed HRP by RAD51% despite the germline *BRCA1* c.5467G>A variant of this patient being classified as likely pathogenic. No genomic information was available for this tumour, but this proficient status suggested that the germline variant might not be the oncogenic driver of this tumour.

Of note, the T133 and T121 tumours showed HRP by RAD51% from germline carriers of (likely) benign variants showed high GIS. This discrepancies between HRR tests revealed that GIS might be capturing genomic instability from causes beyond HRR alterations such as high replication stress [291](**Table 16**).

Table 16. Germline and tumour characteristics of the unreviewed cohort cases.

Germline			Tumour									
Patient	Gene ^a	Variant	ClinVar class ^b	Evidence ^c	Classification	ID	Sample ^d	RAD51 %	BRCA1 %	GIS	gsLOH (BRCA1/2 or PALB2) ^e	Other HRR mutated genes ^f
113	BRCA1	c.398G>A p.(Arg133His)	CON	BS1_P+BP1_S+BS3+BP5_M	Benign	T116	OC(Sx)	7	0	56	BRCA1, BRCA2	PAT-BRCA1 gsLOH c.213-2A>C, (no germline), VUS-FANCL non-gsLOH c.1096_1099dup (unknown germline)
142	BRCA1	Del ex10-14 p.(Val199_Tyr1863delinsCysLeu)	NR	PVS1_M+BP5_M	VUS	T145	BC(Bx)	1	79	54	-	VUS-CDK12
143	BRCA1	c.791_794del p.(Ser264Metfs*33)	PAT	PVS1_M(RNA)+PP4_M	VUS	T146	BC(Bx)	58	89	NA	NA	NA
114	BRCA1	c.932C>G p.(Pro311Arg)	CON	PM2_P+BP1_S	Likely benign	T117	BC(Bx)	22	71	15	BRCA2	PAT-ATM gsLOH c.8283_8284del (germline)
112	BRCA1	c.1655G>A p.(Gly552Asp)	CON	PM2_P+BP1_S	Likely benign	T115	BC(Bx)	48	61	F	BRCA1	VUS- ATM non-gsLOH c.4396C>G (germline)
127	BRCA1 BRCA2	c.2123C>A p.(Ser708Tyr) c.6560C>T p.(Pro2187Leu)	CON CON	BS1+BP1_S+BS3+PP4_P BS1+BP1_S	Benign Benign	T130	BC(Bx)	3	5	57	BRCA1	VUS-RAD54L non-gsLOH c.1487G>C (unknown germline)
115	BRCA1	c.3247A>G p.(Met1083Val)	CON	BP1_S	Likely benign	T118	BC(BX)	18	36	7	BRCA1, PALB2	-
87	BRCA1	c.5152+3A>C	CON	PM2_P+PVS1(RNA)+PS3	Pathogenic	T90	BC(BX)	0	0	NA	NA	NA
86	BRCA1	c.5311_5332+1del p.(Pro1771Ilefs*15)	PAT	PVS1+PMT5_S	Pathogenic	T89	BC(BX)	0	1	NA	NA	NA
84	BRCA1	c.5467G>A p.(Ala1823Thr)	L PAT	PM2_P+PP3+PS3+PP4_M	Likely pathogenic	T87	BC(Bx)	20	38	NA	NA	NA

Germline					Tumour							
Patient	Gene ^a	Variant	ClinVar class ^b	Evidence ^c	Classification	ID	Sample ^d	RAD51 %	BRCA1 %	GIS	gsLOH (BRCA1/2 or PALB2) ^f	Other HRR mutated genes ^g
130	BRCA2	c.800G>A p.(Gly267Glu)	CON	BS1_P+BP1_STR+BP5_VS	Benign	T133	BC(Bx)	19	83	53	BRCA1, PALB2	VUS-BARD1 non-gsLOH c.1690C>G (unknown germline), PAT-RAD51B gsLOH del exons 7-9 (unknown germline)
118	BRCA2	c.1645A>G p.(Lys549Glu)	CON	PM2_P+BP1_S+BP5_P	Likely benign	T121	BC(Bx)	47	80	45	BRCA1, BRCA2, PALB2	-
85	BRCA2	c.2004del p.(Lys669Asnfs*13)	NR	PVS1+PMT5_S	Pathogenic	T88	BC(Bx)	2	83	48	BRCA2	-
134	BRCA2	c.2799A>G p.(Thr933=)	L BEN	PM2_P+BP1_ST	Likely benign	T137	BC(Sx)	9	2	16	BRCA1	VUS-CDK12 gsLOH c.1942A>G (unknown germline)
121	BRCA2	c.3032C>G p.(Thr1011Arg)	CON	BS1_P+BP1_S+PP4_P	Likely benign	T124	BC(Bx)	16	49	C	C	C
128	BRCA2	c.3083A>C p.(Lys1028Thr)	CON	PM2_P+BP1_S	Likely benign	T131	BC(Bx)	24	76	F	F	F
119	BRCA2	c.3216A>T p.(Leu1072Phe)	NR	PM2_P+BP1_S	Likely benign	T122	BC(Sx)	7	60	6	-	-
132	BRCA2	c.3310A>C p.(Thr1104Pro)	CON	BP1_S+BP5_S	Benign	T135	BC(Bx)	14	63	10	-	-
	BRCA2	c.3503T>A p.(Met1168Lys)	CON	BP1_S	Likely benign							
117	BRCA2	c.3398C>T p.(Pro1133Leu)	CON	PM2_P+BP1_S+BP5_P	Likely benign	T120	OC(Sx)	16	47	21	BRCA1, BRCA2, PALB2	-
129	BRCA2	c.4316C>A p.(Ala1439Asp)	CON	BS1_P+BP1_S	Likely benign	T132	BC(Bx)	18	54	F	F	F
135	BRCA2	c.4928T>C p.(Val1643Ala)	CON	BS1_P+BP1_S+BS3	Benign	T138	BC(Bx)	26	86	14	-	-
122	BRCA2	c.6268C>T p.(His2090Tyr)	CON	BP1_S	Likely benign	T125	BC(Sx)	3	23	16	-	VUS-CHEK2 gsLOH c.295C>G (no germline)

Germline					Tumour							
Patient	Gene ^a	Variant	ClinVar class ^b	Evidence ^c	Classification	ID	Sample ^d	RAD51 %	BRCA1 %	GIS	gsLOH (BRCA1/2 or PALB2) ^f	Other HRR mutated genes ^g
120	BRCA2	c.6382A>G p.(Lys2128Glu)	CON	BP1_S	Likely benign	T123	OC(Sx)	5	8	10	-	-
	BRCA2	c.6700T>C p.(Phe2234Leu)	CON	BP1_S	Likely benign							
133	BRCA2	c.7022G>A p.(Arg2341His)	VUS	BP1_S+BP5_P	Likely benign	T136	BC(Sx)	22	31	F	F	F
116	BRCA2	c.7522G>A p.(Gly2508Ser)	CON	BA1+PP3+BP5_P	Benign	T119	BC(Bx)	19	24	31	BRCA2, PALB2	PAT-ATM gsLOH c.331+2dup (no germline), VUS-ATM gsLOH c.2635A>G (germline)
83	BRCA2	c.7985C>G p.(Thr2662Arg)	VUS	PM2_P+PVS1_S(RNA)	Likely pathogenic	T86	BC(Bx)	5	57	47	-	VUS-RAD51B gsLOH c.728A>G (unknown germline), VUS-RAD54L non-gsLOH c.1672C>T (unknown germline)
131	BRCA2	c.8117A>G p.(Asn2706Ser)	CON	BS1+BP4	Likely benign	T134	BC(Sx)	59	NE	30	BRCA1, BRCA2	VUS-ATM gsLOH large rearrangement (unknown germline), PAT-CHEK2 gsLOH c.433C>T (unknown germline)
88	BRCA2	c.8847del p.(Lys2950Argfs*26)	NR	PVS1+PMT5_S	Pathogenic	T91	BC(Sx)	1	57	NA	NA	NA
136	BRCA2	c.9410C>T p.(Thr3137Ile)	VUS	PM2_P+BP4	VUS	T139	OC(Sx)	15	63	31	BRCA1, BRCA2	PAT-CDK12 c.1804C>T gsLOH (unknown germline), VUS-CHEK1 non-gsLOH c.1148A>C (unknown germline)
138	PALB2	c.101G>A p.(Arg34His)	VUS	BP1	VUS	T141	BC(Bx)	3	12	F	F	F
125	PALB2	c.629C>T p.(Pro210Leu)	CON	BA1+BP1	Benign	T128	BC(Sx)	7	18	27	BRCA1	-
124	PALB2	c.925A>G p.(Ile309Val)	L BEN	BA1+BP1	Benign	T127	BC(Bx)	4	6	58	BRCA1	VUS-ATM non-gsLOH large rearrangement (unknown germline), VUS-RAD51B non-gsLOH c.332G>A (unknown germline)

Patient	Germline		Tumour									
	Gene ^a	Variant	ClinVar class ^b	Evidence ^c	Classification	ID	Sample ^d	RAD51 %	BRCA1 %	GIS	gsLOH (BRCA1/2 or PALB2) ^f	Other HRR mutated genes ^g
123	PALB2	c.925A>G p.(Ile309Val)	L BEN	BA1+BP1	Benign	T126	BC(Bx)	28	89	3	BRCA1	-
126	PALB2	c.1861C>A p.(Pro621Thr)	CON	BA1+BP1	Likely benign	T129	BC(Sx)	2	41	17	BRCA1, BRCA2	-
137	PALB2	c.2093T>C p.(Leu698Pro)	L BEN	PM2_P+BP1	VUS	T140	BC(Bx)	31	79	0	BRCA2	-
140	PALB2	c.2401G>A p.(Asp801Asn)	VUS	BP1	VUS	T143	BC(Sx)	23	25	17	-	VUS- ATM non-gsLOH large rearrangement (unknown germline)
141	PALB2	c.2425A>G p.(Thr809Ala)	VUS	PM2_P+BP1	VUS	T144	BC(Sx)	14	65	3	-	-
139	PALB2	c.2612A>G p.(Asp871Gly)	VUS	BP1 + PP3	VUS	T142	BC(Sx)	3	47	5	BRCA1, BRCA2	VUS-CDK12 gsLOH large rearrangement (unknown germline)

^aReference transcripts: BRCA1(NM_007294.3), BRCA2 (NM_000059.3), PALB2 (NM_024675.3). ^bClinvar reported classification. CON (conflicting); multiple submitters conflicting interpretations; VUS(uncertain), L PAT (likely pathogenic), PAT (pathogenic), L BEN (likely benign), BEN (benign); single or multiple submitters no conflicts; NR: not reported. ^cEvidence for variant classification. Detailed evidence application in supplementary tables 27-32. Contradictory evidence was combined based on point system [119]. P: supporting, M: moderate, S: strong, VS: very strong. ^dTumour type: BC (breast cancer) or OC (ovarian cancer) and sample type indicated: Bx (biopsy) or Sx (surgery). ^eGIS: Genomic Instability Score. F: failed, C: cancelled due to low DNA. ^fGsLOH: gene-specific loss of heterozygosity. Germline mutated gene highlighted in bold. ^gMutations identified in tumour samples, and classification reported by Philipps-Universität Marburg laboratories. Germline variant status indicated. HRR: homologous recombination repair. NE: non-evaluable, NA: Not available.

At this point, we had only analysed whether tumour HRR status informed about the pathogenicity of the germline variants. However, it had not yet been established how to use this evidence for variant classification. To determine the feasibility of introducing this type of evidence in variant classification, it was essential to first assess the performance of the RAD51% test in distinguishing between pathogenic and benign germline variants, and then to estimate the strength level at which this evidence could be applied.

To proceed with these analyses, we decided to enlarge the benign and pathogenic cohorts of previously classified variants with the 30 tumours from the “unreviewed” cohort that we classified after applying VCEP current guidelines (detailed in **Figure 56**).

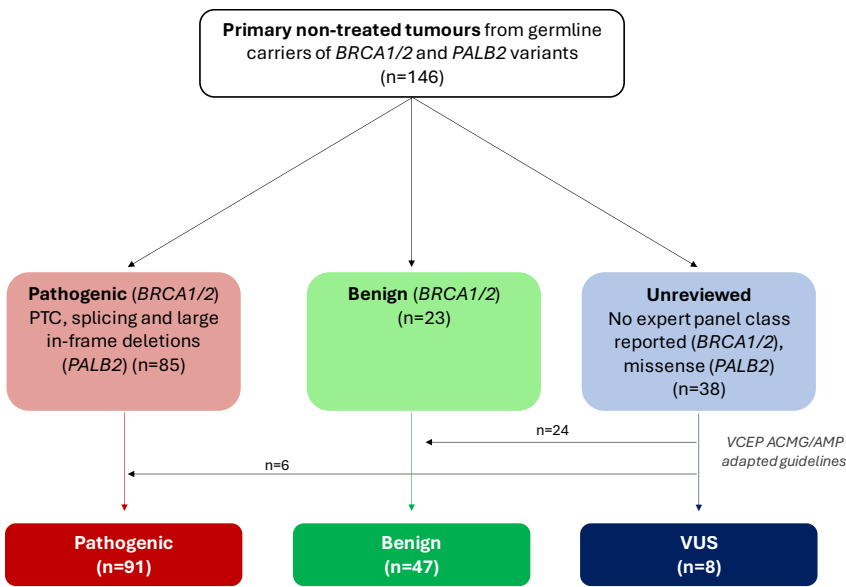


Figure 56. Flow diagram of the cohorts used to validate the RAD51 assay for germline variant classification.

From the 146 evaluable tumours, 85 were from carriers of germline pathogenic variants, 23 from benign of *BRCA1/2* variants and 38 were carriers of variants not classified by an expert panel in *BRCA1/2* or missense variants in *PALB2*. After classifying all 41 variants present in 38 patients from this last group, a total of 26 variants (24 tumours) were classified as benign and 6 variants (6 tumours) were classified as pathogenic. *n*: sample size, *PTC*: premature termination codon, *VCEP*: variant curator expert panel, *ACMG/AMP*: American

College of Medical Genetics and Genomics and the Association for Molecular Pathology, VUS: variant of unknown significance.

Clinical Characteristics of Pathogenic and Benign Cohorts

Patients clinicopathological characteristics are summarized in **Table 17** and detailed in **Supplementary Table 22**. Of the 135 patients included in the final pathogenic or benign cohorts (**Figure 56**), 119 (88%) were women. Additionally, no statistically significant differences were observed between the mean age at diagnosis between the two groups (**Table 17**).

One hundred thirty-two tumours out of the 138 analysed were BC (96%), being the invasive ductal carcinoma (IDC) the most frequent lesion in both groups (pathogenic: 94%, benign: 88%) (**Table 17**). Regarding molecular subtypes, ER+ tumours were the most prevalent, comprising 61% tumours from the pathogenic cohort and 66% of tumours from the benign. The pathogenic cohort was enriched with triple-negative (TN) tumours (37%) compared to the benign group (22%), which in contrast was enriched with HER2+ tumours (13%) (**Table 17**). Additionally, most of BCs were intermediate-high grade (G2-3), representing approximately the 90% of tumours in each cohort (**Table 17**).

OCs represented around 5% of total samples successfully scored for RAD51%, and high-grade serous carcinoma was the most frequent histotype (50%, 3/6), followed by clear cell carcinomas (33%, 2/6) (**Table 17**).

Table 17. Patients' characteristics in pathogenic and benign cohorts.

Characteristic	Pathogenic (91 tumours from 88 patients) n(%)	Benign (47 tumours from 47 patients) n(%)	
Sex			
Male	2 (2.3%)	1 (2.9%) ^a	P = 0.85
Female	85 (97.7%)	34 (97.1%) ^a	P = 0.85
NA	1	12	
Age at diagnosis (mean, range)			
	45.8 (25-77)	48.5 (22-74) ^b	P = 0.25
Germline mutated gene			
BRCA1	38 (43.2%)	-	
BRCA2	32 (36.4%)	-	
PALB2	18 (20.4%)	-	
Breast cancer			
DCIS	0 (0%)	1 (2.3%) ^a	P = 0.1525
IDC	84 (94.4%)	38 (88.4%) ^a	P = 0.2233
ILC	4 (4.5%)	0 (0%) ^a	P = 0.1593
Others ^c	1 (1.12%)	4 (9.30%)	P = 0.0209
Molecular subtypes			
TN	34 (38.6%)	7 (22.0%) ^a	P = 0.0913
ER+	54 (61.4%)	21 (65.5 %) ^a	P = 0.6829
HER2+ (ER-, PR-)	0 (0.0%)	4 (12.5%)	P = 0.0008
NA	1	11	
Grade			
I	2 (2.5%)	4 (11.4%) ^a	P = 0.0500
II	28 (35.4%)	14 (40.0%) ^a	P = 0.6400
III	49 (62.0%)	17 (48.6%) ^a	P = 0.1833
NA	10	8	
Ovarian cancer			
Hystotypes			
High-grade serous	1 (100.0%)	2 (50.0%) ^a	P=0.41
Clear cell		2 (50.0%) ^a	P=0.41
NA	1		

^aNo statistically significant difference between groups: Chi-square. ^bNo statistically significant differences between groups: t-test. ^c"Others" includes mucinous BC, metaplastic BC, and papillary BC.

n: sample size, NA: not available, DCIS: ductal carcinoma in situ, IDC: invasive ductal carcinoma, ILC: invasive lobular carcinoma, TN: triple negative breast cancer, ER+: oestrogen receptor positive, HER2+ (ER-, PR- and HER2+).

Prevalence of Functional HRD in Pathogenic and Benign Cohorts

After including the unreviewed variants that we classified as (likely) pathogenic and (likely) benign into the pathogenic and benign cohorts (**Figure 56**), the HRD prevalence in tumours from germline pathogenic variant carriers (91%, 83/91)

was significantly higher than in the benign (30%, 14/47) ($p < 0.0001$) and VUS (38%, 3/8) ($p = 0.008$) cohorts (**Figure 57** and **Supplementary Table 22**).

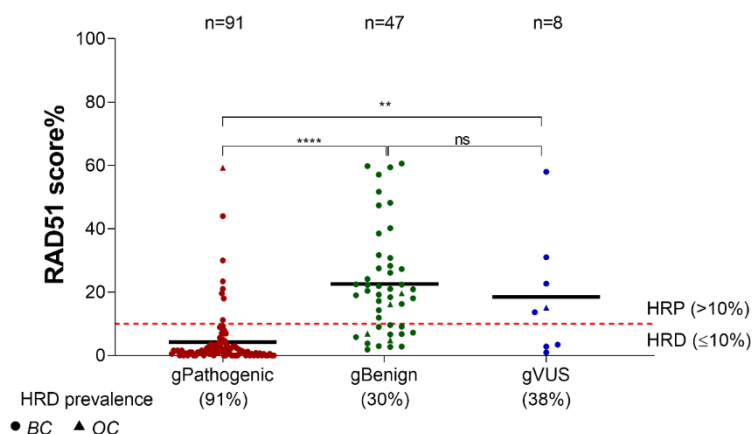


Figure 57. RAD51% score distribution in breast and ovarian tumour samples from patients carrying germline pathogenic, benign and VUS classified variants.

The percentage of geminin-positive tumour cells with at least 5 *foci* of RAD51 is represented. Dashed line indicates test cutoff (RAD51 $\leq 10\%$), and each dot represents a single tumour. Kruskal-Wallis non-parametric test and Dunn's test used for statistical analyses (ns $p > 0.05$, ** $p \leq 0.01$, **** $p \leq 0.0001$). *n*: number of data points, *g*: germline. HRD: homologous recombination deficiency, HRP: homologous recombination proficiency.

Regarding the 14 HRR deficient tumours in the benign cohort, 50% (7/14) of cases showed low BRCA1% (**Figure 58 A** and **Supplementary Table 22**), suggesting possible somatic *BRCA1* alterations (somatic loss of function variant or promoter hypermethylation). Indeed, among these cases was included the above-mentioned T116, which carried a germline variant reclassified as benign but a somatic *BRCA1* alteration (**Table 16**). In contrast, 98% of HRR deficient *BRCA2* and *PALB2* pathogenic tumours (39 of 40 tumours with evaluable BRCA1%) showed high levels of BRCA1%, supporting that the HRD observed was likely due to the germline pathogenic variants and not related to *BRCA1* somatic alterations (**Figure 58 B** and **Supplementary Table 22**).

Focusing on germline *BRCA1* pathogenic tumours, 24 out of 35 (69%) of HRR deficient cases also presented low BRCA1% (**Figure 58 B** and **Supplementary Table 22**). Interestingly, the remaining 11 tumours with low RAD51% indicating HRD, showed high BRCA1% scores. Eight out of these 11 (73%) cases involved tumours from germline carriers of nonsense and frameshift variants in exon 11 (**Figure 58 B** and **Supplementary Table 22**). The BRCA1 foci detected are likely to be in-frame BRCA1 proteins, encoded by alternative isoforms lacking part or all of exon 11 ($\Delta 11q$, $\Delta 11$, $\Delta 9_{-11}$) [176,292]. These isoforms bypass the premature termination codons introduced by nonsense and frameshift variants and maintain all clinically relevant domains of BRCA1. However, the low RAD51% detected in these tumours suggests a limited function of these proteins in HRR.

Notably, 9% of tumours (8/91) in the pathogenic cohort showed high RAD51% scores, compatible with HRP (**Figure 58 B** and **Supplementary Table 22**). Thus, suggesting that germline variants might not be the driver mutations of cancer or the likely existence of hypomorphic proteins that partially kept HR function [248,292,293].

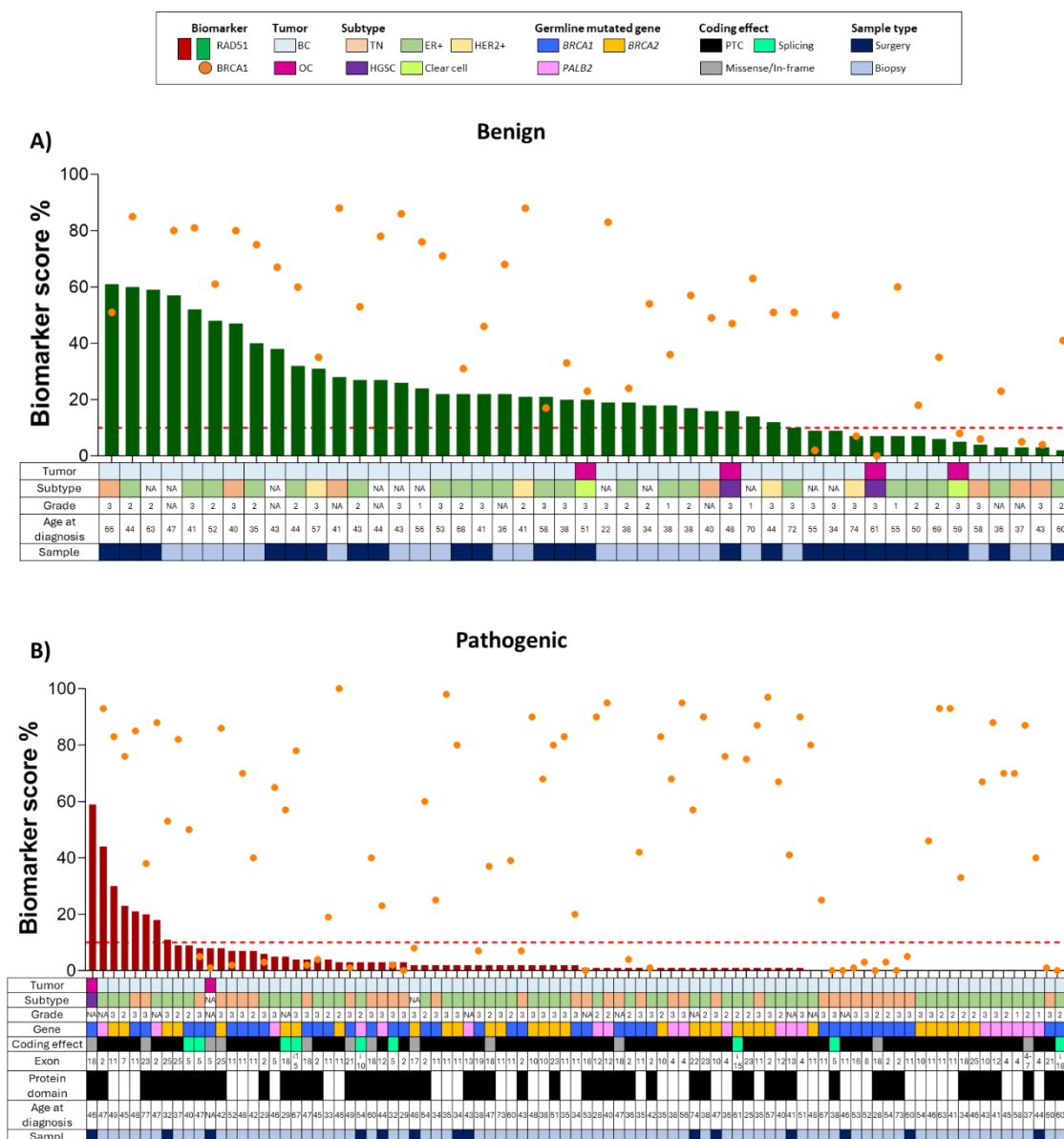


Figure 58. Summary of clinical and molecular features of the tumours from germline carriers of benign (A) and pathogenic (B) variants.

Waterfall with the RAD51% (bars) and BRCA1% scores (orange dots) for each sample. Dashed line indicates biomarker cutoff (10% for both). Table contains information regarding tumour type, subtype, histopathological grade, age at diagnosis and sample type. Additionally, for pathogenic tumours (B), germline mutated gene, germline variant exon and protein domain location, and variant coding effect are indicated. For benign cohort, patients were carriers of multiple *BRCA1/2* benign variants and were not included in the graph but are detailed in supplementary table 1. *BC*: breast cancer, *OC*: ovarian cancer, *TN*: triple negative breast cancer, *ER+*: BC with oestrogen receptor (ER) positive status, *HER2+*: BC with HER2 positive status but no expression of ER and/or progesterone receptor (PR), *HGSC*: high-grade serous carcinoma, *PTC*: premature termination codon.

Moreover, in the pathogenic cohort, no significant differences were observed between carriers of germline variants in different genes (**Figure 59**), or between carriers of variants with different coding effects (**Figure 60**).

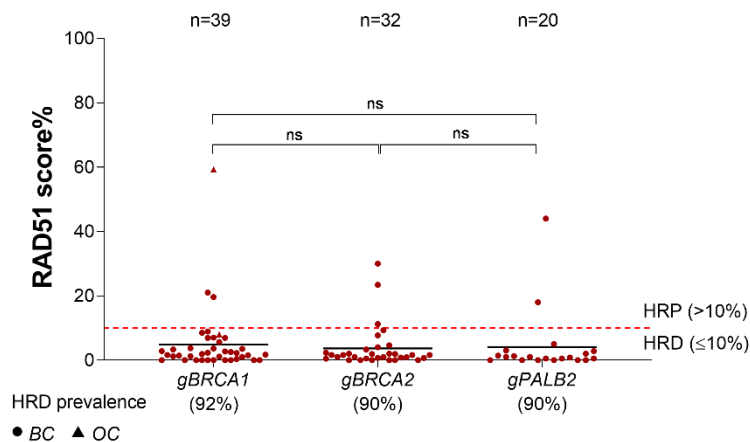


Figure 59. RAD51% score distribution among tumours from patients carrying germline pathogenic variants in *BRCA1*, *BRCA2* and *PALB2*.

The percentage of geminin-positive tumour cells with at least 5 *foci* of RAD51 is represented. RAD51 score $\leq 10\%$ indicates HRD. No statistically significant differences between groups were observed by Kruskal-Wallis non-parametric and post-hoc Dunn's multiple comparisons tests ($p > 0.05$). Tumour type indicated by dot shape. Each dot represents a single tumour. Sample size and HRD prevalence indicated for each group. *n*: number of data points, *ns*: non-significant, *BC*: breast cancer, *OC*: ovarian cancer, *HRD*: homologous recombination deficiency, *HRP*: homologous recombination proficiency, *g*: germline.

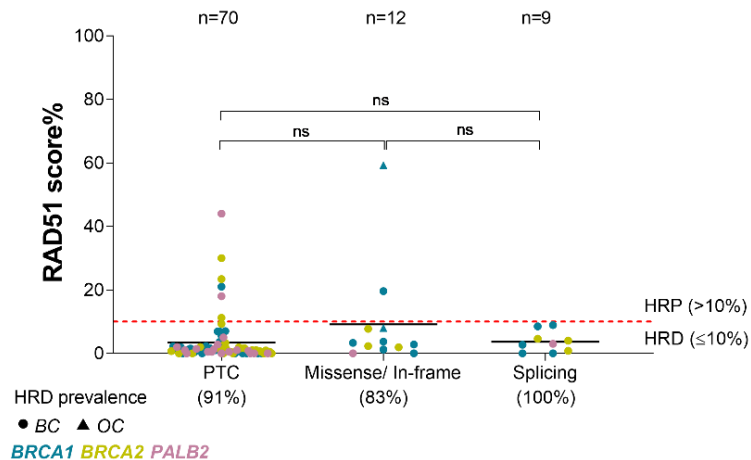


Figure 60. RAD51% score distribution among tumours from germline carriers of premature termination codon, missense/in-frame deletions or insertions, and splicing pathogenic variants.

The percentage of geminin-positive tumour cells with at least 5 foci of RAD51 is represented. RAD51 score $\leq 10\%$ indicates HRD. No statistically significant differences between groups were observed by Kruskal-Wallis non-parametric and post-hoc Dunn's multiple comparisons tests ($p > 0.05$). Tumour type indicated by dot shape. Each dot represents a single tumour. Sample size and HRD prevalence indicated for each group. Splicing group includes the c.211A>G (*BRCA1*) [294] and c.7985C>G (*BRCA2*) [161] missense variants due to their complete aberrant splicing effect. *n*: number of data points, *Ns*: non-significant, *BC*: breast cancer, *OC*: ovarian cancer, *HRD*: homologous recombination deficiency, *HRP*: homologous recombination proficiency, *g*: germline, *PTC*: premature stop codon variant (nonsense or frameshift).

Furthermore, germline carriers of the same variant also showed similar results thus suggesting a high reproducibility of the RAD51% test (**Figure 61**).

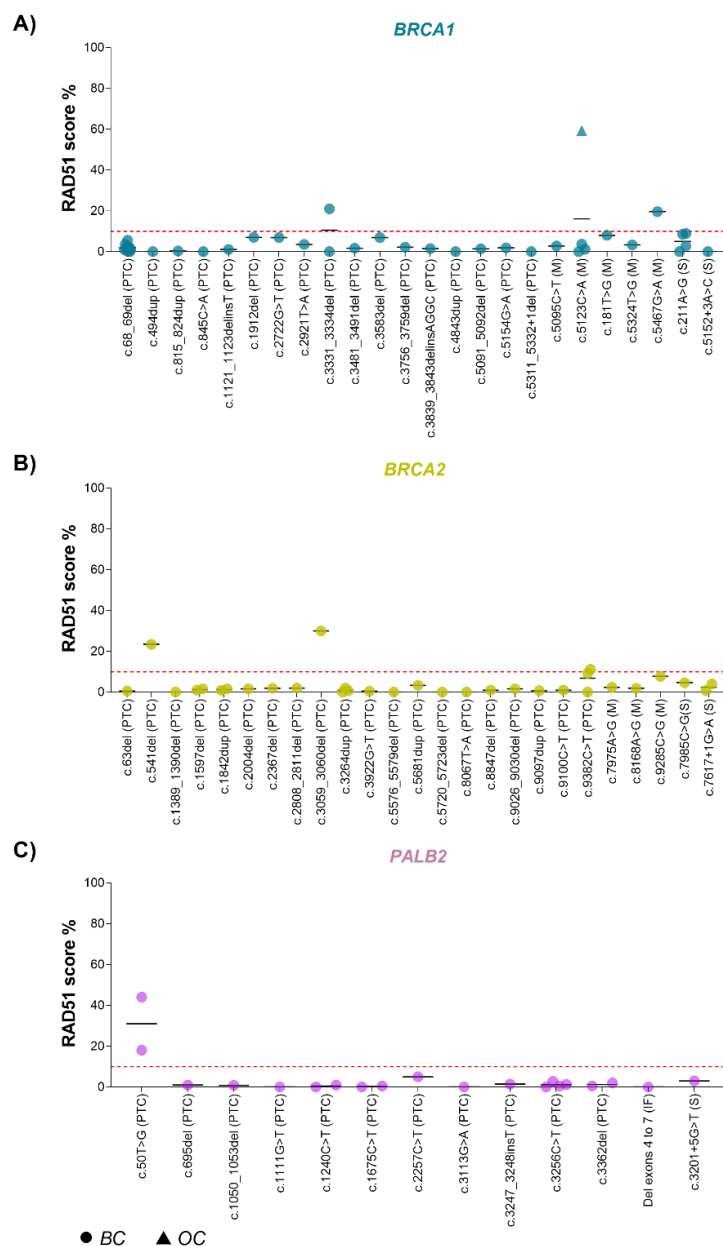


Figure 61. RAD51% score distribution across tumours from patients carrying distinct pathogenic variants in *BRCA1*(A), *BRCA2*(B) and *PALB2*(C).

The percentage of geminin-positive tumour cells with at least 5 foci of RAD51 is represented. RAD51 score $\leq 10\%$ indicates HRD. Tumour type indicated by dot shape. Each dot represents a single tumour. BC: breast cancer, OC: ovarian cancer, HRD: homologous recombination deficiency, HRP: homologous recombination proficiency, PTC: premature stop codon variant (nonsense or frameshift), M: missense, IF: in-frame, S: splicing.

RAD51% Performance in Predicting Germline Variant Status

Receiver Operating Characteristic (ROC) analysis revealed a high predictive value of RAD51% in distinguishing benign from pathogenic variants, with an Area Under the Curve (AUC) of 0.91 (**Figure 62 A**). The pre-established 10% threshold demonstrated a sensitivity of 91.2% and a specificity of 70.2% (**Figure 62 A**). When only BC samples were considered, RAD51% performance improved, achieving an AUC of 0.93, with a sensitivity of 92.1% and a specificity of 72.1% at the 10% cutoff (**Figure 62 B**). However, we observed that a cutoff of 5.7% provided a more balanced sensitivity and specificity in both scenarios (maximum *J* obtained from ROC analysis) (**Figure 62**).

Overall, the 5.7% cutoff demonstrated both higher specificity (87.2–88%) and a higher positive predictive value (PPV, 0.93–0.94) compared to the 10% cutoff (PPV of 0.86–0.87), reducing the probability of incorrectly considering a benign variant as pathogenic based on HRR status by RAD51% (**Figure 62**). This suggested that the test cutoff could be readjusted for classifying germline variants, as the established 10% (based on PARPi response [247]) might overestimate the pathogenicity of the variants.

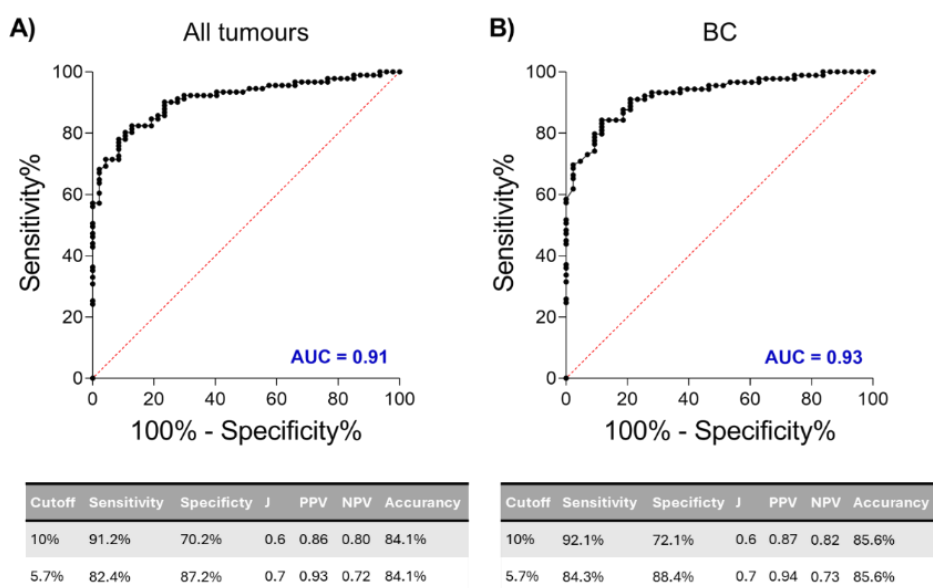


Figure 62. Receiver-operating characteristic curves (ROC) of RAD51% performance in predicting germline variant status.

In **A)** ROC considering all tumour samples from the pathogenic (n=91) and benign (n=47) groups. **B)** RAD51% performance exclusively considering breast tumours (pathogenic n=89, benign =43). *AUC: area under the curve. BC: breast cancer, J: Youden index, PPV: positive predictive value, NPV: negative predictive value.*

RAD51% LR Estimates for Tumour-Phenotype Evidence in Germline Variant Classification

This part of the thesis aimed to assess whether tumour RAD51% can predict pathogenicity or benignity for *BRCA1/2* and *PALB2* variants and be used as phenotype evidence in germline variant classification. With this purpose, LR estimates were calculated considering HR status by RAD51% (cutoff of 10%) in tumours from pathogenic and benign variant carriers and then refined by considering tumours and patients' characteristics.

Under ACMG/AMP system [118], HRD by RAD51% provided supporting evidence in favour of pathogenicity (LR: 3.06, 95%CI:1.97-4.77) and HRP yielded

moderate evidence against (LR: 0.12, 95%CI: 0.06-0.25) regardless of tumour type (breast or ovarian) (**Table 18**).

However, when LR estimations were separately calculated for breast and ovarian samples, we observed non-informative evidence for OC (LR: 1, 95%CI:0.18-5.46) due to the limited number of tumours in our cohorts (**Table 18**). For that reason, we excluded OC for the rest of the estimations combining other characteristics. Focusing on LR for BC, we observed that HRD provided supporting evidence in favour of pathogenicity (LR: 3.30 95%CI:2.03-5.36) and HRP provided moderate evidence against (LR: 0.11 95%CI:0.05-0.23) (**Table 18**).

Furthermore, we also refined LR analyses by considering germline mutated gene, but for all three genes evidence in favour or against pathogenicity reached the same evidence strength as when all genes were considered together for estimation (**Table 18**).

However, LR estimations changed when considering age at diagnosis, since HRD in tumours from patients diagnosed before age 50 reached moderate evidence in favour of pathogenicity (LR: 5.88, 95%CI: 2.38-14.54) and evidence was uninformative when HRD was detected in tumours from patients diagnosed at age 50 or older (LR:2.04, 95%CI: 1.23-3.40) (**Table 18**).

Likewise, when tumour grade was considered, evidence in favour of pathogenicity was upgraded to moderate for HRD in grade II BC (LR: 4.50, 95%CI: 1.65-12.30), and evidence against pathogenicity was upgraded to strong for HRP in grade II BC (LR:0.045, 95%CI:0.01-0.32) (**Table 18**). No changes were observed for grade III and for grade I, LR could not be calculated due to the low number of cases (**Table 18**).

Lastly, when BC molecular subtypes were considered to estimations, HRD in TN tumours reached uninformative evidence in favour of pathogenicity (LR:2.2, 95%CI: 0.93-5.19), while HRP derived evidence against pathogenicity was not changed from moderate level (LR:0.10, 95%CI: 0.02-0.46) (**Table 18**). No changes were observed for ER+ tumours and no estimates could be calculated for HER+ tumours given the low number of cases (**Table 18**).

All above mentioned characteristics (gene, age at diagnosis, grade and subtype) were also combined but not all generated sub-groups provided sufficient occurrences for informative comparisons, and for those with sufficient occurrences, calculations were based in a limited number of cases thus reaching in most cases non-significant evidence (**Supplementary Table 33**, **Supplementary Table 34**, **Supplementary Table 35** and **Supplementary Table 36**).

Overall, this data supports the use of RAD51% evidence considering only the age of BC onset as summarised in **Table 19**. This held by the fact that this variable, showed a significant change on pathogenic evidence (**Table 18**) and only one value from the total 140 samples used was missing (**Supplementary Table 25**).

Specifically, this RAD51% evidence fits in the PP4 and BP5 specifications. According to Richards et al. (2015) [99] PP4 was defined as evidence capturing patient's phenotype highly specific for the disease, and our results demonstrate that tumour HRD by RAD51% is highly specific (**Figure 62**) of tumours from carriers of pathogenic variants. We chose the BP5 code even though the general guidelines did not define it as evidence that would capture phenotypic features from tumours. We decided this in accordance with the ENIGMA *BRCA1/2* VCEP specifications [122], which propose the use of this code to capture evidence

against pathogenicity for data types that are not directly captured by existing ACMG/AMP codes.

Table 18. Estimated likelihood ratios for predicting *BRCA1/2* and *PALB2* mutation status by tumour HRR status by RAD51%.

		HRD by RAD51 test	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength
All tumours		Yes (HRD)	83 (91.2)	14 (29.8)	3.06 (1.97-4.77)	Pathogenic supporting
		No (HRP)	8 (8.8)	33 (70.2)	0.12 (0.06-0.25)	Benign moderate
By tumour type	BC	Yes (HRD)	82 (92.1)	12 (27.9)	3.30 (2.03-5.36)	Pathogenic supporting
		No (HRP)	7 (7.9)	31 (72.1)	0.11 (0.05-0.23)	Benign moderate
	OC	Yes (HRD)	2 (50.0)	2 (50.0)	1 (0.18-5.46)	Non-informative
		No (HRP)	2 (50.0)	2 (50.0)	1 (0.18-5.46)	Non-informative
By age at diagnosis (only BC)	<50	Yes (HRD)	57 (90.5)	4 (15.4)	5.88 (2.38-14.54)	Pathogenic moderate
		No (HRP)	6 (9.5)	22 (84.6)	0.11 (0.05-0.25)	Benign moderate
	≥50	Yes (HRD)	25 (96.2)	8 (47.1)	2.04 (1.23-3.40)	Non-informative
		No (HRP)	1 (3.8)	9 (52.9)	0.07 (0.01-0.52)	Benign moderate
By germline mutated gene (only BC)	<i>BRCA1</i>	Yes (HRD)	35 (94.6)	12 (27.9)	3.39 (2.08-5.51)	Pathogenic supporting
		No (HRP)	2 (5.4)	31 (72.1)	0.07 (0.02-0.29)	Benign moderate
	<i>BRCA2</i>	Yes (HRD)	29 (90.6)	12 (27.9)	3.25 (1.98-5.32)	Pathogenic supporting
		No (HRP)	3 (9.4)	31 (72.1)	0.13 (0.04-0.39)	Benign moderate
	<i>PALB2</i>	Yes (HRD)	18 (90.0)	12 (27.9)	3.23 (1.95-5.33)	Pathogenic supporting
		No (HRP)	2 (10.0)	31 (72.1)	0.14 (0.04-0.52)	Benign moderate
By tumour grade (only BC)	I	Yes (HRD)	2 (100.0)	1 (25.0)	4.00 (0.73-21.84)	Non-informative
		No (HRP)	0 (0.0)	3 (75.0)	-	-
	II	Yes (HRD)	27 (96.4)	3 (21.4)	4.50 (1.65-12.30)	Pathogenic moderate
		No (HRP)	1 (3.6)	11 (78.6)	0.045 (0.01-0.32)	Benign strong
	III	Yes (HRD)	45 (91.8)	6 (35.3)	2.60 (1.36-4.98)	Pathogenic supporting
		No (HRP)	4 (8.2)	11 (64.7)	0.13 (0.05-0.34)	Benign moderate
By molecular subtype (only BC)	TN	Yes (HRD)	32 (94.1)	3 (42.9)	2.20 (0.93-5.19)	Non-informative
		No (HRP)	2 (5.9)	4 (57.1)	0.10 (0.02-0.46)	Benign moderate
	ER+	Yes (HRD)	49 (90.7)	6 (28.6)	3.18 (1.61-6.28)	Pathogenic supporting
		No (HRP)	5 (9.3)	15 (71.4)	0.13 (0.05-0.31)	Benign moderate
	HER2+ (ER-, PR-)	Yes (HRD)	0	1 (25.0)	-	-
		No (HRP)	0	3 (75.0)	-	-

LR>1: association with pathogenic variant, pathogenic evidence; LR<1: no association with pathogenic variant, benign evidence. Evidence strength was measured using Bayesian modelling of ACMG/AMP rules (see Materials and Methods); Benign supporting ($0.23 \leq LR < 0.48$), Benign moderate ($0.05 \leq LR < 0.23$), Pathogenic supporting ($2.08 \leq LR < 4.30$), non-informative ($0.48 \leq LR < 2.08$). LR estimates reaching informative ACMG/AMP strengths at a statistically significant CI (i.e., not spanning 1), are highlighted in bold. n: sample size, LR: Likelihood ratio; CI: Confidence Interval, ACMG/AMP: American College of Medical Genetics/Association for Molecular Pathology. TN: triple negative, ER: oestrogen receptor, PR: progesterone receptor.

Table 19. Proposed utilization of RAD51% evidence for variant classification.

	HRD by RAD51 test	LR (95%CI)	ACMG/AMP strength	Application in variant classification
<i>BC <50 years old</i>	Yes (HRD)	5.88 (2.38-14.54)	Pathogenic moderate	PP4_M*
	No (HRP)	0.11 (0.05-0.25)	Benign moderate	BP5_M*
<i>BC ≥50 years old</i>	Yes (HRD)	2.04 (1.23-3.40)	Non-informative	Combine this tumour LR with multifactorial data
	No (HRP)	0.07 (0.01-0.52)	Benign moderate	BP5_M*

* If multifactorial data is available, RAD51% LR should be combined with all LRs already calculated for the specific variant. Then BP5/PP4 codes should be applied according to ACMG/AMP strength alignment as detailed in **Table 8**. BC: breast cancer, HRD: homologous recombination deficiency, HRP: homologous recombination proficiency, P: supporting, M: moderate.

Results from different tumours of germline carriers of the same variant should be combined by multiplication to increase code strengths [118]. Thus, for BP5 evidence, the number of breast tumours can be counted, whereby HRP in one breast tumour diagnosed before age 50 meets BP5 moderate evidence, two or more meet very strong. Likewise, for HRP in late-onset cases, one single tumour meets BP5_moderate, two meet strong and three or more very strong. Regarding PP4 evidence, HRD in one breast tumour diagnosed before age 50 provides moderate evidence in favour of pathogenicity, two is strong and four or more reach very strong. For HRD in tumours diagnosed at age of 50 or older, PP4 requires two different tumours to reach supporting strength, three for moderate, five for strong and nine or more for very strong. However, evidence derived from one single tumour can be also used for variant classification if combined with other LRs from a multifactorial quantitative likelihood analysis already available. Indeed, this recommendation is not limited to HRD in late-onset tumours, if multifactorial likelihood analysis is available for a given variant, the RAD51% LR

should be combined with the other LRs, and the resulting LR must be aligned to ACMG/AMP strength according to **Table 8** for BP5/PP4 proper consideration.

In case of contradictory results between tumours of germline carriers of the same variant (i.e. HRD and HRP results for a specific variant) we propose to combine all RAD51% LRs from different tumours and use the value of the resulting LR for BP5/PP4 code strength.

VUS Classification by Adding RAD51% Tumour Evidence

Following the recommendations shown in **Table 19**, we applied evidence derived from HRR status by tumour RAD51% to classify the eight variants from the initial unreviewed cohort classified as VUS under the current ACMG/AMP adapted guides (**Table 16**).

Three out of the eight tumours displayed HRD by RAD51%, and all cases were breast tumours diagnosed before age 50. Interestingly, there was a multifactorial analysis for the *BRCA1* in-frame deletion of exons 10 to 14 (**Supplementary Table 31**), hence we combined the HRD-RAD51% LR of 5.88 with those obtained in Li et al. (2020)[276], together with LRs derived from tumour histopathological features [198]. The resulting LR of 0.44 was uninformative and did not allow to change VUS classification of this variant (**Table 20**). For the c.101G>A and c.2612A>G *PALB2* variants, the HRD by RAD51 was captured by the PP4 code (**Table 20**). However, combined with all other specification met for both germline variants did not achieve a different classification, and remained as VUS.

For the c.9410C>T *BRCA2* variant we could not apply evidence derived from RAD51% as it was an OC sample (**Table 20**).

Lastly, the remaining four breast tumours showed HRP by RAD51%. BP5_M specification was applied for the interpretation of the *PALB2* c.2093T>C,

c.2401G>A and c.2425A>G variants, which reached likely benign interpretation thanks to the new tumour evidence. On contrast, for the *BRCA1* c.791_794del, the 0.11 LR resulting from the combination of RAD51% LR with the LRs obtained in Li et al. (2020)[276], and the LR derived from tumour histopathological features [198] (**Supplementary Table 31**) was uninformative. Thus, this variant remained as a VUS (**Table 20**).

Table 20. Final classification of VUS by adding RAD51% tumour evidence.

Patient	Germline		Tumour				All evidence ^g		Final classification	
	Gene ^a	Variant	Clinvar class ^b	Classification ^c	ID	Sample ^d	Clinical information ^e	RAD51 %		Evidence ^f
142	BRCA1	Del ex10-14 p.(Val199_Tyr1863delinsCysLeu)	NR	VUS	T145	BC(Bx)	Diagnosis age: 45, G2, TN	1 (HRD)	PP4_M*, Multifactorial LR available: 0.07x 5.88 = 0.44 → uninformative	PVS1_M VUS
143	BRCA1	c.791_794del p.(Ser264Metfs*33)	PAT	VUS	T146	BC(Bx)	Diagnosis age: 38, G3, TN	58 (HRP)	BP5_M*, Multifactorial LR available: 9.29 x 0.11 = 1.02 → uninformative	PVS1_M(RNA) VUS
136	BRCA2	c.9410C>T p.(Thr3137Ile)	VUS	VUS	T139	OC(Sx)	Diagnosis age: 62, G3	15 (HRP)	NA	PM2_P(1)+ BP4(-1) = 0 VUS
138	PALB2	c.101G>A p.(Arg34His)	VUS	VUS	T141	BC(Bx)	Diagnosis age: 30, NA, ER+	3 (HRD)	PP4_M	BP1(-1)+ PP4_M (2) = 1 VUS
137	PALB2	c.2093T>C p.(Leu698Pro)	L BEN	VUS	T140	BC(Bx)	Diagnosis age: 57, G2, ER+	31 (HRP)	BP5_M	PM2_SUP(1)+BP1(-1)+ BP5_M (-2) = -2 Likely benign
140	PALB2	c.2401G>A p.(Asp801Asn)	VUS	VUS	T143	BC(Sx)	Diagnosis age: 41, NA, ER+	23 (HRP)	BP5_M	BP1+BP5_M Likely benign
141	PALB2	c.2425A>G p.(Thr809Ala)	VUS	VUS	T144	BC(Sx)	Diagnosis age: 35, G3, ER+	14 (HRP)	BP5_M	PM2_SUP(1)+ BP1(-1)+ BP5_M(-2) = -2 Likely benign
139	PALB2	c.2612A>G p.(Asp871Gly)	VUS	VUS	T142	BC(Sx)	Diagnosis age: 26, G3, ER+	3 (HRD)	PP4_M	BP1(-1) + PP3(1)+ PP4_M (2) = 2 VUS

^a Reference transcript: *BRCA1*(NM_007294.3), *BRCA2* (NM_000059.3), *PALB2* (NM_024675.3). ^b ClinVar reported classification: VUS (Uncertain), PAT (pathogenic), LBEN (likely benign), NR: not reported. ^c Classification without considering tumour RAD51%. ^d Sample: BC (breast cancer), OC (ovarian cancer), Bx (biopsy) and Sx (surgery). ^e Clinical information: age at diagnosis, tumour grade and tumour type (TN: triple negative, ER+: oestrogen receptor positive). ^f If multifactorial data was available (detailed in Supplementary table 7), RAD51% derived LR (in bold) was combined with all LRs already calculated for the specific variant. Then BP5/PP4 codes were applied according to ACMG/AMP strength alignment as detailed in Table 1. ^g Contradictory criteria were combined based on point system [119]. HRD: homologous recombination deficiency, HRP: homologous recombination proficiency, NA: Non-applicable, M: moderate, P: supporting.

DISCUSSION

Part 1

The contribution of deep intronic variants to HBOC disease is not well known due to their location in poorly screened regions, but their potential effect on transcript splicing including intron sequences in mature RNA may be clinically significant [133,253]. For this reason, the identification and subsequent RNA characterization of this type of variants should be considered when conventional genetic analysis focused on coding regions and exon/intron boundaries does not lead to the identification of pathogenic variants [295].

However, the identification of deep intronic variants is challenging due to the lack of specific *in silico* pipelines [149]. Recently, some published studies using SpliceAI, a deep learning-derived algorithm [168], suggest its utility to identify with high efficiency intronic and exonic variants creating or enhancing cryptic splice sites and leading to splicing alterations [150,296–298]. Nevertheless, to our knowledge, only small datasets of deep intronic variants have been used to test the performance of the SpliceAI tool for identifying this type of variants [149,168,172,297].

Our work provides a large dataset of deep intronic variants that are clinically relevant, as they were tested in a clinical setting using blood, mini, or midigene assays and is well balanced with 133 altering and 100 non-altering splicing events. Hence, this data can be used as a positive control training set for further improvements of computational prediction tools. With this data, we confirmed that SpliceAI with a threshold of ≥ 0.05 has an optimal predictive value in the identification of spliceogenic deep intronic variants, obtaining a MCC of 0.86. Interestingly, Riepe et al. (2021) [297] with an optimized SpliceAI cut-off score of 0.18, also showed a high performance of 0.84 MCC for predicting 81 deep intronic variants in the *ABCA4* gene, that are also included in our literature

dataset. The authors further demonstrated that SpliceAI was the best tool for these 81 deep intronic variants compared with other deep-learning based algorithms [297]. In this line, Jaganathan et al. (2019) [168] in their SpliceAI development article, demonstrated that by applying SpliceAI with a cut-off of ≥ 0.5 to GTEx RNAseq data, it achieved a sensitivity of 71% when the variants were near exons (82 variants, overlapping exons or ≤ 50 nt from exon-intron boundaries), but dropped to 41% when the variants were in deep intronic regions (37 variants, >50 nt from exons). In sum, our study together with the two last works mentioned above, support that a low SpliceAI threshold is needed to especially detect deep intronic splice-altering variants. In our experimental dataset with clinical variants, using our optimized ≥ 0.05 threshold, SpliceAI attained a performance of 0.62 MCC, predicting all five experimentally confirmed spliceogenic variants, but with a low specificity of 71.42% due to a high number of false positives (**Table 13**).

Besides creating or enhancing cryptic splice sites, the intronic variants can lead to the inclusion of pseudoexons by creating or disrupting intronic SREs. Our evaluation of the prediction of splice alteration through SRE involvement of both deep intronic (**Table 9**) and exonic (**Table 10**) variants pointed out that for SpliceAI, it is more challenging to predict the impact of this type of variants. This is possibly due to the fact that the deep learning network approach used for SpliceAI development was not able to account for the SREs, denoting that the performance of SpliceAI can still be improved. To note, this was the first study testing the prediction capacity of a deep learning method such as SpliceAI of exonic variants disrupting SREs leading to exon skipping.

To supplement lower SpliceAI performance for detecting SREs altering variants, we added the ESRseq, which has a high capacity to recognize this type of

variants [172], obtaining an increase of sensitivity but a lower specificity (**Table 11**). We reasoned that this limitation was due to the fact that ESRseq evaluates on a hexamer local level, without accounting for a SREs landscape that defines a region to be included as a pseudoexon. This prompted us to characterize the landscape of SREs in pseudoexons using SRE scores, showing that the relation of the SRE landscape between the pseudoexon and flanking introns is similar to that of canonical exons, but less defined (**Figure 31** and **Figure 32**). In contrast, the ESRseq developers in Ke et al. (2021) [134] reported that the pseudoexons did not present a different balance of SRE concerning the adjacent intronic regions. This discrepancy could be due to the fact that the pseudoexons analysed in the above-mentioned work were theoretically defined, without an experimental RNA evaluation, as the intronic sequences had lengths between 50 and 250 nt and consensus values based on the Shapiro-Senapathy algorithm, of ≥ 75 for 3' splice sites and ≥ 78 for 5' splice sites, and were located beyond 100 bp from the exons [134]. Instead, we collected 133 pseudoexons from literature, experimentally confirmed using patient, mini, or midigene-derived RNA. Notably, similar findings to our results using approaches other than ESRseq tools have been recently reported, in 42 pseudoexons experimentally validated in the *DMD* gene, showing a smaller density of exonic splicing enhancers (ESEs) together with a higher density of exonic splicing silencers (ESSs) compared to canonical exons, which suggested that the pseudoexons presented a weaker exon profile in terms of SREs [299]. Interestingly, these differences have also been observed between alternative and canonical exons. The effect of variants altering the balance of SREs appears to be greater in alternative exons, which have fewer redundant enhancer elements, compared to constitutive ones [300]. Therefore, we suggest that deep intronic variants that strengthen an enhancer or

even decrease a silencer will have a greater chance of being spliceogenic provided they are located in intron regions with an exon-like SRE landscape, similar to what happens in alternative exons.

Given the role of an exon-like SRE landscape in the inclusion of pseudoexons, we combined the $\Delta\text{SpliceAI}$ 0.05 and 0.63 ΔESRseq optimized thresholds, and the absolute difference of SRE balance between the regions before and after a variant. This last value, with the variants compiled from literature, indicates clear differences between splicing altering variants and those without any effect (**Figure 33**) and helps to identify spliceogenic variants, with high sensitivity and specificity (**Table 11**). The combination was assayed in a tumour RNA-seq dataset (**Supplementary Table 3**) [259] obtaining a sensitivity of 86%. Moreover, applying the same *in silico* combination to a set of deep intronic variants from a cohort of HBOC patients allowed us to identify five spliceogenic variants that were predicted by SpliceAI to activate a cryptic splice site nearby, while assaying eight false positives (**Supplementary Table 4**).

The low accuracy of the *in silico* strategy combining $\Delta\text{SpliceAI}$, ΔESRseq , and SRE landscape using ESRseq scores can be explained by two reasons. First, our literature dataset only contains 16 splice altering variants by affecting a regulatory element. Second, the splicing aberrations of the literature and experimental datasets were assayed using RT-PCR, which has an intrinsic bias towards smaller amplicons. This is especially relevant in the case of pseudoexons as they generate larger fragments than normal transcripts and which can also be less expressed if they cause a premature termination codon that would make them targetable for nonsense-mediated decay.

Additionally, it is worth stating that the main purpose of this specific part of the thesis (**Part 1**) was not the clinical classification of the variants experimentally assessed but rather to investigate how to improve the *in silico* identification of deep intronic splicing variants. Our qualitative analysis only allowed the detection of aberrant transcripts. Additional splicing assays measuring variants' contribution to normal transcript expression, as well as co-segregation data or other functional assays are further needed to classify the spliceogenic variants detected [178,301].

Overall, the results obtained with the three sets of deep intronic variants (literature, tumour, and experimental) demonstrated that SpliceAI alone is able to identify variants causing pseudoexons and that the addition of ESRseq increases the number of false positives. Moreover, our use of ESRseq values to map the SRE balance in canonical exons and pseudoexons differentiating exonic landscapes from intronic ones suggest that this approach might be systematically used to identify exon-like landscapes in introns of HBOC and Lynch genes, thus helping to interpret whether an intronic variant makes a region much more exonizable.

Part 2

In **Part 2**, we conducted a comprehensive analysis of 166 variants affecting *BRCA1* exon 18 since several variants showed conflicting interpretations of pathogenicity among laboratories depending on if *in vitro* splicing analyses were or not were performed. Additionally, the skipping of this exon generates an in-frame transcript coding for a BRCA1 protein with unknown function, which also diffculted variants interpretation Our results allowed us to identify variants that induced altered splicing, characterize the severity of the splicing alteration, determine the tolerable levels of exon skipping and corresponding full-length (FL) levels required to maintain BRCA1's tumour suppressor activity and provide assertive clinical classifications for 152 variants.

We centralized mRNA splicing profiling in a single laboratory to ensure that all patient samples would be analysed uniformly. However, RNA was obtained from different sources, and this may have led to slight differences observed in carriers of the same variant, such as the c.5152+2dup. RNA from short-term lymphocyte cultures treated with puromycin to inhibit NMD showed increased levels of the PTC $\Delta 17-18$ transcript, whereas levels were lower in RNA whole-blood samples from two other carriers (**Figure 38 and Supplementary Table 13**). Another variant showing major discrepancies between carriers was the c.5123C>A. We observed ~18% of exon skipping in whole-blood and Peripheral Blood Mononuclear Cells (PBMCs) RNA from two different carriers, and around 2% of skipping in LCL RNA from another carrier. These differences in exon 18 skipping rate for c.5123C>A variant depending on the RNA source have been also previously reported [302–304].

Normal splicing profiles were observed in RNA obtained from LCLs [302] or from cultured lymphocytes treated with puromycin [304], but partial skipping of exon 18 was detected when RNA was extracted directly from fresh lymphocytes [305]. This variant is predicted to eliminate an exonic splicing enhancer (ESE) and create an exonic splicing silencer (ESS) according to Δ ESRseq score (-2.6) (**Supplementary Table 9**), and it has been proven that this variant induces exon skipping by increasing the binding of splicing repressors [303]. These differences in splicing patterns depending on cell type might be due to the differential expression of factors involved in SRE recognition [306–308]. Additionally, this variant also decreased exon 18 inclusion in minigenes from Millevoi et al. (2010)[303], but this result was not observed in our construct. Of note, Millevoi et al. (2010)[303] used a *BRCA1* minigene comprising exons 17–19 and shortened introns 17 and 18, and those minigenes were spliced in *in vitro* reactions containing HeLa nuclear extracts [159]. Our minigenes did not include exon 17 and were transfected into HeLa cells, hence the non-concordant result might be due to the technical differences between the minigene reporter assays. We also detected differences among carriers of c.5078_5080del, wherein we detected both r.5078_5080del and Δ 18 transcripts. Notably, exon 18 skipping was more pronounced in the PBMC RNA sample of one carrier (Δ 18=56.2% splicing fraction) compared to RNA from LCLs from the other carrier (Δ 18=32.4% splicing fraction). We consider that the increased level of Δ 18 in PBMCs might be due to the comparatively lower *BRCA1* expression in these cells, with 0.8 transcripts per million (TPM) in whole blood versus 19 TPM in LCLs (<https://gtexportal.org/home/gene/BRCA1>). It's plausible that PCR amplification preferences for smaller fragments could have led to an overestimation of exon 18 skipping due to the lower basal *BRCA1* mRNA levels in the PBMC sample.

Nevertheless, we cannot rule out that these differences might be also result from interindividual variability due to different genetic background in each carrier. Additionally, slight differences in $\Delta 18$ levels between carriers of the same variant were also observed although splicing analyses done using the same or very similar sample type (i.e. whole-blood and PBMCs). Specifically, we observe this phenomenon for carriers of c.5075-53C>T ($\Delta 18$ splicing fraction 2.3%, 9.4%, and 10.2%), c.5095C>T ($\Delta 18$ splicing fraction 0.2%, 1.2%, and 10.6%), c.5096G>A ($\Delta 18$ splicing fraction 0% and 6.1%), c.5100A>G ($\Delta 18$ splicing fraction 0% and 12.6%), and c.5117G>C ($\Delta 18$ splicing fraction 0% and 11.8%) (**Figure 38, Supplementary Table 9**). We might attribute the differences to individual genetic background, as previous works reported that blood tissue showed a high level of genetic modulation of alternative splicing[309]. Nevertheless, these intra-individual splicing differences did not result in an incongruent variant interpretation.

A limitation with patient samples was the challenge of achieving the highest level of evidence for PVS1 allele-specific approaches. This requires the presence of an exonic variant in heterozygosity, which may not always be attainable. Furthermore, the collaborative efforts of an international consortium like ENIGMA have enabled the analysis of a substantial number of variants in patients, although this may not always be feasible. In both scenarios, minigenes offer a valuable solution. In this study, minigenes allowed increase of the number of tested variants from 32 to 166, and as observed in Tubeuf et al.(2020) [188]and Klift et al. (2015) [310] , we confirmed that these reporter assays are highly predictive of human variant allele splicing profiles (**Figure 41**). However, we faced incongruent PVS1 code application between results for three splicing-altering variants located outside acceptor or donor conserved regions (**Figure 41**

and **Supplementary Table 17**: c.5080G>T, c.5078_5080del and c.5152+7A>G. Goina et al. (2008) [160] demonstrated that the c.5080G>T variant induces skipping of exon 18 by creating a splicing silencer, which is recognized by the splicing repressors hnRNPA1/A2 and DAZAP. Accordingly, Δ ESRseq (-3.68) predicted the creation of an exonic splicing silencer (ESS) (**Supplementary Table 9**). Likewise, the c.5078, c.5079 and c.5080 positions are predicted to be part of an ESE located near the canonical acceptor site, (**Figure 31**), and the c.5152+7A>G is located outside the donor region (-3 to +6) but placed in an ISS, which is predicted to be disrupted by this variant (c.5152+7 WT ESRseq -2.83, Δ ESRseq=0.75). These findings suggest that the minigene construct used is not accurately pinpointing all variants affecting SREs in exon 18 of *BRCA1*. Still, certain variants outside acceptor (-4/+1) and donor (-3/+6) splice sites of this exon, predicted by SpliceAI to alter splicing (probably by affecting SREs), showed intermediate levels of exon skipping in the minigene assay as expected, such as c.5123C>G (Δ 18=27%), c.5076del (Δ 18=21%), c.5080G>A (Δ 18=21%), (**Figure 40**). Confrontation with patient RNA-derived data, including allele-specific expression, suggests: (i) that minigene construct used in this project, as well as the mESC system may underestimate the severity of certain *BRCA1* exon 18 variants, especially of those leading to partial splicing defects in patient's RNA, and/or (ii) that tissue type and/or patients' genetic background and may influence the spliceogenicity of such variants.

Regarding in silico predictions, SpliceAI showed a specificity and sensitivity higher than 75% in both patients and minigenes, being higher in this last model probably due to the higher number of variants tested and because patients could carry additional variants that can influence on the splicing of exon 18. Although, this tool was not capable of identifying five splicing altering variants (false

negative cases) with a $\Delta\text{SpliceAI}$ score ≤ 0.1 and $\text{FL} < 99\%$ in minigenes or $\Delta 18 > 3.84\%$ in patients. Two of them led to minor splicing alterations, the c.5112del and c.5125G>A, which showed $\sim 97\%$ of FL in minigenes. The three remaining variants were the delins c.5077_5079delinsTTCATTCTGCA, c.5077_5080delinsTTGATTCTGC, and c.5077_5080delinsTTCATTCTGC, which showed $\sim 3\%$ of FL in minigenes (i.e. $\sim 97\%$ of $\Delta 18$), with the last variant showing a complete exon 18 skipping in patients. Noticing the higher prevalence of delins variants among false negatives, we estimated SpliceAI's accuracy based on variant type. Our findings revealed that while SpliceAI an accuracy range of 88.9% to 100% for deletions, insertions, duplications, and substitutions, its accuracy was to 0% for delins variants (**Table 21**). This suggests that SpliceAI is less reliable in predicting spliceogenicity for complex delins. Consequently, we emphasize the need for experimental assessment of splicing impact for these variants, regardless of computational predictions, at least in this exon.

Table 21. $\Delta\text{SpliceAI}$ accuracy based on variant type

Variant type	Variants with SpliceAI prediction (≥ 0.2 or ≤ 0.1)	True positives (TP) + True negatives (TN)	False positives (FP) + False negatives (FN)	Accuracy (TP+TN)/(TP+TN+FP+FN)
Deletion	18	3+13	1+1	88.9%
Insertion	2	2+0	0+0	100.0%
Delins	3	0+0	0+3	0.0%
Duplication	3	1+2	0+0	100.0%
Substitution	101	30+66	4+1	95.0%

Only minigene results were considered for calculations. Splicing alteration= full-length $< 99.0\%$, No splicing alteration= full-length $\geq 99.0\%$. Delins: deletion insertion variant.

Furthermore, our experimental results reinforce that splicing *in vitro* assessment is crucial for the accurate classification of predicted spliceogenic variants, but

splicing alterations interpretation displays higher complexity since variants can affect physiological isoform expression or induce leaky splicing defects of resulting truncating or in-frame transcripts. Defining thresholds for incomplete splicing is complex and requires expert knowledge of the disease susceptibility gene, and a large amount of experimental splicing data from control (well-known pathogenic and benign) variants on specific gene [170].

ENIGMA as *BRCA1/2* VCEP included a pre-established thresholds for determining which levels of aberrant transcripts should be considered harmful for gene function [122]. Specifically, the minimum level of 30% of functional transcripts to define a non-clinically relevant splicing alteration was determined by correlating splicing profiles of carriers of *BRCA1* c.[594-2A>C; 641A>G], with multifactorial clinical evidence against pathogenicity [194]. *BRCA1* c.[594-2A>C; 641A>G] carriers showed expression of ~70% of $\Delta 10$, a PTC transcript predicted to be degraded by NMD [145], and ~30% of the in-frame $\Delta 9-10$ transcript, predicted to encode a BRCA1 protein that retains tumour suppressor activity, as it preserves all the clinically relevant domains.

Following the VCEP recommendations, we applied these cutoffs using our splicing data for variant interpretation. However, these thresholds were not based on the abundance of in-frame transcripts that encode BRCA1 proteins lacking clinically important functional domains, such as the BRCT domain encoded by exon 18. Our findings using a mESC-based functional assay, which accurately reflects patient splicing profiles (**Supplementary figure 1**), revealed that the elimination of exon 18, or variants resulting in its complete exclusion, generate an in-frame $\Delta 18$ transcript encoding a non-functional BRCA1 protein (**Figure 43 and Supplementary figure 2**). Furthermore, as this model system

provides a splicing assay with direct measure of protein function, we could establish the $\Delta 18$ threshold for pathogenicity. We estimated that a $\Delta 18$ transcript level of up to 34%, combined with at least 59% full-length (FL) expression, is tolerable, as the HDR activity associated with these transcripts proportions falls within the functional range of known benign variants (**Figure 45**). However, these results are not in line with the previously proposed tolerability thresholds, which allows up to 70% PTC-NMD transcripts alongside a minimum of 30% rescue isoforms [194]. We suggest that the reduced tolerability of in-frame $\Delta 18$ (up to per-allele 34%) to maintain BRCA1's tumour suppressor activity in comparison to PTC-NMD $\Delta 10$ (up to per-allele 70%) [194] could be due to a dominant-negative effect. Indeed, two missense variants (c.5117G>A and c.5123C>A) located in exon 18 demonstrated a dominant negative effect in comparison to PTC variants [311], this supporting that exon 18 disruption could display this effect and that PTC-NMD transcripts could be more tolerable than the ones coding for altered proteins.

These discrepancies in tolerability ranges between in-frame and PTC-NMD isoforms, prompt us to question whether BP7_strong(RNA) and PVS1(RNA) capturing splicing observed impact should be adapted accordingly. We propose that per-allele levels of an in-frame event affecting a functional domain of BRCA1 lower than 34% (based on mESC **Figure 45 A**) together with a minimum level of 59% of FL (**Figure 45 B**) or confirmed functional transcript would met BP7_strong(RNA). On contrast, the already proposed levels of less than 70% of LoF transcripts together with a minimum of 30% of FL/assumed functional transcript should be applied when the LoF is a PTC-NMD transcript (**Figure 63**).

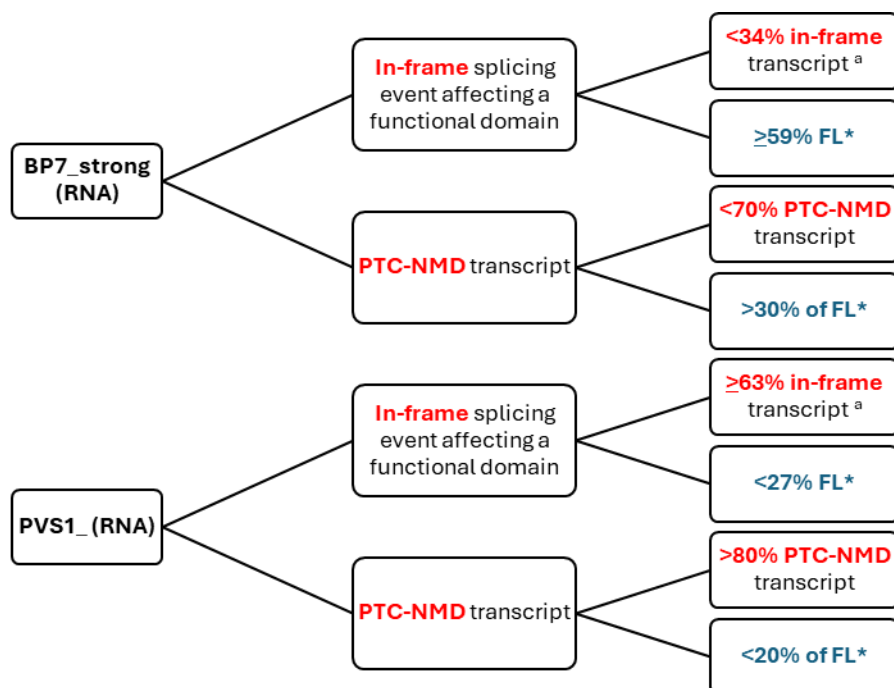


Figure 63. Proposed code adaptation for in-frame splicing events targeting critical *BRCA1* domains based on mESC HDR capacity and splicing profiles.

All % refers to per-allele expression levels.

^aThe remaining 10% correspond to other splice events assumed not rescue. ^bFL: full-length or assumed functional transcripts

According to the proposed PVS1 decision tree for *BRCA1*, all in-frame splicing alterations targeting a region critical to protein function (RING, Coiled-coil and BRCT domains), met PVS1 [122]. Considering our lineal regression analyses in mESC (**Figure 45**), those variants leading to a per-allele expression equal or greater than 63% of $\Delta 18$, together with FL (or functional rescue transcripts) lower than 27% would meet PVS1(RNA) (**Figure 63**). Although, PVS1 strength would depend also on whether per-allele expression levels have been observed in patients' samples or constructs such as minigenes, we propose decrease the PVS1 established threshold (**Figure 23**) from 80% to 63% for in-frame LoF transcripts (**Figure 63**) [122].

Despite further validation with other transcripts within this category is needed (such as $\Delta 5$, $\Delta 20$, or $\Delta 21,22$), our results constitute as a proof of concept about how splicing observed evidence could be adapted considering the nature of all splicing events that compose an mRNA profile.

In addition, we acknowledge that our mESC functional approach included a limited number of synonymous and intronic variants inducing partial exon 18 skipping. Increasing the number of variants with an intermediate effect could further refine the splicing threshold for maintaining BRCA1's tumour suppressor activity. However, our analysis using Findlay saturation genome editing assays revealed similar tolerable thresholds ($\Delta 18 < 39\%$ and $FL \geq 61\%$, **Figure 47**). Furthermore, RNA splicing results from carriers of intronic, or synonymous variants classified as (likely) benign showed maximum per-allele $\Delta 18$ and FL values around 25% and 60%, respectively (**Supplementary Table 17**), consistent with these tolerable thresholds.

Finally, we demonstrated that by using the new VCEP *BRCA1/2* guidelines we were able to decrease the percentage of VUS and unclassified variants from 45.7% to 18.3%, and this further decreased to 8.5% by adding experimental RNA splicing results (**Figure 49 and Supplementary Table 22**). Similarly, Kamps-Hughes et al. (2023)[265], observed that by integrating splicing results into variant classification the ratio of VUS reclassified as pathogenic/likely pathogenic was around 0.4% and around 6% were reinterpreted as benign/likely benign.

Of the 14 variants that remained as VUS despite our comprehensive analysis, the minigene experiments showing a complete splicing effect were not sufficient to upgrade the classification as likely pathogenic of

c.5077_5080delinsTTGATTCTGC, c.5152+3_5152+4insG and c.5152+4_5152+5insT variants. These variants could benefit from mRNA testing in patient samples to confirm whether they impose a complete splicing effect as observed in the minigene experiments (**Supplementary Table 22**). Reviewing previously published RNA splicing data, we did not find any reports of carriers for the c.5152+3_5152+4insG and c.5152+4_5152+5insT variants. The only exception was the c.5077_5080delinsTTGATTCTGC variant (**Supplementary Table 24**). Houdayer et al. (2012)[184] reported a similar variant identified as c.5077_5080del4ins10, which leads to exon 18 skipping. However, the nomenclature "del4ins10" is ambiguous, making it unclear if it corresponds to c.5077_5080delinsTTGATTCTGC. Therefore, we decided not to consider it.

The missense VUS c.5086G>T, c.5090G>A, c.5096G>C, c.5114T>G, c.5128G>A, and c.5150T>A presented ambiguous functional evidence (**Supplementary Table 18**) and/or contradictory specifications (**Supplementary Table 22**). Large-scale genetic studies could aid in determining if these inconclusive results might translate to reduced cancer risk. Indeed, intermediate risk of breast and ovarian cancer has been confirmed for other *BRCA1* variants such as c.5096G>A (here classified as likely pathogenic). This variant also exhibits ambiguous functional abrogation results [277], meet both benign and pathogenic evidence (**Supplementary Table 22**), and data from 129 *BRCA1* c.5096G>A families revealed that the cumulative risk of BC and OC by age 70 was 20% and 6%, respectively [312].

Additional information about personal and family history and co-occurrence with other variants might be also useful for the interpretation of c.5122G>A, c.5123C>G, and c.5138T>C variants that despite meeting two pathogenic

evidence criteria (PS3 and PP3), were classified as uncertain, as well as for the c.5139_5141del variant, which did not meet any criteria.

Lastly, the c.5152+7A>G was classified as uncertain despite all splicing and functional results obtained in this study due to the amount of contradictory and ambiguous results at all levels. Interestingly, all changes at c.5152+7 position were loss of function in Findlay et al. (2018)[271], whereas no other changes at other +7 positions in *BRCA1* resulted in loss of function. Further studies assessing splicing profiles in HAP1 c.5152+7A>G cells and other cell models are needed to explain contradictory results observed with our mESC functional assay. Additionally, three carriers were evaluated by different methodologies, and we could not ensure whether the observed differences in exon 18 skipping levels between them were due to biological variations (e.g. the presence of *cis* additional intronic variants that affect exon 18 splicing) or technical discrepancies (**Table 15**). Although the co-occurrence with a well-known pathogenic variant in one carrier without a FA phenotype suggests evidence against pathogenicity for the c.5152+7A>G variant, individuals carrying hypomorphic *BRCA1* variants may not fit the criteria for applying FA-related evidence. Previously, Keupp et al. (2019) [313] reported a patient with some FA phenotype features but no chromosomal fragility, carrying *BRCA1* founder c.181T>G p.(Cys61Gly) pathogenic variant and the reduced penetrance hypomorphic c.5096G>A p.(Arg1699Gln) *in trans*. Similarly, Block et al. (2024)[314] reported a male carrying the well-known pathogenic variant c.2475del and an exon 20 tandem duplication *in trans*, but with an apparently normal clinical phenotype and no *BRCA1*-associated cancer. For the c.5152+7A>G variant, the interpretation differs when the FA phenotype is considered in the combined multifactorial data versus considering the data

separately, resulting in classifications of likely benign or uncertain, respectively. This equivocal classification, along with the other contradictory specifications (**Table 15**), suggests that the variant could be an intermediate risk allele. To confirm whether a small amount of FL *BRCA1* transcript is still produced from the variant allele, allele-specific assays must be performed, as this may underlie the potential reduced penetrance phenotype. Additionally, clinical studies in a large cohort of carriers are needed to determine its associated cancer-risk.

Regarding the limitations of our study, we note the potential for splicing fraction overestimation of smaller transcripts due to the semi-quantitative nature and related to the use of Taq polymerase, which tends to preferentially amplify smaller amplicons. This bias could mask a slight contribution of the variant allele to full-length expression.

Considering our findings collectively, we can conclude that exon 18 skipping is not a rescue event and per-allele levels up to 34% are tolerable to maintain *BRCA1*'s tumour suppressor activity, suggesting that this in-frame transcript is less tolerated than PTC-NMD transcripts. Furthermore, our results demonstrate that applying recently developed *BRCA1/2* VCEP specifications significantly reduced the rate of VUS in this exon from 45.7% to 8.5%, being experimental mRNA testing essential in solving uncertainty for approximately 10% of variants.

Part 3

In this part of the thesis, we aimed to validate RAD51% score as tumour-derived evidence for germline variant classification. With this purpose, we analysed RAD51, BRCA1 and γ H2AX *foci* formation in primary untreated breast and ovarian tumours from germline carriers of variants in *BRCA1/2* and *PALB2*.

Samples were divided into three different cohorts depending on germline variant classification and review status on ClinVar: pathogenic, benign and unreviewed. No differences were detected in the distribution of γ H2AX *foci* among carriers of germline pathogenic, benign or unreviewed variants, confirming that the HRR pathway was activated similarly across all cohorts (**Figure 53**). This demonstrates the feasibility of the RAD51% test in FFPE untreated primary tumour samples, since it does not require DNA damage induction, unlike other tumour-based HRR functional assays [246,315].

The current ACMG/AMP gene-adapted guides enabled us to classify 75% (30/40) of the previously unreviewed variants as either (likely) pathogenic or (likely) benign. Thanks to this, we increased the number of variants classified in both the pathogenic and benign categories to calibrate the use of tumour evidence for variant classification (**Figure 56**).

Functional HRD by low RAD51% was evidenced in 30% of the benign cohort samples (14/47, **Figure 58 A** and **Supplementary Table 25**). This HRD rate closely aligns with the 25% prevalence observed in tumors from *BRCA1/2*-negative patients, as reported by Feng et al. (2023)[316], and it is also comparable to the 16% HRD rate noted by Nones et al. (2019)[317] in another breast cancer cohort of non-*BRCA1/2/PALB2* carriers. Half of our HRR deficient cases from benign cohort patients (7/14, 2 OC and 5 BC, **Figure 58 A** and

Supplementary Table 25) showed low *BRCA1 foci* detection which is compatible with somatic alterations or epigenetic silencing of *BRCA1* gene. Indeed, it has been reported that *BRCA1* somatic mutations occur in approximately 5%–7% of ovarian carcinomas, while a reduced *BRCA1* expression due to promoter hypermethylation is observed in 5 to 30% of epithelial ovarian cancers [290,318].

The two ovarian cancer cases from our benign cohort that exhibited HRD are consistent with these findings. One case (T116) showed a somatic potentially pathogenic *BRCA1* variant, while the other (T123) had low *BRCA1*% levels (**Table 16**). This suggests that their HRD status, as indicated by RAD51, may be due to somatic alterations in *BRCA1* rather than their respective germline variants (**Table 16**). The five remaining HRD tumours from the benign cohort with low *BRCA1*% were BC, with 60% of them (3/5) being TN (**Figure 58 A**). Approximately 21% of TN breast tumours from non-germline *BRCA1* carriers show *BRCA1* promoter hypermethylation [319], consistent with the higher proportion of this tumour type in our HRD tumours from the benign cohort with low *BRCA1*%. The remaining 50% (7/14) of tumours with presence of *BRCA1 foci* but HRD by RAD51% (**Figure 58 A**) could likely be due to somatic alterations in other HRR genes beyond *BRCA1* [289].

In the pathogenic cohort, functional HRD was evidenced in 91% of samples. However, 9% of tumours (8/91) showed HRP (**Figure 58 B**), suggesting that germline variants might not be the driver mutations or the likely existence of hypomorphic proteins [292]. These hypomorphic proteins could be the result of alternative RNA splicing events. Specifically, tumour HRP was detected in one sample from a germline carrier of the c.541del frameshift variant in exon 7 of *BRCA2* (**Supplementary Table 25**), which coding effect could be bypassed by

the naturally occurring alternative in-frame splicing event $\Delta 4,7$. This isoform had been detected expressed in blood and BC cell lines but no in normal breast tissue, and encodes for a protein lacking 105 residues that preserved HR BRCA2 functionality [177] . Similarly, RAD51 *foci* were also detected in two different breast (synchronic right and left) tumours from one patient carrying the c.50T>G nonsense variant in exon 2 of *PALB2* (**Supplementary Table 25**). In this case, the $\Delta 2p6$ naturally occurring splicing event, lacking the first 6 nucleotides of exon 2 [320], could bypass the premature stop codon introduced. This in-frame event is expressed in blood, breast and ovary normal tissue, and encodes for a protein lacking two residues from the coiled-coil (CC) PALB2 domain [320], but its functionality is still unknown. One tumour from a germline carrier of the c.3331_3334del frameshift variant in exon 11 of *BRCA1* also showed high levels of RAD51 as well as BRCA1 *foci*. The premature stop codon might be evaded by alternative events affecting exon 11, such as the naturally alternative $\Delta 11q$ isoform. This isoform is expressed in breast cancer cell lines, normal breast tissue and blood [176][292] and encodes for a BRCA1 protein that lacks residues from 264 to 1366, preserving RING, CC and BRCT functional domains. Nonetheless, our results in nine tumours from germline carriers of truncating variants in exon 11, suggest that despite the expression of BRCA1 exon 11 variants (indicated by the BRCA1 *foci* detection), these tumours are classified as HRD by RAD51 *foci* (**Figure 58 B, Supplementary Table 25**). Mechanistically, Wang et al. (2016) [292] reported that $\Delta 11q$ -BRCA1 protein generates BRCA1 and RAD51 *foci* after irradiation in BC cell lines harbouring mutations in exon 11 but BRCA1 interacts less efficiently with functional protein partners. In addition, it was also observed that PARPi had a greater impact on the proliferation rate of exon 11 mutated cell lines in comparison to *BRCA1* WT cells, but over time

proliferation gradually increased, corresponding to increased expression of BRCA1-Δ11q. We hypothesize that in primary untreated human tumours, there is not treatment pressure affecting the expression of BRCA1-Δ11q, and the basal levels of this protein may not be sufficient for proficient HR. Therefore, our results demonstrate that RAD51% in primary untreated samples captures this HRD and it might be useful for predicting variant pathogenicity even for variants located in exon 11 [28].

Despite the presence of HRD in tumours from the benign cohort and HRP in cases in the pathogenic group, the ROC analysis yielded an AUC of 0.9 for RAD51%, indicating its high accuracy in distinguishing between pathogenic and benign carriers (**Figure 62**). Thus, indicating that RAD51% is highly predictive of germline *BRCA1/2* and *PALB2* pathogenic/benign status. Genomic scars and signatures had also demonstrated being highly accurate in identifying deficient *BRCA1/2* tumours [232–234,249]. For instance, the HRDetect, a genomic test based on six mutational signatures, identified *BRCA1/2* deficient tumours with an AUC of 0.98 [233]. Similarly, COSMIC mutational signature 3 has been widely shown to be associated with *BRCA1/2* and *PALB2* pathogenic variants [234,236]. While these results support the utility of such tests for variant classification, examples of practical application remain limited, and there is no clear guidance on integrating this evidence following ACMG/AMP guidelines. Polak et al. (2017)[234] proposed a framework for enhancing variant classification considering levels of signature 3 together with biallelic inactivation. Likewise, Yoshida et al. (2021) [249] tested their BRCAness model based on 116 genomic variables obtained from whole exome sequencing (WGS) for VUS reclassification, identifying one variant as potentially pathogenic and five as potentially benign from an initial cohort of 337 VUS. Regarding functional HRD,

the first attempt to use RAD51% as an additional tool for variant classification was in a previous work of our group, in which the HRD detected in two tumours from two different carriers supported the pathogenicity the c.3201+5G>T *PALB2* variant [183].

To our knowledge, our study is the first approach to standardize the use of tumour HRR status by RAD51% as tumour evidence against and favour of pathogenicity for HBOC high-risk genes under the ACMG/AMP specifications. Regarding our results according to tumour type, the estimated LR_s revealed non-informative evidence for ovarian samples (**Table 18**), probably due to the low number of cases (**Table 17**), together with the high prevalence of somatic *BRCA1* alterations in the benign cohort (**Figure 58 A**). Oppositely, breast's LR_s alignment to ACMG/AMP evidence strengths showed supporting evidence in favour and moderate against variants' pathogenicity (**Table 18**).

Of note, pathogenic evidence strength was upgraded to moderate when HRD was detected in tumours diagnosed before age 50, while a weak association was observed in HRR deficient tumours diagnosed at age 50 or older (LR of 2.04, 95%CI 1.23-3.40). For these last group, LR fell into the uninformative range ($0.48 \leq \text{LR} < 2.08$), but the association between HRD in favour of variant pathogenicity ($\text{LR} > 1$) was statistically significant (i.e., 95% CI not spanning 1) and therefore may be suitable for inclusion as tumour LR evidence in multifactorial likelihood modelling established for *BRCA1/2* by ENIGMA [106,117]. This weak association in patients diagnosed at an older age could be attributed to the increased proportion of HRR deficient cases identified in the benign cohort (47.1% in the late onset versus the 15.4% in early onset cases, **Table 18**). This observation implies that the prevalence of sporadic cases with HRD could be higher in late-

onset breast tumours, rendering the detection of HRD in patients with germline VUS less informative.

The histopathological features currently used in multifactorial models for *BRCA1/2* variant classification are more informative for assessing *BRCA1* pathogenicity than for *BRCA2* [198]. Our HRR findings expand the range of informative tumour evidence for both *BRCA2* and *PALB2*. This is because no significant differences were observed between tumours from germline variant carriers of *BRCA1*, *BRCA2*, and *PALB2* (**Figure 59**), leading to similar LRs estimates across these genes when no additional features were considered (**Table 18**).

LRs estimates considering molecular subtypes in BCs revealed non-informative association between HRD and variant pathogenicity in TN tumours. This is due to the high prevalence of HRD in the benign cohort TN cases (42.9%) (**Table 18**), probably due to *BRCA1* somatic alterations (**Figure 58 A**), in comparison to ER+ tumours (28.6%) (**Table 18**). This observation differs from the findings of Spurdle et al. (2014) [198], which showed that TN phenotype was highly predictive of *BRCA1* germline pathogenic mutation status, and modestly predictive of *BRCA2* status in cases diagnoses at age of 50 or older. This discrepancy is likely because we are considering HR status, and it had been reported that HRD is more prevalent in sporadic TN tumours than in ER+ cases [288,316].

Previous studies pointed that ER-negative breast tumours, such as TN are more likely to present high-grade [321]. In line with this, we observed that approximately half of grade III tumours were TN (31/66), while most of grade II tumours were non-TN cases (36/42) (**Supplementary Table 25**). Hence, grade III tumours showed HRD (LR of 2.13 95%CI:1.36-4.98) and HRP (LR of 0.13

95%CI:0.05-0.34), LRs being close to those observed for TN cases (HRD LR of 2.2 CI95%:0.93-59 and HRP LR of 0.10 95%CI 0.02-046). Contrarily, the HRD prevalence in the grade II benign cohort was lower (21.4%) (**Table 18**), probably due to the low number of TN cases in this group (**Supplementary Table 33**). This reduced HRD prevalence led to an LR that moderately supports pathogenicity (**Table 18**). Accordingly, the high HRP proportion within this same group (grade II benign) was translated into an LR aligned with strong evidence against pathogenicity.

Despite that several informative LR were found in refinement estimations, the small number of cases within some sub-groups made us recommend just focus on age at diagnosis in BC (LR estimations for HRD/HRP according to <50/≥50 age of onset) (**Table 19, Table 17, Supplementary Table 33 - 36**). Further patients and data will validate these results and would aid in determining in which subgroups RAD51% is not actually informative and in which provides strongest evidence in favour and against pathogenicity.

We propose that RAD51 tumour evidence could be used under BP5 and PP4 codes, or it be included in multifactorial likelihood model for *BRCA1/2* when available, as exemplified with *BRCA1* in-frame exon 10-14 deletion and c.791_794del variants (**Table 20**). We also demonstrated the potential use of this new evidence for VUS classification as our recommendations enabled the classification of three out of the eight VUS as likely benign (**Table 20**). For the remaining five VUS the HR status from different tumours of germline carriers of these variants would aid in their interpretations, since all existing evidence have been considered.

Despite the significant findings, we acknowledge the following limitations. RAD51% is technically challenging. The major causes of non-evaluability were weak geminin staining and background noise (36/54) (**Supplementary Table 26**). These problems might be explained by preanalytical variables including the fixation process. FFPE samples are the most accessible source of patients' tumour, but they show limitations for IF assays. The increased time to fixation is the main factor affecting IF evaluability (personal communication to Experimental Therapeutics Group from VHIO). Additionally, formalin fixation affects tissue antigens by the formation of methylene bridges that modify proteins [322] leading to poor antibody reactivity possibly explained by the poor-to-moderate fixation quality of large routine diagnostic samples [323]. In addition, manual scoring of RAD51 can be time- consuming, and tumour areas selected for quantification are reader dependent, hence inter-reader scoring variability can be increased by tumour heterogeneity [324].

Regarding patient selection, our criteria was based on a suspicion of HBOC, as patients were selected according to germline variants identified by panel testing. However, we cannot rule out the presence of pathogenic variants in the benign cohort patients, in regions not analysed by current panel testing. Furthermore, the benign cohort accounts with lower sample size than pathogenic cohort, which could affect the sensitivity and specificity estimations. For mitigating it, we are considering including more breast cancer patients with negative panel results, including patients without any benign variant identified. Another option is to assemble a new independent more balanced cohort of pathogenic and benign cases to validate our results using a two-stage approach, with our current cohort serving as the discovery stage.

Although no statistical differences were observed among HRD prevalence in pathogenic carriers of PTC, missense/in-frame and splicing variant carriers (**Figure 60**), increasing the number of missense variants would confirm if they exhibited the same HRD prevalence and association with pathogenicity as PTC variants did.

Of note, *PALB2* nonsense, frameshift splicing, and large in-frame deletion variants included in the pathogenic cohort were not initially reviewed by an Expert Panel given that such a panel was not yet established at the start of this study. We included these variants to compare RAD51% performance across the three HBOC high-risk genes and evaluate its potential in *PALB2* variant classification. Since no differences were observed in the prevalence of HRD among pathogenic variants in these genes (**Figure 59**), we assumed that *PALB2* variants were truly pathogenic and considered their HRR status in the consecutive LR estimates.

Lastly, the RAD51% cutoff that we used for LR estimations was based on PARPi response under the assumption of synthetic lethality in HRR deficient tumours. Nonetheless, our results suggested that this HRD threshold may be lower for variant classification. This finding suggests that a different strength of evidence may be reached by using this lower threshold; however, it still needs to be validated in independent cohorts (**Figure 62**).

Overall, our results suggest that HRR by RAD51 *foci* detection in BC primary untreated tumour samples could be used for variant classification, considering the age of onset a key contributor to evidence strength refinement.

INTEGRATED DISCUSSION

Since the clinical management of HBOC families relies on identifying germline pathogenic variants, the primary aim of this thesis was to enhance the classification of variants of unknown significance, enabling a greater number of patients to benefit from surveillance programs, risk-reducing strategies as well as from specific therapeutic options. With this goal, we focus on exploring how to improve the detection of potentially splicing altering variants on poorly explored deep intronic regions (**Part 1**), on assessing how essential is mRNA experimental validation for variant classification (**Part 2**) and on unravelling the potential usage of tumour information for germline variant classification (**Part 3**).

The Future in HBOC Genetic Testing: Whole Genome Sequencing and Need for Guidance on Deep Intronic Regions

In the last few years, whole-genome sequencing (WGS) has been reported to significantly increase the diagnostic yield of germline genetic testing [325,326]. Additionally, while the cost of WGS was initially high, advances in second-generation chips and improved chemistry have made its pricing comparable to many other clinical diagnostic procedures [327]. Consequently, genetic testing is expected to shift from gene panel testing to WGS, leading to a higher detection rate of deep intronic variants, which, without additional evidence remain as variants of uncertain significance (VUS).

In silico tools play an essential role in prioritizing variants that are more likely to be spliceogenic, as experimentally validating the potential effect of all detected variants is currently impractical. Although several studies have demonstrated that the SpliceAI computational tool outperforms other free-available tools

[170,297,328,329], its sensitivity for detecting spliceogenic deep intronic variants was originally reported to be only 41% using a 0.5 maximum Δ SpliceAI score threshold [168]. Our results (**Part 1**) in a dataset of 233 deep intronic variants, showed an improved sensitivity of ~94% (**Table 9**) by using a lower 0.05 threshold. Regarding the performance of this threshold in our experimental dataset (**Supplementary Table 4**), even greater sensitivity was observed (100%), but with a high rate of false positives leading to a lower specificity (71.42%) in comparison to the literature dataset (92.00%). Of note, for the five spliceogenic deep intronic variants correctly identified by Δ SpliceAI in our experimental dataset (**Supplementary Table 4**), paired donor–acceptor splice site predictions were observed, (i.e Δ SpliceAI prediction indicates the exact splice sites that would be used in the variant sequence). This observation was also done by Qian et al. (2021)[298] in deep intronic variants in inherited retinal diseases. Accordingly, Canson et al. (2023)[330] developed “SpliceAI-10K calculator”, an algorithm to interpret the splicing defects induced by SNV using SpliceAI ± 4999 pb predictions, in which paired acceptor-donor prediction was considered for the identification of variants creating pseudoexons. However, its performance has only been tested using the same “deep intronic literature dataset” that we generated and used for defining the 0.05 threshold (**Supplementary Table 1**). Therefore, an independent dataset is required to rigorously validate the accuracy of the proposed SpliceAI-10K calculator pipeline for deep intronic variants. However, data on splicing experimental results is often dispersed across various publications and datasets.

In our work, detailed in **Part 1**, and during the process of validation of our own proposed pipeline in 2021, we identified a significant gap in the availability of a freely accessible repository of splicing data for such validation purposes. Now

this need has been addressed by Sullivan et al. (2024)[143], who have created the SpliceVarDB, a comprehensive database of experimentally validated splicing variants that is free to access and can be extensively used for validating *in silico* tools or even for variant interpretation.

In **Part 1** we also pointed that Δ SpliceAI was less accurate identifying spliceogenic variants affecting SREs in deep intronic regions (**Table 9**). We also observed that variants creating pseudoexons were placed in regions with a SREs balance like that observed in canonical exons after comparing the abundance of SREs using ESRseq scores [134](**Figure 31**). However, our proposal to enhance the pipeline by integrating SpliceAI estimations with ESRseq predictions, accounting for both alterations in SREs (through Δ ESRseq) and SREs distribution (via absolute differences), did not improve the performance compared to using Δ SpliceAI alone. Despite an increase in sensitivity (95.49% vs 93.98%), the pipeline resulted in a higher rate of false positives, reflected in significantly lower specificity (69.00% Vs 92.00%) (**Table 11**).

Similarly, in a dataset of 374 pathogenic deep intronic variants located in 180 genes, Kurosawa et al. (2023)[331] reported a 90% sensitivity for Δ SpliceAI 0.2 threshold with ~50% (18/36) of false negative variants located outside splicing motifs regions of the pseudoexons (acceptor -20/+3, donor: -3+6), probably by affecting SREs. To enhance the detection of deep intronic variants affecting SREs, they proposed to incorporate, in addition to the Δ SpliceAI score, the values computed by SpliceAI for splice sites in the presence of the variant prior to subtracting the reference sequence score (referred to by the authors as “raw scores”) [331].

They observed that these raw scores for SRE's deep intronic variants were higher than Δ SpliceAI values (mean of spliceAI raw score 0.18 vs mean Δ SpliceAI of 0.06). This lower Δ SpliceAI values rely on the fact that reference and raw SpliceAI scores (computed for the sequence containing the variant) were similar values. The high reference scores computed by SpliceAI suggest that these regions were more likely to be exonizable, this in line with our findings that SRE's deep intronic spliceogenic variants are in regions with high potential to be wrongly included in mRNA (**Figure 31**). This finding prompted them to include also SpliceAI variant scores in PDIVAS tool, which showed a MCC of 0.88 in comparison to our MCC 0.83 pipeline including Δ SpliceAI and ESRseq calculations (**Table 9**)[331]. Although PDIVAS includes several *in silico* tools (Δ SpliceAI, SpliceAI variant score, MES and ConSplice), some variants affecting specific *cis*-splicing elements, such as branch point, might be miss predicted [332], as no specific dedicated tool, such as Branchpointer or BPP have been considered to improve deep intronic predictions [174].

To our knowledge, a comprehensive *in silico* predictor that identifies all potential *cis*-acting elements contributing to the variant-induced pseudoexons within deep intronic regions is still lacking [287,331].

SpliceAI in Variant Classification: Can *In Silico* Evidence Be Equally Considered for All Variants?

In **Part 1** we focused mainly on how to enhance the *in silico* detection of spliceogenic variants in deep intronic regions, as this bioinformatic evidence regardless the variant is exonic or intronic, can be also included in germline variant classification under BP4 and PP3 codes.

The ClinGen SVI Splicing Subgroup has recently proposed a process for calibrating splicing prediction score thresholds for computational tools (**Figure 64**)[170].

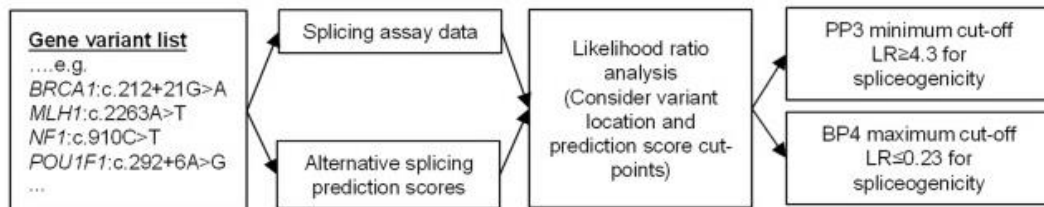


Figure 64. Process for calibrating splicing prediction score thresholds for *in silico* tools according to the ClinGen SVI splicing subgroup recommendations.

LR equal or greater than 4.3 aligned with moderate strength in favour of spliceogenicity, while LR lower or equal to 0.23 aligned with moderate evidence against spliceogenicity. LR: *likelihood ratio*. Adapted from Walker et al. (2023) [170].

As an example, they investigated the optimal threshold values for $\Delta\text{SpliceAI}$ by comparing the $\Delta\text{SpliceAI}$ scores for over 2,000 variants outside splice site dinucleotides across 8 genes (*BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *NF1* and *POU1F1*) with associated *in vitro* splicing assay outcomes. Based on this analysis they defined that the optimal $\Delta\text{SpliceAI}$ threshold for assigning PP3 to variants outside the donor/acceptor $\pm 1,2$ dinucleotides was ≥ 0.2 , which equated to a moderate level of evidence for spliceogenicity, while the optimal threshold for assigning BP4 was determined to be ≤ 0.1 as it equated to moderate level of evidence for non-spliceogenicity [170]. This calibration was based on a limited number of genes, but the obtained thresholds are proposed for generic use of PP3 and BP4, given that the mechanism of RNA splicing is not gene-specific [170].

We sought to validate the usage of both thresholds for PP3 and BP4 application, using data from the datasets from **Part 1**: 233 deep intronic variants from

literature (**Supplementary Table 1**) and 1161 exonic variants affecting SREs (**Supplementary Table 2**); and from **Part 2**: 166 variants in exon 18 of *BRCA1*, focusing only on minigenes result (**Supplementary Table 9**) given that all included variants have been analysed in this system and showed strong correlation with patients' results (**Figure 41**).

In our deep intronic dataset, which comprised variants in 80 different genes, applying the SVI splicing subgroup proposed cutoffs of ≥ 0.2 and ≤ 0.1 , resulted in LR_s corresponding to strong evidence for spliceogenicity and to moderate for non-spliceogenicity (**Table 22 A**). These results are in line with the SVI splicing subgroup's analysis only focused on intronic variants outside splice sites (-3,+6/-4,+1) which also showed strong evidence for spliceogenicity and moderate against [170].

Interestingly, when we assessed the evidence strength of our proposed 0.05 threshold, evidence for non-spliceogenicity kept the moderate strength while PP3 drop into moderate category, due to the increase of false positives (**Table 22 B**). These results demonstrate that our 0.05 threshold, apart from being considered for variant prioritization for RNA testing, might be also used for PP3/BP4 assignment on deep intronic variants.

Table 22. Likelihood ratio analysis of the Δ SpliceAI proposed thresholds from (A) Walker et al. (2023) [170] and (B) those obtained in Part 1 for 233 deep intronic variants (Supplementary Table 1).**A)**

SpliceAI	Non-spliceogenic	Proportion	Spliceogenic	Proportion	LR*
≥ 0.2	3	3/100=3%	110	110/133=82.7%	82.7/3=27.6→Strong spliceogenicity
0.2-0.1	4	4/100=4%	10	10/133=7.5%	7.5/4=1.9→ Uninformative
≤ 0.1	93	93/100=93%	13	13/133=9.8%	9.8/93=0.10→Moderate non-spliceogenicity
	100		133		

LR: likelihood ratio. *For LR derived strength see **Table 8****B)**

SpliceAI	Non-spliceogenic	Proportion	Spliceogenic	Proportion	LR*
≥ 0.05	8	8/100=8.0%	125	125/133=94.0%	94/8=11.75→Moderate spliceogenicity
<0.05	92	92/100=92.0%	8	8/133=6.0%	6/92=0.06→ Moderate non-spliceogenicity
	100		133		

LR: likelihood ratio. *For LR derived strength see **Table 8**.

Interestingly, results from the LRs for the proposed thresholds in exonic variants affecting SREs (**Supplementary Table 2**), suggested that only ≥ 0.2 should be considered for PP3 application as LRs estimates align with moderate evidence for spliceogenicity. In contrast, LRs for non-spliceogenicity resulted uninformative (**Table 23**). Similarly, Walker et al. (2023)[170] estimates for exonic variants outside splice regions achieved moderate strength for spliceogenicity and supporting against, in comparison to the moderate evidence for non-spliceogenicity observed when all variants were considered together.

Table 23. Likelihood ratio analysis of the Δ SpliceAI proposed thresholds from Walker et al. (2023) [170] for 1161 exonic variants affecting SREs (Supplementary Table 2).

	Non-SpliceAI	spliceogenic	Proportion	Spliceogenic	Proportion	LR*
≥ 0.2	49	49/801=6.1%	124	124/346=35.8%	35.8/6.1=5.9	$\rightarrow$ Moderate spliceogenicity
0.2-0.1	105	105/801=13.1%	72	72/346=20.8%	20.8/13.1=1.6	$\rightarrow$ Uninformative
≤ 0.1	647	621/801=80.8%	150	150/346=43.3%	43.3/80.8=0.54	$\rightarrow$ Uninformative
	801		346			

LR: likelihood ratio. *For LR derived strength see **Table 8**.

Regarding estimates considering the Δ SpliceAI predictions under the ≥ 0.2 and ≤ 0.1 thresholds and the results from 166 variants affecting *BRCA1* exon 18 (**Supplementary Table 9**), we observed that 0.2 cutoff reached strong evidence for spliceogenicity while 0.1 reached moderate strength for non-spliceogenicity (**Table 24**), this validating the calibration done in Walker et al. (2023)[170] and supporting the recommendation from the ENIGMA *BRCA1/2* VCEP guidelines to use PP3 and BP4 based on Δ SpliceAI 0.2 and 0.1 thresholds [122].

Table 24. Likelihood ratio analysis of the Δ SpliceAI proposed thresholds from Walker et al. (2023) [170] for 166 variants in exon 18 of *BRCA1* (Supplementary Table 9).

	Non-SpliceAI	spliceogenic	Proportion	Spliceogenic	Proportion	LR*
≥ 0.2	5	5/121=4.1%	36	36/45=80.0%	80/4.1=19.5	$\rightarrow$ Strong spliceogenicity
0.2-0.1	35	35/121=28.9%	4	4/45=8.9%	8.9/28.9=0.31	$\rightarrow$ Supporting non-spliceogenicity
≤ 0.1	81	81/121=66.9%	5	5/45=11.1%	11.1/66.9=0.17	$\rightarrow$ Moderate non-spliceogenicity
	121		45			

LR: likelihood ratio. *For LR derived strength see **Table 8**.

Overall, our results validated the usage of $\Delta\text{SpliceAI} \geq 0.2$ and ≤ 0.1 cutoffs in variant classification under PP3 and BP4 codes for most of variants. However, its usage for exonic variants affecting SREs seems not optimal (**Table 23** and [170]).

Of note, as mentioned in Walker et al. (2023)[170], despite SpliceAI LRs achieve moderate or even strong evidence weights according to ACMG/AMP alignment [118], *in silico* tools predict spliceogenicity, which is not synonymous of pathogenicity. Both PP3 and BP4 evidence should be applied at supporting level following a conservative approach, since until experimental validation is performed confirming the nature of splicing event(s) and its severity (referring to rate of normal and aberrant transcripts), pathogenicity cannot be yet bioinformatically determined.

Relevance of mRNA Experimental Testing in Variant Classification

According to VCEP *BRCA1/2*-ENIGMA recommendations for mRNA testing [122], assays should:

- Be performed on germline tissue samples (blood, cultured lymphocytes, lymphoblastoid cell lines, etc)
- Include mRNA from the same tissue type for reference controls (5-10), given that transcripts identified at similar levels in controls are naturally occurring isoforms and not mRNA aberrations
- Consider the sequencing of the FL transcript for the variant allele (if exonic), or a common polymorphism in *cis* (if variant is intronic), to assess if variant allele contributes to production of wild-type transcript
- Consider the predicted potential aberrant transcripts and the naturally occurring isoforms to ensure optimal primer design/location to capture splicing effect

- Ponder the usage of NMD inhibitors for the interpretation of relative levels of transcripts expected to be degraded by NMD pathway

Results from specific construct-based mRNA assays, such as minigenes, can be considered to aid variant interpretation if i) the transcript profile observed for WT sequence from the construct adequately represents the transcript profile observed for WT sequence from blood-based assays, ii) the construct-based mRNA assay provides estimates of allele-specific extent of aberration, and/or ii) results from the construct mRNA assay are not materially different to those observed for construct-based assays for another spliceogenic variant that has been classified using clinical information.

Regarding experimental assessment of variants in exon 18 of *BRCA1* (**Part 2**), we followed all the above-mentioned recommendations, this allowing to apply PVS1(RNA) and BP7_strong(RNA) at full weights considering patients and minigene results, which had an effect on the interpretation of 34% (54/166) of variants (**Figure 49**).

In contrast, the qualitative mRNA assessment used for the experimental dataset in **Part 1** did not provide sufficient information to interpret the five deep intronic variants affecting splicing (**Figure 34**). Based on the general recommendations from the ClinGen SVI Splicing Subgroup [170], further analyses are recommended to determine whether these effects are complete and to assess if the resulting transcripts impact clinically relevant domains. Such analyses would support a more accurate variant classification.

Multiple studies have demonstrated the value of RNA testing, based on RT-PCR and minigenes, for unequivocal variant interpretation

[157,161,178,182,183,286,333–336]. However, we acknowledge that our strategy based on RT-PCR and minigenes in **Part 2**, despite have been used extensively [150,185–188], involves several time-consuming steps that hinder its widespread in clinical diagnostics laboratory [126]. It is also important to note that our results are based on the quantification of splicing events limited to the vicinity of exon 18, and we may be overlooking much more complex events that would be detected with other methodologies.

High-throughput RNA-seq offers both qualitative and quantitative insights, enabling the identification of aberrant splicing, gene expression, and monoallelic expression. Unlike targeted RT-PCR, RNA-seq is not limited by primer design or gene annotations, allowing for a broader detection of abnormal transcripts [337]. However, several studies suggest that a read depth of approximately 5 transcripts per million (TPM) is needed for RNA-seq to be diagnostically informative; consequently, RT-PCR remains the preferred method for studying low-expressed genes [181,337,338]. *BRCA1*, along with other HBOC risk genes, is expressed at low levels in blood (**Table 25**). Therefore, the RT-PCR approach used in **Part 2** was the most suitable, given that approximately 86% (83 out of 97) of patient mRNA samples (carriers and controls) were derived from blood.

Table 25. Gene expression in whole blood and lymphoblastoid cell lines.

Gene	Whole blood (TPM)	LCLs (TPM)
BRCA1	0.9	19.8
BRCA2	0.2	11.2
PALB2	0.8	13.6
RAD51C	0.7	11.8
RAD51D	1.1	8.6
ATM	3.5	21.0
CHEK2	1.3	14.5
BRIP1	0.1	9.3
BARD1	0.7	17.3
TP53	7.7	73.0
PTEN	40.3	62.5
MLH1	6.6	48.5
MSH6	2.3	26.3
MSH2	1.4	34.6
CDH1	0.2	2.5

LCL: Lymphoblastoid Cell lines, TPM: transcripts per million. Data source: Genotype-Tissue Expression (GTEx) Project, <https://www.gtexportal.org/home/>

A promising alternative for the low-expressed genes are targeted short and long-read RNA sequencing (RNAseq) [339]. Landrith et al. (2020)[340] demonstrated that mRNA splicing profiles of 18 tumour suppressor genes by capture RNAseq using whole-blood samples increased by 9.1% the detection of pathogenic variants [264]. Likewise, Kamps-Hughes et al. (2023) [265] reclassified uncertain variants in a cohort of 20,317 patients, 0.4% of VUS as (likely) pathogenic and 5.9% as (likely) benign by incorporating results from blood targeted-RNAseq. Regarding targeted long-read RNAseq, Gulsuner et al. (2024)[341] identified eight deep intronic spliceogenic variants by long-read RNAseq in *BRCA1*, *PALB2* and *ATM*.

Potential Usage and Limitations of Tumour Information for Germline Variant Classification

The evidence-based approach taken to consider relevance of each of the original ACMG/AMP codes for curation variants in *BRCA1/2* made by ENIGMA as VCEP, demonstrated the value of detailed gene-specific adaptations to assist variant curation and resolve discordances and uncertainty in variant classification. This was demonstrated by the classification of the ENIGMA-VCEP pilot dataset composed of 40 variants, in which variant uncertainty decreased from 32.5% to 5%, by increasing (likely) pathogenic classification from 35% to 50% and (likely) benign from 32.5% to 45% [122]. We observed similar results after applying the same criteria to our 166 variants in exon 18 of *BRCA1* (**Part 2**), for which uncertainty decreased from 45.7% to 8.5% while (likely) pathogenic class increased from 36.6% to 49.4% and (likely) benign from 17.7% to 42.1% (**Figure 50**). Furthermore, we also were able to classify most of the “unreviewed cohort” *BRCA1/2* germline variants from **Part 3** (n=32) under the same rules, being the prevalence of uncertain variants reduced from 87.5% to 9.4%, and reported (likely) pathogenic and (likely) benign increased from 9.4% to 18.8% and from 3.1% to 72%, respectively (**Table 16**).

Beyond *BRCA1/2*, several efforts have been made to define gene specifications of the ACMG/AMP rules for the classification of variants in the non-*BRCA1/2*, non-MMR DNA repair associated breast, ovarian and pancreatic cancer predisposition genes, such as *PALB2*. We used the current *PALB2* VCEP-guides to classify the germline *PALB2* variants (n= 8) from the “unreviewed cohort” from **Part 3** of this thesis. After our classification, the decrease in VUS was not as significant as it was for *BRCA1/2*, given that it dropped from 75% to 62.5%, while the prevalence of benign slightly increased from 25% to 37.5% (**Table 16**). This

lower efficacy for classifying variants might be due to the fact that several ACMG/AMP original specifications cannot be currently considered for *PALB2*. For instance, this is the case for functional PS3/BS3, and PP3/BP4 *in silico* codes due to the lack of a reference set of known benign and pathogenic missense variants needed for gene-specific calibration of evidence types.

Despite demonstrating that current VCEP adaptations decrease the rate of VUS, additional evidence is still needed to solve uncertainty. Indeed, other new evidence is arising such, a recent calibration analysis of a large-scale breast cancer case sequencing dataset which has shown that co-observation of a *BRCA1*, *BRCA2*, or *PALB2* rare variant with a pathogenic variant in another breast cancer panel gene can provide supporting evidence against pathogenicity [342]. In **Part 3**, we propose another new evidence from a poorly explored source in germline variant classification: primary tumours.

Tumour information, including somatic variant data and biomarkers, had been proposed for assessing the pathogenicity of germline variants by the ClinGen Germline/Somatic Variant Subcommittee (GSVS). Some VCEPs, such as those that defined classification specifications for *TP53* or *VHL* included somatic mutational hotspots to capture evidence in favour of pathogenicity under PM1 code [250]. Likewise, the GSVS also proposed the usage of tumour biomarkers as evidence for pathogenicity, such as mutational signature 3 for *BRCA1/2*. Indeed, in Polak et al. (2017)[234] proposed a framework for using high levels of signature 3 for VUS classification. However, currently only histopathological tumour features are considered in multifactorial clinical evidence for variant classification for *BRCA1/2* [198,199], and any tumour characteristic has not been yet considered for *PALB2*. Other VCEPs such as InSiGHT ClinGen

Hereditary Colorectal Cancer/Polyposis Variant Curation Expert Panel, have already incorporated tumour features into the classification of germline variants. These features, include the expression of mismatch repair (MMR) genes assessed via immunohistochemistry, as well as the detection of microsatellite instability (MSI) determined through genetic or genomic tumour tests, in colorectal and endometrial cancer tumours [343].

Our study demonstrated that HRR status by RAD51 *foci* detection in primary breast tumours is highly accurate in distinguishing between benign and pathogenic germline *BRCA1/2* and *PALB2* carriers (**Figure 62**). According to the Bayesian framework of ACMG/AMP guides, we corroborated the potential utility of this evidence in variant classification (**Table 19**).

Of note, our LR results regarding the detection of HRP in breast tumour (independently of age at diagnosis) allows BP5_moderate code consideration (**Table 18**). Thanks to the ENIGMA/VCEP agreement to use the point system [119] to justify the expansion of benign code combinations, to include benign codes at moderate strength level, our results regarding evidence against pathogenicity can be used for variant classification. This suggests that identifying a single tumour with HRP could classify a variant as likely benign (-2 points), without requiring additional evidence. However, in the current version of *BRCA1/2* guidelines (v1.1) it is pointed out that likely benign classification can only be assigned based on one Strong Benign code if multiple types of evidence contribute to the code assigned. Hence, despite that one single HRP tumour would score -2 points, and the variant would be classified as likely benign according to Bayesian analysis, tumour features still cannot be used as unique evidence for unequivocal classification.

Despite uninformative results obtained for OC cases due to the low number of cases, HRD by RAD51% has been also observed in prostate tumours from germline carriers of pathogenic *BRCA2*, *BRCA1* and *PALB2* variants [344]. This supports that HRD is highly specific of germline pathogenic variant carriers and suggests that HRR status by RAD51% can be informative of germline variant classification beyond breast tumours.

The RAD51 test is technically less complex than NGS-based tests, requires less amount of sample, is less costly and can capture tumour heterogeneity and changes in HRR status during tumour evolution [245]. However, it is subjected to preanalytical variables that limit the handling of surgical specimens, it requires reader-training and further cross-validation studies about its feasibility for widespread usage are pending. Our results in biopsies from the unreviewed cohort (**Part 3**) showed a great concordance (86%) between HRR status by GIS and RAD51% (**Figure 55 C**). Similarly, Llop-Guevara et al. (2021) [280] observed an 87% concordance between both HRR tests, in a different cohort of TN breast tumours. These findings suggest that GIS may be as useful as RAD51% in germline classification. Likewise, other HRR mutational scars-based tests, such as HRDetect [233], mutational signature 3 [234] or CHORD [232], could be valuable alternatives once calibrated according to ACMG/AMP standards [118].

Exploring other tumour information apart from HRR status could be also useful itself, or even might be used to refine RAD51%, as age at diagnosis does (**Table 18**). Santana dos Santos et al. (2022)[345] assessed the potential use of LOH to *BRCA1* variant classification and observed that in tumours from pathogenic variant carriers the prevalence of LOH of the WT allele was 65.4% while in benign

carriers it was 22.2%, thus suggesting that LOH data can be a complementary information for *BRCA1* variant classification.

Despite the potential usages of tumour-derived information for variant classification, the main limitation is that only variant carriers that have developed cancer are informative. An alternative line of emerging evidence could involve exploring specific phenotypic features of germline carriers using normal patient-derived cells. Tan et al. (2017) [346] reported that formaldehyde at no-toxic concentration selectively depletes *BRCA2* via proteasomal degradation, which in cells heterozygous for truncating *BRCA2* variants caused genome instability. Similarly, other two works demonstrated that homologous recombination-deficient cells in *BRCA/FANC* pathway are hypersensitive to formaldehyde [347,348]. These findings suggest that sensitivity to aldehydes could be a phenotypic trait specific from (at least) *BRCA2* carriers, that could be explored using cells from heterozygous variant carriers. Alternatively, several studies have indicated that PHA-stimulated lymphocytes from *BRCA1/2* pathogenic variant carriers may exhibit enhanced *in vitro* radiosensitivity estimated by using G2 micronuclei (MN) and G2 chromosomal assays. However, these reports showed inconsistent evidence suggesting that these approaches are not likely to be useful for the differentiate carriers from non-carriers (reviewed in [349]).

The Relevance of ENIGMA Collaborative Efforts

The ENIGMA, (<https://enigmaconsortium.org/>) consortium was created in 2009. The overall aim of this consortium is to facilitate variant classification by pooling genetic, clinical, histopathological, functional and splicing data from a network of laboratories (further information in sections **2.3** and **2.4**)[117].

Several projects within the ENIGMA consortium demonstrated the added value of collaboration between laboratories, and across disciplines, to collate and interpret information from clinical testing laboratories to consolidate patient management [176,177,179,256,312,350–352].

Focusing on exon 18 project (**Part 2**), if we had limited this project to Vall d’Hebron Hospital only seven unique variants would have been tested using patients’ samples (**Figure 65**). At the national level (laboratories from Spain), the number of variants would have been increased to 18 unique variants (**Figure 65**). Thanks to the project proposal into ENIGMA, we achieved a total of 32 unique variants analysed in carriers, together with minigene splicing data generated by another ENIGMA collaborator in these 32 variants and 134 more (**Figure 65**). Additionally, the mESC functional assay performed by another collaborator for 18 out of the 166 splicing assessed variants, allowed us to estimate the tolerability range of exon 18 skipping. Overall, these pooled data allowed to classify 152 variants as (likely) benign and (likely) pathogenic under the current VCEP specifications (**Figure 50**).

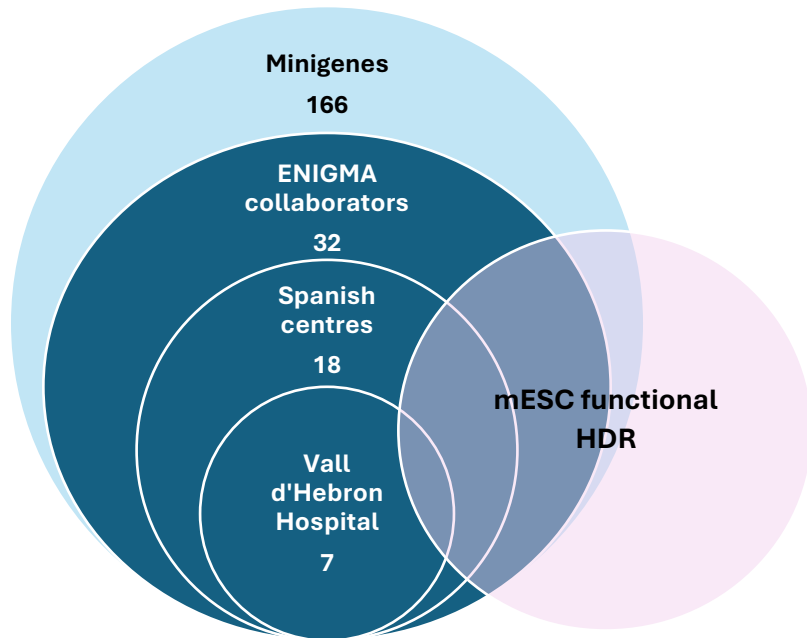


Figure 65. Number of unique *BRCA1* exon 18 analysed variants.

However, 14 variants remained uncertain despite the efforts made to classify them. Especially highlighting those made for the c.5152+7A>G, for which splicing, functional, and multifactorial quantitative model were not enough to get a conclusive classification (**Table 15, Supplementary Table 22**).

The multifactorial likelihood approach provides a quantitative estimate of pathogenicity by assessing measures from clinical data (co-segregation with disease, co-occurrence with a pathogenic variant in the same gene, reported family history, tumour pathology, etc) of variant carriers relative to characteristics observed for classical high-risk variants. Despite more than 500 *BRCA1/2* variants have been classified by this approach [256], this model is designed to distinguish high-risk pathogenic variants from variants with no or little clinical significance, and it is likely that additional methods are required to detect and validate *BRCA1/2* rare variants associated with more modest risk

than the average penetrance reported for well-known pathogenic variants in these genes [277]. This was exemplified by the c.5096G>A variant, which displayed ambiguous behaviour among several functional assays and has been confirmed, by large genetic studies, that this ambiguous effect is translated to intermediate risk for BC and OC [277,312]. Similarly, the c.5152+7A>G presented ambiguous behaviour among two functional studies (**Table 15**). Therefore, initiating a new collaborative project within the ENIGMA consortium dedicated to compiling additional splicing, genetic and clinical data from carriers will be valuable for the definitive classification of c.5152+7A>G.

Regarding germline variants from the patients included in the “pathogenic cohort” of **Part 3 (Supplementary Table 25)**, there was an overrepresentation of *BRCA1/2* Spanish most common variants, given that so far only 2 out the 91 included samples were from patients from other countries (one from Italy and one from Singapore). In detail for *BRCA1*, the four most frequent variants in our cohort were c.68_69del (21.1%, 8/38), c.211A>G (10.5% 4/38), c.5123C>A (10.5% 4/38), c.3331_3334del (5.3%, 2/38). Whereas the two most common *BRCA2* variants in our cohort were the c.3264dup (6.3%, 2/32) and the c.9382C>T (6.3%, 2/32). These frequencies were similar to those observed in a worldwide study of 29,700 families with *BRCA1* or *BRCA2* mutations, in which c.211A>G was the most common *BRCA1* mutation among Spanish families, followed by the c.68_69del, c.5123C>A, c.3331_3334del, while c.3264dup was the most frequent *BRCA2* variant in Spanish families [124]. Regarding *PALB2*, the most frequent variant in our pathogenic cohort was the c.3256C>T variant (16.6%, 3/18), which was also detected in 7 out of 60,466 cases of BRIDGES study versus only in one out of the 53,641 controls [37]. This amount of frequent

pathogenic variants among Spanish population allowed us to assess RAD51% reproducibility among carriers of the same variants.

On the other hand, most of these repeated variants, along with the unique ones, were truncating (77.1%, 54/70), which limits our understanding of whether RAD51% evidence is equally informative for missense variants (**Supplementary Table 25**). Since the ENIGMA project is still ongoing and additional samples from worldwide collaborators are expected, the number of missense variants could be increased to address if RAD51% could be equally informative for this variant type. Furthermore, we also plan to gather an independent cohort to validate the results obtained in samples analysed to date.

Collectively, the findings presented in this thesis on the improvement of detecting spliceogenic variants in deep intronic regions, the demonstration of the value of mRNA splicing assays for optimal variant interpretation, and the proposal of a new tumour-derived evidence, will significantly impact the clinical management of HBOC families. Notably, much of this work was accomplished through the collaborative efforts of the global expert network ENIGMA, which maximized the information gained from each sequenced individual and variant interpreted. This collaborative approach provides a robust framework for interpreting germline mutations and brings us closer to resolving variants of uncertain significance.

CONCLUSIONS

Part 1

1. SpliceAI predicts splicing altering deep intronic variants with high performance. However, its lower accuracy for detecting variants that impact SREs, including intronic variants that introduce pseudoexons and exonic variants that induce exon skipping, underscores the need for complementary approaches to capture the full range of spliceogenic variants.
2. Pseudoexons exhibit an SREs landscape similar to canonical exons, suggesting that certain intronic regions are predisposed to be included in mature RNA.
3. Systematically mapping of SREs in intronic regions could provide a framework to better identify regions susceptible to pseudoexonization.

Part 2

4. Experimental splicing assessment is essential for accurate variant interpretation, as evidenced by the reclassification of 34% of variants in *BRCA1* exon 18 and surrounding intronic regions when splicing results were incorporated with other available evidence.
5. The high concordance between minigene assays and patient results reinforces the utility of this construct as an optimal alternative when patient samples are unavailable. However, this system underestimates the severity of certain variants outside splice sites; for these cases,

additional verification with patient mRNA may be warranted to ensure accurate classification.

6. Given the lower reliability of SpliceAI predictions for complex delins variants compared to its high accuracy for most single-nucleotide variants, delins variants should undergo experimental splicing validation, regardless of *in silico* prediction, to ensure accurate interpretation.
7. Variant-expressing mESCs mirrored patient and minigene splicing profiles and allowed to confirm that exon 18 skipping, despite being an in-frame event, does not encode for a functional protein.
8. Exon 18 skipping appears to be less tolerable than transcripts degraded by NMD, suggesting that the level of splice perturbation required to impair *BRCA1* function may depend on the nature of the resulting loss-of-function transcript.

Part 3

9. The RAD51 test can serve as reliable tumour-derived evidence for classifying germline variants since it is highly accurate in differentiating between carriers of pathogenic and benign variants, it is reproducible across tumours from carriers of the same variant and no differences have been observed among variants with different coding effects.

- 10.** The established 10% cutoff of the RAD51 test may be adjusted to a lower threshold for variant classification, but further validation is required.
- 11.** Tumour sequencing and BRCA1 *foci* detection help uncover underlying causes of HRD in tumours from germline benign variant carriers.
- 12.** Likelihood ratio estimates based on HRR status by RAD51 test in breast tumours from germline carriers of pathogenic and benign variants, demonstrated that both HRD and HRP are informative for variant classification.
- 13.** The age of onset of breast tumours should be considered when applying RAD51 evidence for germline variant classification, as it significantly modifies the weight of HRD evidence.
- 14.** The proposed recommendations for applying RAD51 tumour-derived evidence allowed the reclassification of several VUS as likely benign, demonstrating its utility in enhancing variant classification.
- 15.** Validating the LR estimates in independent cohorts, will enhance RAD51 test robustness as new evidence for germline variant classification.

REFERENCES

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2024**, *74*, 229–263, doi:10.3322/caac.21834.
2. Łukasiewicz, S.; Czezelewski, M.; Forma, A.; Baj, J.; Sitarz, R.; Stanisławek, A. Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers (Basel)* **2021**, *13*, doi:10.3390/cancers13174287.
3. Narod, S.A. Breast Cancer in Young Women. *Nat Rev Clin Oncol* **2012**, *9*, 460–470, doi:10.1038/nrclinonc.2012.102.
4. Hennessey, B.T.; Coleman, R.L.; Markman, M. Ovarian Cancer Epidemiology and Risk Factors. *www.thelancet.com* **2009**, *374*, doi:10.1016/S0140.
5. Matulonis, U.A.; Sood, A.K.; Fallowfield, L.; Howitt, B.E.; Sehouli, J.; Karlan, B.Y. Ovarian Cancer. *Nat Rev Dis Primers* **2016**, *2*, 1–22, doi:10.1038/nrdp.2016.61.
6. Stewart, C.; Ralyea, C.; Lockwood, S. Ovarian Cancer: An Integrated Review. *Semin Oncol Nurs* **2019**, *35*, 151–156, doi:10.1016/j.soncn.2019.02.001.
7. Collaborative Group on Hormonal Factors in Breast Cancer Breast Cancer and Hormone Replacement Therapy: Collaborative of Data from 51 Epidemiological Studies of 52 705 with Breast Cancer and 108 411 Women without Breast. *Lancet* **1997**, *350*, 1047–1059, doi:10.1016/S0140-6736(97)08233-0.
8. Rodriguez, C.; Patel, A. V; Calle, E.E.; Jacob, E.J.; Thun, M.J. Estrogen Replacement Therapy and Ovarian Cancer Mortality in a Large Prospective Study of US Women. *JAMA* **2001**, *285*, 1460–1465, doi:10.1001/jama.285.11.1460.
9. Lacey, J. V; Mink, P.J.; Lubin, J.H.; Sherman, M.E.; Troisi, R.; Hartge, P.; Schatzkin, A.; Schairer, C. Menopausal Hormone Replacement Therapy and Risk of Ovarian Cancer. *JAMA* **2002**, *288*, 334–341, doi:0.1001/jama.288.3.334.
10. Nielsen, F.C.; Van Overeem Hansen, T.; Sørensen, C.S. Hereditary Breast and Ovarian Cancer: New Genes in Confined Pathways. *Nature Reviews Cancer* **2016**, *16*, 9–16.

- 599–612,
doi:10.1038/NRC.2016.72.
11. Toss, A.; Tomasello, C.; Razzaboni, E.; Contu, G.; Grandi, G.; Cagnacci, A.; Schilder, R.J.; Cortesi, L. Hereditary Ovarian Cancer: Not Only BRCA 1 and 2 Genes. *Biomed Res Int* **2015**, *2015*, 341723, doi:10.1155/2015/341723.
 12. Wendt, C.; Margolin, S. Identifying Breast Cancer Susceptibility Genes – a Review of the Genetic Background in Familial Breast Cancer. *Acta Oncol (Madr)* **2019**, *58*, 135–146, doi:10.1080/0284186X.2018.1529428.
 13. Nagy, R.; Sweet, K.; Eng, C. Highly Penetrant Hereditary Cancer Syndromes. *Oncogene* **2004**, *23*, 6445–6470, doi:10.1038/SJ.ONC.1207714.
 14. Sinn, H.P.; Kreipe, H. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care* **2013**, *8*, 149–154, doi:10.1159/000350774.
 15. Harbeck, N.; Penault-Llorca, F.; Cortes, J.; Gnant, M.; Houssami, N.; Poortmans, P.; Ruddy, K.; Tsang, J.; Cardoso, F. Breast Cancer. *Nat Rev Dis Primers* **2019**, *5*, doi:10.1038/s41572-019-0111-2.
 16. Elston, C.W.; Ellis, O. Pathological Prognostic Factors in Breast Cancer. I. The Value of Histological Grade in Breast Cancer: Experience from a Large Study with Long-Term Follow-Up. *Histopathology* **1991**, *19*, 403–410, doi:10.1111/j.1365-2559.1991.tb00229.x.
 17. Koh, J.; Kim, M.J. Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage. *Korean J Radiol* **2019**, *20*, 69–82, doi:10.3348/kjr.2018.0231.
 18. Perou, C.M.; Sørile, T.; Eisen, M.B.; Van De Rijn, M.; Jeffrey, S.S.; Renshaw, C.A.; Pollack, J.R.; Ross, D.T.; Johnsen, H.; Aksten, L.A.; et al. Molecular Portraits of Human Breast Tumours. *Nature* **2000**, *406*, 747–752, doi:10.1038/35021093.
 19. Parker, J.S.; Mullins, M.; Cheung, M.C.U.; Leung, S.; Voduc, D.; Vickery, T.; Davies, S.; Fauron, C.; He, X.; Hu, Z.; et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J Clin Oncol* **2009**, *27*, 1160–1167, doi:10.1200/JCO.2008.18.1370.
 20. Loibl, S.; André, F.; Bachelot, T.; Barrios, C.H.; Bergh, J.; Burstein,

- H.J.; Cardoso, M.J.; Carey, L.A.; Dawood, S.; Del Mastro, L.; et al. Early Breast Cancer: ESMO Clinical Practice Guideline for Diagnosis, Treatment and Follow-up 5 Behalf of the ESMO Guidelines Committee. **2024**, *46*, doi:10.1016/j.
21. Nielsen, T.O.; Hsu, F.D.; Jensen, K.; Cheang, M.; Karaca, G.; Hu, Z.; Hernandez-Boussard, T.; Livasy, C.; Cowan, D.; Dressler, L.; et al. Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma. *Clinical Cancer Research* **2004**, *10*, 5367–5374, doi:10.1158/1078-0432.CCR-04-0220.
 22. Blows, F.M.; Driver, K.E.; Schmidt, M.K.; Broeks, A.; van Leeuwen, F.E.; Wesseling, J.; Cheang, M.C.; Gelmon, K.; Nielsen, T.O.; Blomqvist, C.; et al. Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. *PLoS Med* **2010**, *7*, doi:10.1371/journal.pmed.1000279.
 23. Hugh, J.; Hanson, J.; Cheang, M.C.U.; Nielsen, T.O.; Perou, C.M.; Dumontet, C.; Reed, J.; Krajewska, M.; Treilleux, I.; Rupin, M.; et al. Breast Cancer Subtypes and Response to Docetaxel in Node-Positive Breast Cancer: Use of an Immunohistochemical Definition in the BCIRG 001 Trial. *Journal of Clinical Oncology* **2009**, *27*, 1168–1176, doi:10.1200/JCO.2008.18.1024.
 24. Cheang, M.C.U.; Chia, S.K.; Voduc, D.; Gao, D.; Leung, S.; Snider, J.; Watson, M.; Davies, S.; Bernard, P.S.; Parker, J.S.; et al. Ki67 Index, HER2 Status, and Prognosis of Patients with Luminal B Breast Cancer. *J Natl Cancer Inst* **2009**, *101*, 736–750, doi:10.1093/jnci/djp082.
 25. Lheureux, S.; Braunstein, M.; Oza, A.M. Epithelial Ovarian Cancer: Evolution of Management in the Era of Precision Medicine. *CA Cancer J Clin* **2019**, *69*, 280–304, doi:10.3322/caac.21559.
 26. Perez-Fidalgo, J.A.; Gálvez-Montosa, F.; Guerra, E.M.; Madariaga, A.; Manzano, A.; Martin-Lorente, C.; Rubio-Pérez, M.J.; Alarcón, J.; Barretina-Ginesta, M.P.; Gaba, L. SEOM–GEICO Clinical Guideline on Epithelial Ovarian Cancer (2023). *Clinical and Translational Oncology* **2024**, doi:10.1007/s12094-024-03531-3.

27. Prat, J.; Belhadj, H.; Berek, J.; Bermudez, A.; Bhatla, N.; Cain, J.; Denny, L.; Fujiwara, K.; Hacker, N.; Åvall-Lundqvist, E.; et al. Figo's Staging Classification for Cancer of the Ovary, Fallopian Tube, and Peritoneum: Abridged Republication. *J Gynecol Oncol* **2015**, *26*, 87–89, doi:10.3802/jgo.2015.26.2.87.
28. Sessa, C.; Balmaña, J.; Bober, S.L.; Cardoso, M.J.; Colombo, N.; Curigliano, G.; Domchek, S.M.; Evans, D.G.; Fischerova, D.; Harbeck, N.; et al. Risk Reduction and Screening of Cancer in Hereditary Breast-Ovarian Cancer Syndromes: ESMO Clinical Practice Guideline☆. *Ann Oncol* **2023**, *34*, 33–47, doi:10.1016/j.annonc.2022.10.004.
29. Mavaddat, N.; Antoniou, A.C.; Easton, D.F.; Garcia-Closas, M. Genetic Susceptibility to Breast Cancer. *Mol Oncol* **2010**, *4*, 174–191, doi:10.1016/j.molonc.2010.04.011.
30. Woodward, E.R.; Van Veen, E.M.; Gareth Evans, D. From BRCA1 to Polygenic Risk Scores: Mutation-Associated Risks in Breast Cancer-Related Genes. *Breast Care* **2021**, *16*, 202–213, doi:10.1159/000515319.
31. Paul Broca *Traité Des Tumeurs*; Asselin, P.; Paris, 1866;
32. Hall, J.M.; Lee, M.K.; Newman, B.; Morrow, J.E.; Anderson, L.A.; Huey, B.; King, M.-C. Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21. *Science* (1979) **1990**, *250*, 1684–1689, doi:10.1126/SCIENCE.2270482.
33. Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P.A.; Harshman, K.; Tavtigian, S.; Liu, Q.; Cochran, C.; Bennett, L.M.; Ding, W.; et al. A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1. *Science* **1994**, *266*, 66–71, doi:10.1126/SCIENCE.7545954.
34. Wooster, R.; Neuhausen, S.L.; Mangion, J.; Quirk, Y.; Ford, D.; Collins, N.; Nguyen, K.; Seal, S.; Tran, T.; Averill, D.; et al. Localization of a Breast Cancer Susceptibility Gene, BRCA2, to Chromosome 13q12-13. *Science* **1994**, *265*, 2088–2090, doi:10.1126/SCIENCE.8091231.
35. Roy, R.; Chun, J.; Powell, S.N. BRCA1 and BRCA2: Different Roles in a Common Pathway of Genome Protection. *Nat Rev Cancer* **2011**, *12*, 68–78, doi:10.1038/nrc3181.
36. Kuchenbaecker, K.B.; Hopper, J.L.; Barnes, D.R.; Phillips, K.A.;

- Mooij, T.M.; Roos-Blom, M.J.; Jervis, S.; van Leeuwen, F.E.; Milne, R.L.; Andrieu, N.; et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* **2017**, *317*, 2402–2416, doi:10.1001/JAMA.2017.7112.
37. Breast Cancer Association Consortium; Dorling, L.; Carvalho, S.; Allen, J.; González-Neira, A.; Luccarini, C.; Wahlström, C.; Pooley, K.A.; Parsons, M.T.; Fortuno, C.; et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* **2021**, *384*, 428–439, doi:10.1056/NEJMOA1913948.
 38. Rowlands, C.F.; Allen, S.; Balmaña, J.; Domchek, S.M.; Evans, D.G.; Hanson, H.; Hoogerbrugge, N.; James, P.A.; Nathanson, K.; Robson, M.; et al. Population-Based Germline Breast Cancer Gene Association Studies and Meta-Analysis to Inform Wider Mainstream Testing. *Ann Oncol* **2024**, *35*, doi:10.1016/J.ANNONC.2024.07.244.
 39. Yoshida, R. Hereditary Breast and Ovarian Cancer (HBOC): Review of Its Molecular Characteristics, Screening, Treatment, and Prognosis. *Breast Cancer* **2021**, *28*, 1167–1180, doi:10.1007/s12282-020-01148-2.
 40. Sokolova, A.; Johnstone, K.J.; McCart Reed, A.E.; Simpson, P.T.; Lakhani, S.R. Hereditary Breast Cancer: Syndromes, Tumour Pathology and Molecular Testing. *Histopathology* **2023**, *82*, 70–82, doi:10.1111/HIS.14808.
 41. Schon, K.; Tischkowitz, M. Clinical Implications of Germline Mutations in Breast Cancer: TP53. *Breast Cancer Res Treat* **2018**, *167*, 417–423, doi:10.1007/S10549-017-4531-Y/METRICS.
 42. Nieuwenhuis, M.H.; Kets, C.M.; Murphy-Ryan, M.; Yntema, H.G.; Evans, D.G.; Colas, C.; Møller, P.; Hes, F.J.; Hodgson, S. V.; Olderode-Berends, M.J.W.; et al. Cancer Risk and Genotype-Phenotype Correlations in PTEN Hamartoma Tumor Syndrome. *Fam Cancer* **2014**, *13*, 57–63, doi:10.1007/s10689-013-9674-3.
 43. Bubien, V.; Bonnet, F.; Brouste, V.; Hoppe, S.; Barouk-Simonet, E.; David, A.; Edery, P.; Bottani, A.; Layet, V.; Caron, O.; et al. High Cumulative Risks of Cancer in Patients with PTEN Hamartoma Tumour Syndrome. *J Med Genet* **2013**, *50*, 255–263, doi:10.1136/JMEDGENET-2012-101339.

44. Tan, M.H.; Mester, J.L.; Ngeow, J.; Rybicki, L.A.; Orloff, M.S.; Eng, C. Lifetime Cancer Risks in Individuals with Germline PTEN Mutations. *Clinical Cancer Research* **2012**, *18*, 400, doi:10.1158/1078-0432.CCR-11-2283.
45. Lim, W.; Olschwang, S.; Keller, J.J.; Westerman, A.M.; Menko, F.H.; Boardman, L.A.; Scott, R.J.; Trimbath, J.; Giardiello, F.M.; Gruber, S.B.; et al. Relative Frequency and Morphology of Cancers in STK11 Mutation Carriers. *Gastroenterology* **2004**, *126*, 1788–1794, doi:10.1053/j.gastro.2004.03.014.
46. Hansford, S.; Kaurah, P.; Li-Chang, H.; Woo, M.; Senz, J.; Pinheiro, H.; Schrader, K.A.; Schaeffer, D.F.; Shumansky, K.; Zogopoulos, G.; et al. Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol* **2015**, *1*, 23–32, doi:10.1001/JAMAONCOL.2014.168.
47. Woodward, E.R.; van Veen, E.M.; Forde, C.; Harkness, E.F.; Byers, H.J.; Ellingford, J.M.; Burghel, G.J.; Schlech, H.; Bowers, N.L.; Wallace, A.J.; et al. Clinical Utility of Testing for PALB2 and CHEK2 c.1100delC in Breast and Ovarian Cancer. *Genetics in Medicine* **2021**, *23*, 1969–1976, doi:10.1038/s41436-021-01234-6.
48. Couch, F.J.; Nathanson, K.L.; Offit, K. Two Decades after BRCA: Setting Paradigms in Personalized Cancer Care and Prevention. *Science* (1979) **2014**, *343*, 1466–1470, doi:10.1126/SCIENCE.1251827.
49. Rahman, N.; Seal, S.; Thompson, D.; Kelly, P.; Renwick, A.; Elliott, A.; Reid, S.; Spanova, K.; Barfoot, R.; Chagtai, T.; et al. PALB2, Which Encodes a BRCA2-Interacting Protein, Is a Breast Cancer Susceptibility Gene. *Nat Genet* **2007**, *39*, 165–167, doi:10.1038/NG1959.
50. Yang, X.; Leslie, G.; Doroszuk, A.; Schneider, S.; Allen, J.; Decker, B.; Dunning, A.M.; Redman, J.; Scarth, J.; Plaskocinska, I.; et al. Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *J Clin Oncol* **2020**, *38*, 674–685, doi:10.1200/JCO.19.01907.
51. Renwick, A.; Thompson, D.; Seal, S.; Kelly, P.; Chagtai, T.; Ahmed, M.; North, B.; Jayatilake, H.; Barfoot, R.; Spanova, K.; et al. ATM Mutations That Cause Ataxia-Telangiectasia Are Breast

- Cancer Susceptibility Alleles. *Nat Genet* **2006**, 38, 873–875, doi:10.1038/NG1837.
52. Meijers-Heijboer, H.; Van den Ouweland, A.; Klijn, J.; Wasielewski, M.; De Shoo, A.; Oldenburg, R.; Hollestelle, A.; Houben, M.; Crepin, E.; Van Veghel-Plandsoen, M.; et al. Low-Penetrance Susceptibility to Breast Cancer Due to CHEK2(*)1100delC in Noncarriers of BRCA1 or BRCA2 Mutations. *Nat Genet* **2002**, 31, 55–59, doi:10.1038/NG879.
 53. Weischer, M.; Nordestgaard, B.G.; Pharoah, P.; Bolla, M.K.; Nevanlinna, H.; Van't Veer, L.J.; Garcia-Closas, M.; Hopper, J.L.; Hall, P.; Andrulis, I.L.; et al. CHEK2*1100delC Heterozygosity in Women with Breast Cancer Associated with Early Death, Breast Cancer-Specific Death, and Increased Risk of a Second Breast Cancer. *J Clin Oncol* **2012**, 30, 4308–4316, doi:10.1200/JCO.2012.42.7336.
 54. Meindl, A.; Hellebrand, H.; Wiek, C.; Erven, V.; Wappenschmidt, B.; Niederacher, D.; Freund, M.; Lichtner, P.; Hartmann, L.; Schaal, H.; et al. Germline Mutations in Breast and Ovarian Cancer Pedigrees Establish RAD51C as a Human Cancer Susceptibility Gene. *Nat Genet* **2010**, 42, 410–414, doi:10.1038/NG.569.
 55. Loveday, C.; Turnbull, C.; Ramsay, E.; Hughes, D.; Ruark, E.; Frankum, J.R.; Bowden, G.; Kalmyrzaev, B.; Warren-Perry, M.; Snape, K.; et al. Germline Mutations in RAD51D Confer Susceptibility to Ovarian Cancer. *Nat Genet* **2011**, 43, 879–882, doi:10.1038/NG.893.
 56. Yang, X.; Song, H.; Leslie, G.; Engel, C.; Hahnen, E.; Auber, B.; Horváth, J.; Kast, K.; Niederacher, D.; Turnbull, C.; et al. Ovarian and Breast Cancer Risks Associated With Pathogenic Variants in RAD51C and RAD51D. *J Natl Cancer Inst* **2020**, 112, 1242–1250, doi:10.1093/JNCI/DJAA030.
 57. Seal, S.; Thompson, D.; Renwick, A.; Elliott, A.; Kelly, P.; Barfoot, R.; Chagtai, T.; Jayatilake, H.; Ahmed, M.; Spanova, K.; et al. Truncating Mutations in the Fanconi Anemia J Gene BRIP1 Are Low-Penetrance Breast Cancer Susceptibility Alleles. *Nat Genet* **2006**, 38, 1239–1241, doi:10.1038/NG1902.
 58. Rofes, P.; Valle, J. Del; Torres-Esquius, S.; Feliubadaló, L.; Stradella, A.; Moreno-Cabrera, J.M.; López-Doriga, A.; Munté, E.; De Cid, R.; Campos, O.; et al. BARD1 Pathogenic Variants Are

- Associated with Triple-Negative Breast Cancer in a Spanish Hereditary Breast and Ovarian Cancer Cohort. *Genes* **2021**, *Vol. 12*, Page 150 **2021**, *12*, 150, doi:10.3390/GENES12020150.
59. Xiao, X.; Melton, D.W.; Gourley, C. Mismatch Repair Deficiency in Ovarian Cancer — Molecular Characteristics and Clinical Implications. *Gynecol Oncol* **2014**, *132*, 506–512, doi:10.1016/J.YGYNO.2013.12.003.
 60. Sharif, S.; Moran, A.; Huson, S.M.; Iddenden, R.; Shenton, A.; Howard, E.; Evans, D.G.R. Women with Neurofibromatosis 1 Are at a Moderately Increased Risk of Developing Breast Cancer and Should Be Considered for Early Screening. *J Med Genet* **2007**, *44*, 481–484, doi:10.1136/JMG.2007.049346.
 61. Seminog, O.O.; Goldacre, M.J. Risk of Benign Tumours of Nervous System, and of Malignant Neoplasms, in People with Neurofibromatosis: Population-Based Record-Linkage Study. *Br J Cancer* **2013**, *108*, 193–198, doi:10.1038/BJC.2012.535.
 62. Lee, K.; Seifert, B.A.; Shimelis, H.; Ghosh, R.; Crowley, S.B.; Carter, N.J.; Doonanco, K.; Foreman, A.K.; Ritter, D.I.; Jimenez, S.; et al. Clinical Validity Assessment of Genes Frequently Tested on Hereditary Breast and Ovarian Cancer Susceptibility Sequencing Panels. *Genetics in Medicine* **2019**, *21*, 1497–1506, doi:10.1038/s41436.
 63. Hu, C.; Hart, S.N.; Gnanaolivu, R.; Huang, H.; Lee, K.Y.; Na, J.; Gao, C.; Lilyquist, J.; Yadav, S.; Boddicker, N.J.; et al. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *New England Journal of Medicine* **2021**, *384*, 440–451, doi:10.1056/NEJMoa2005936.
 64. Overview | Familial Breast Cancer: Classification, Care and Managing Breast Cancer and Related Risks in People with a Family History of Breast Cancer | Guidance | NICE Available online: <https://www.nice.org.uk/guidance/cg164> (accessed on 27 August 2024).
 65. Easton, D.F.; Pharoah, P.D.P.; Antoniou, A.C.; Tischkowitz, M.; Tavtigian, S. V; Nathanson, K.L.; Devilee, P.; Meindl, A.; Couch, F.J.; Southey, M.; et al. Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk. *New England Journal of Medicine* **2015**, *372*, 2243–2257, doi:10.1056/NEJMSr1501341.

66. Yanes, T.; Young, M.A.; Meiser, B.; James, P.A. Clinical Applications of Polygenic Breast Cancer Risk: A Critical Review and Perspectives of an Emerging Field. *Breast Cancer Research* **2020**, *22*, 1–10, doi:10.1186/s13058-020-01260-3.
67. Mars, N.; Widén, E.; Kerminen, S.; Meretoja, T.; Pirinen, M.; della Briotta Parolo, P.; Palta, P.; Havulinna, A.; Elliott, A.; Shcherban, A.; et al. The Role of Polygenic Risk and Susceptibility Genes in Breast Cancer over the Course of Life. *Nature Communications* **2020**, *11*, 1–9, doi:10.1038/s41467-020-19966-5.
68. González-Santiago, S.; Ramón y Cajal, T.; Aguirre, E.; Alés-Martínez, J.E.; Andrés, R.; Balmaña, J.; Graña, B.; Herrero, A.; Llorca, G.; González-del-Alba, A.; et al. SEOM Clinical Guidelines in Hereditary Breast and Ovarian Cancer (2019). *Clinical and Translational Oncology* **2020**, *22*, 193–200, doi:10.1007/S12094-019-02262-0/TABLES/2.
69. Beitsch, P.D.; Whitworth, P.W.; Hughes, K.; Patel, R.; Rosen, B.; Compagnoni, G.; Baron, P.; Simmons, R.; Smith, L.A.; Grady, I.; et al. Underdiagnosis of Hereditary Breast Cancer: Are Genetic Testing Guidelines a Tool or an Obstacle? *J Clin Oncol* **2019**, *37*, 453–460, doi:10.1200/JCO.18.01631.
70. Henkel, J.; Laner, A.; Locher, M.; Wohlfarth, T.; Neitzel, B.; Becker, K.; Neuhaus, T.; Abicht, A.; Steinke-Lange, V.; Holinski-Feder, E. Diagnostic Yield and Clinical Relevance of Expanded Germline Genetic Testing for Nearly 7000 Suspected HBOC Patients. *European Journal of Human Genetics* **2023**, *31*, 925–930, doi:10.1038/s41431-023-01380-2.
71. Laduca, H.; Stuenkel, A.J.; Dolinsky, J.S.; Keiles, S.; Tandy, S.; Pesaran, T.; Chen, E.; Gau, C.L.; Palmaer, E.; Shoaepour, K.; et al. Utilization of Multigene Panels in Hereditary Cancer Predisposition Testing: Analysis of More than 2,000 Patients. *Genet Med* **2014**, *16*, 830–837, doi:10.1038/GIM.2014.40.
72. Bonache, S.; Esteban, I.; Moles-Fernández, A.; Tenés, A.; Duran-Lozano, L.; Montalbán, G.; Bach, V.; Carrasco, E.; Gadea, N.; López-Fernández, A.; et al. Multigene Panel Testing beyond BRCA1/2 in Breast/Ovarian Cancer Spanish Families and Clinical Actionability of Findings. *J Cancer Res Clin Oncol* **2018**,

- 144, 2495–2513,
doi:10.1007/S00432-018-2763-9.
73. Balaguer, F.; Balmaña, J.; Bellosillo, B.; Brunet, J.; Lázaro, C.; Montoro, J.; Puig, J.A.; Salvador, H.; Tuset, E. Determinacions Del Perfil Genètic de Les Síndromes Hereditàries de Càncer En l'adult i Pediatria 2023.
74. Mandelker, D.; Donoghue, M.; Talukdar, S.; Bandlamudi, C.; Srinivasan, P.; Vivek, M.; Jezdic, S.; Hanson, H.; Snape, K.; Kulkarni, A.; et al. Germline-Focussed Analysis of Tumour-Only Sequencing: Recommendations from the ESMO Precision Medicine Working Group. *Ann Oncol* **2019**, *30*, 1221–1231, doi:10.1093/ANNONC/MDZ136.
75. DeLeonardis, K.; Hogan, L.; Cannistra, S.A.; Rangachari, D.; Tung, N. When Should Tumor Genomic Profiling Prompt Consideration of Germline Testing? *J Oncol Pract* **2019**, *15*, 465–473, doi:10.1200/JOP.19.00201.
76. Robson, M.; Offit, K. Clinical Practice. Management of an Inherited Predisposition to Breast Cancer. *N Engl J Med* **2007**, *357*, 154–162, doi:10.1056/NEJMcp071286.
77. Ang, J.; Yap, T.A.; Fong, P.; Carden, C.P.; Tan, D.S.; Hanwell, J.; Carmichael, J.; De Bono, J.S.; Gore, M.E.; Kaye, S.B. Preliminary Experience with the Use of Chemotherapy (CT) Following Treatment with Olaparib, a Poly(ADP-Ribose) Polymerase Inhibitor (PARPi), in Patients with BRCA1/2-Deficient Ovarian Cancer (BDOC). *Journal of Clinical Oncology* **2010**, *28*, 5041, doi:10.1200/jco.2010.28.15_suppl.5041.
78. Moran, A.; O'Hara, C.; Khan, S.; Shack, L.; Woodward, E.; Maher, E.R.; Laloo, F.; Evans, D.G.R. Risk of Cancer Other than Breast or Ovarian in Individuals with BRCA1 and BRCA2 Mutations. *Fam Cancer* **2012**, *11*, 235–242, doi:10.1007/S10689-011-9506-2.
79. Breast Cancer Linkage Consortium Cancer Risks in BRCA2 Mutation Carriers. *JNCI: Journal of the National Cancer Institute* **1999**, *91*, 1310–1316, doi:10.1093/JNCI/91.15.1310.
80. Hall, M.J.; Bernhisel, R.; Hughes, E.; Larson, K.; Rosenthal, E.T.; Singh, N.A.; Lancaster, J.M.; Kurian, A.W. Germline Pathogenic Variants in the Ataxia Telangiectasia Mutated (ATM) Gene Are Associated with High

- and Moderate Risks for Multiple Cancers. *Cancer Prev Res (Phila)* **2021**, *14*, 433–440, doi:10.1158/1940-6207.CAPR-20-0448.
81. Li, X.; You, R.; Wang, X.; Liu, C.; Xu, Z.; Zhou, J.; Yu, B.; Xu, T.; Cai, H.; Zou, Q. Effectiveness of Prophylactic Surgeries in BRCA1 or BRCA2 Mutation Carriers: A Meta-Analysis and Systematic Review. *Clin Cancer Res* **2016**, *22*, 3971–3981, doi:10.1158/1078-0432.CCR-15-1465.
 82. Heemskerk-Gerritsen, B.A.M.; Rookus, M.A.; Aalfs, C.M.; Ausems, M.G.E.M.; Collée, J.M.; Jansen, L.; Kets, C.M.; Keymeulen, K.B.M.I.; Koppert, L.B.; Meijers-Heijboer, H.E.J.; et al. Improved Overall Survival after Contralateral Risk-Reducing Mastectomy in BRCA1/2 Mutation Carriers with a History of Unilateral Breast Cancer: A Prospective Analysis. *Int J Cancer* **2015**, *136*, 668–677, doi:10.1002/IJC.29032.
 83. Xiao, Y.L.; Wang, K.; Liu, Q.; Li, J.; Zhang, X.; Li, H.Y. Risk Reduction and Survival Benefit of Risk-Reducing Salpingo-Oophorectomy in Hereditary Breast Cancer: Meta-Analysis and Systematic Review. *Clin Breast Cancer* **2019**, *19*, e48–e65, doi:10.1016/J.CLBC.2018.09.011.
 84. King, M.C.; Wieand, S.; Hale, K.; Lee, M.; Walsh, T.; Owens, K.; Tait, J.; Ford, L.; Dunn, B.K.; Costantino, J.; et al. Tamoxifen and Breast Cancer Incidence among Women with Inherited Mutations in BRCA1 and BRCA2: National Surgical Adjuvant Breast and Bowel Project (NSABP-P1) Breast Cancer Prevention Trial. *JAMA* **2001**, *286*, 2251–2256, doi:10.1001/JAMA.286.18.2251.
 85. Phillips, K.A.; Milne, R.L.; Rookus, M.A.; Daly, M.B.; Antoniou, A.C.; Peock, S.; Frost, D.; Easton, D.F.; Ellis, S.; Friedlander, M.L.; et al. Tamoxifen and Risk of Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *J Clin Oncol* **2013**, *31*, 3091–3099, doi:10.1200/JCO.2012.47.8313.
 86. Huber, D.; Seitz, S.; Kast, K.; Emons, G.; Ortmann, O. Use of Oral Contraceptives in BRCA Mutation Carriers and Risk for Ovarian and Breast Cancer: A Systematic Review. *Arch Gynecol Obstet* **2020**, *301*, 875–884, doi:10.1007/S00404-020-05458-W/TABLES/2.
 87. Scully, R.; Panday, A.; Elango, R.; Willis, N.A. DNA Double-Strand Break Repair-Pathway Choice in

- Somatic Mammalian Cells. *Nat Rev Mol Cell Biol* **2019**, *20*, 698–714, doi:10.1038/S41580-019-0152-0.
88. Farmer, H.; McCabe, H.; Lord, C.J.; Tutt, A.H.J.; Johnson, D.A.; Richardson, T.B.; Santarosa, M.; Dillon, K.J.; Hickson, I.; Knights, C.; et al. Targeting the DNA Repair Defect in BRCA Mutant Cells as a Therapeutic Strategy. *Nature* **2005**, *434*, 917–921, doi:10.1038/NATURE03445.
 89. Poggio, F.; Bruzzone, M.; Ceppi, M.; Pondé, N.F.; La Valle, G.; Del Mastro, L.; De Azambuja, E.; Lambertini, M. Platinum-Based Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer: A Systematic Review and Meta-Analysis. *Ann Oncol* **2018**, *29*, 1497–1508, doi:10.1093/ANNONC/MDY127.
 90. Tutt, A.; Tovey, H.; Cheang, M.C.U.; Kernaghan, S.; Kilburn, L.; Gazinska, P.; Owen, J.; Abraham, J.; Barrett, S.; Barrett-Lee, P.; et al. A Randomised Phase III Trial of Carboplatin Compared with Docetaxel in BRCA1/2 Mutated and Pre-Specified Triple Negative Breast Cancer “BRCAness” Subgroups: The TNT Trial. *Nat Med* **2018**, *24*, 628, doi:10.1038/S41591-018-0009-7.
 91. Helleday, T. The Underlying Mechanism for the PARP and BRCA Synthetic Lethality: Clearing up the Misunderstandings. *Mol Oncol* **2011**, *5*, 387, doi:10.1016/J.MOLONC.2011.07.001.
 92. Neiger, H.E.; Siegler, E.L.; Shi, Y. Breast Cancer Predisposition Genes and Synthetic Lethality. *International Journal of Molecular Sciences* **2021**, *Vol. 22*, Page 5614 **2021**, *22*, 5614, doi:10.3390/IJMS22115614.
 93. Lynparza | European Medicines Agency (EMA) Available online: <https://www.ema.europa.eu/en/medicines/human/EPAR/lynparza> (accessed on 12 August 2024).
 94. Rubraca | European Medicines Agency (EMA) Available online: <https://www.ema.europa.eu/en/medicines/human/EPAR/rubraca> (accessed on 12 August 2024).
 95. Zejula | European Medicines Agency (EMA) Available online: <https://www.ema.europa.eu/en/medicines/human/EPAR/zejula> (accessed on 12 August 2024).
 96. Akeega | European Medicines Agency (EMA) Available online: <https://www.ema.europa.eu/en/medicines/human/EPAR/akeega> (accessed on 12 August 2024).
 97. Talzenna | European Medicines Agency (EMA) Available online: <https://www.ema.europa.eu/en/>

- medicines/human/EPAR/talzenna (accessed on 12 August 2024).
98. Rehm, H.L.; Alaimo, J.T.; Aradhya, S.; Bayrak-Toydemir, P.; Best, H.; Brandon, R.; Buchan, J.G.; Chao, E.C.; Chen, E.; Clifford, J.; et al. The Landscape of Reported VUS in Multi-Gene Panel and Genomic Testing: Time for a Change. *Genetics in Medicine* **2023**, *25*, 100947, doi:10.1016/J.GIM.2023.100947.
 99. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E.; 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, doi:10.1038/GIM.2015.30.
 100. Millot, G.A.; Carvalho, M.A.; Caputo, S.M.; Vreeswijk, M.P.G.; Brown, M.A.; Webb, M.; Rouleau, E.; Neuhausen, S.L.; Hansen, T. v O.; Galli, A.; et al. A Guide for Functional Analysis of BRCA1 Variants of Uncertain Significance. *Hum Mutat* **2012**, *33*, 1526–1537, doi:10.1002/humu.22150.
 101. Fowler, D.M.; Rehm, H.L. Will Variants of Uncertain Significance Still Exist in 2030? *The American Journal of Human Genetics* **2024**, *111*, 5–10, doi:10.1016/J.AJHG.2023.11.005.
 102. Federici, G.; Soddu, S. Variants of Uncertain Significance in the Era of High-Throughput Genome Sequencing: A Lesson from Breast and Ovary Cancers. *Journal of Experimental and Clinical Cancer Research* **2020**, *39*, 1–12, doi:10.1186/S13046-020-01554-6/FIGURES/1.
 103. Cotton, R.G.H.; Scriver, C.R. Proof of “Disease Causing” Mutation. *Hum Mutat* **1998**, *12*, 1–3.
 104. Kazazian, H.H.; Boehm, C.D.; Seltzer, W.K. ACMG Recommendations for Standards for Interpretation of Sequence Variations. *Genetics in Medicine* **2000**, *2*, 302–303, doi:10.1097/00125817-200009000-00009.
 105. Richards, C.S.; Bale, S.; Bellissimo, D.B.; Das, S.; Grody, W.W.; Hegde, M.R.; Lyon, E.; Ward, B.E. ACMG Recommendations for Standards for Interpretation and Reporting of Sequence Variations: Revisions 2007. *Genetics in Medicine* **2008**, *10*, 294–300, doi:10.1097/GIM.0B013E31816B5CAE.

106. Goldgar, D.E.; Easton, D.F.; Deffenbaugh, A.M.; Monteiro, A.N.A.; Tavtigian, S. V.; Couch, F.J. Integrated Evaluation of DNA Sequence Variants of Unknown Clinical Significance: Application to BRCA1 and BRCA2. *Am J Hum Genet* **2004**, *75*, 535–544, doi:10.1086/424388.
107. Chenevix-Trench, G.; Healey, S.; Lakhani, S.; Waring, P.; Cummings, M.; Brinkworth, R.; Deffenbaugh, A.M.; Burbidge, L.A.; Pruss, D.; Judkins, T.; et al. Genetic and Histopathologic Evaluation of BRCA1 and BRCA2 DNA Sequence Variants of Unknown Clinical Significance. *Cancer Res* **2006**, *66*, 2019–2027, doi:10.1158/0008-5472.CAN-05-3546.
108. Easton, D.F.; Deffenbaugh, A.M.; Pruss, D.; Frye, C.; Wenstrup, R.J.; Allen-Brady, K.; Tavtigian, S. V.; Monteiro, A.N.A.; Iversen, E.S.; Couch, F.J.; et al. A Systematic Genetic Assessment of 1,433 Sequence Variants of Unknown Clinical Significance in the BRCA1 and BRCA2 Breast Cancer-Predisposition Genes. *Am J Hum Genet* **2007**, *81*, 873–883, doi:10.1086/521032.
109. Goldgar, D.E.; Easton, D.F.; Byrnes, G.B.; Spurdle, A.B.; Iversen, E.S.; Greenblatt, M.S. Genetic Evidence and Integration of Various Data Sources for Classifying Uncertain Variants into a Single Model. *Hum Mutat* **2008**, *29*, 1265–1272, doi:10.1002/HUMU.20897.
110. Gómez García, E.B.; Oosterwijk, J.C.; Timmermans, M.; van Asperen, C.J.; Hogervorst, F.B.L.; Hoogerbrugge, N.; Oldenburg, R.; Verhoef, S.; Dommering, C.J.; Ausems, M.G.E.M.; et al. A Method to Assess the Clinical Significance of Unclassified Variants in the BRCA1 and BRCA2 Genes Based on Cancer Family History. *Breast Cancer Res* **2009**, *11*, doi:10.1186/BCR2223.
111. Spearman, A.D.; Sweet, K.; Zhou, X.P.; McLennan, J.; Couch, F.J.; Toland, A.E. Clinically Applicable Models to Characterize BRCA1 and BRCA2 Variants of Uncertain Significance. *J Clin Oncol* **2008**, *26*, 5393–5400, doi:10.1200/JCO.2008.17.8228.
112. Spurdle, A.B.; Lakhani, S.R.; Healey, S.; Parry, S.; Da Silva, L.M.; Brinkworth, R.; Hopper, J.L.; Brown, M.A.; Babikyan, D.; Chenevix-Trench, G.; et al. Clinical Classification of BRCA1 and BRCA2 DNA Sequence Variants: The Value of Cytokeratin Profiles and Evolutionary Analysis--a Report

- from the KConFab Investigators. *J Clin Oncol* **2008**, 26, 1657–1663, doi:10.1200/JCO.2007.13.2779.
113. Walker, L.C.; Whiley, P.J.; Couch, F.J.; Farrugia, D.J.; Healey, S.; Eccles, D.M.; Lin, F.; Butler, S.A.; Goff, S.A.; Thompson, B.A.; et al. Detection of Splicing Aberrations Caused by BRCA1 and BRCA2 Sequence Variants Encoding Missense Substitutions: Implications for Prediction of Pathogenicity. *Hum Mutat* **2010**, 31, doi:10.1002/HUMU.21267.
 114. Whiley, P.J.; Guidugli, L.; Walker, L.C.; Healey, S.; Thompson, B.A.; Lakhani, S.R.; Da Silva, L.M.; Investigators, K.; Tavtigian, S. V.; Goldgar, D.E.; et al. Splicing and Multifactorial Analysis of Intronic BRCA1 and BRCA2 Sequence Variants Identifies Clinically Significant Splicing Aberrations up to 12 Nucleotides from the Intron/Exon Boundary. *Hum Mutat* **2011**, 32, 678–687, doi:10.1002/HUMU.21495.
 115. Tavtigian, S. V.; Byrnes, G.B.; Goldgar, D.E.; Thomas, A. Classification of Rare Missense Substitutions, Using Risk Surfaces, with Genetic- and Molecular-Epidemiology Applications. *Hum Mutat* **2008**, 29, 1342–1354, doi:10.1002/HUMU.20896.
 116. Plon, S.E.; Eccles, D.M.; Easton, D.; Foulkes, W.D.; Genuardi, M.; Greenblatt, M.S.; Hogervorst, F.B.L.; Hoogerbrugge, N.; Spurdle, A.B.; Tavtigian, S. V. Sequence Variant Classification and Reporting: Recommendations for Improving the Interpretation of Cancer Susceptibility Genetic Test Results. *Hum Mutat* **2008**, 29, 1282–1291, doi:10.1002/HUMU.20880.
 117. Spurdle, A.B.; Healey, S.; Devereau, A.; Hogervorst, F.B.L.; Monteiro, A.N.A.; Nathanson, K.L.; Radice, P.; Stoppa-Lyonnet, D.; Tavtigian, S.; Wappenschmidt, B.; et al. ENIGMA - Evidence-Based Network for the Interpretation of Germline Mutant Alleles: An International Initiative to Evaluate Risk and Clinical Significance Associated with Sequence Variation in BRCA1 and BRCA2 Genes. *Hum Mutat* **2012**, 33, 2, doi:10.1002/HUMU.21628.
 118. Tavtigian, S. V.; Greenblatt, M.S.; Harrison, S.M.; Nussbaum, R.L.; Prabhu, S.A.; Boucher, K.M.; Biesecker, L.G. Modeling the ACMG/AMP Variant Classification Guidelines as a Bayesian Classification

- Framework. *Genet Med* **2018**, 20, 1054–1060, doi:10.1038/gim.2017.210.
119. Tavtigian, S. V.; Harrison, S.M.; Boucher, K.M.; Biesecker, L.G. Fitting a Naturally Scaled Point System to the ACMG/AMP Variant Classification Guidelines. *Hum Mutat* **2020**, 41, 1734, doi:10.1002/HUMU.24088.
 120. Rehm, H.L.; Berg, J.S.; Brooks, L.D.; Bustamante, C.D.; Evans, J.P.; Landrum, M.J.; Ledbetter, D.H.; Maglott, D.R.; Martin, C.L.; Nussbaum, R.L.; et al. ClinGen — The Clinical Genome Resource. *N Engl J Med* **2015**, 372, 2235, doi:10.1056/NEJMSR1406261.
 121. Rivera-Muñoz, E.A.; Milko, L. V.; Harrison, S.M.; Azzariti, D.R.; Kurtz, C.L.; Lee, K.; Mester, J.L.; Weaver, M.A.; Currey, E.; Craigen, W.; et al. ClinGen Variant Curation Expert Panel Experiences and Standardized Processes for Disease and Gene-Level Specification of the ACMG/AMP Guidelines for Sequence Variant Interpretation. *Hum Mutat* **2018**, 39, 1614–1622, doi:10.1002/HUMU.23645.
 122. Parsons, M.T.; Hoya, M. de la; Richardson, M.E.; Tadini, E.; Anderson, M.; Berkofsky-Fessler, W.; Caputo, S.M.; Chan, R.C.; Cline, M.S.; Feng, B.-J.; et al. Evidence-Based Recommendations for Gene-Specific ACMG/AMP Variant Classification from the ClinGen ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel. *The American Journal of Human Genetics* **2024**, 0, doi:10.1016/J.AJHG.2024.07.013.
 123. Rhine, C.L.; Cygan, K.J.; Soemedi, R.; Maguire, S.; Murray, M.F.; Monaghan, S.F.; Fairbrother, W.G. Hereditary Cancer Genes Are Highly Susceptible to Splicing Mutations. *PLoS Genet* **2018**, 14, e1007231, doi:10.1371/JOURNAL.PGEN.1007231.
 124. Rebbeck, T.R.; Friebel, T.M.; Friedman, E.; Hamann, U.; Huo, D.; Kwong, A.; Olah, E.; Olopade, O.I.; Solano, A.R.; Teo, S.H.; et al. Mutational Spectrum in a Worldwide Study of 29,700 Families with BRCA1 or BRCA2 Mutations. *Hum Mutat* **2018**, 39, 593–620, doi:10.1002/HUMU.23406.
 125. Spurdle, A.B.; Couch, F.J.; Hogervorst, F.B.L.; Radice, P.; Sinilnikova, O.M. Prediction and Assessment of Splicing Alterations: Implications for Clinical Testing. *Hum Mutat*

- 2008**, 29, 1304–1313, doi:10.1002/humu.20901.
126. Lord, J.; Baralle, D. Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Front Genet* **2021**, 12, 689892, doi:10.3389/fgene.2021.689892.
 127. Burset, M.; Seledtsov, I.A.; Solovyev, V. V. Analysis of Canonical and Non-Canonical Splice Sites in Mammalian Genomes. *Nucleic Acids Res* **2000**, 28, 4364–4375, doi:10.1093/NAR/28.21.4364.
 128. Sibley, C.R.; Blazquez, L.; Ule, J. Lessons from Non-Canonical Splicing. *Nature Reviews Genetics* 2016 17:7 **2016**, 17, 407–421, doi:10.1038/nrg.2016.46.
 129. Parada, G.E.; Munita, R.; Cerda, C.A.; Gysling, K. A Comprehensive Survey of Non-Canonical Splice Sites in the Human Transcriptome. *Nucleic Acids Res* **2014**, 42, 10564–10578, doi:10.1093/NAR/GKU744.
 130. Abramowicz, A.; Gos, M. Splicing Mutations in Human Genetic Disorders: Examples, Detection, and Confirmation. *J Appl Genet* **2018**, 59, 253–268, doi:10.1007/s13353-018-0444-7.
 131. Mercer, T.R.; Clark, M.B.; Andersen, S.B.; Brunck, M.E.; Haerty, W.; Crawford, J.; Taft, R.J.; Nielsen, L.K.; Dinger, M.E.; Mattick, J.S. Genome-Wide Discovery of Human Splicing Branchpoints. *Genome Res* **2015**, 25, 290–303, doi:10.1101/GR.182899.114.
 132. Cartegni, L.; Chew, S.L.; Krainer, A.R. Listening to Silence and Understanding Nonsense: Exonic Mutations That Affect Splicing. *Nature Reviews Genetics* 2002 3:4 **2002**, 3, 285–298, doi:10.1038/nrg775.
 133. Zhang, X.H.F.; Chasin, L.A. Computational Definition of Sequence Motifs Governing Constitutive Exon Splicing. *Genes Dev* **2004**, 18, 1241–1250, doi:10.1101/GAD.1195304.
 134. Ke, S.; Shang, S.; Kalachikov, S.M.; Morozova, I.; Yu, L.; Russo, J.J.; Ju, J.; Chasin, L.A. Quantitative Evaluation of All Hexamers as Exonic Splicing Elements. *Genome Res* **2011**, 21, 1360–1374, doi:10.1101/gr.119628.110.
 135. Baralle, D.; Baralle, M. Splicing in Action: Assessing Disease Causing Sequence Changes. *J Med Genet* **2005**, 42, 737–748, doi:10.1136/JMG.2004.029538.
 136. Ule, J.; Blencowe, B.J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* **2019**, 76,

- 329–345,
doi:10.1016/J.MOLCEL.2019.09.017.
137. Wahl, M.C.; Will, C.L.; Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **2009**, *136*, 701–718, doi:10.1016/J.CELL.2009.02.009.
 138. Baralle, F.E.; Giudice, J. Alternative Splicing as a Regulator of Development and Tissue Identity. *Nat Rev Mol Cell Biol* **2017**, *18*, 437–451, doi:10.1038/NRM.2017.27.
 139. Liu, Q.; Fang, L.; Wu, C. Alternative Splicing and Isoforms: From Mechanisms to Diseases. *Genes (Basel)* **2022**, *13*, doi:10.3390/GENES13030401.
 140. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* **2008**, *456*, 470–476, doi:10.1038/nature07509.
 141. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing. *Nat Genet* **2008**, *40*, 1413–1415, doi:10.1038/NG.259.
 142. Wright, C.J.; Smith, C.W.J.; Jiggins, C.D. Alternative Splicing as a Source of Phenotypic Diversity. *Nature Reviews Genetics* **2022** *23:11* **2022**, *23*, 697–710, doi:10.1038/s41576-022-00514-4.
 143. Sullivan, P.J.; Quinn, J.M.W.; Wu, W.; Pinese, M.; Cowley, M.J.; Cowley, M.J. SpliceVarDB: A Comprehensive Database of Experimentally Validated Human Splicing Variants. *The American Journal of Human Genetics* **2024**, *0*, doi:10.1016/J.AJHG.2024.08.002.
 144. Lázaro, C.; Lerner-Ellis, J.; Spurdle, A. *Clinical DNA Variant Interpretation: Theory and Practice*; Academic Press, Ed.; Elsevier Science, 2021; ISBN 9780128205204.
 145. Chang, Y.F.; Imam, J.S.; Wilkinson, M.F. The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu Rev Biochem* **2007**, *76*, 51–74, doi:10.1146/annurev.biochem.76.050106.093909.
 146. Wang, G.S.; Cooper, T.A. Splicing in Disease: Disruption of the Splicing Code and the Decoding Machinery. *Nat Rev*

- Genet* **2007**, 8, 749–761, doi:10.1038/NRG2164.
147. Kelemen, O.; Convertini, P.; Zhang, Z.; Wen, Y.; Shen, M.; Falaleeva, M.; Stamm, S. Function of Alternative Splicing. *Gene* **2013**, 514, 1–30, doi:10.1016/J.GENE.2012.07.083.
 148. Ward, A.J.; Cooper, T.A. The Pathobiology of Splicing. *J Pathol* **2010**, 220, 152–163, doi:10.1002/PATH.2649.
 149. Canson, D.; Glubb, D.; Spurdle, A.B. Variant Effect on Splicing Regulatory Elements, Branchpoint Usage, and Pseudoexonization: Strategies to Enhance Bioinformatic Prediction Using Hereditary Cancer Genes as Exemplars. *Hum Mutat* **2020**, 41, 1705–1721, doi:10.1002/HUMU.24074.
 150. Wai, H.A.; Lord, J.; Lyon, M.; Gunning, A.; Kelly, H.; Cibir, P.; Seaby, E.G.; Spiers-Fitzgerald, K.; Lye, J.; Ellard, S.; et al. Blood RNA Analysis Can Increase Clinical Diagnostic Rate and Resolve Variants of Uncertain Significance. *Genetics in Medicine* **2020**, 22, 1005–1014, doi:10.1038/s41436-020-0766-9.
 151. Moles-Fernandez, A.; Duran-Lozano, L.; Montalban, G.; Bonache, S.; Lopez-Perolio, I.; Menendez, M.; Santamarina, M.; Behar, R.; Blanco, A.; Carrasco, E.; 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, doi:10.3389/fgene.2018.00366.
 152. Dhir, A.; Buratti, E. Alternative Splicing: Role of Pseudoexons in Human Disease and Potential Therapeutic Strategies. *FEBS J* **2010**, 277, 841–855, doi:10.1111/J.1742-4658.2009.07520.X.
 153. Sironi, M.; Menozzi, G.; Riva, L.; Cagliani, R.; Comi, G.P.; Bresolin, N.; Giorda, R.; Pozzoli, U. Silencer Elements as Possible Inhibitors of Pseudoexon Splicing. *Nucleic Acids Res* **2004**, 32, 1783–1791, doi:10.1093/NAR/GKH341.
 154. Pozzoli, U.; Sironi, M. Silencers Regulate Both Constitutive and Alternative Splicing Events in Mammals. *Cellular and Molecular Life Sciences* **2005**, 62, 1579–1604, doi:10.1007/S00018-005-5030-6/METRICS.
 155. Vaz-Drago, R.; Custódio, N.; Carmo-Fonseca, M. Deep Intronic Mutations and Human Disease. *Hum Genet* **2017**, 136, 1093–1111,

- doi:10.1007/S00439-017-1809-4.
156. Romano, M.; Buratti, E.; Baralle, D. Role of Pseudoexons and Pseudointrons in Human Cancer. *Int J Cell Biol* **2013**, *2013*, 810572, doi:10.1155/2013/810572.
 157. Montalban, G.; Bonache, S.; Moles-Fernández, A.; Gisbert-Beamud, A.; Tenés, A.; Bach, V.; Carrasco, E.; López-Fernández, A.; Stjepanovic, N.; Balmaña, J.; et al. Screening of BRCA1/2 Deep Intronic Regions by Targeted Gene Sequencing Identifies the First Germline BRCA1 Variant Causing Pseudoexon Activation in a Patient with Breast/Ovarian Cancer. *J Med Genet* **2019**, *56*, 63–74, doi:10.1136/jmedgenet-2018-105606.
 158. Paganil, F.; Burattil, E.; Stuani, C.; Bendix, R.; Dörk, T.; Baralle, F.E. A New Type of Mutation Causes a Splicing Defect in ATM. *Nat Genet* **2002**, *30*, 426–429, doi:10.1038/NG858.
 159. Liu, H.X.; Cartegni, L.; Zhang, M.Q.; Krainer, A.R. A Mechanism for Exon Skipping Caused by Nonsense or Missense Mutations in BRCA1 and Other Genes. *Nat Genet* **2001**, *27*, 55–58, doi:10.1038/83762.
 160. Goina, E.; Skoko, N.; Pagani, F. Binding of DAZAP1 and HnRNPA1/A2 to an Exonic Splicing Silencer in a Natural BRCA1 Exon 18 Mutant. *Mol Cell Biol* **2008**, *28*, 3850–3860, doi:10.1128/mcb.02253-07.
 161. Fraile-Bethencourt, E.; Díez-Gómez, B.; Velásquez-Zapata, V.; Acedo, A.; Sanz, D.J.; Velasco, E.A. Functional Classification of DNA Variants by Hybrid Minigenes: Identification of 30 Spliceogenic Variants of BRCA2 Exons 17 and 18. *PLoS Genet* **2017**, *13*, doi:10.1371/JOURNAL.PGEN.1006691.
 162. Hartmann, L.; Theiss, S.; Niederacher, D.; Schaal, H. Diagnostics of Pathogenic Splicing Mutations: Does Bioinformatics Cover All Bases? *Frontiers in Bioscience* **2008**, *13*, 3252–3272, doi:10.2741/2924/PDF.
 163. Rowlands, C.F.; Baralle, D.; Ellingford, J.M. Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. *Cells* **2019**, *Vol. 8*, Page 1513 **2019**, *8*, 1513, doi:10.3390/CELLS8121513.
 164. Cartegni, L.; Wang, J.; Zhu, Z.; Zhang, M.Q.; Krainer, A.R. ESEfinder: A Web Resource to Identify Exonic Splicing

- Enhancers. *Nucleic Acids Res* **2003**, *31*, 3568–3571, doi:10.1093/nar/gkg616.
165. Desmet, F.O.; Hamroun, D.; Lalande, M.; Collod-B  roud, G.; Claustres, M.; B  roud, C. Human Splicing Finder: An Online Bioinformatics Tool to Predict Splicing Signals. *Nucleic Acids Res* **2009**, *37*, e67–e67, doi:10.1093/NAR/GKP215.
 166. Ohno, K.; Takeda, J.I.; Masuda, A. Rules and Tools to Predict the Splicing Effects of Exonic and Intronic Mutations. *Wiley Interdiscip Rev RNA* **2018**, *9*, e1451, doi:10.1002/WRNA.1451.
 167. Leman, R.; Gaildrat, P.; Gac, G.L.; Ka, C.; Fichou, Y.; Audrezet, M.P.; Caux-Moncoutier, V.; Caputo, S.M.; Boutry-Kryza, N.; L  one, M.; et al. Novel Diagnostic Tool for Prediction of Variant Spliceogenicity Derived from a Set of 395 Combined in Silico/in Vitro Studies: An International Collaborative Effort. *Nucleic Acids Res* **2018**, *46*, 7913–7923, doi:10.1093/NAR/GKY372.
 168. Jaganathan, K.; Kyriazopoulou Panagiotopoulou, S.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **2019**, *176*, 535–548.e24, doi:10.1016/j.cell.2018.12.015.
 169. Rowlands, C.; Thomas, H.B.; Lord, J.; Wai, H.A.; Arno, G.; Beaman, G.; Sergouniotis, P.; Gomes-Silva, B.; Campbell, C.; Gossan, N.; et al. Comparison of in Silico Strategies to Prioritize Rare Genomic Variants Impacting RNA Splicing for the Diagnosis of Genomic Disorders. *Sci Rep* **2021**, *11*, doi:10.1038/S41598-021-99747-2.
 170. Walker, L.C.; Hoya, M. de la; Wiggins, G.A.R.; Lindy, A.; Vincent, L.M.; Parsons, M.T.; Canson, D.M.; Bis-Brewer, D.; Cass, A.; Tchourbanov, A.; et al. 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* **2023**, *110*, 1046–1067, doi:10.1016/J.AJHG.2023.06.002.
 171. Rosenberg, A.B.; Patwardhan, R.P.; Shendure, J.; Seelig, G. Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. *Cell* **2015**, *163*, 698–711, doi:10.1016/j.cell.2015.09.054.

172. Tubeuf, H.; Charbonnier, C.; Soukarieh, O.; Blavier, A.; Lefebvre, A.; Dauchel, H.; Frebourg, T.; Gaildrat, P.; Martins, A. Large-Scale Comparative Evaluation of User-Friendly Tools for Predicting Variant-Induced Alterations of Splicing Regulatory Elements. *Hum Mutat* **2020**, *41*, 1811–1829, doi:10.1002/humu.24091.
173. Zhang, Q.; Fan, X.; Wang, Y.; Sun, M.A.; Shao, J.; Guo, D. BPP: A Sequence-Based Algorithm for Branch Point Prediction. *Bioinformatics* **2017**, *33*, 3166–3172, doi:10.1093/BIOINFORMATICS/BTX401.
174. Leman, R.; Tubeuf, H.; Raad, S.; Tournier, I.; Derambure, C.; Lanos, R.; Gaildrat, P.; Castelain, G.; Hauchard, J.; Killian, A.; et al. Assessment of Branch Point Prediction Tools to Predict Physiological Branch Points and Their Alteration by Variants. *BMC Genomics* **2020**, *21*, 1–12, doi:10.1186/s12864-020-6484-5.
175. Yeo, G.; Holste, D.; Kreiman, G.; Burge, C.B. Variation in Alternative Splicing across Human Tissues. *Genome Biol* **2004**, *5*, 1–15, doi:10.1186/gb-2004-5-10-r74.
176. Colombo, M.; Blok, M.J.; Wilely, P.; Santamariñ, M.; Gutiérrez-Enríquez, S.; Romero, A.; Garre, P.; Becker, A.; Smith, L.D.; De Vecchi, G.; et al. Comprehensive Annotation of Splice Junctions Supports Pervasive Alternative Splicing at the BRCA1 Locus: A Report from the ENIGMA Consortium. *Hum Mol Genet* **2014**, *23*, 3666–3680, doi:10.1093/hmg/ddu075.
177. Fackenthal, J.D.; Yoshimatsu, T.; Zhang, B.; de Garibay, G.R.; Colombo, M.; De Vecchi, G.; Ayoub, S.C.; Lal, K.; Olopade, O.I.; Vega, A.; et al. Naturally Occurring BRCA2 Alternative MRNA Splicing Events in Clinically Relevant Samples. *J Med Genet* **2016**, *53*, 548–558, doi:10.1136/jmedgenet-2015-103570.
178. Montalban, G.; Bonache, S.; Moles-Fernández, A.; Gadea, N.; Tenés, A.; Torres-Esquius, S.; Carrasco, E.; Balmaña, J.; Díez, O.; Gutiérrez-Enríquez, S. Incorporation of Semi-Quantitative Analysis of Splicing Alterations for the Clinical Interpretation of Variants in BRCA1 and BRCA2 Genes. *Hum Mutat* **2019**, *40*, 2296–2317, doi:10.1002/humu.23882.
179. Wilely, P.J.; de la Hoya, M.; Thomassen, M.; Becker, A.; Brandao, R.; Pedersen, I.S.; Montagna, M.; Menendez, M.;

- Quiles, F.; Gutierrez-Enriquez, S.; et al. Comparison of MRNA Splicing Assay Protocols across Multiple Laboratories: Recommendations for Best Practice in Standardized Clinical Testing. *Clin Chem* **2014**, *60*, 341–352, doi:10.1373/clinchem.2013.210658.
180. Walker, L.C.; Whiley, P.J.; Houdayer, C.; Hansen, T.V.O.; Vega, A.; Santamarina, M.; Blanco, A.; Fachal, L.; Southey, M.C.; Lafferty, A.; et al. Evaluation of a 5-Tier Scheme Proposed for Classification of Sequence Variants Using Bioinformatic and Splicing Assay Data: Inter-Reviewer Variability and Promotion of Minimum Reporting Guidelines. *Hum Mutat* **2013**, *34*, 1424–1431, doi:10.1002/HUMU.22388.
181. Bournazos, A.M.; Riley, L.G.; Bommireddipalli, S.; Ades, L.; Akesson, L.S.; Al-Shinnag, M.; Alexander, S.I.; Archibald, A.D.; Balasubramaniam, S.; Berman, Y.; et al. Standardized Practices for RNA Diagnostics Using Clinically Accessible Specimens Reclassifies 75% of Putative Splicing Variants. *Genetics in Medicine* **2022**, *24*, 130–145, doi:10.1016/j.gim.2021.09.001.
182. Montalban, G.; Fraile-Bethencourt, E.; López-Perolio, I.; Pérez-Segura, P.; Infante, M.; Durán, M.; Alonso-Cerezo, M.C.; López-Fernández, A.; Diez, O.; de la Hoya, M.; et al. Characterization of Spliceogenic Variants Located in Regions Linked to High Levels of Alternative Splicing: BRCA2 c.7976+5G > T as a Case Study. *Hum Mutat* **2018**, *39*, 1155–1160, doi:10.1002/HUMU.23583.
183. Duran-Lozano, L.; Montalban, G.; Bonache, S.; Moles-Fernández, A.; Tenés, A.; Castroviejo-Bermejo, M.; Carrasco, E.; López-Fernández, A.; Torres-Esquius, S.; Gadea, N.; et al. Alternative Transcript Imbalance Underlying Breast Cancer Susceptibility in a Family Carrying PALB2 c.3201+5G>T. *Breast Cancer Res Treat* **2019**, *174*, 543–550, doi:10.1007/s10549-018-05094-8.
184. Houdayer, C.; Caux-Moncoutier, V.; Krieger, S.; Barrois, M.; Bonnet, F.; Bourdon, V.; Bronner, M.; Buisson, M.; Coulet, F.; Gaildrat, P.; et al. Guidelines for Splicing Analysis in Molecular Diagnosis Derived from a Set of 327 Combined in Silico/in Vitro Studies on BRCA1 and BRCA2 Variants. *Hum Mutat* **2012**, *33*,

- 1228–1238,
doi:10.1002/HUMU.22101.
185. Gaildrat, P.; Killian, A.; Martins, A.; Tournier, I.; Frébourg, T.; Tosi, M. Use of Splicing Reporter Minigene Assay to Evaluate the Effect on Splicing of Unclassified Genetic Variants. *Methods Mol Biol* **2010**, *653*, 249–257, doi:10.1007/978-1-60761-759-4_15.
 186. Sharma, N.; Sosnay, P.R.; Ramalho, A.S.; Douville, C.; Franca, A.; Gottschalk, L.B.; Park, J.; Lee, M.; Vecchio-Pagan, B.; Raraigh, K.S.; et al. Experimental Assessment of Splicing Variants Using Expression Minigenes and Comparison with In Silico Predictions. *Hum Mutat* **2014**, *35*, 1249–1259, doi:10.1002/HUMU.22624.
 187. Dionnet, E.; Defour, A.; Da Silva, N.; Salvi, A.; Lévy, N.; Krahn, M.; Bartoli, M.; Puppo, F.; Gorokhova, S. Splicing Impact of Deep Exonic Missense Variants in CAPN3 Explored Systematically by Minigene Functional Assay. *Hum Mutat* **2020**, *41*, 1797–1810, doi:10.1002/HUMU.24083.
 188. Tubeuf, H.; Caputo, S.M.; Sullivan, T.; Rondeaux, J.; Krieger, S.; Caux-Moncoutier, V.; Hauchard, J.; Castelain, G.; Fievèt, A.; Meulemans, L.; et al. Calibration of Pathogenicity Due to Variant-Induced Leaky Splicing Defects by Using BRCA2 Exon 3 as a Model System. *Cancer Res* **2020**, *80*, 3593–3605, doi:10.1158/0008-5472.CAN-20-0895.
 189. Kremer, L.S.; Bader, D.M.; Mertes, C.; Kopajtich, R.; Pichler, G.; Iuso, A.; Haack, T.B.; Graf, E.; Schwarzmayer, T.; Terrile, C.; et al. Genetic Diagnosis of Mendelian Disorders via RNA Sequencing. *Nature Communications 2017 8:1* **2017**, *8*, 1–11, doi:10.1038/ncomms15824.
 190. Cummings, B.B.; Marshall, J.L.; Tukiainen, T.; Lek, M.; Donkervoort, S.; Foley, A.R.; Bolduc, V.; Waddell, L.B.; Sandaradura, S.A.; O’Grady, G.L.; et al. Improving Genetic Diagnosis in Mendelian Disease with Transcriptome Sequencing. *Sci Transl Med* **2017**, *9*, doi:10.1126/scitranslmed.aal5209.
 191. Brandão, R.D.; Mensaert, K.; López-Perolio, I.; Tserpelis, D.; Xenakis, M.; Lattimore, V.; Walker, L.C.; Kvist, A.; Vega, A.; Gutiérrez-Enríquez, S.; et al. Targeted RNA-Seq Successfully Identifies Normal and Pathogenic Splicing Events in Breast/Ovarian Cancer Susceptibility and Lynch

- Syndrome Genes. *Int J Cancer* **2019**, *145*, 401–414, doi:10.1002/IJC.32114.
192. Abou Tayoun, A.N.; Pesaran, T.; DiStefano, M.T.; Oza, A.; Rehm, H.L.; Biesecker, L.G.; Harrison, S.M. Recommendations for Interpreting the Loss of Function PVS1 ACMG/AMP Variant Criterion. *Hum Mutat* **2018**, *39*, 1517–1524, doi:10.1002/HUMU.23626.
 193. Colombo, M.; Lòpez-Perolio, I.; Meeks, H.D.; Caleca, L.; Parsons, M.T.; Li, H.; De Vecchi, G.; Tudini, E.; Foglia, C.; Mondini, P.; et al. The BRCA2 c.68-7T > A Variant Is Not Pathogenic: A Model for Clinical Calibration of Spliceogenicity. *Hum Mutat* **2018**, *39*, 729–741, doi:10.1002/HUMU.23411.
 194. de la Hoya, M.; Soukarieh, O.; Lòpez-Perolio, I.; Vega, A.; Walker, L.C.; van Ierland, Y.; Baralle, D.; Santamariña, M.; Lattimore, V.; Wijnen, J.; et al. Combined Genetic and Splicing Analysis of BRCA1 c.[594-2A>C; 641A>G] Highlights the Relevance of Naturally Occurring in-Frame Transcripts for Developing Disease Gene Variant Classification Algorithms. *Hum Mol Genet* **2016**, *25*, 2256, doi:10.1093/HMG/DDW094.
 195. Knudson, A.G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc Natl Acad Sci U S A* **1971**, *68*, 820–823, doi:10.1073/PNAS.68.4.820.
 196. Rebbeck, T.R.; Mitra, N.; Wan, F.; Sinilnikova, O.M.; Healey, S.; McGuffog, L.; Chenevix-Trench, G.; Easton, D.F.; Antoniou, A.C.; Nathanson, K.L.; et al. Association of Type and Location of BRCA1 and BRCA2 Mutations With Risk of Breast and Ovarian Cancer. *JAMA* **2015**, *313*, 1347–1361, doi:10.1001/JAMA.2014.5985.
 197. Loibl, S.; Poortmans, P.; Morrow, M.; Denkert, C.; Curigliano, G. Breast Cancer. *The Lancet* **2021**, *397*, 1750–1769, doi:10.1016/S0140-6736(20)32381-3.
 198. Spurdle, A.B.; Couch, F.J.; Parsons, M.T.; McGuffog, L.; Barrowdale, D.; Bolla, M.K.; Wang, Q.; Healey, S.; Schmutzler, R.K.; Wappenschmidt, B.; et al. Refined Histopathological Predictors of BRCA1 and BRCA2 mutation Status: A Large-Scale Analysis of Breast Cancer Characteristics from the BCAC, CIMBA, and ENIGMA Consortia. *Breast Cancer Res* **2014**, *16*, doi:10.1186/S13058-014-0474-Y.

199. O'Mahony, D.G.; Ramus, S.J.; Southey, M.C.; Meagher, N.S.; Hadjisavvas, A.; John, E.M.; Hamann, U.; Imyanitov, E.N.; Andrulis, I.L.; Sharma, P.; et al. Ovarian Cancer Pathology Characteristics as Predictors of Variant Pathogenicity in BRCA1 and BRCA2. *Br J Cancer* **2023**, *128*, 2283–2294, doi:10.1038/S41416-023-02263-5.
200. Hashizume, R.; Fukuda, M.; Maeda, I.; Nishikawa, H.; Oyake, D.; Yabuki, Y.; Ogata, H.; Ohta, T. The RING Heterodimer BRCA1-BARD1 Is a Ubiquitin Ligase Inactivated by a Breast Cancer-Derived Mutation. *Journal of Biological Chemistry* **2001**, *276*, 14537–14540, doi:10.1074/jbc.C000881200.
201. Witus, S.R.; Stewart, M.D.; Klevit, R.E. The BRCA1/BARD1 Ubiquitin Ligase and Its Substrates. *Biochem J* **2021**, *478*, 3467, doi:10.1042/BCJ20200864.
202. Joukov, V.; Chen, J.; Fox, E.A.; Green, J.B.A.; Livingston, D.M. Functional Communication between Endogenous BRCA1 and Its Partner, BARD1, during *Xenopus Laevis* Development. *Proc Natl Acad Sci U S A* **2001**, *98*, 12078–12083, doi:10.1073/pnas.211427098.
203. Fabbro, M.; Rodriguez, J.A.; Baer, R.; Henderson, B.R. BARD1 Induces BRCA1 Intranuclear Foci Formation by Increasing RING-Dependent BRCA1 Nuclear Import and Inhibiting BRCA1 Nuclear Export. *Journal of Biological Chemistry* **2002**, *277*, 21315–21324, doi:10.1074/JBC.M200769200.
204. Sy, S.M.H.; Huen, M.S.Y.; Chen, J. PALB2 Is an Integral Component of the BRCA Complex Required for Homologous Recombination Repair. *Proc Natl Acad Sci U S A* **2009**, *106*, 7155–7160, doi:10.1073/pnas.0811159106.
205. Zhang, F.; Ma, J.; Wu, J.; Ye, L.; Cai, H.; Xia, B.; Yu, X. PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. *Curr Biol* **2009**, *19*, 524–529, doi:10.1016/J.CUB.2009.02.018.
206. Zhang, F.; Fan, Q.; Ren, K.; Andreassen, P.R. PALB2 Functionally Connects the Breast Cancer Susceptibility Proteins BRCA1 and BRCA2. *Mol Cancer Res* **2009**, *7*, 1110–1118, doi:10.1158/1541-7786.MCR-09-0123.
207. Xu, B.; Kim, S.; Kastan, M.B. Involvement of Brca1 in S-Phase and G₂-Phase Checkpoints after Ionizing Irradiation. *Mol Cell Biol* **2001**, *21*, 3445–3450,

- doi:10.1128/mcb.21.10.3445-3450.2001.
208. Cortez, D.; Wang, Y.; Qin, J.; Elledge, S.J. Requirement of ATM-Dependent Phosphorylation of Brca1 in the DNA Damage Response to Double-Strand Breaks. *Science* **1999**, *286*, 1162–1166, doi:10.1126/SCIENCE.286.5442.1162.
 209. Cantor, S.B.; Bell, D.W.; Ganesan, S.; Kass, E.M.; Drapkin, R.; Grossman, S.; Wahrer, D.C.R.; Sgroi, D.C.; Lane, W.S.; Haber, D.A.; et al. BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function. *Cell* **2001**, *105*, 149–160, doi:10.1016/S0092-8674(01)00304-X.
 210. Yu, X.; Wu, L.C.; Bowcock, A.M.; Aronheim, A.; Baer, R. The C-Terminal (BRCT) Domains of BRCA1 Interact in Vivo with CtIP, a Protein Implicated in the CtBP Pathway of Transcriptional Repression. *J Biol Chem* **1998**, *273*, 25388–25392, doi:10.1074/JBC.273.39.25388.
 211. Toland, A.E.; Andreassen, P.R. DNA Repair-Related Functional Assays for the Classification of BRCA1 and BRCA2 Variants: A Critical Review and Needs Assessment. *J Med Genet* **2017**, *54*, 721–731, doi:10.1136/JMEDGENET-2017-104707.
 212. Whiley, P.J.; Parsons, M.T.; Leary, J.; Tucker, K.; Warwick, L.; Dopita, B.; Thorne, H.; Lakhani, S.R.; Goldgar, D.E.; Brown, M.A.; et al. Multifactorial Likelihood Assessment of BRCA1 and BRCA2 Missense Variants Confirms That BRCA1:C.122A>G(p.His41Arg) Is a Pathogenic Mutation. *PLoS One* **2014**, *9*, doi:10.1371/JOURNAL.PONE.0086836.
 213. Woods, N.T.; Baskin, R.; Golubeva, V.; Jhuraney, A.; De-Gregoriis, G.; Vaclova, T.; Goldgar, D.E.; Couch, F.J.; Carvalho, M.A.; Iversen, E.S.; et al. Functional Assays Provide a Robust Tool for the Clinical Annotation of Genetic Variants of Uncertain Significance. *NPJ Genom Med* **2016**, *1*, doi:10.1038/NPJGENMED.2016.1.
 214. Lindor, N.M.; Guidugli, L.; Wang, X.; Vallée, M.P.; Monteiro, A.N.A.; Tavtigian, S.; Goldgar, D.E.; Couch, F.J. A Review of a Multifactorial Probability-Based Model for Classification of BRCA1 and BRCA2 Variants of Uncertain Significance (VUS). *Hum Mutat* **2012**, *33*, 8–21, doi:10.1002/HUMU.21627.

215. Dines, J.N.; Shirts, B.H.; Slavin, T.P.; Walsh, T.; King, M.C.; Fowler, D.M.; Pritchard, C.C. Systematic Misclassification of Missense Variants in BRCA1 and BRCA2 “Coldspots.” *Genetics in Medicine* **2020**, *22*, 825, doi:10.1038/S41436-019-0740-6.
216. Xia, B.; Sheng, Q.; Nakanishi, K.; Ohashi, A.; Wu, J.; Christ, N.; Liu, X.; Jasin, M.; Couch, F.J.; Livingston, D.M. Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. *Mol Cell* **2006**, *22*, 719–729, doi:10.1016/j.molcel.2006.05.022.
217. Bignell, G.; Micklem, G.; Stratton, M.R.; Ashworth, A.; Wooster, R. The BRC Repeats Are Conserved in Mammalian BRCA2 Proteins. *Hum Mol Genet* **1997**, *6*, 53–58, doi:10.1093/HMG/6.1.53.
218. Chen, P.L.; Chen, C.F.; Chen, Y.; Xiao, J.; Sharp, Z.D.; Lee, W.H. The BRC Repeats in BRCA2 Are Critical for RAD51 Binding and Resistance to Methyl Methanesulfonate Treatment. *Proc Natl Acad Sci U S A* **1998**, *95*, 5287–5292, doi:10.1073/pnas.95.9.5287.
219. Yang, H.; Li, Q.; Fan, J.; Holloman, W.K.; Pavletich, N.P. The BRCA2 Homologue Brh2 Nucleates RAD51 Filament Formation at a DsDNA-SsDNA Junction. *Nature* **2005**, *433*, 653–657, doi:10.1038/NATURE03234.
220. Li, J.; Zou, C.; Bai, Y.; Wazer, D.E.; Band, V.; Gao, Q. DSS1 Is Required for the Stability of BRCA2. *Oncogene* **2006**, *25*, 1186–1194, doi:10.1038/SJ.ONC.1209153.
221. Kristensen, C.N.; Bystol, K.M.; Li, B.; Serrano, L.; Brenneman, M.A. Depletion of DSS1 Protein Disables Homologous Recombinational Repair in Human Cells. *Mutat Res* **2010**, *694*, 60–64, doi:10.1016/J.MRFMMM.2010.08.007.
222. Davies, O.R.; Pellegrini, L. Interaction with the BRCA2 C-Terminus Protects RAD51–DNA Filaments from Disassembly by BRC Repeats. *Nat Struct Mol Biol* **2007**, *14*, 475, doi:10.1038/NSMB1251.
223. Jensen, R.B.; Carreira, A.; Kowalczykowski, S.C. Purified Human BRCA2 Stimulates RAD51-Mediated Recombination. *Nature* **2010**, *467*:7316 **2010**, *467*, 678–683, doi:10.1038/nature09399.
224. Moynahan, M.E.; Jasin, M. Mitotic Homologous Recombination Maintains

- Genomic Stability and Suppresses Tumorigenesis. *Nat Rev Mol Cell Biol* **2010**, *11*, 196–207, doi:10.1038/NRM2851.
225. Yuan, S.-S.F.; Lee, S.-Y.; Chen, G.; Song, M.; Tomlinson, G.E.; Lee, E.Y.-H.P. BRCA2 Is Required for Ionizing Radiation-Induced Assembly of Rad51 Complex in Vivo 1. *Cancer Res* **1999**, *59*, 3547–3551.
 226. Caputo, S.M.; Léone, M.; Damiola, F.; Ehlen, A.; Carreira, A.; Gaidrat, P.; Martins, A.; Brandão, R.D.; Peixoto, A.; Vega, A.; et al. Full In-Frame Exon 3 Skipping of BRCA2 Confers High Risk of Breast and/or Ovarian Cancer. *Oncotarget* **2018**, *9*, 17334–17348, doi:10.18632/ONCOTARGET.24671.
 227. Wu, S.; Zhou, J.; Zhang, K.; Chen, H.; Luo, M.; Lu, Y.; Sun, Y.; Chen, Y. Molecular Mechanisms of PALB2 Function and Its Role in Breast Cancer Management. *Front Oncol* **2020**, *10*, 301, doi:10.3389/FONC.2020.00301.
 228. Boonen, R.A.C.M.; Rodrigue, A.; Stoepker, C.; Wiegant, W.W.; Vroiling, B.; Sharma, M.; Rother, M.B.; Celosse, N.; Vreeswijk, M.P.G.; Couch, F.; et al. Functional Analysis of Genetic Variants in the High-Risk Breast Cancer Susceptibility Gene PALB2. *Nature Communications* **2019** *10*:1 **2019**, *10*, 1–15, doi:10.1038/s41467-019-13194-2.
 229. Jackson, S.P.; Bartek, J. The DNA-Damage Response in Human Biology and Disease. *Nature* **2009** *461*:7267 **2009**, *461*, 1071–1078, doi:10.1038/nature08467.
 230. Guidugli, L.; Carreira, A.; Caputo, S.M.; Ehlen, A.; Galli, A.; Monteiro, A.N.A.; Neuhausen, S.L.; Hansen, T.V.O.; Couch, F.J.; Vreeswijk, M.P.G. Functional Assays for Analysis of Variants of Uncertain Significance in BRCA2. *Hum Mutat* **2014**, *35*, 151–164, doi:10.1002/HUMU.22478.
 231. Kinner, A.; Wu, W.; Staudt, C.; Iliakis, G. γ -H2AX in Recognition and Signaling of DNA Double-Strand Breaks in the Context of Chromatin. *Nucleic Acids Res* **2008**, *36*, 5678, doi:10.1093/NAR/GKN550.
 232. Nguyen, L.; W. M. Martens, J.; Van Hoeck, A.; Cuppen, E. Pan-Cancer Landscape of Homologous Recombination Deficiency. *Nature Communications* **2020** *11*:1 **2020**, *11*, 1–12, doi:10.1038/s41467-020-19406-4.
 233. Davies, H.; Glodzik, D.; Morganella, S.; Yates, L.R.;

- Staaf, J.; Zou, X.; Ramakrishna, M.; Martin, S.; Boyault, S.; Sieuwerts, A.M.; et al. HRDetect Is a Predictor of BRCA1 and BRCA2 Deficiency Based on Mutational Signatures. *Nat Med* **2017**, *23*, 517–525, doi:10.1038/NM.4292.
234. Polak, P.; Kim, J.; Braunstein, L.Z.; Karlic, R.; Haradhavala, N.J.; Tiao, G.; Rosebrock, D.; Livitz, D.; Kübler, K.; Mouw, K.W.; et al. A Mutational Signature Reveals Alterations Underlying Deficient Homologous Recombination Repair in Breast Cancer. *Nat Genet* **2017**, *49*, 1476, doi:10.1038/NG.3934.
235. Sfeir, A.; Symington, L.S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci* **2015**, *40*, 701–714, doi:10.1016/J.TIBS.2015.08.006.
236. Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C.; et al. Landscape of Somatic Mutations in 560 Breast Cancer Whole-Genome Sequences. *Nature* **2016**, *534*, 47–54, doi:10.1038/NATURE17676.
237. Fong, P.C.; Boss, D.S.; Yap, T.A.; Tutt, A.; Wu, P.; Mergui-Roelvink, M.; Mortimer, P.; Swaisland, H.; Lau, A.; O'Connor, M.J.; et al. Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. *N Engl J Med* **2009**, *361*, 123–134, doi:10.1056/NEJMOA0900212.
238. Murai, J.; Pommier, Y. BRCAness, Homologous Recombination Deficiencies, and Synthetic Lethality. *Cancer Res* **2023**, *83*, 1173–1174, doi:10.1158/0008-5472.CAN-23-0628.
239. Stewart, M.D.; Vega, D.M.; Arend, R.C.; Baden, J.F.; Barbash, O.; Beaubier, N.; Collins, G.; French, T.; Ghahramani, N.; Hinson, P.; et al. Homologous Recombination Deficiency: Concepts, Definitions, and Assays. *Oncologist* **2022**, *27*, 167–174, doi:10.1093/oncolo/oyab053.
240. Abkevich, V.; Timms, K.M.; Hennessy, B.T.; Potter, J.; Carey, M.S.; Meyer, L.A.; Smith-McCune, K.; Broadus, R.; Lu, K.H.; Chen, J.; et al. Patterns of Genomic Loss of Heterozygosity Predict Homologous Recombination Repair Defects in Epithelial Ovarian Cancer. *Br J Cancer* **2012**, *107*, 1776–1782, doi:10.1038/BJC.2012.451.
241. Birkbak, N.J.; Wang, Z.C.; Kim, J.Y.; Eklund, A.C.; Li, Q.; Tian, R.; Bowman-Colin, C.; Li, Y.; Greene-Colozzi, A.; Dirk Iglehart, J.; et al. Telomeric

- Allelic Imbalance Indicates Defective DNA Repair and Sensitivity to DNA-Damaging Agents. *Cancer Discov* **2012**, *2*, 366–375, doi:10.1158/2159-8290.CD-11-0206.
242. Popova, T.; Manié, E.; Rieunier, G.; Caux-Moncoutier, V.; Tirapo, C.; Dubois, T.; Delattre, O.; Sigal-Zafrani, B.; Bollet, M.; Longy, M.; et al. Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-like Breast Carcinomas with BRCA1/2 Inactivation. *Cancer Res* **2012**, *72*, 5454–5462, doi:10.1158/0008-5472.CAN-12-1470.
243. Barili, V.; Ambrosini, E.; Bortesi, B.; Minari, R.; De Sensi, E.; Cannizzaro, I.R.; Taiani, A.; Michiara, M.; Sikokis, A.; Boggiani, D.; et al. Genetic Basis of Breast and Ovarian Cancer: Approaches and Lessons Learnt from Three Decades of Inherited Predisposition Testing. *Genes* **2024**, *Vol. 15, Page 219* **2024**, *15*, 219, doi:10.3390/GENES15020219.
244. Wicks, A.J.; Krastev, D.B.; Pettitt, S.J.; Tutt, A.N.J.; Lord, C.J. Opinion: PARP Inhibitors in Cancer—What Do We Still Need to Know? *Open Biol* **2022**, *12*, 220118, doi:10.1098/RSOB.220118.
245. Pellegrino, B.; Herencia-Roper, A.; Llop-Guevar, A.; Pedretti, F.; Moles-Fernández, A.; Viaplana, C.; Villacampa, G.; Guzmán, M.; Rodríguez, O.; Grueso, J.; et al. Preclinical In Vivo Validation of the RAD51 Test for Identification of Homologous Recombination-Deficient Tumors and Patient Stratification. *Cancer Res* **2022**, *82*, 1646–1657, doi:10.1158/0008-5472.CAN-21-2409.
246. Naipal, K.A.T.; Verkaik, N.S.; Ameziane, N.; Van Deurzen, C.H.M.; Brugge, P. Ter; Meijers, M.; Sieuwerts, A.M.; Martens, J.W.; O'Connor, M.J.; Vrieling, H.; et al. Functional Ex Vivo Assay to Select Homologous Recombination-Deficient Breast Tumors for PARP Inhibitor Treatment. *Clin Cancer Res* **2014**, *20*, 4816–4826, doi:10.1158/1078-0432.CCR-14-0571.
247. Castroviejo-Bermejo, M.; Cruz, C.; Llop-Guevara, A.; Gutiérrez-Enríquez, S.; Ducy, M.; Ibrahim, Y.H.; Gris-Oliver, A.; Pellegrino, B.; Bruna, A.; Guzmán, M.; et al. A RAD 51 Assay Feasible in Routine Tumor Samples Calls PARP Inhibitor Response beyond BRCA Mutation. *EMBO Mol Med* **2018**, *10*, 23, doi:10.15252/emmm.201809172.

248. Cruz, C.; Castroviejo-Bermejo, M.; Gutiérrez-Enríquez, S.; Llop-Guevara, A.; Ibrahim, Y.H.; Gris-Oliver, A.; Bonache, S.; Morancho, B.; Bruna, A.; Rueda, O.M.; et al. RAD51 Foci as a Functional Biomarker of Homologous Recombination Repair and PARP Inhibitor Resistance in Germline BRCA-Mutated Breast Cancer. *Annals of Oncology* **2018**, *29*, 1203–1210, doi:10.1093/ANNONC/MDY099.
249. Yoshida, R.; Hagio, T.; Kaneyasu, T.; Gotoh, O.; Osako, T.; Tanaka, N.; Amino, S.; Yaguchi, N.; Nakashima, E.; Kitagawa, D.; et al. Pathogenicity Assessment of Variants for Breast Cancer Susceptibility Genes Based on BRCAness of Tumor Sample. *Cancer Sci* **2021**, *112*, 1310–1319, doi:10.1111/CAS.14803.
250. Walsh, M.F.; Ritter, D.I.; Kesserwan, C.; Sonkin, D.; Chakravarty, D.; Chao, E.; Ghosh, R.; Kemel, Y.; Wu, G.; Lee, K.; et al. Integrating Somatic Variant Data and Biomarkers for Germline Variant Classification in Cancer Predisposition Genes. *Hum Mutat* **2018**, *39*, 1542–1552, doi:10.1002/HUMU.23640.
251. Shapiro, M.B.; Senapathy, P. RNA Splice Junctions of Different Classes of Eukaryotes: Sequence Statistics and Functional Implications in Gene Expression. *Nucleic Acids Res* **1987**, *15*, 7155–7174, doi:10.1093/NAR/15.17.7155.
252. Yeo, G.; Burge, C.B. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J Comput Biol* **2004**, *11*, 377–394, doi:10.1089/1066527041410418.
253. Erkelenz, S.; Theiss, S.; Otte, M.; Widera, M.; Peter, J.O.; Schaal, H. Genomic HEXploring Allows Landscaping of Novel Potential Splicing Regulatory Elements. *Nucleic Acids Res* **2014**, *42*, 10681–10697, doi:10.1093/nar/gku736.
254. Dhir, A.; Buratti, E.; Van Santen, M.A.; Lührmann, R.; Baralle, F.E. The Intronic Splicing Code: Multiple Factors Involved in ATM Pseudoexon Definition. *EMBO J* **2010**, *29*, 749–760, doi:10.1038/EMBOJ.2009.397.
255. Xie, Z.; Tang, L.; Xie, Z.; Sun, C.; Shuai, H.; Zhou, C.; Liu, Y.; Yu, M.; Zheng, Y.; Meng, L.; et al. Splicing Characteristics of Dystrophin Pseudoexons and Identification of a Novel Pathogenic Intronic Variant in the DMD Gene. *Genes (Basel)* **2020**, *11*, 1–13, doi:10.3390/GENES11101180.

256. Parsons, M.T.; Tudini, E.; Li, H.; Hahnen, E.; Wappenschmidt, B.; Feliubadaló, L.; Aalfs, C.M.; Agata, S.; Aittomäki, K.; Alducci, E.; et al. Large Scale Multifactorial Likelihood Quantitative Analysis of BRCA1 and BRCA2 Variants: An ENIGMA Resource to Support Clinical Variant Classification. *Hum Mutat* **2019**, *40*, 1557–1578, doi:10.1002/humu.23818.
257. Caputo, S.M.; Golmard, L.; Léone, M.; Damiola, F.; Guillaud-Bataille, M.; Revillion, F.; Rouleau, E.; Derive, N.; Buisson, A.; Basset, N.; et al. Classification of 101 BRCA1 and BRCA2 Variants of Uncertain Significance by Cosegregation Study: A Powerful Approach. *The American Journal of Human Genetics* **2021**, *108*, 1907–1923, doi:10.1016/J.AJHG.2021.09.003.
258. Wappenschmidt, B.; Becker, A.A.; Hauke, J.; Weber, U.; Engert, S.; Köhler, J.; Kast, K.; Arnold, N.; Rhiem, K.; Hahnen, E.; et al. Analysis of 30 Putative BRCA1 Splicing Mutations in Hereditary Breast and Ovarian Cancer Families Identifies Exonic Splice Site Mutations That Escape In Silico Prediction. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0050800.
259. Jung, H.; Lee, K.S.; Choi, J.K. Comprehensive Characterisation of Intronic Mis-Splicing Mutations in Human Cancers. *Oncogene* **2021**, *40*, 1347–1361, doi:10.1038/s41388-020-01614-3.
260. Bolisetty, M.T.; Beemon, K.L. Splicing of Internal Large Exons Is Defined by Novel Cis-Acting Sequence Elements. *Nucleic Acids Res* **2012**, *40*, 9244–9254, doi:10.1093/NAR/GKS652.
261. Feliubadaló, L.; López-Fernández, A.; Pineda, M.; Díez, O.; del Valle, J.; Gutiérrez-Enríquez, S.; Teulé, A.; González, S.; Stjepanovic, N.; Salinas, M.; et al. Opportunistic Testing of BRCA1, BRCA2 and Mismatch Repair Genes Improves the Yield of Phenotype Driven Hereditary Cancer Gene Panels. *Int J Cancer* **2019**, *145*, 2682–2691, doi:10.1002/IJC.32304.
262. Leman, R.; Parfait, B.; Vidaud, D.; Girodon, E.; Pacot, L.; Le Gac, G.; Ka, C.; Ferec, C.; Fichou, Y.; Quesnelle, C.; et al. SPiP: Splicing Prediction Pipeline, a Machine Learning Tool for Massive Detection of Exonic and Intronic Variant Effects on mRNA Splicing. *Hum Mutat* **2022**, *43*, 2308–2323, doi:10.1002/humu.24491.

263. Feng, B.-J. PERCH: A Unified Framework for Disease Gene Prioritization. *OFFICIAL JOURNAL Hum Mutat* **2017**, *38*, 243–251, doi:10.1002/humu.23158.
264. Horton, C.; Cass, A.; Conner, B.R.; Hoang, L.; Zimmermann, H.; Abualkheir, N.; Burks, D.; Qian, D.; Molparia, B.; Vuong, H.; et al. Mutational and Splicing Landscape in a Cohort of 43,000 Patients Tested for Hereditary Cancer. *NPJ Genom Med* **2022**, *7*, doi:10.1038/S41525-022-00323-Y.
265. Kamps-Hughes, N.; Carlton, V.E.H.; Fresard, L.; Osazuwa, S.; Starks, E.; Vincent, J.J.; Albritton, S.; Nussbaum, R.L.; Nykamp, K. A Systematic Method for Detecting Abnormal MRNA Splicing and Assessing Its Clinical Impact in Individuals Undergoing Genetic Testing for Hereditary Cancer Syndromes. *Journal of Molecular Diagnostics* **2023**, *25*, 156–167, doi:10.1016/j.jmoldx.2022.12.002.
266. Soukarieh, O.; Gaildrat, P.; Hamieh, M.; Drouet, A.; Baert-Desurmont, S.; Frébourg, T.; Tosi, M.; Martins, A. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet* **2016**, *12*, doi:10.1371/JOURNAL.PGEN.1005756.
267. Ho, S.N.; Hunt, H.D.; Horton, R.M.; Pullen, J.K.; Pease, L.R. Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction. *Gene* **1989**, *77*, 51–59, doi:10.1016/0378-1119(89)90358-2.
268. Mesman, R.L.S.; Calléja, F.M.G.R.; Hendriks, G.; Morolli, B.; Misovic, B.; Devilee, P.; van Asperen, C.J.; Vrieling, H.; Vreeswijk, M.P.G. The Functional Impact of Variants of Uncertain Significance in BRCA2. *Genetics in Medicine* **2019**, *21*, 293–302, doi:10.1038/s41436-018-0052-2.
269. Bouwman, P.; Van Der Gulden, H.; Van Der Heijden, I.; Drost, R.; Klijn, C.N.; Prasetyanti, P.; Pieterse, M.; Wientjens, E.; Seibler, J.; Hogervorst, F.B.L.; et al. A High-Throughput Functional Complementation Assay for Classification of BRCA1 Missense Variants. **2013**, doi:10.1158/2159-8290.CD-13-0094.
270. Bouwman, P.; van der Heijden, I.; van der Gulden, H.; de Bruijn, R.; Braspenning, M.E.; Moghadasi, S.; Wessels, L.F.A.; Vreeswijk, M.P.G.; Jonkers, J. Functional Categorization of

- BRCA1 Variants of Uncertain Clinical Significance in Homologous Recombination Repair Complementation Assays. *Clin Cancer Res* **2020**, 26, 4559–4568, doi:10.1158/1078-0432.CCR-20-0255.
271. Findlay, G.M.; Daza, R.M.; Martin, B.; Zhang, M.D.; Leith, A.P.; Gasperini, M.; Janizek, J.D.; Huang, X.; Starita, L.M.; Shendure, J. Accurate Classification of BRCA1 Variants with Saturation Genome Editing. *Nature* **2018**, 562, 217–222, doi:10.1038/s41586-018-0461-z.
272. Petitalot, A.; Dardillac, E.; Jacquet, E.; Nhiri, N.; Guirouilh-Barbat, J.; Julien, P.; Bouazzaoui, I.; Bonte, D.; Feunteun, J.; Schnell, J.A.; et al. Combining Homologous Recombination and Phosphopeptide-Binding Data to Predict the Impact of BRCA1 BRCT Variants on Cancer Risk. *Molecular Cancer Research* **2019**, 17, 54–69, doi:10.1158/1541-7786.MCR-17-0357.
273. Vanessa Fernandes, X.C.; Golubeva, V.A.; Giuliano Di Pietro, X.; Shields, C.; Amankwah, K.; Nepomuceno, T.C.; Giuliana de Gregoriis, X.; Renata V Abreu, X.B.; Harro, C.; Gomes, T.T.; et al. Impact of Amino Acid Substitutions at Secondary Structures in the BRCT Domains of the Tumor Suppressor BRCA1: Implications for Clinical Annotation. *J. Biol. Chem* **2019**, 294, 5980–5992, doi:10.1074/jbc.RA118.005274.
274. Nepomuceno, T.C.; Dos Santos, A.P.P.; Fernandes, V.C.; Elias, A.B.R.; Gomes, T.T.; Suarez-Kurtz, G.; Iversen, E.S.; Couch, F.J.; Monteiro, A.N.A.; Carvalho, M.A. Assessment of Small In-Frame Indels and C-Terminal Nonsense Variants of BRCA1 Using a Validated Functional Assay. **2022**, 12, doi:10.1038/s41598-022-20500-4.
275. Brnich, S.E.; Abou Tayoun, A.N.; Couch, F.J.; Cutting, G.R.; Greenblatt, M.S.; Heinen, C.D.; Kanavy, D.M.; Luo, X.; McNulty, S.M.; Starita, L.M.; et al. Recommendations for Application of the Functional Evidence PS3/BS3 Criterion Using the ACMG/AMP Sequence Variant Interpretation Framework. *Genome Med* 2019, 12.
276. Li, H.; LaDuca, H.; Pesaran, T.; Chao, E.C.; Dolinsky, J.S.; Parsons, M.; Spurdle, A.B.; Polley, E.C.; Shimelis, H.; Hart, S.N.; et al. Classification of Variants of Uncertain Significance in BRCA1 and

- BRCA2 Using Personal and Family History of Cancer from Individuals in a Large Hereditary Cancer Multigene Panel Testing Cohort. *Genet Med* **2020**, 22, 701–708, doi:10.1038/S41436-019-0729-1.
277. Spurdle, A.B.; Whaley, P.J.; Thompson, B.; Feng, B.; Healey, S.; Brown, M.A.; Pettigrew, C.; Van Asperen, C.J.; Ausems, M.G.E.M.; Kattentidt-Mouravieva, A.A.; et al. BRCA1 R1699Q Variant Displaying Ambiguous Functional Abrogation Confers Intermediate Breast and Ovarian Cancer Risk. *J Med Genet* **2012**, 49, 525–532, doi:10.1136/JMEDGENET-2012-101037.
 278. Thompson, D.; Easton, D.F.; Goldgar, D.E. A Full-Likelihood Method for the Evaluation of Causality of Sequence Variants from Family Data. *Am J Hum Genet* **2003**, 73, 652–655, doi:10.1086/378100.
 279. Belman, S.; Parsons, M.T.; Spurdle, A.B.; Goldgar, D.E.; Feng, B.J. Considerations in Assessing Germline Variant Pathogenicity Using Cosegregation Analysis. *Genet Med* **2020**, 22, 2052–2059, doi:10.1038/S41436-020-0920-4.
 280. Llop-Guevara, A.; Loibl, S.; Villacampa, G.; Vladimirova, V.; Schneeweiss, A.; Karn, T.; Zahm, D.-M.; Herencia-Ropero, A.; Jank, P.; Van Mackelenbergh, M.; et al. Association of RAD51 with Homologous Recombination Deficiency (HRD) and Clinical Outcomes in Untreated Triple-Negative Breast Cancer (TNBC): Analysis of the GeparSixto Randomized Clinical Trial. **2021**, doi:10.1016/j.annonc.2021.09.003.
 281. Llop-Guevara, A.; Loibl, S.; Villacampa, G.; Vladimirova, V.; Schneeweiss, A.; Karn, T.; Zahm, D.M.; Herencia-Ropero, A.; Jank, P.; van Mackelenbergh, M.; et al. Association of RAD51 with Homologous Recombination Deficiency (HRD) and Clinical Outcomes in Untreated Triple-Negative Breast Cancer (TNBC): Analysis of the GeparSixto Randomized Clinical Trial. *Annals of Oncology* **2021**, 32, 1590–1596, doi:10.1016/j.annonc.2021.09.003.
 282. Telli, M.L.; Timms, K.M.; Reid, J.; Hennessy, B.; Mills, G.B.; Jensen, K.C.; Szallasi, Z.; Barry, W.T.; Winer, E.P.; Tung, N.M.; et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with

- Triple-Negative Breast Cancer. *Clin Cancer Res* **2016**, 22, 3764–3773, doi:10.1158/1078-0432.CCR-15-2477.
283. Denkert, C.; Romey, M.; Swedlund, B.; Hattesohl, A.; Teply-Szymanski, J.; Kommoss, S.; Kaiser, K.; Staebler, A.; du Bois, A.; Grass, A.; et al. Homologous Recombination Deficiency as an Ovarian Cancer Biomarker in a Real-World Cohort Validation of Decentralized Genomic Profiling. *The Journal of Molecular Diagnostics* **2022**, 24, 1254–1263, doi:10.1016/j.jmoldx.2022.09.004.
 284. Vallée, M.P.; Di Sera, T.L.; Nix, D.A.; Paquette, A.M.; Parsons, M.T.; Bell, R.; Hoffman, A.; Hogervorst, F.B.L.; Goldgar, D.E.; Spurdle, A.B.; et al. Adding In Silico Assessment of Potential Splice Aberration to the Integrated Evaluation of BRCA Gene Unclassified Variants. *Hum Mutat* **2016**, 37, 627, doi:10.1002/HUMU.22973.
 285. Feliubadaló, L.; Domènech-Vivó, J.; Menéndez, M.; Llop-Guevara, A.; Tornero, E.; Moles-Fernández, A.; de Miguel, M.; Potrony, M.; Santamariña, M.; Sanz, J.; et al. P13.079.A Clinical, Splicing and Functional Analysis of a Predicted Truncating BRCA1 Variant That Alters Splicing, Increasing an in-Frame $\Delta 11q$ Naturally Occurring Isoform. Abstracts from the 55th European Society of Human Genetics (ESHG) Conference: Hybrid Posters. *European Journal of Human Genetics* **2023**, 31:1 **2023**, 31, 345–709, doi:10.1038/s41431-023-01338-4.
 286. Valenzuela-Palomo, A.; Bueno-Martínez, E.; Sanoguera-Miralles, L.; Lorca, V.; Fraile-Bethencourt, E.; Esteban-Sánchez, A.; Gómez-Barrero, S.; Carvalho, S.; Allen, J.; García-Álvarez, A.; et al. Splicing Predictions, Minigene Analyses, and ACMG-AMP Clinical Classification of 42 Germline PALB2 Splice-Site Variants. *J Pathol* **2022**, 256, 321–334, doi:10.1002/PATH.5839.
 287. Moles-Fernández, A.; Domènech-Vivó, J.; Tenés, A.; Balmaña, J.; Diez, O.; Gutiérrez-Enríquez, S. Role of Splicing Regulatory Elements and In Silico Tools Usage in the Identification of Deep Intronic Splicing Variants in Hereditary Breast/Ovarian Cancer Genes. *Cancers (Basel)* **2021**, 13, doi:10.3390/cancers13133341.
 288. Torres-Esquius, S.; Llop-Guevara, A.; Gutiérrez-Enríquez, S.; Romey, M.; Teulé, À.; Llorc,

- G.; Herrero, A.; Sánchez-Henarejos, P.; Vallmajó, A.; González-Santiago, S.; et al. Prevalence of Homologous Recombination Deficiency Among Patients With Germline RAD51C/D Breast or Ovarian Cancer. *JAMA Netw Open* **2024**, 7, e247811–e247811, doi:10.1001/JAMANETWORKOP.EN.2024.7811.
289. Turner, N.C. Signatures of DNA-Repair Deficiencies in Breast Cancer. *New England Journal of Medicine* **2017**, 377, 2490–2492, doi:10.1056/NEJMCIBR1710161/SUPPL_FILE/NEJMCIBR1710161_DISCLOSURES.PDF.
 290. Xu, L.; Liddell, B.; Nesic, K.; Geissler, F.; Ashwood, L.M.; Wakefield, M.J.; Scott, C.L.; Waddell, N.; Kondrashova, O. High-Level Tumour Methylation of BRCA1 and RAD51C Is Required for Homologous Recombination Deficiency in Solid Cancers. *NAR Cancer* **2024**, 6, 33, doi:10.1093/NARCAN/ZCAE033.
 291. Chen, M.; Linstra, R.; van Vugt, M.A.T.M. Genomic Instability, Inflammatory Signaling and Response to Cancer Immunotherapy. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2022**, 1877, 188661, doi:10.1016/J.BBCAN.2021.188661.
 292. Wang, Y.; Bernhardt, A.J.; Cruz, C.; Krais, J.J.; Nacson, J.; Nicolas, E.; Peri, S.; Van Der Gulden, H.; Van Der Heijden, I.; O'Brien, S.W.; et al. The BRCA1-Δ11q Alternative Splice Isoform Bypasses Germline Mutations and Promotes Therapeutic Resistance to PARP Inhibition and Cisplatin. *Cancer Res* **2016**, 76, 2778–2790, doi:10.1158/0008-5472.CAN-16-0186.
 293. Drost, R.; Dhillon, K.K.; Van Der Gulden, H.; Van Der Heijden, I.; Brandsma, I.; Cruz, C.; Chondronasiou, D.; Castroviejo-Bermejo, M.; Boon, U.; Schut, E.; et al. BRCA1185delAG Tumors May Acquire Therapy Resistance through Expression of RING-Less BRCA1. *J Clin Invest* **2016**, 126, 2903–2918, doi:10.1172/JCI70196.
 294. Santos, C.; Peixoto, A.; Rocha, P.; Vega, A.; Soares, M.J.; Cerveira, N.; Bizarro, S.; Pinheiro, M.; Pereira, D.; Rodrigues, H.; et al. Haplotype and Quantitative Transcript Analyses of Portuguese Breast/Ovarian Cancer Families with the BRCA1 R71G Founder Mutation of Galician Origin. *Fam Cancer* **2009**, 8, 203–208, doi:10.1007/S10689-008-9229-1.

295. Sakaguchi, N.; Suyama, M. In Silico Identification of Pseudo-Exon Activation Events in Personal Genome and Transcriptome Data. *RNA Biol* **2021**, *18*, 382–390, doi:10.1080/15476286.2020.1809195.
296. Rentzsch, P.; Schubach, M.; Shendure, J.; Kircher, M. CADD-Splice—Improving Genome-Wide Variant Effect Prediction Using Deep Learning-Derived Splice Scores. *Genome Med* **2021**, *13*, 1–12, doi:10.1186/s13073-021-00835-9.
297. Riepe, T. V.; Khan, M.; Roosing, S.; Cremers, F.P.M.; 't Hoen, P.A.C. Benchmarking Deep Learning Splice Prediction Tools Using Functional Splice Assays. *Hum Mutat* **2021**, *42*, 799–810, doi:10.1002/HUMU.24212.
298. Qian, X.; Wang, J.; Wang, M.; Igelman, A.D.; Jones, K.D.; Li, Y.; Wang, K.; Goetz, K.E.; Birch, D.G.; Yang, P.; et al. Identification of Deep-Intronic Splice Mutations in a Large Cohort of Patients With Inherited Retinal Diseases. *Front Genet* **2021**, *12*, 647400, doi:10.3389/fgene.2021.647400.
299. Xie, Z.; Tang, L.; Xie, Z.; Sun, C.; Shuai, H.; Zhou, C.; Liu, Y.; Yu, M.; Zheng, Y.; Meng, L.; et al. Splicing Characteristics of Dystrophin Pseudoexons and Identification of a Novel Pathogenic Intronic Variant in the DMD Gene. *Genes (Basel)* **2020**, *11*, 1–13, doi:10.3390/genes11101180.
300. Baeza-Centurion, P.; Miñana, B.; Valcárcel, J.; Lehner, B. Mutations Primarily Alter the Inclusion of Alternatively Spliced Exons. *Elife* **2020**, *9*, 1–74, doi:10.7554/eLife.59959.
301. Anczuków, O.; Buisson, M.; Léoñe, M.; Coutanson, C.; Lasset, C.; Calender, A.; Sinilnikova, O.M.; Mazoyer, S. BRCA2 Deep Intronic Mutation Causing Activation of a Cryptic Exon: Opening toward a New Preventive Therapeutic Strategy. *Clinical Cancer Research* **2012**, *18*, 4903–4909, doi:10.1158/1078-0432.CCR-12-1100.
302. Lovelock, P.K.; Healey, S.; Au, W.; Sum, E.Y.M.; Tesoriero, A.; Wong, E.M.; Hinson, S.; Brinkworth, R.; Bekessy, A.; Diez, O.; et al. Genetic, Functional, and Histopathological Evaluation of Two C-terminal BRCA1 Missense Variants. *J Med Genet* **2006**, *43*, 74, doi:10.1136/JMG.2005.033258.
303. Millevoi, S.; Bernat, S.; Telly, D.; Fouque, F.; Gladieff, L.; Favre, G.; Vagner, S.; Toulas, C. The

- c.5242C>A BRCA1 Missense Variant Induces Exon Skipping by Increasing Splicing Repressors Binding. *Breast Cancer Res Treat* **2010**, *120*, 391–399, doi:10.1007/S10549-009-0392-3.
304. Quiles, F.; Menéndez, M.; Tornero, E.; del Valle, J.; Teulé, À.; Palanca, S.; Izquierdo, A.; Gómez, C.; Campos, O.; Santamaria, R.; et al. Investigating the Effect of 28 BRCA1 and BRCA2 Mutations on Their Related Transcribed MRNA. *Breast Cancer Res Treat* **2016**, *155*, 253–260, doi:10.1007/s10549-015-3676-9.
 305. Sanz, D.J.; Acedo, A.; Infante, M.; Durán, M.; Pérez-Cabornero, L.; Esteban-Cardenosa, E.; Lastra, E.; Pagani, F.; Miner, C.; Velasco, E.A. A High Proportion of DNA Variants of BRCA1 and BRCA2 Is Associated with Aberrant Splicing in Breast/Ovarian Cancer Patients. *Clinical Cancer Research* **2010**, *16*, 1957–1967, doi:10.1158/1078-0432.CCR-09-2564.
 306. Grosso, A.R.; Gomes, A.Q.; Barbosa-Morais, N.L.; Caldeira, S.; Thorne, N.P.; Grech, G.; von Lindern, M.; Carmo-Fonseca, M. Tissue-Specific Splicing Factor Gene Expression Signatures. *Nucleic Acids Res* **2008**, *36*, 4823, doi:10.1093/NAR/GKN463.
 307. Dvinge, H. Regulation of Alternative MRNA Splicing: Old Players and New Perspectives. *FEBS Lett* **2018**, *592*, 2987–3006, doi:10.1002/1873-3468.13119.
 308. Wang, Z.; Burge, C.B. Splicing Regulation: From a Parts List of Regulatory Elements to an Integrated Splicing Code. *RNA* **2008**, *14*, 802, doi:10.1261/RNA.876308.
 309. Amoah, K.; Hsiao, Y.-H.E.; Hoon Bahn, J.; Sun, Y.; Burghard, C.; Tan, B.X.; Yang, E.-W.; Xiao, X. Allele-Specific Alternative Splicing and Its Functional Genetic Variants in Human Tissues. **2021**, doi:10.1101/gr.265637.120.
 310. van der Klift, H.M.; Jansen, A.M.L.; Steenstraten, N. van der; Bik, E.C.; Tops, C.M.J.; Devilee, P.; Wijnen, J.T. Splicing Analysis for Exonic and Intronic Mismatch Repair Gene Variants Associated with Lynch Syndrome Confirms High Concordance between Minigene Assays and Patient Rna Analyses. *Mol Genet Genomic Med* **2015**, *3*, 327–345, doi:10.1002/MGG3.145.

311. Vaclová, T.; Woods, N.T.; Megías, D.; Gomez-Lopez, S.; Setién, F.; García Bueno, J.M.; Macías, J.A.; Barroso, A.; Urioste, M.; Esteller, M.; et al. Germline Missense Pathogenic Variants in the BRCA1 BRCT Domain, p.Gly1706Glu and p.Ala1708Glu, Increase Cellular Sensitivity to PARP Inhibitor Olaparib by a Dominant Negative Effect. *Hum Mol Genet* **2016**, *25*, 5287–5299, doi:10.1093/HMG/DDW343.
312. Moghadasi, S.; Meeks, H.D.; Vreeswijk, P.M.; Janssen, L.A.M.; Borg, Å.; Ehrencrona, H.; Paulsson-Karlsson, Y.; Wappenschmidt, B.; Engel, C.; Gehrig, A.; et al. The BRCA1 c.5096G>A p.Arg1699Gln (R1699Q) Intermediate Risk Variant: Breast and Ovarian Cancer Risk Estimation and Recommendations for Clinical Management from the ENIGMA Consortium. *J Med Genet* **2018**, *55*, 15–20, doi:10.1136/JMEDGENET-2017-104560.
313. Keupp, K.; Hampp, S.; Hübbel, A.; Maringa, M.; Kostezka, S.; Rhiem, K.; Waha, A.; Wappenschmidt, B.; Pujol, R.; Surrallés, J.; et al. Biallelic Germline BRCA1 Mutations in a Patient with Early Onset Breast Cancer, Mild Fanconi Anemia-like Phenotype, and No Chromosome Fragility. *Mol Genet Genomic Med* **2019**, *7*, doi:10.1002/MGG3.863.
314. Block, I.; Mateu-Regué, À.; Do, T.T.N.; Miceikaite, I.; Sdogati, D.; Larsen, M.J.; Hao, Q.; Nielsen, H.R.; Boonen, S.E.; Skytte, A.B.; et al. Male with an Apparently Normal Phenotype Carrying a BRCA1 Exon 20 Duplication in Trans to a BRCA1 Frameshift Variant. *Breast Cancer Research* **2024**, *26*, 1–14, doi:10.1186/s13058-023-01755-9.
315. van Wijk, L.M.; Kramer, C.J.H.; Vermeulen, S.; Ter Haar, N.T.; de Jonge, M.M.; Kroep, J.R.; de Kroon, C.D.; Gaarenstroom, K.N.; Vrieling, H.; Bosse, T.; et al. The Rad51-Ffpe Test; Calibration of a Functional Homologous Recombination Deficiency Test on Diagnostic Endometrial and Ovarian Tumor Blocks. *Cancers (Basel)* **2021**, *13*, 2994, doi:10.3390/CANCERS13122994/S1.
316. Feng, C.; Zhang, Y.; Wu, F.; Li, J.; Liu, M.; Lv, W.; Li, C.; Wang, W.; Tan, Q.; Xue, X.; et al. Relationship between Homologous Recombination Deficiency and Clinical Features of Breast Cancer Based on Genomic Scar Score. *The Breast* **2023**, *69*, 392–400,

doi:10.1016/j.breast.2023.04.002.

317. Nones, K.; Johnson, J.; Newell, F.; Patch, A.M.; Thorne, H.; Kazakoff, S.H.; De Luca, X.M.; Parsons, M.T.; Ferguson, K.; Reid, L.E.; et al. Whole-Genome Sequencing Reveals Clinically Relevant Insights into the Aetiology of Familial Breast Cancers. *Annals of Oncology* **2019**, *30*, 1071, doi:10.1093/ANNONC/MDZ132.
318. Moschetta, M.; George, A.; Kaye, S.B.; Banerjee, S. BRCA Somatic Mutations and Epigenetic BRCA Modifications in Serous Ovarian Cancer. **2016**, doi:10.1093/annonc/mdw142.
319. Brianese, R.C.; Nakamura, K.D. de M.; de Almeida, F.G. dos S.R.; Ramalho, R.F.; Barros, B.D. de F.; e Ferreira, E.N.; Formiga, M.N. da C.; de Andrade, V.P.; de Lima, V.C.C.; Carraro, D.M. BRCA1 Deficiency Is a Recurrent Event in Early-Onset Triple-Negative Breast Cancer: A Comprehensive Analysis of Germline Mutations and Somatic Promoter Methylation. *Breast Cancer Res Treat* **2018**, *167*, 803–814, doi:10.1007/S10549-017-4552-6.
320. Lopez-Perolio, I.; Leman, R.; Behar, R.; Lattimore, V.; Pearson, J.F.; Castéra, L.; Martins, A.; Vaur, D.; Goardon, N.; Davy, G.; et al. Alternative Splicing and ACMG-AMP-2015-Based Classification of PALB2 Genetic Variants: An ENIGMA Report. *J Med Genet* **2019**, *56*, 453–460, doi:10.1136/JMEDGENET-2018-105834.
321. Putti, T.C.; Abd El-Rehim, D.M.; Rakha, E.A.; Paish, C.E.; Lee, A.H.S.; Pinder, S.E.; Ellis, I.O. Estrogen Receptor-Negative Breast Carcinomas: A Review of Morphology and Immunophenotypical Analysis. *Modern Pathology* **2005** *18*:1 **2004**, *18*, 26–35, doi:10.1038/modpathol.3800255.
322. Metz, B.; Kersten, G.F.A.; Hoogerhout, P.; Brugghe, H.F.; Timmermans, H.A.M.; De Jong, A.; Meiring, H.; Ten Hove, J.; Hennink, W.E.; Crommelin, D.J.A.; et al. Identification of Formaldehyde-Induced Modifications in Proteins: Reactions with Model Peptides. *J Biol Chem* **2004**, *279*, 6235–6243, doi:10.1074/JBC.M310752200.
323. Compadre, A.J.; van Biljon, L.N.; Valentine, M.C.; Llop-Guevara, A.; Graham, E.; Fashemi, B.; Herencia-Ropero, A.; Kotnik, E.N.; Cooper, I.; Harrington, S.P.; et al. RAD51 Foci as a Biomarker Predictive of Platinum

- Chemotherapy Response in Ovarian Cancer. *Clin Cancer Res* **2023**, 29, 2466–2479, doi:10.1158/1078-0432.CCR-22-3335.
324. Kramer, C.J.H.; Llop-Guevara, A.; Yaniz-Galende, E.; Pellegrino, B.; ter Haar, N.T.; Herencia-Ropero, A.; Campanini, N.; Musolino, A.; Bosse, T.; Leary, A.; et al. RAD51 as a Biomarker for Homologous Recombination Deficiency in High-grade Serous Ovarian Carcinoma: Robustness and Interobserver Variability of the RAD51 Test. *J Pathol Clin Res* **2023**, 9, 442, doi:10.1002/CJP2.336.
 325. Lionel, A.C.; Costain, G.; Monfared, N.; Walker, S.; Reuter, M.S.; Hosseini, S.M.; Thiruvahindrapuram, B.; Merico, D.; Jobling, R.; Nalpathamkalam, T.; et al. Improved Diagnostic Yield Compared with Targeted Gene Sequencing Panels Suggests a Role for Whole-Genome Sequencing as a First-Tier Genetic Test. *Genet Med* **2018**, 20, 435–443, doi:10.1038/GIM.2017.119.
 326. Davidson, A.L.; Dressel, U.; Norris, S.; Canson, D.M.; Glubb, D.M.; Fortuno, C.; Hollway, G.E.; Parsons, M.T.; Vidgen, M.E.; Holmes, O.; et al. The Clinical Utility and Costs of Whole-Genome Sequencing to Detect Cancer Susceptibility Variants—a Multi-Site Prospective Cohort Study. *Genome Med* **2023**, 15, 1–16, doi:10.1186/s13073-023-01223-1.
 327. Bagger, F.O.; Borgwardt, L.; Jespersen, A.S.; Hansen, A.R.; Bertelsen, B.; Kodama, M.; Nielsen, F.C. Whole Genome Sequencing in Clinical Practice. *BMC Medical Genomics* **2024**, 17:1 **2024**, 17, 1–16, doi:10.1186/S12920-024-01795-W.
 328. Jang, W.; Park, J.; Chae, H.; Kim, M. Comparison of In Silico Tools for Splice-Altering Variant Prediction Using Established Spliceogenic Variants: An End-User's Point of View. *Int J Genomics* **2022**, 2022, doi:10.1155/2022/5265686.
 329. Ha, C.; Kim, J.W.; Jang, J.H. Performance Evaluation of Spliceai for the Prediction of Splicing of Nf1 Variants. *Genes (Basel)* **2021**, 12, 1308, doi:10.3390/GENES12091308/S1.
 330. Canson, D.M.; Davidson, A.L.; de la Hoya, M.; Parsons, M.T.; Glubb, D.M.; Kondrashova, O.; Spurdle, A.B. SpliceAI-10k Calculator for the Prediction of Pseudoexonization, Intron Retention, and Exon Deletion. *Bioinformatics* **2023**, 39,

doi:10.1093/BIOINFORMATICS/BTAD179.

331. Kurosawa, R.; Iida, K.; Ajiro, M.; Awaya, T.; Yamada, M.; Kosaki, K.; Hagiwara, M. PDIVAS: Pathogenicity Predictor for Deep-Intronic Variants Causing Aberrant Splicing. *BMC Genomics* **2023**, *24*, 1–14, doi:10.1186/s12864-023-09645-2.
332. Keegan, N.P.; Wilton, S.D.; Fletcher, S. Analysis of Pathogenic Pseudoexons Reveals Novel Mechanisms Driving Cryptic Splicing. *Front Genet* **2022**, *12*, doi:10.3389/FGENE.2021.806946.
333. Fraile-Bethencourt, E.; Valenzuela-Palomo, A.; Díez-Gómez, B.; Goina, E.; Acedo, A.; Buratti, E.; Velasco, E.A. Mis-Splicing in Breast Cancer: Identification of Pathogenic BRCA2 Variants by Systematic Minigene Assays. *J Pathol* **2019**, *248*, 409–420, doi:10.1002/PATH.5268.
334. Rofes, P.; Menéndez, M.; González, S.; Tornero, E.; Gómez, C.; Vargas-Parra, G.; Montes, E.; Salinas, M.; Solanes, A.; Brunet, J.; et al. Improving Genetic Testing in Hereditary Cancer by RNA Analysis: Tools to Prioritize Splicing Studies and Challenges in Applying American College of Medical Genetics and Genomics Guidelines. *Journal of Molecular Diagnostics* **2020**, *22*, 1453–1468, doi:10.1016/j.jmoldx.2020.09.007.
335. Conner, B.R.; Hernandez, F.; Souders, B.; Landrith, T.; Boland, C.R.; Karam, R. RNA Analysis Identifies Pathogenic Duplications in MSH2 in Patients With Lynch Syndrome. *Gastroenterology* **2019**, *156*, 1924–1925.e4, doi:10.1053/j.gastro.2019.01.248.
336. Ryu, J.S.; Lee, H.Y.; Cho, E.H.; Yoon, K.A.; Kim, M.K.; Joo, J.; Lee, E.S.; Kang, H.S.; Lee, S.; Lee, D.O.; et al. Exon Splicing Analysis of Intronic Variants in Multigene Cancer Panel Testing for Hereditary Breast/Ovarian Cancer. *Cancer Sci* **2020**, *111*, 3912–3925, doi:10.1111/CAS.14600.
337. Jaramillo Oquendo, C.; Wai, H.A.; Rich, W.I.; Bunyan, D.J.; Thomas, N.S.; Hunt, D.; Lord, J.; Douglas, A.G.L.; Baralle, D. Identification of Diagnostic Candidates in Mendelian Disorders Using an RNA Sequencing-Centric Approach. *Genome Med* **2024**, *16*, 110, doi:10.1186/s13073-024-01381-w.

338. Wai, H.A.; Constable, M.; Drewes, C.; Davies, I.C.; Svobodova, E.; Dempsey, E.; Saggar, A.; Homfray, T.; Mansour, S.; Douzgou, S.; et al. Short Amplicon Reverse Transcription-Polymerase Chain Reaction Detects Aberrant Splicing in Genes with Low Expression in Blood Missed by Ribonucleic Acid Sequencing Analysis for Clinical Diagnosis. *Hum Mutat* **2022**, *43*, 963–970, doi:10.1002/HUMU.24378.
339. Sedlazeck, F.J.; Lee, H.; Darby, C.A.; Schatz, M.C. Piercing the Dark Matter: Bioinformatics of Long-Range Sequencing and Mapping. *Nat Rev Genet* **2018**, *19*, 329–346, doi:10.1038/s41576-018-0003-4.
340. Landrith, T.; Li, B.; Cass, A.A.; Conner, B.R.; LaDuca, H.; McKenna, D.B.; Maxwell, K.N.; Domchek, S.; Morman, N.A.; Heinlen, C.; et al. Splicing Profile by Capture RNA-Seq Identifies Pathogenic Germline Variants in Tumor Suppressor Genes. *npj Precision Oncology* **2020**, *4*, 1–9, doi:10.1038/s41698-020-0109-y.
341. Gulsuner, S.; AbuRayyan, A.; Mandell, J.B.; Lee, M.K.; Bernier, G. V.; Norquist, B.M.; Pierce, S.B.; King, M.-C.; Walsh, T. Long-Read DNA and CDNA Sequencing Identify Cancer-Predisposing Deep Intronic Variation in Tumor-Suppressor Genes. *Genome Res* **2024**, doi:10.1101/GR.279158.124.
342. Davidson, A.L.; Michailidou, K.; Parsons, M.T.; Fortuno, C.; Bolla, M.K.; Wang, Q.; Dennis, J.; Naven, M.; Abubakar, M.; Ahearn, T.U.; et al. Co-Observation of Germline Pathogenic Variants in Breast Cancer Predisposition Genes: Results from Analysis of the BRIDGES Sequencing Dataset. *Am J Hum Genet* **2024**, *111*, doi:10.1016/J.AJHG.2024.07.004.
343. Plazzer, J.P.; Macrae, F.; Yin, X.; Thompson, B.A.; Farrington, S.M.; Currie, L.; Lagerstedt-Robinson, K.; Frederiksen, J.H.; Hansen, T. van O.; Graversen, L.; et al. Mismatch Repair Gene Specifications to the ACMG/AMP Classification Criteria: Consensus Recommendations from the InSiGHT ClinGen Hereditary Colorectal Cancer / Polyposis Variant Curation Expert Panel. *medRxiv* **2024**, *10*, 2024.05.13.24307108, doi:10.1101/2024.05.13.24307108.
344. Carreira, S.; Porta, N.; Arce-Gallego, S.; Seed, G.; Llop-Guevara, A.; Bianchini, D.;

- Rescigno, P.; Paschalis, A.; Bertan, C.; Baker, C.; et al. Biomarkers Associating with PARP Inhibitor Benefit in Prostate Cancer in the TOPARP-B Trial. *Cancer Discov* **2021**, *11*, 2812–2827, doi:10.1158/2159-8290.CD-21-0007.
345. Santana dos Santos, E.; Spurdle, A.B.; Carraro, D.M.; Briaux, A.; Southey, M.; Torrezan, G.; Petitalot, A.; Leman, R.; Lafitte, P.; Meseure, D.; et al. Value of the Loss of Heterozygosity to BRCA1 Variant Classification. *NPJ Breast Cancer* **2022**, *8*, doi:10.1038/S41523-021-00361-2.
346. Tan, S.L.W.; Chadha, S.; Liu, Y.; Gabasova, E.; Perera, D.; Ahmed, K.; Constantinou, S.; Renaudin, X.; Lee, M.; Aebersold, R.; et al. A Class of Environmental and Endogenous Toxins Induces BRCA2 Haploinsufficiency and Genome Instability. *Cell* **2017**, *169*, 1105–1118.e15, doi:10.1016/j.cell.2017.05.010.
347. Ridpath, J.R.; Nakamura, A.; Tano, K.; Luke, A.M.; Sonoda, E.; Arakawa, H.; Buerstedde, J.-M.; Gillespie, D.A.F.; Sale, J.E.; Yamazoe, M.; et al. Cells Deficient in the FANC/BRCA Pathway Are Hypersensitive to Plasma Levels of Formaldehyde. *Cancer Res* **2007**, *67*, 11117–11122, doi:10.1158/0008-5472.CAN-07-3028.
348. Noda, T.; Takahashi, A.; Kondo, N.; Mori, E.; Okamoto, N.; Nakagawa, Y.; Ohnishi, K.; Zdzienicka, M.Z.; Thompson, L.H.; Helleday, T.; et al. Repair Pathways Independent of the Fanconi Anemia Nuclear Core Complex Play a Predominant Role in Mitigating Formaldehyde-Induced DNA Damage. *Biochem Biophys Res Commun* **2011**, *404*, 206–210, doi:10.1016/j.bbrc.2010.11.094.
349. Sadeghi, F.; Asgari, M.; Matloubi, M.; Ranjbar, M.; Karkhaneh Yousefi, N.; Azari, T.; Zaki-Dizaji, M. Molecular Contribution of BRCA1 and BRCA2 to Genome Instability in Breast Cancer Patients: Review of Radiosensitivity Assays. *Biol Proced Online* **2020**, *22*, 1–28, doi:10.1186/s12575-020-00133-5.
350. Stolarova, L.; Kleiblova, P.; Zemankova, P.; Stastna, B.; Janatova, M.; Soukupova, J.; Achatz, M.I.; Ambrosone, C.; Apostolou, P.; Arun, B.K.; et al. ENIGMA CHEK2gether Project: A Comprehensive Study Identifies Functionally Impaired CHEK2 Germline Missense Variants Associated with Increased Breast Cancer Risk. *Clinical*

- Cancer Research* **2023**, *29*, 3037–3050, doi:10.1158/1078-0432.CCR-23-0212.
351. Thomassen, M.; Blanco, A.; Montagna, M.; Hansen, T.V.O.; Pedersen, I.S.; Gutiérrez-Enríquez, S.; Menéndez, M.; Fachal, L.; Santamariña, M.; Steffensen, A.Y.; et al. Characterization of BRCA1 and BRCA2 Splicing Variants: A Collaborative Report by ENIGMA Consortium Members. *Breast Cancer Res Treat* **2012**, *132*, 1009–1023, doi:10.1007/s10549-011-1674-0.
 352. Spurdle, A.B.; Greville-Heygate, S.; Antoniou, A.C.; Brown, M.; Burke, L.; de la Hoya, M.; Domchek, S.; Dörk, T.; Firth, H. V.; Monteiro, A.N.; et al. Towards Controlled Terminology for Reporting Germline Cancer Susceptibility Variants: An ENIGMA Report. *J Med Genet* **2019**, *56*, 347–357, doi:10.1136/JMEDGENET-2018-105872.
 353. Zhao, W.; Wiese, C.; Kwon, Y.; Hromas, R.; Sung, P. The BRCA Tumor Suppressor Network in Chromosome Damage Repair by Homologous Recombination. *Annu Rev Biochem* **2019**, *88*, 221–245, doi:10.1146/ANNUREV-BIOCHEM-013118-111058.
 354. Fuh, K.; Mullen, M.; Blachut, B.; Stover, E.; Konstantinopoulos, P.; Liu, J.; Matulonis, U.; Khabele, D.; Mosammaparast, N.; Vindigni, A. Homologous Recombination Deficiency Real-Time Clinical Assays, Ready or Not? *Gynecol Oncol* **2020**, *159*, 877–886, doi:10.1016/J.YGYNO.2020.08.035.
 355. Fortuno, C.; Mester, J.; Pesaran, T.; Weitzel, J.N.; Dolinsky, J.; Yussuf, A.; McGoldrick, K.; Garber, J.E.; Savage, S.A.; Khincha, P.P.; et al. Suggested Application of HER2+ Breast Tumor Phenotype for Germline TP53 Variant Classification within ACMG/AMP Guidelines. *Hum Mutat* **2020**, *41*, 1555–1562, doi:10.1002/HUMU.24060.

APPENDIX I

Supplementary Tables Index

Supplementary Table 1. Deep intronic variants collected from literature (literature dataset).	321
Supplementary Table 2. Exonic variants altering splicing by affecting SREs and causing exon skipping and exonic non-splice altering variants.	333
Supplementary Table 3. Deep intronic splice-altering variants detected and characterised in tumour.	335
Supplementary Table 4. Deep intronic experimentally assessed variants.	343
Supplementary Table 5. List of HBOC risk and Lynch genes and their respective NCBI reference transcripts used to annotate the ESRseq scores from canonical exons and adjacent intronic nucleotides	346
Supplementary Table 6. Primers and PCR conditions for characterizing the splicing effect of deep intronic variants using RNA from patients (whole blood).	347
Supplementary Table 7. ΔESRseq threshold optimization (sensitivity + specificity).	349
Supplementary Table 8. Performance of ΔSpliceAI alone and sequentially combined with ESRseq tool for 233 intronic variants.	351
Supplementary Table 9. Related in silico, splicing and functional data from the 166 variants included in Part 2.	352
Supplementary Table 10. Detailed information from patients' samples.	365
Supplementary Table 11. Primers and PCR conditions.	367
Supplementary Table 12. Splicing events detected by semi-quantitative analysis in patients and mESC RNA samples.	376
Supplementary Table 13. Complete semi-quantitative mRNA splicing profiling of variant carriers and controls by capillary electrophoresis of fluorescent amplicons, resulting from RT-PCR using primers in exon 16 and exon 22 of BRCA1.	378
Supplementary Table 14. ΔSpliceAI(4999bp) performance in BRCA1 exon 18 variants analysed in patients (A) and minigenes (B).	378
Supplementary Table 15. Semi-quantitative mRNA splicing profiling of minigene constructs.	379
Supplementary Table 16. Complete semi-quantitative mRNA splicing profiling of mESC by capillary electrophoresis of fluorescent amplicons, resulting from RT-PCR using primers in exon 16 and exon 22 of BRCA1.	386
Supplementary Table 17. In silico, coding effect, variant position and experimental mRNA evidence for variant classification.	387
Supplementary Table 18. Functional data for BRCA1 exon 18 variants' classification.	402
Supplementary Table 19. Variant allele frequency data for BRCA1 exon 18 variants' classification.	402
Supplementary Table 20. Case-control data from BRIDGES study [37] for BRCA1 exon 18 variants' classification.	403

Supplementary Table 21. Evidence code assignment based on multifactorial likelihood clinical data.	404
Supplementary Table 22. Variant classification based on evidence available for the 166 variants included in the Part 2.	406
Supplementary Table 23. Clinical, tumoral, familial and multifactorial data obtained from carriers of BRCA1 c.5152+7A>G.	407
Supplementary Table 24. Published splicing results from variants included in Part 2.	409
Supplementary Table 25. Clinical and molecular characteristics of patients and tumours evaluated in Part 3.	413
Supplementary Table 26. Non-evaluable and excluded tumour samples.	424
Supplementary Table 27. In silico evidence for unreviewed variants' classification.	427
Supplementary Table 28. Splicing observed impact evidence for unreviewed variants' classification.	430
Supplementary Table 29. Allele frequency evidence for unreviewed variants' classification.	431
Supplementary Table 30. Functional evidence for unreviewed variants' classification.	431
Supplementary Table 31. Multifactorial likelihood clinical data for unreviewed variants' classification.	432
Supplementary Table 32. Case-control data from BRIDGES study for unreviewed variants' classification.	434
Supplementary Table 33. Likelihood ratio association analyses of RAD51% HRR status based on BC molecular subtype, tumour grade and age at diagnosis.	435
Supplementary Table 34. Likelihood ratio association analyses of RAD51% HRR status in tumours from germline BRCA1 pathogenic variant carriers based on BC molecular subtype, tumour grade and age at diagnosis.	437
Supplementary Table 35. Likelihood ratio association analyses of RAD51% HRR status in tumours from germline BRCA2 pathogenic variant carriers based on BC molecular subtype, tumour grade and age at diagnosis.	439
Supplementary Table 36. Likelihood ratio association analyses of RAD51% HRR status in tumours from germline PALB2 pathogenic variant carriers based on BC molecular subtype, tumour grade and age at diagnosis.	441

Supplementary Figures Index

<i>Supplementary figure 1. Correlation analysis between FL levels observed in mESC and Patients' FL per-allele fraction (A) and FL in minigenes (B).</i>	<i>443</i>
<i>Supplementary figure 2. Western blot analysis of protein samples from mESC expressing WT BRCA1 or BRCA1 exon 18 variants.....</i>	<i>443</i>
<i>Supplementary figure 3. Representative immunofluorescence images of two breast tumours.</i>	<i>444</i>

Supplementary Table 1. Deep intronic variants collected from literature (literature dataset).

This table is an adapted version of ST1 from Moles-Fernández et al. (2021) [287].

Comment	Splicing effect	ESRseq										ASpicelal (4999)		ASpicelal (50)		ESRseq scores intron acceptor		ESRseq scores exonized region		ESRseq scores intron donor		Abs. Difference	ESRseq scores pre variant		ESRseq scores post variant		
		Gene	Transcript	Variant	Reference	AG X	AG MA	AG DG	AG MA	AG DG	AG MA	Δ	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area
Splicing Regulatory Element	YES	ABCA4	NM_000350.2	c.1937+37C>G	PMID: 32307445	0.99	0.99	-0.21	-0.25	42.70	0.43	100	44.18	1.23	36	-69.50	-0.70	100	2.20	34.25	1.63	-51.45	-0.57				
	YES	ABCA4	NM_000350.2	c.1937+435C>G	PMID: 30643219	0.00	0.00	-1.88	3.17	-61.96	-0.62	100	32.94	0.25	134	-44.65	-0.45	100	0.53	-15.96	-0.18	31.73	0.35				
	YES	ABCA4	NM_000350.2	c.1938-514A>G	PMID: 32307445	0.70	0.69	-1.06	-0.88	-29.99	-0.30	100	-4.58	-0.04	109	-83.04	-0.83	100	0.67	6.23	0.07	-54.09	-0.60				
	YES	ABCA4	NM_000350.2	c.1938-619A>G	PMID: 31397521	0.79	0.69	0.51	-2.87	-72.34	-0.72	100	-46.57	-0.27	174	-12.96	-0.13	100	0.25	-17.88	-0.20	4.62	0.05				
Splicing Regulatory Element	YES	ABCA4	NM_000350.2	c.1938-621G>A	PMID: 32307445	0.28	0.28	-0.32	-3.00	-72.34	-0.72	100	-48.24	-0.28	174	-12.08	-0.12	100	0.28	-20.39	-0.23	5.18	0.06				
	YES	ABCA4	NM_000350.2	c.2919-826T>A	PMID: 31397521	0.56	0.56	-0.27	-0.93	-10.71	-0.11	100	18.67	0.14	133	7.41	0.07	100	0.10	-2.57	-0.03	6.43	0.07				
	YES	ABCA4	NM_000350.2	c.3050+370C>T	PMID: 31397521	0.91	0.91	0.38	-2.19	-73.11	-0.73	100	65.02	0.32	205	-79.85	-0.80	100	0.89	20.27	0.23	-59.43	-0.66				
	YES	ABCA4	NM_000350.2	c.3863-1064A>G	PMID: 32307445	0.00	0.00	-0.67	-2.27									0.63	-13.79	-0.15	-70.37	-0.78					
Splicing Regulatory Element	YES	ABCA4	NM_000350.2	c.4352+61G>A	PMID: 31397521	0.71	0.71	-1.48	1.48	36.23	0.36	100	31.67	0.56	57	26.52	0.27	100	0.56	30.22	0.67	10.08	0.11				
	YES	ABCA4	NM_000350.2	c.4539+2001G>A	PMID: 30670881	0.16	0.00	-0.58	3.02	-62.02	-0.62	100	-23.06	-0.07	345	-26.73	-0.27	100	0.06	-5.33	-0.06	0.44	0.00				
	YES	ABCA4	NM_000350.2	c.4539+2028C>T	PMID: 30670881	0.03	0.01	0.28	-0.43	-62.02	-0.62	100	-43.77	-0.13	345	-26.73	-0.27	100	0.16	9.08	0.10	-5.72	-0.06				
	YES	ABCA4	NM_000350.2	c.4539+2064C>T	PMID: 30670881	0.31	0.31	0.08	0.37	-62.02	-0.62	100	-38.99	-0.11	345	-26.73	-0.27	100	0.32	8.05	0.09	-20.88	-0.23				
Splicing Regulatory Element	YES	ABCA4	NM_000350.2	c.4634+741A>G	PMID: 32307445	0.48	0.48	-0.37	-2.24	-13.09	-0.13	100	-18.05	-0.14	127	-47.53	-0.48	100	0.51	4.18	0.05	-41.71	-0.46				
	YES	ABCA4	NM_000350.2	c.5196+1056A>G	PMID: 23918662	0.61	0.61	-3.95	1.95	-100.55	-1.01	100	-8.89	-0.05	177	-61.49	-0.61	100	0.80	29.49	0.33	-42.95	-0.48				
	YES	ABCA4	NM_000350.2	c.5196+1137G>A	PMID: 23918662	0.15	0.15	1.39	-0.23	-84.35	-0.84	100	7.27	0.10	73	-73.88	-0.74	100	0.88	-86.15	-0.96	-6.84	-0.08				
	YES	ABCA4	NM_000350.2	c.570+1798A>G	PMID: 32307445	0.98	0.98	1.96	-3.67	-46.94	-0.47	100	-11.07	-0.17	65	-59.54	-0.60	100	0.21	-34.71	-0.39	-53.47	-0.59				
YES	ABCA4	NM_000350.2	c.6148-84A>T	PMID: 32307445	0.57	0.57	1.52	-0.99	-14.03	-0.14	100	28.89	0.13	221	-24.55	-0.29	84	0.57	13.15	0.15	-37.87	-0.42					

Comment	Splicing effect	Gene	Transcript	Variant	Reference	ESRseq										ESRseq scores				ESRseq scores				Abs. Difference	ESRseq scores pre variant		ESRseq scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
						(50)										intron acceptor				exonized region					intron donor				scores pre variant		scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
						Δspliceal (4999)					Δspliceal (50)					Area		Size		Norm. SRF area		Area			Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area

Comment	Splicing effect	ΔspliceAI (50)										ESRseq										ESRseq scores intron acceptor				ESRseq scores exonized region				ESRseq scores intron donor				Abs. Difference		ESRseq scores pre variant		ESRseq scores post variant	
		ΔspliceAI (499)										ESRseq										ESRseq scores intron acceptor				ESRseq scores exonized region				ESRseq scores intron donor				Abs. Difference		ESRseq scores pre variant		ESRseq scores post variant	
		AG X	AG MA	AG DG	AG MA	AG X	WT	Δ	Area	Size	Norm. SRF area	AG X	AG MA	AG DG	AG MA	AG X	WT	Δ	Area	Size	Norm. SRF area	AG X	AG MA	AG DG	AG MA	AG X	WT	Δ	Area	Size	Norm. SRF area	RE area pre)	(Norm.SR E Area Post)	Area	Norm. SRF area	Area	Norm. SRF area		
YES	CNGB ₃	NM_019098.4	c.1663-2137C>T		PMID: 31544997	0.30	0.30	0.39	-2.88	-153.84	-1.54	100	8.75	0.09	99	-60.66	-0.61	100	0.54	1.31	0.01	-47.54	-0.53																
YES	CNGB ₃	NM_019098.4	c.852+4751A>T		PMID: 32881472	0.60	0.60	-1.16	0.39	-85.88	-0.86	100	122.19	0.41	301	61.77	0.62	100	1.50	-95.70	-1.06	39.07	0.43																
YES	COL2 _{A1}	NM_001844.4	c.86-50C>G		PMID: 32196734	0.67	0.67	-1.46	-0.34	-47.56	-0.48	100	-15.97	-0.33	49	45.81	0.46	100	0.28	-37.85	-0.42	-12.72	-0.14																
YES	COL4 _{A5}	NM_033380.2	c.1948+894C>G		PMID: 25183659	0.86	0.86	-0.79	-0.08	-46.02	-0.46	100	-41.18	-0.39	106	-76.59	-0.77	100	0.02	-30.48	-0.34	-28.81	-0.32																
YES	COL4 _{A5}	NM_033380.2	c.4528+1754T>G		PMID: 25183659	0.99	0.99	-3.27	0.60	-82.29	-0.82	100	-8.42	-0.10	84	-66.95	-0.67	100	0.27	-14.84	-0.16	-39.53	-0.44																
YES	COL4 _{A5}	NM_033380.2	c.4529-345A>G		PMID: 25183659	0.93	0.93	-0.60	-1.71	-61.62	-0.62	100	30.19	0.36	84	-51.35	-0.51	100	0.73	19.66	0.22	-46.06	-0.51																
YES	COL4 _{A5}	NM_033380.2	c.609+875G>T		PMID: 25183659	0.91	0.91	0.66	-3.62	-90.92	-0.91	100	-16.71	-0.24	70	-35.55	-0.36	100	0.42	7.32	0.08	-30.53	-0.34																
YES	CTDP1	NM_004715.4	c.863+389C>T		PMID: 14517542	0.13	0.13	0.10	-2.03	-107.13	-1.07	100	-6.21	-0.07	95	-76.98	-0.77	100	0.72	-12.50	-0.14	-77.46	-0.86																
YES	CYBB	NM_000397.3	c.675-1157A>G		PMID: 16516412; PMID: 11566256	0.73	0.73	-2.97	1.27	-93.80	-0.94	100	11.48	0.21	56	-41.67	-0.42	100	0.13	-25.14	-0.28	-13.62	-0.15																
YES	DBT	NM_001918.2	c.1018-550A>G		PMID: 9621512	0.82	0.79	-0.84	-1.87	-141.45	-1.41	100	19.71	0.16	126	-71.04	-0.71	100	0.69	2.75	0.03	-59.63	-0.66																
YES	DICER ₁	NM_177438.3	c.5364+1187T>G		PMID: 29469200	0.96	0.96	-1.62	-1.74	-132.38	-1.32	100	0.29	0.00	132	-129.96	-1.30	100	1.46	-114.36	-1.27	16.90	0.19																
YES	DMD	NM_004006.1	c.3787-843C>A		PMID: 20485447; PMID: 14659407; PMID: 17041906; PMID: 19040728; PMID: 10094556; PMID: 18059005	0.56	0.56	-0.66	-1.17	-83.48	-0.83	100	-64.19	-0.54	119	-84.47	-0.84	100	0.20	-56.27	-0.63	-74.31	-0.83																
YES	DMD	NM_004006.1	c.9225-285A>G		PMID: 20485447; PMID: 12754707; PMID: 18059005	0.76	0.76	-0.26	-1.40	-30.87	-0.31	100	9.45	0.16	58	-7.08	-0.07	100	0.14	-5.99	-0.07	6.72	0.07																
YES	DMD	NM_004006.1	c.93+5590T>A		PMID: 20485447; PMID: 12754707	0.92	0.92	-0.79	-1.90	-90.01	-0.90	100	-33.38	-0.25	132	-84.21	-0.84	100	0.76	-79.07	-0.88	-11.01	-0.12																
YES	DMD	NM_004006.1	c.9807+2714C>T		PMID: 20485447; PMID: 14659407; PMID: 17041906; PMID: 19040728; PMID: 10094556; PMID: 18059005	0.66	0.66	0.47	-2.32	-34.41	-0.34	100	-14.94	-0.12	121	-58.37	-0.58	100	0.51	-14.42	-0.16	-60.65	-0.67																

Comment	Splicing effect	ESRseq										ESRseq scores				ESRseq scores				Abs. Difference	ESRseq scores pre variant		ESRseq scores post variant									
		(50) ΔspliceAI										intron acceptor				exonized region					intron donor											
		ΔspliceAI (4999)	ΔspliceAI (50)	WT		Δ		Area		Size		Norm. SRF area		Area		Size		Norm. SRF area			Area		Size		Norm. SRF area		Area		Size			
AG	AG	DG	DG	MA	MA	X	X	WT	Δ	Area	Size	Area	Size	Area	Size	Area	Size	Area	Size	Area	Size	Area	Size	Area	Size	Area	Size	Area	Size			
YES	DPVD	NM_000110.3	c.1129-5923C>G	PMID: 20803296		0.77	0.77	0.23	-0.88	-53.01	-0.53	100	-12.83	-0.29	44	-46.66	-0.47	100	0.25	RE area pre)-	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size	
YES	DYSF	NM_003494.4	c.4886+1249G>T	PMID: 25493284		0.26	0.26	0.35	-2.99	-41.33	-0.41	100	-46.33	-0.26	177	-29.03	-0.29	100	0.22	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	F7	NG_009262.1	c.571+78G>A	PMID: 31273093		0.65	0.65	0.86	-1.80	28.44	0.28	100	5.92	0.07	79	-22.92	-0.23	100	0.22	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	F7	NG_009262.1	c.572-392C>G	PMID: 31273093		0.53	0.53	-0.64	-0.29	3.25	0.03	100	47.12	0.16	289	-33.16	-0.33	100	0.75	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	F8	NM_000132.3	c.143+1567A>G	PMID: 32812322		0.54	0.54	0.00	-1.52	-133.49	-1.33	100	-39.99	-0.21	191	-122.72	-1.23	100	1.00	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	F8	NM_000132.3	c.1443+602A>G	PMID: 32812322		0.88	0.88	2.20	-2.52	-76.54	-0.77	100	-113.00	-0.43	262	-81.47	-0.81	100	0.79	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	F8	NM_000132.3	c.2113+1153G>C	PMID: 33094873		0.33	0.33	-3.61	4.42	-119.35	-1.19	100	-30.24	-0.27	113	-58.19	-0.58	100	1.23	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	F8	NM_000132.3	c.2113+601G>A	PMID: 32812322		0.14	0.14	-1.04	1.02	-107.40	-1.07	100	34.77	0.28	122	-69.73	-0.70	100	1.02	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	F8	NM_000132.3	c.5220-8563A>G	PMID: 33094873		0.71	0.71	-1.32	-1.18	-33.78	-0.34	100	18.66	0.15	124	-100.48	-1.00	100	0.39	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	FECH	NM_000140.4	c.315-48T>C	PMID: 31273344		0.17	0.17	-1.58	1.85	-84.27	-0.84	100	-49.66	-0.79	63	58.82	0.41	144	0.66	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	FERMT1	NM_017671.4	c.1139+740G>A	PMID: 26083552		0.86	0.86	1.95	-2.44	-73.42	-0.73	100	-13.35	-0.11	124	-83.84	-0.84	100	0.78	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	FGG	NM_021870.2	c.667-320A>T	PMID: 17854317; PMID: 19551918		0.78	0.78	0.61	-2.51	-34.26	-0.34	100	-17.20	-0.23	75	-91.79	-0.92	100	0.50	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	GALT	NM_000155.3	c.1059+390A>G	PMID: 32748411		0.15	0.15	-0.49	-0.36	-57.54	-0.58	100	-29.44	-0.16	187	-40.53	-0.41	100	0.39	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	GBA	NM_000157.4	c.999+242C>A	ClinVar VCV000689500.1 (accessed May 6, 2021)		0.45	0.45	-0.56	-1.60	23.68	0.24	100	-41.63	-0.17	239	-39.77	-0.40	100	0.00	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	GLA	NM_000169.2	c.639+919G>A	PMID: 27595546; PMID: 27779620		0.22	0.00	-1.23	-0.50	-26.72	-0.27	100	-38.14	-0.67	57	-60.71	-0.61	100	0.32	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	GPR143	NM_000273.2	c.659-131T>G	PMID: 26061757; PMID: 16550551		0.91	0.91	-0.02	0.18	-110.24	-1.10	100	22.79	0.56	41	6.02	0.06	100	0.84	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	GPR179	NM_0010043	c.903+343G>A	PMID: 32881472		0.86	0.86	1.90	-2.47	-22.54	-0.23	100	51.52	0.26	198	-8.15	-0.08	100	0.07	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	HADH	NM_005327.5	c.636+471G>T	PMID: 26316438; PMID: 23273570		0.85	0.85	2.25	-4.08	-71.12	-0.71	100	12.03	0.14	84	-69.85	-0.70	100	0.57	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size
YES	HSPG2	NM_005529.5	c.574+481C>T	PMID: 16927315		0.89	0.89	0.95	-1.49	-92.16	-0.92	100	16.40	0.13	130	49.86	0.50	100	0.23	Norm. SRF	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td><td>Area<td>Size</td></td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td><td>Area<td>Size</td></td></td>	Size	Area <td>Size</td> <td>Area<td>Size</td></td>	Size	Area <td>Size</td>	Size

Comment	Splicing effect	Gene	Transcript	Variant	Reference	ESRseq										ESRseq scores intron acceptor	ESRseq scores exonized region	ESRseq scores intron donor	Abs. Differenc e	ESRseq scores pre variant	ESRseq scores post variant		
						ASpliceal (499)					ASpliceal (50)												
						AG MA X	AG DG MA X	WT	Δ	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area							Area	Size
YES	KRT17	NM_004912.3	c.262+132_262+133del	PMID: 24251678	0.82	0.00	0.00	1.19	-92.14	-0.92	100	100	-51.31	-0.52	99	-84.12	-0.84	100	0.65	-82.40	-0.92	-23.97	-0.27
YES	MLC1	NM_015166.3	c.895-226T>G	PMID: 22552818	0.73	0.50	-0.04	0.67	38.45	0.38	100	246	45.62	0.19	246	-24.11	-0.24	100	0.33	5.60	0.06	-24.11	-0.27
YES	MTM1	NM_000252.2	c.1468-577A>G	PMID: 33505229	0.79	0.79	0.58	-0.66	-116.49	-1.16	100	48	-24.34	-0.51	48	-18.26	-0.18	100	0.84	-90.01	-1.00	-14.55	-0.16
YES	MUT	NM_000255.4	c.1956+3691C>A	PMID: 17966092	0.47	0.47	1.23	-1.43	-69.34	-0.69	100	76	5.17	0.07	76	-86.29	-0.86	100	0.83	-12.09	-0.13	-87.03	-0.97
YES	MYBP C3	NM_000256.3	c.1224-52G>A	PMID: 32163302	0.99	0.99	0.95	-1.48	-31.00	-0.31	100	50	47.21	0.94	50	-32.96	-0.33	100	0.00				
YES	NDUF S7	NM_024407.5	c.17-1167C>G	Lebon et al., 2007 PMID: 17604671	0.31	0.27	1.19	-1.39	-33.36	-0.33	100	122	23.58	0.19	122	-54.67	-0.55	100	0.71	17.30	0.19	-46.50	-0.52
YES	NF1	NM_0010424 92.2	c.888+651T>A	Svaasand et al., Hereditary Genet 2015, 4:3 DOI: 10.4172/2161-1041.1000152	0.66	0.66	-0.49	-3.03	-105.90	-1.06	100	132	-30.08	-0.23	132	-76.79	-0.77	100	0.63	-84.95	-0.94	-28.64	-0.32
Splicing Regulatory Element	YES	NOTC H3	NM_000435.2 c.341-26_341-24del	PMID: 23587639	0.28	0.00	-1.37	0.33	21.30	0.21	100	76	-15.18	-0.20	76	45.71	0.46	100	0.93	-16.97	-0.46	3.81	0.48
YES	NPC1	NM_000271.5	c.1554-1009G>A	PMID: 19718781	0.47	0.38	0.86	-2.23	-44.89	-0.45	100	194	48.63	0.25	194	-9.54	-0.10	100	0.57	54.22	0.60	2.98	0.03
YES	NPHP 3	NM_153240.5	c.671-996C>G	PMID: 33024573	0.15	0.15	0.39	-2.02	-75.64	-0.76	100	163	19.42	0.12	163	-54.84	-0.55	100	0.17	20.06	0.22	4.91	0.05
YES	OCRL	NM_000276.4	c.238+4701A>G	PMID: 27790796	0.97	0.95	-0.31	-2.95	-52.84	-0.53	100	66	-8.46	-0.13	66	2.00	0.02	100	0.60	-21.71	-0.24	31.90	0.35
YES	OFD1	NM_003611.2	c.935+706A>G	PMID: 22619378	0.61	0.39	-1.90	0.19	-65.68	-0.66	100	62	-23.19	-0.37	62	-24.89	-0.25	100	0.33	-57.45	-0.64	-27.70	-0.31
YES	OPA1	NG_008805.2	c.610+360G>A	PMID: 27874857	0.33	0.36	2.03	-2.42	-68.40	-0.68	100	93	9.92	0.11	93	-15.65	-0.16	100	0.79	-68.98	-0.77	2.51	0.03
YES	OPA1	NM_130837.2	c.610+364G>A	PMID: 24970096	0.36	0.36	1.98	-2.15	-63.41	-0.63	100	93	6.52	0.07	93	-15.65	-0.16	100	0.83	-69.85	-0.78	5.06	0.06
YES	OTC	NM_000531.6	c.540+265G>A	PMID: 18204299	0.66	0.65	0.46	-1.28	-73.78	-0.74	100	135	-9.43	-0.07	135	-110.15	-1.10	100	0.75	-65.76	-0.73	1.70	0.02
YES	P3H1 (LEPR Et)	NM_022356.3	c.1224-80G>A	PMID: 32235935	0.62	0.62	-2.07	-2.02	36.86	0.37	100	174	35.97	0.21	174	91.36	0.91	100	0.59	54.54	0.69	6.62	0.10
YES	PCCB	NM_0011780	c.603+462A>G	PMID: 17966092	0.01	0.00	-0.16	1.07	-2.87	-0.03	100	72	-55.08	-0.76	72	-90.99	-0.91	100	0.71	-75.10	-0.83	-11.22	-0.12
YES	PKHD 1	NM_138694.3	c.7350+653A>G	PMID: 19021639	0.00	0.00	0.27	1.11	-96.44	-0.96	100	116	-34.27	-0.30	116	-71.63	-0.72	100	0.26	-100.99	-1.12	-77.61	-0.86

Comment	Splicing effect	Gene	Transcript	Variant	Reference	ΔspliceAI (4999)										ESRseq										ESRseq scores				ESRseq scores				ESRseq scores				ESRseq scores pre variant				ESRseq scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
						ΔspliceAI (999)					ΔspliceAI (50)					ESRseq					intron acceptor					exonized region					intron donor					Abs. Difference				ESRseq scores pre variant				ESRseq scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
						AG MA	AG DG	AG MA	AG DG	Δ	WT	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area	Size	Norm. SRF area	Area

Comment	Splicing effect	Δspliceal (4999)										Δspliceal (50)										ESRseq										ESRseq scores intron acceptor				ESRseq scores exonized region				ESRseq scores intron donor				Abs. Difference				ESRseq scores pre variant				ESRseq scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
		AG DG MA X										AG DG MA X										WT Δ										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size Norm. SRF area										Area Size									

Comment	Splicing effect	Gene	Transcript	Variant	Reference	ESRseq										ΔspliceAI (499)		ΔspliceAI (50)		ESRseq scores intron acceptor				ESRseq scores exonized region				ESRseq scores intron donor				Abs. Difference		ESRseq scores pre variant		ESRseq scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
						AG X	AG MA	AG DG	AG MA	AG DG	AG X	Δ	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area	Area	Size	Norm. SRE area

Comment	Splicing effect	Gene	Transcript	Variant	Reference	ESRseq										ESRseq scores intron acceptor	ESRseq scores exonized region	ESRseq scores intron donor	Abs. Difference	ESRseq scores pre variant	ESRseq scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
						ΔspliceAI (4999)					ΔspliceAI (50)											ESRseq scores intron acceptor					ESRseq scores exonized region					ESRseq scores intron donor																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
						AG X	AG MA	AG DG	AG MA	AG DG	WT	Δ	Area	Size	Norm. SRE area							ESRseq scores intron acceptor	Area	Size	Norm. SRE area	ESRseq scores exonized region	Area	Size	Norm. SRE area	ESRseq scores intron donor	Area	Size	Norm. SRE area	ESRseq scores intron donor																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
NO	ABCA4	NM_000350.2	c.442+1311C>T	PMID: 32307445		0.00	0.00	0.70	-0.43																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					

Comment	Splicing effect	ΔspliceAI (999)										ΔspliceAI (50)										ESRseq			ESRseq scores intron acceptor			ESRseq scores exonized region			ESRseq scores intron donor			Abs. Difference	ESRseq scores pre variant		ESRseq scores post variant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
		AG					DG					MA					WT					Δ			Norm. SRE area			Norm. SRE area			Norm. SRE area			e	Norm. SRE area		Norm. SRE area																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Comment	Splicing effect	Gene	Transcript	Variant	Reference	ESRseq										ESRseq scores intron donor	ESRseq scores pre variant	ESRseq scores post variant		
						ΔspliceAI (4999)					ESRseq scores intron acceptor								ESRseq scores exonized region	Abs. Difference
						AG X	AG MA	AG DG	AG MA	Δ	Area	Size	Norm. SRE area	Area	Size					
NO	NO	BRCA1	NM_007294.3	c.4676-284_4676-283insAGTGAACA	PMID: 30472649	0.02	0.00	-1.26	-1.39						0.49	-62.34	-0.62	-97.43	-0.97	
NO	NO	BRCA1	NM_007294.3	c.4987-635A>G	PMID: 30472649	0.00	0.00	-1.13	0.73						1.47	-166.97	-1.67	-54.87	-0.55	
NO	NO	BRCA1	NM_007294.3	c.5075-1061T>A	PMID: 30472649	0.00	0.00	-3.06	2.60						0.27	-10.01	-0.10	-29.95	-0.30	
NO	NO	BRCA1	NM_007294.3	c.5075-1445G>C	PMID: 30472649	0.00	0.00	2.76	-2.25						0.20	-26.18	-0.26	-25.22	-0.25	
NO	NO	BRCA1	NM_007294.3	c.5075-1544G>C	PMID: 30472649	0.00	0.00	-2.39	2.72						0.30	-58.57	-0.59	-16.64	-0.17	
NO	NO	BRCA1	NM_007294.3	c.5278-1708_5278-1697del	PMID: 30472649	0.00	0.00	-0.65	0.55						0.42	-16.79	-0.17	-82.35	-0.82	
NO	NO	BRCA1	NM_007294.3	c.5278-801T>G	PMID: 30472649	0.00	0.00	-1.27	1.76						0.36	-19.46	-0.19	4.35	0.04	
NO	NO	BRCA1	NM_007294.3	c.671-260C>T	PMID: 30472649	0.00	0.00	0.12	-1.49						0.17	-91.27	-0.91	-	-1.03	
NO	NO	BRCA1	NM_007294.3	c.81-1360delA	PMID: 30472649	0.01	0.00	-1.78	-0.03						0.35	-61.07	-0.61	-49.03	-0.49	
NO	NO	BRCA2	NM_000059.3	c.1909+1343T>C	PMID: 30472649	0.00	0.00	-2.47	1.39						0.42	-45.88	-0.46	-77.74	-0.78	
NO	NO	BRCA2	NM_000059.3	c.316+1053C>T	PMID: 30472649	0.00	0.00	-1.00	-1.64						0.22	-62.28	-0.62	-52.19	-0.52	
NO	NO	BRCA2	NM_000059.3	c.316+1669A>G	PMID: 30472649	0.00	0.00	-0.56	0.33						0.70	-43.32	-0.43	-81.88	-0.82	
NO	NO	BRCA2	NM_000059.3	c.316+2196A>G	PMID: 30472649	0.00	0.00	0.39	-1.19						0.19	-65.84	-0.66	-36.41	-0.36	
NO	NO	BRCA2	NM_000059.3	c.316+368A>G	PMID: 30472649	0.04	0.04	-1.72	0.79						0.15	-53.57	-0.54	-76.98	-0.77	
NO	NO	BRCA2	NM_000059.3	c.317-339C>T	PMID: 30472649	0.00	0.00	0.38	1.31						0.51	-4.83	-0.05	-57.86	-0.58	
NO	NO	BRCA2	NM_000059.3	c.-39-137_-39-135del	PMID: 30472649	0.03	0.00	-1.84	1.40						0.41	-43.48	-0.43	-11.50	-0.12	
NO	NO	BRCA2	NM_000059.3	c.425+33A>G	PMID: 30472649	0.00	0.00	-1.42	0.25						0.92	-44.35	-1.64	-86.07	-0.86	
NO	NO	BRCA2	NM_000059.3	c.632-1101G>A	PMID: 30472649	0.01	0.00	-0.01	1.61						0.47	-32.30	-0.32	-59.87	-0.60	
NO	NO	BRCA2	NM_000059.3	c.632-1136dup	PMID: 30472649	0.01	0.00	0.60	0.06						0.50	-68.82	-0.69	-14.46	-0.14	
NO	NO	BRCA2	NM_000059.3	c.632-943dup	PMID: 30472649	0.01	0.00	-1.09	0.27						0.36	-42.55	-0.43	-23.72	-0.24	
NO	NO	BRCA2	NM_000059.3	c.6841+1018A>G	PMID: 30472649	0.00	0.00	-1.13	0.71						0.19	-10.66	-0.11	-15.03	-0.15	
NO	NO	BRCA2	NM_000059.3	c.6937+302C>T	PMID: 30472649	0.00	0.00	0.76	0.09						0.02	-67.08	-0.67	-72.98	-0.73	

Comment	Splicing effect	ΔSpliceAI										ESRseq			ESRseq scores			ESRseq scores			ESRseq scores			ESRseq scores pre variant			ESRseq scores post variant		
		(4999)					(50)								Intron acceptor			exonized region			intron donor			Abs. Difference					
		AG	DG	MA	X	AG	DG	MA	X	WT	Δ	Area	Size	Area	Norm. SRE area	Size	Area	Norm. SRE area	Size	Area	Norm. SRE area	Size	Area	Norm. SRE area	Size	Area	Norm. SRE area	Size	Area
NO	BRCA2	NM_000059.3	c.7007+2200C>T	PMID: 30472649	0.00	0.00	0.00	0.00	1.46	-0.07																			
NO	BRCA2	NM_000059.3	c.7008-1289A>C	PMID: 30472649	0.01	0.00	-0.36	-0.47																					
NO	BRCA2	NM_000059.3	c.7008-1374G>T	PMID: 30472649	0.00	0.00	-0.92	-0.61																					
NO	BRCA2	NM_000059.3	c.7436-182G>T	PMID: 30472649	0.00	0.00	-2.30	-0.31																					
NO	BRCA2	NM_000059.3	c.7805+973C>G	PMID: 30472649	0.00	0.00	-0.79	0.31																					
NO	BRCA2	NM_000059.3	c.8331+1543C>T	PMID: 30472649	0.00	0.00	-1.49	-1.04																					
NO	BRCA2	NM_000059.3	c.8331+2157A>T	PMID: 30472649	0.00	0.00	-2.41	0.43																					
NO	BRCA2	NM_000059.3	c.8332-13T>G	PMID: 30472649	0.17	0.17	-2.70	-1.25																					
NO	BRCA2	NM_000059.3	c.8332-2638C>T	PMID: 30472649	0.00	0.00	-0.34	-3.01																					
NO	BRCA2	NM_000059.3	c.8332-2898A>G	PMID: 30472649	0.00	0.00	0.00	0.57																					
NO	BRCA2	NM_000059.3	c.8633-1008G>A	PMID: 30472649	0.00	0.00	0.61	-0.61																					
NO	BRCA2	NM_000059.3	c.8754+759A>G	PMID: 30472649	0.00	0.00	-0.66	1.52																					
NO	BRCA2	NM_000059.3	c.9256+256A>G	PMID: 30472649	0.00	0.00	-0.73	-0.19																					
NO	BRCA2	NM_000059.3	c.9257-3591G>T	PMID: 30472649	0.00	0.00	0.50	-1.71																					
NO	BRCA2	NM_000059.3	c.9257-4843C>T	PMID: 30472649	0.00	0.00	-0.34	-2.62																					
NO	BRCA2	NM_000059.3	c.9257-882T>C	PMID: 30472649	0.00	0.00	-1.02	1.72																					
NO	BRCA2	NM_000059.3	c.9501+599T>C	PMID: 30472649	0.00	0.00	-1.49	1.07																					

ΔSpliceAI (4999) / ΔSpliceAI (50): ΔScore ranges from 0 to 1 and can be interpreted as the probability that the variant affects splicing at any position around it. For each variant, SpliceAI analyse within a window (+/- 4999 or +/- 50 in this case) to see how the variant affects the probabilities of different positions in the pre-mRNA of being splice acceptors or donors. ESRseq difference in absolute values: differences of normalized SRE area (norm.SRE area) between previous and posterior variant sequences. AG= Acceptor Gain; AL= Acceptor Loss; DG=Donor Gain; DL= Donor Loss; MAX AG DG= maximum value in AG or DG, Norm: normalized.

Supplementary Table 2. Exonic variants altering splicing by affecting SREs and causing exon skipping and exonic non-splice altering variants.
 Dataset retrieved from Tubeuf et al., 2020[172].
 This table is an adapted version of ST2 from Moles-Fernández et al. (2021)[287]. There are only shown 20 splicing altering and 20 non-splice altering variants in BRCA1.

Exon skipping	Gene	Transcript	Variant (cno men)	ΔSpliceAI(4999)			ΔSpliceAI(50)		
				MAX AG/DG	MAX AL/DL	MAX	MAX AG/DG	MAX AL/DL	MAX
YES	BRCA1	NM_007294.3	c.224A>T	0.55	0.62	0.55	0.55	0.45	0.45
YES	BRCA1	NM_007294.3	c.230C>A	0.04	0.28	0.04	0.04	0.24	0.24
YES	BRCA1	NM_007294.3	c.230C>G	0	0.24	0	0.24	0	0.2
YES	BRCA1	NM_007294.3	c.231G>T	0	0.54	0	0.54	0	0.54
YES	BRCA1	NM_007294.3	c.692C>T	0	0.04	0	0.04	0	0.01
YES	BRCA1	NM_007294.3	c.693G>A	0	0.1	0	0.1	0	0.05
YES	BRCA1	NM_007294.3	c.825C>T	0.02	0	0.02	0	0.02	0
YES	BRCA1	NM_007294.3	c.693G>C	0	0.04	0	0.04	0	0.01
YES	BRCA1	NM_007294.3	c.693G>T	0	0.06	0	0.06	0	0.02
YES	BRCA1	NM_007294.3	c.732T>C	0.05	0	0	0	0	0
YES	BRCA1	NM_007294.3	c.688G>C	0.02	0.04	0.02	0.04	0.02	0.01
YES	BRCA1	NM_007294.3	c.690G>T	0	0.03	0	0.03	0	0
YES	BRCA1	NM_007294.3	c.691A>C	0.02	0	0	0	0	0
YES	BRCA1	NM_007294.3	c.692C>A	0	0.02	0	0.02	0	0.01
YES	BRCA1	NM_007294.3	c.694G>C	0	0.11	0	0.11	0	0.03
YES	BRCA1	NM_007294.3	c.694G>T	0.01	0.12	0.01	0.12	0.01	0.04
YES	BRCA1	NM_007294.3	c.5078C>T	0.01	0.21	0.01	0.21	0.01	0.21
YES	BRCA1	NM_007294.3	c.5080G>A	0	0.21	0	0.21	0	0.21
YES	BRCA1	NM_007294.3	c.5080G>T	0	0.41	0	0.41	0	0.4
YES	BRCA1	NM_007294.3	c.5123C>A	0	0.22	0	0.22	0	0.22
NO	BRCA1	NM_007294.3	c.216C>A	0.02	0.04	0.02	0.04	0.02	0.04
NO	BRCA1	NM_007294.3	c.224A>G	0.1	0	0.1	0	0.1	0
NO	BRCA1	NM_007294.3	c.225A>C	0.01	0.01	0.01	0.01	0	0.01
NO	BRCA1	NM_007294.3	c.225A>G	0	0.03	0	0.03	0	0.03
NO	BRCA1	NM_007294.3	c.225A>T	0	0.19	0	0.19	0	0.19

Exon skipping	Gene	Transcript	Variant (cno men)	ΔSpliceAI(4999)			ΔSpliceAI(50)		
				MAX AG/DG	MAX AL/DL	MAX AG/DG	MAX AL/DL	MAX AG/DG	MAX AL/DL
NO	BRCA1	NM_007294.3	c.226A>C	0.16	0	0.14	0		
NO	BRCA1	NM_007294.3	c.226A>G	0.46	0.11	0.46	0.06		
NO	BRCA1	NM_007294.3	c.226A>T	0.04	0	0.04	0		
NO	BRCA1	NM_007294.3	c.227G>A	0.06	0.01	0.03	0.01		
NO	BRCA1	NM_007294.3	c.227G>C	0.18	0.01	0.16	0.01		
NO	BRCA1	NM_007294.3	c.227G>T	0.03	0.01	0.01	0.01		
NO	BRCA1	NM_007294.3	c.229A>C	0.07	0.01	0.06	0.01		
NO	BRCA1	NM_007294.3	c.230C>T	0	0.37	0	0.31		
NO	BRCA1	NM_007294.3	c.231G>A	0	0.05	0	0.01		
NO	BRCA1	NM_007294.3	c.231G>C	0	0.16	0	0.09		
NO	BRCA1	NM_007294.3	c.232A>G	0.06	0	0.04	0		
NO	BRCA1	NM_007294.3	c.232A>T	0.05	0.14	0.05	0.14		
NO	BRCA1	NM_007294.3	c.233G>A	0	0.06	0	0.06		
NO	BRCA1	NM_007294.3	c.233G>C	0.09	0	0.09	0		
NO	BRCA1	NM_007294.3	c.233G>T	0.01	0.16	0.01	0.15		

ΔSpliceAI (4999) / ΔSpliceAI (50): ΔScore ranges from 0 to 1 and can be interpreted as the probability that the variant affects splicing at any position around it. For each variant, SpliceAI analyse within a window (+/- 4999 or +/- 50) to see how the variant affects the probabilities of different positions in the pre-mRNA of being splice acceptors or donors. AG: Acceptor Gain, AL: Acceptor Loss, DG: Donor Gain, DL: Donor Loss, MAX AG DG: maximum value in AG or DG.

Supplementary Table 3. Deep intronic splice-altering variants detected and characterised in tumour.

Data retrieved from *Jung et al. (2021)[259]*.

This table is an adapted version of S13 from *Moles-Fernández et al. (2021) [287]*.

ΔSpliceAI (4999)													
Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX			MAX			ΔESRseq ^a	DS_DL ^a	Abs. difference ^{a,b}
					AG	DG	DL	DS_AG	DS_AL	DS_DG			
1	PEX10	2338369	C	A	0.18	0.05	0.18	0.05	0	0.01	-	-	-
1	ERRFI1	8075064	G	A	0.42	0.04	0.06	0.01	0.42	0.04	-	-	-
1	ENO1	8938033	G	A	0.53	0.04	0.03	0.01	0.53	0.04	-	-	-
1	PLEKHM2	16059869	C	T	0.22	0.02	0.22	0.02	0	0	-	-	-
1	TMEM54	33360582	C	G	0.74	0.89	0.74	0.89	0.01	0	-	-	-
1	NFYC	41229163	C	T	0.7	0.03	0.69	0	0.7	0.03	-	-	-
1	ZFYVE9	52774660	C	T	0.19	0.01	0.11	0.01	0.19	0.01	-	-	-
1	RABGGTB	76252612	T	C	0.01	0	0.01	0	0	0	1.52	0.22	0.22
1	MOV10	113242108	G	A	0.04	0	0.04	0	0	0	1.70	0.23	0.23
1	TARS2	150479279	C	T	0.05	0.03	0.05	0.03	0	0	-	-	-
1	C1orf56	151022052	G	T	0.17	0.05	0.17	0.05	0.03	0.01	-	-	-
1	NES	156643796	C	A	0.18	0.01	0.17	0	0.18	0.01	-	-	-
1	GLUL	182355813	C	T	0.28	0	0.28	0	0.22	0	-	-	-
1	SNRPE	203831806	C	G	0.27	0.02	0.07	0.02	0.27	0.01	-	-	-
1	PFKFB2	207233335	T	C	0	0.01	0	0.01	0	0.01	0.18	0.40	0.40
1	INTS7	212180229	T	C	0.2	0.02	0.07	0.02	0.2	0.01	-	-	-
1	EGLN1	231505310	A	C	0.04	0	0.02	0	0.04	0	-0.42	0.69	0.69
2	NCOA1	24954705	T	C	0.01	0	0	0	0.01	0	0.03	1.05	1.05
2	TMEM214	27263151	G	T	0.21	0.05	0.21	0.05	0.01	0	-	-	-
2	EIF2B4	27592535	G	C	0.71	0.13	0.54	0.01	0.71	0.13	-	-	-
2	SOS1	39346993	G	C	0.61	0.05	0.37	0	0.61	0.05	-	-	-
2	EML4	42553173	A	C	0.12	0.01	0.12	0.01	0	0	-	-	-
2	MDH1	63827551	C	T	0.09	0	0.09	0	0.04	0	-	-	-
2	UGP2	64111423	A	C	0.25	0	0.21	0	0.25	0	-	-	-
2	SPR	73115237	G	A	0.15	0.05	0.02	0.01	0.15	0.05	-	-	-

ΔSpliceAI(4999)													
Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX AG	MAX AL	DS_AG	DS_AL	DS_DG	DS_DL	ΔESRseq ^a	Abs. difference ^{a,b}	
2	RPL31	101619794	G	C	0.31	0.01	0.27	0.01	0.31	0.01	-	-	
2	RPL31	101620767	G	A	0.06	0.01	0	0	0.06	0.01	-	-	
2	ZC3H8	112999006	C	G	0.08	0	0.08	0	0.08	0	-	-	
2	PSD4	113955280	G	C	0.02	0.2	0.02	0.2	0.01	0.03	-	-	
2	SPOPL	139312034	G	A	0	0	0	0	0	0	1.96	0.73	
2	GORASP2	171805024	T	G	0.64	0.03	0	0	0.64	0.03	-	-	
2	NOP58	203157652	A	G	0.97	0.13	0	0	0.97	0.13	-	-	
2	ZFAND2B	220073405	G	A	0.35	0.06	0	0.01	0.35	0.06	-	-	
2	SGPP2	223398709	G	T	0.5	0.01	0.37	0	0.5	0.01	-	-	
2	SH3BP4	235945099	A	G	0.4	0	0.4	0	0.33	0	-	-	
2	GPC1	241405335	C	A	0.65	0.13	0.22	0	0.65	0.13	-	-	
3	NUP210	13362654	T	C	0.06	0.02	0.01	0.01	0.06	0.02	-	-	
3	RAB5A	20020604	G	A	0.1	0.01	0.04	0.01	0.1	0	-	-	
3	GLB1	33058892	G	T	0.6	0.08	0.53	0	0.6	0.08	-	-	
3	GOLGA4	37375925	A	G	0.34	0.06	0.34	0	0.34	0.06	-	-	
3	NBEAL2	47031282	G	A	0.74	0.04	0.74	0.04	0.61	0	-	-	
3	APEH	49719055	C	A	0.66	0.29	0.36	0.01	0.66	0.29	-	-	
3	SENP7	101050803	A	C	0.75	0.1	0.02	0	0.75	0.1	-	-	
3	TIMMDC1	119218896	C	T	0.42	0.01	0.22	0	0.42	0.01	-	-	
3	NCK1	136665160	A	G	0.54	0.04	0	0.01	0.54	0.04	-	-	
3	IL20RB	136701582	C	G	0.43	0	0.35	0	0.43	0	-	-	
3	ACTL6A	179293210	C	A	0	0.15	0	0.15	0	0.14	-	-	
3	IGF2BP2	185375702	G	A	0	0.11	0	0.08	0	0.11	-	-	
4	TRMT44	8477242	C	T	0.06	0.01	0.06	0.01	0	0.01	-	-	
4	SEL1L3	25836711	C	T	0.22	0.02	0.01	0.01	0.22	0.02	-	-	
4	TMEM156	38970886	T	G	0.32	0	0.24	0	0.32	0	-	-	
5	MYO10	16681101	C	G	0.1	0.06	0.09	0.06	0.1	0	-	-	

ΔSpliceAI(4999)

Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX AG	MAX AL	DS_AG	DS_AL	DS_DG	DS_DL	ΔESRseq ^a	Abs. difference ^{a,b}
5	DROSHA	31467600	T	C	0.83	0.2	0.45	0	0.83	0.2	-	-
5	PAPD4	78910355	T	A	0.17	0	0.16	0	0.17	0	-	-
5	NREP	111068304	G	A	0.86	0.03	0.68	0.01	0.86	0.03	-	-
5	NDFIP1	141526805	C	T	0.15	0	0.15	0	0.14	0	-	-
6	RREB1	7211296	C	G	0.85	0.24	0.03	0	0.85	0.24	-	-
6	MBOAT1	20117591	T	C	0.74	0.02	0.74	0	0.72	0.02	-	-
6	LTB	31549785	G	C	0.12	0.03	0.04	0.01	0.12	0.03	-	-
6	FRS3	41739367	G	T	0.65	0.01	0.19	0.01	0.65	0.01	-	-
6	MAP3K7	91256017	C	A	0.19	0.02	0.19	0.02	0	0	-	-
6	MICAL1	109768860	C	T	0.24	0.05	0.01	0.04	0.24	0.05	-	-
6	WDR27	170061799	T	C	0.51	0.01	0.34	0	0.51	0.01	-	-
7	TMEM184A	1595683	C	A	0.07	0.06	0	0	0.07	0.06	-	-
7	WIPI2	5240397	G	A	0.17	0.04	0.15	0	0.17	0.04	-	-
7	GPNMB	23294005	A	T	0.18	0	0.18	0	0.02	0	-	-
7	RP9	33138788	T	C	0.37	0.05	0.25	0.01	0.37	0.05	-	-
7	MLXIPL	73018226	G	A	0.06	0	0.06	0	0.03	0	-	-
7	GTF2IRD1	74009430	G	C	0.21	0.02	0	0	0.21	0.02	-	-
7	ARPC1B	98991929	G	A	0.02	0.02	0.02	0.02	0	0.01	-1.87	0.27
7	ATP5J2-PTCD1	99056055	G	A	0.05	0.01	0.05	0.01	0.02	0.01	-	-
7	MCM7	99691532	G	A	0.11	0	0.11	0	0.05	0	-	-
7	POT1	124532129	C	G	0.52	0	0.35	0	0.52	0	-	-
7	POT1	124532780	T	C	0.41	0	0.36	0	0.41	0	-	-
7	KRBA1	149428305	C	A	0.74	0.03	0.48	0.02	0.74	0.03	-	-
7	INSIG1	155099062	G	A	0.07	0	0.07	0	0.06	0	-	-
8	RPS20	56986296	T	C	0.01	0.01	0	0	0.01	0.01	NA	NA
8	NCOA2	71168777	T	C	0.36	0	0.36	0	0.28	0	-	-
8	RMDN1	87520679	T	C	0.66	0.12	0	0.01	0.66	0.12	-	-

ΔSpliceAI(4999)

Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX			DS_AG	DS_AL	DS_DG	DS_DL	ΔESRseq _a	Abs. difference _{a,b}
					AG	DL	AL						
8	RPL30	99057714	C	G	0.01	0.03	0.01	0	0.01	0.03	-2.14	0.89	
8	EBAG9	110574146	A	C	0.57	0.01	0.57	0.01	0.46	0	-	-	
8	EEF1D	144676049	C	A	0.1	0.01	0	0.01	0.1	0.01	-	-	
8	SHARPIN	145157654	G	A	0.67	0.13	0.01	0	0.67	0.13	-	-	
8	MROH1	145242443	G	C	0.78	0.01	0.76	0.01	0.78	0	-	-	
8	TONSL	145654959	C	G	0.18	0	0.18	0	0.15	0	-	-	
9	PLIN2	19119164	A	G	0.01	0	0	0	0.01	0	2.40	0.31	
9	RP11-145E5.5	22008383	A	C	0	0.01	0	0.01	0	0	NA	NA	
9	HINT2	35814670	C	T	0.39	0.15	0.03	0.01	0.39	0.15	-	-	
9	ASS1	133362687	T	C	0.01	0	0	0	0.01	0	0.33	0.83	
9	SURF1	136219256	T	C	0.8	0.83	0.04	0.03	0.8	0.83	-	-	
9	WDR5	137001975	C	T	0.84	0.21	0.18	0	0.84	0.21	-	-	
9	MAN1B1	139992541	T	G	0.1	0	0.06	0	0.1	0	-	-	
9	PNPLA7	140373788	C	A	0.01	0.01	0.01	0	0	0.01	0.40	0.21	
10	WDR37	1171009	G	A	0.33	0.09	0	0	0.33	0.09	-	-	
10	CUL2	35320605	C	G	0.59	0.51	0.59	0.51	0	0.03	-	-	
10	NCOA4	51580064	G	A	0.59	0.01	0.4	0.01	0.59	0	-	-	
10	ZWINT	58118021	T	A	0.44	0.14	0.08	0.01	0.44	0.14	-	-	
10	MMS19	99240647	G	C	0.46	0.03	0.05	0	0.46	0.03	-	-	
10	SCD	102111029	G	T	0.08	0.02	0.08	0	0.06	0.02	-	-	
10	WDR11	122611080	A	G	0.04	0.16	0	0	0.04	0.16	-	-	
10	UROS	127477946	G	T	0.03	0.09	0.03	0.09	0.01	0.01	-	-	
11	EPS8L2	706569	C	T	0.01	0.02	0	0.02	0.01	0.01	-0.15	0.56	
11	RPLP2	810086	C	A	0.63	0.02	0.01	0	0.63	0.02	-	-	
11	CTSD	1782096	C	T	0.14	0.03	0.01	0.02	0.14	0.03	-	-	
11	RRM1	4143584	C	A	0.95	0.31	0.8	0.08	0.95	0.31	-	-	
11	TAF10	6632766	A	C	0.19	0.07	0.04	0	0.19	0.07	-	-	

ΔSpliceAI(4999)

Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX			MAX			Abs. difference		
					AG	DG	DL	DS_AG	DS_AL	DS_DG	DS_DL	ΔESRseq ^a	ΔESRseq ^{a,b}
11	NUCB2	17317591	C	T	0.18	0.18	0.24	0.18	0.24	0	0	-	-
11	USH1C	17518058	C	A	0.11	0.11	0.08	0.11	0.04	0.08	0.08	-	-
11	USH1C	17518139	G	T	0.11	0.11	0.08	0.04	0.02	0.11	0.08	-	-
11	MYRF	61549474	A	G	0.37	0.37	0.06	0.37	0.06	0.17	0	-	-
11	GANAB	62406615	G	A	0.18	0.18	0.04	0.18	0.04	0	0	-	-
11	RELA	65423477	C	T	0.08	0.08	0.06	0.08	0.06	0	0.01	-	-
11	CORO1B	67207186	G	A	0.44	0.44	0.04	0	0	0.44	0.04	-	-
11	PITPNM1	67269680	C	A	0.65	0.65	0.27	0.65	0.27	0.08	0	-	-
11	ANKRD49	94230225	G	A	0.15	0.15	0.04	0	0	0.15	0.04	-	-
11	C11orf1	111750491	C	T	0.61	0.61	0.16	0.08	0.01	0.61	0.16	-	-
11	RPS25	11888601	G	C	0.71	0.71	0.36	0.05	0.01	0.71	0.36	-	-
12	TPI1	6977481	C	G	0.75	0.75	0.11	0.16	0	0.75	0.11	-	-
12	ERGIC2	29500852	G	C	0.17	0.17	0.01	0.09	0.01	0.17	0	-	-
12	MYL6	56552215	C	T	0.85	0.85	0.24	0.03	0.01	0.85	0.24	-	-
12	TESC	117479874	G	A	0.02	0.02	0.01	0	0.01	0.02	0	-2.38	0.42
12	CLIP1	122845962	C	T	0.23	0.23	0.05	0.08	0.01	0.23	0.05	-	-
13	PAN3	28849497	G	T	0.21	0.21	0	0.21	0	0.12	0	-	-
13	RCBTB2	49073402	A	C	0	0	0.01	0	0.01	0	0	3.79	0.07
13	FARP1	99041790	C	G	0.73	0.73	0.05	0.48	0	0.73	0.05	-	-
14	APEX1	20923547	G	A	0.32	0.32	0.13	0.02	0	0.32	0.13	-	-
14	RNASE4	21154898	T	A	0.67	0.67	0	0.6	0	0.67	0	-	-
14	MAX	65543605	C	T	0.26	0.26	0.02	0.12	0	0.26	0.02	-	-
14	ELMSAN1	74190516	G	T	0.74	0.74	0	0.72	0	0.74	0	-	-
14	LGMN	93172407	T	C	0.24	0.24	0	0.24	0	0.2	0	-	-
14	WARS	100834798	C	G	0.08	0.08	0.01	0.03	0	0.08	0.01	-	-
15	SNRPN	25216551	A	T	0.33	0.33	0.01	0.33	0.01	0.27	0	-	-
15	TJP1	30030372	A	C	0.11	0.11	0	0.09	0	0.11	0	-	-

ΔSpliceAI(4999)

Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX			DS_AG	DS_AL	DS_DG	DS_DL	ΔESRseq ^a	Abs. difference ^{a,b}
					AG	DL	AL						
15	ANXA2	60685645	G	A	0.15	0	0	0.15	0	0.02	0	-	-
16	FAM234A	303855	A	C	0.3	0.01	0.3	0.01	0.3	0.26	0	-	-
16	WDR90	707711	G	A	0.29	0.01	0.29	0.01	0.01	0.01	0	-	-
16	ZC3H7A	11875245	G	A	0.99	0.45	0.14	0.01	0.99	0.45	-	-	-
16	USP31	23118742	T	A	0.33	0.01	0.24	0.01	0.33	0	-	-	-
16	LCMT1	25137870	T	G	0.85	0.01	0.75	0	0.85	0.01	-	-	-
16	IL4R	27372198	C	T	0.67	0.14	0.3	0.02	0.67	0.14	-	-	-
16	SULT1A1	28619444	G	A	0.02	0	0	0	0.02	0	0.35	0.44	0.44
16	ATXN2L	28835888	T	C	0.04	0.02	0	0.02	0.04	0.02	0.02	-0.45	0.68
16	ATXN2L	28843238	T	A	0.03	0.01	0.03	0.01	0.02	0	-0.03	0.19	0.19
16	CORO1A	30195685	C	T	0.01	0.02	0	0.02	0.01	0.01	0.01	-1.01	1.62
16	ZNF48	30407295	C	T	0.63	0.12	0.02	0.02	0.63	0.12	-	-	-
16	ITGAL	30485879	T	C	0.4	0.01	0.4	0	0.16	0.01	-	-	-
16	AARS	70287564	G	C	0.63	0.56	0.04	0.02	0.63	0.56	-	-	-
16	MAP1LC3B	87436071	G	C	0.74	0.04	0.62	0.04	0.74	0	-	-	-
16	TRAPPC2L	88923915	C	T	0.01	0.02	0.01	0.02	0	0.01	-0.93	0.70	0.70
16	ZNF276	89796647	G	C	0.15	0.07	0.03	0.01	0.15	0.07	-	-	-
16	TCF25	89961617	T	G	0.12	0.01	0.02	0.01	0.12	0.01	-	-	-
17	GAS7	9835293	G	A	0	0	0	0	0	0	-0.88	0.13	0.13
17	GGNBP2	34945334	C	G	0.93	0.03	0.9	0.03	0.93	0.03	-	-	-
17	THRA	38245688	C	A	0.44	0.04	0.11	0.04	0.44	0	-	-	-
17	ACLY	40066749	C	A	0.02	0	0.01	0	0.02	0	-0.12	0.46	0.46
17	HOXB6	46675032	T	C	0.51	0.13	0	0.01	0.51	0.13	-	-	-
17	NGFR	47589203	G	T	0.16	0	0.16	0	0.02	0	-	-	-
17	MRC2	60743677	G	C	0.26	0.03	0.01	0.03	0.26	0.03	-	-	-
17	PSMC5	61906744	G	A	0.58	0.07	0.45	0.01	0.58	0.07	-	-	-
17	RGS9	63198301	C	G	0.51	0.02	0	0.02	0.51	0.02	-	-	-

ΔSpliceAI(4999)

Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX AG	MAX AL	DS_AG	DS_AL	DS_DG	DS_DL	ΔESRseq ^a	Abs. difference ^{a,b}
17	CDC42EP4	71282849	C	T	0.41	0	0.39	0	0.41	0	-	-
17	NAT9	72771731	T	C	0.07	0.23	0	0.02	0.07	0.23	-	-
17	ITGB4	73733305	G	A	0.35	0.03	0.13	0.03	0.35	0.01	-	-
17	CCDC57	80066834	C	T	0.25	0.05	0.25	0.05	0.06	0	-	-
19	C19orf24	1277440	C	T	0.34	0	0.01	0	0.34	0	-	-
19	NCLN	3192717	C	T	0.87	0.02	0.02	0.01	0.87	0.02	-	-
19	MAP2K2	4094651	G	T	0.23	0.04	0.23	0.04	0	0	-	-
19	LONP1	5708135	C	G	0.28	0.01	0.28	0.01	0.08	0.01	-	-
19	ZNF317	9263091	G	A	0.14	0	0.14	0	0.08	0	-	-
19	ATG4D	10655359	G	T	0.17	0.25	0.17	0.25	0.02	0	-	-
19	DNM2	10872109	G	A	0.37	0.01	0.37	0.01	0.37	0	-	-
19	ECSIT	11622327	G	A	0.07	0.02	0.07	0.02	0.01	0	-	-
19	SYDE1	15222368	C	T	0.02	0.02	0.02	0.02	0	0	-1.40	0.76
19	KXD1	18671359	C	T	0.16	0.11	0	0	0.16	0.11	-	-
19	UBA52	18682724	G	C	0.03	0.83	0	0	0.03	0.83	-	-
19	CRTC1	18883525	C	T	0	0.03	0	0.03	0	0.01	-0.50	0.01
19	CD22	35836911	T	G	0.61	0.08	0.55	0.02	0.61	0.08	-	-
19	LIG1	48639538	G	A	0.03	0	0.03	0	0	0	-1.50	0.66
19	RPL13A	49991073	G	T	0.42	0.02	0.42	0.01	0.19	0.02	-	-
19	IL4I1	50399629	G	A	0	0.02	0	0	0	0.02	-0.65	1.35
19	JOSD2	51014348	G	A	0.96	0.66	0.07	0	0.96	0.66	-	-
19	EPN1	56199796	C	A	0.05	0.06	0.05	0.06	0	0.02	-	-
20	CBFA2T2	32228368	G	A	0.32	0.09	0.02	0.01	0.32	0.09	-	-
20	RALY	32605229	T	G	0.21	0.12	0.21	0.12	0.05	0	-	-
20	RBM39	34327470	C	A	0.02	0.49	0.02	0.49	0.01	0.26	-	-
20	RALGAPB	37176317	T	C	0.1	0	0.1	0	0.06	0	-	-
20	ZMYND8	45864986	G	A	0.55	0.03	0.01	0.02	0.55	0.03	-	-

ΔSpliceAI (4999)

Chromosome	Gene	Genomic position (GRCh37)	REF	ALT	MAX AG	MAX AL	DS_AG	DS_AL	DS_DG	DS_DL	ΔESRseq ^a	Abs. difference ^b
20	DDX27	47852017	C	T	0.65	0.03	0.56	0.03	0.65	0.02	-	-
20	ZFP64	50785196	C	T	0.04	0	0.03	0	0.04	0	-3.20	0.53
20	SS18L1	60746514	T	C	0.09	0	0.09	0	0.06	0	-	-
20	PPDPF	62152271	G	A	0.42	0.01	0.01	0	0.42	0.01	-	-
20	PTK6	62167808	C	T	0.03	0.27	0	0.01	0.03	0.27	-	-
20	RTEL1	62327053	G	C	0.02	0.11	0.01	0.09	0.02	0.11	-	-
21	BACE2	42612462	G	C	0.05	0.01	0.05	0.01	0.05	0	-	-
21	COL6A1	47422668	C	A	0.68	0.01	0.02	0	0.68	0.01	-	-
22	SLC25A1	19165152	G	A	0.98	0.72	0.02	0.01	0.98	0.72	-	-
22	HIRA	19375501	G	A	0.14	0.02	0.14	0.02	0	0	-	-
22	IFT27	37164226	A	G	0.26	0.01	0.26	0.01	0.07	0.01	-	-
22	PDGFB	39637045	G	T	0	0.01	0	0	0	0.01	-1.16	0.55
22	RPL3	39714377	C	A	0.02	0.01	0.01	0	0.02	0.01	0.21	0.53
22	TRMU	46747917	G	A	0.56	0.09	0.49	0	0.56	0.09	-	-
X	HSD17B10	53461167	G	T	0.48	0.16	0	0	0.48	0.16	-	-
X	HUWE1	53560612	G	A	0.61	0.02	0.61	0.01	0.37	0.02	-	-
X	FGD1	54493671	A	T	0.35	0.03	0.08	0	0.35	0.03	-	-
X	ATRX	76817274	A	G	0.12	0	0.12	0	0.02	0	-	-
X	MORC4	106199190	T	C	0.76	0.48	0.34	0	0.76	0.48	-	-
X	FAM3A	153736984	C	T	0.88	0.87	0.88	0.87	0.01	0.01	-	-

^a ΔESRseq and ESRseq absolute difference in absolute values were exclusively calculated for variants with a ΔSpliceAI score <0.05

^b Abs. difference (difference in absolute values): difference of normalized SRE area between previous and posterior variant sequences.

REF: reference allele, ALT: variant allele, DS: ΔScore, AG: Acceptor Gain, AL: Acceptor Loss, DG: Donor Gain, DL: Donor Loss, MAX AG DG: maximum value in AG or DG, MAX AL DL: maximum value in AL or DL, NA: not available.

Supplementary Table 4. Deep intronic experimentally assessed variants.

This table is an adapted version of ST4 from Moles-Fernández et al. (2021)[287].

Gene	Transcript	Variant	Intron	ΔSpliceAI (4999) ^a	ΔESRseq ^b	Abs. Difference ^c	Splicing outcome ^d	Comment	Population variant frequencies (gnomAD v2.1)	ClinVar review status (2021) ^e
ATM	NM_000051.3	c.331+2432T>A	4	0	4.25	0.36	No alteration		NR	NR
ATM	NM_000051.3	c.1899-123A>G	12	AG 0.74 (- 90bp)/ DG 0.71 (-1bp)	0.63	1.22	PSE	▼12A.1 (from c.1899- 174 or c.1899-177 to c.1899-124) and ▼12A.2 (from c.1899- 213 to c.1899-124)	0.000	NR
ATM	NM_000051.3	c.2466+1552G>C	16	AG 0.93 (3bp)/ DG 0.69 (97bp)	-0.14	1.82	PSE	▼16A(from c.2466+1555 to c.2466+1650)	NR	Likely pathogenic [1]
ATM	NM_000051.3	c.6453-418del	44	0	1.11	0.13	No alteration		0.0117	NR
ATM	NM_000051.3	c.8419-54_8419- 49del	57	0	3.45	0.45	No alteration		0.0015	NR
ATM	NM_000051.3	c.8672-2496G>A	59	0	1.56	0.70	No alteration		0.0068	NR
ATM	NM_000051.3	c.8850+1791G>C	61	0	1.87	0.35	No alteration		0.0114	NR
ATM	NM_000051.3	c.8850+2029A>G	61	AG 0.22 (1bp)/ DG 0.16 (102bp)	-0.19	0.77	PSE	▼61A (from c.8850+2030 to c.8850+2131)	NR	NR
BARD1	NM_000465.3	c.2001+179A>G	10	DG 0.12 (85 bp)	1.23	0.186	No alteration		NR	NR
BRCA2	NM_000059.3	c.9256+4091G>A	24	0	1.44	0.28	No alteration		NR	NR
BRCA2	NM_000059.3	c.9256+2960_925 6+2991dup	24	AG 0.17 (- 97bp)/ DG 0.1 (-7bp)	-5.24	0.36	No alteration		NR	NR
BRCA2	NM_000059.3	c.9257-6359G>C	24	0	2.23	0.43	No alteration		NR	NR
FAM175A	NM_139076.2	c.476+158G>T	5	AG 0.17 (2bp)/ DG 0.22 (-94bp)	0.56	0.59	PSE	▼5A (from c.476+156 to c.476+252)	0.000	NR

Gene	Transcript	Variant	Intron	ΔSpliceAI (4999) ^a	ΔESRseq ^b	Abs. Difference ^c	Splicing outcome ^d	Comment	Population variant frequencies (gnomAD v2.1)	ClinVar review status (2021) ^e
MLH1	NM_000249.3	c.545+25G>A	6	0.000	1.96	0.06	No alteration		0.000	NR
MLH1	NM_000249.3	c.1559-39T>C	13	0	2.34	1.51	No alteration		0.000	NR
MSH2	NM_000251.2	c.1662-22A>G	10	AG 0.15 (1bp)	1.96	1.33	No alteration		NR	NR
MSH2	NM_000251.2	c.2005+42_2005+ 44del	12	0	0.45	0.02	No alteration		0.000	NR
MUTYH	NM_0011284 25.1	c.998-27G>A	11	AG 0.41 (- 4bp)/ DG 0.05 (-215bp)	0.34	1.00	Intron retention	▼12A (from c.998- 230 c.998)	0.001	Likely benign Likely benign 1 1
NF1	NM_0010424 92.2	c.1846-123G>A	16	0	1.49	0.26	No alteration		NR	NR
NF1	NM_0010424 92.2	c.4110+681_4110 +693del	30	0	3.83	0.23	No alteration		0.004	NR
NF1	NM_0010424 92.2	c.4174-578C>A	31	AG 0.08 (- 79bp)/ DG 0.12 (19bp)	2.46	0.09	No alteration		0.030	NR
NF1	NM_0010424 92.2	c.8113+49C>T	55	DG 0.18 (-2bp)	-1.75	1.20	No alteration		0.002	NR
PTEN	NM_000314.4	c.79+1277G>A	1	0	2.53	0.25	No alteration		0.003	NR
PTEN	NM_000314.4	c.801+23G>A	7	0	2.62	0.41	No alteration		0.000	Likely benign Likely benign
PTEN	NM_000314.4	c.80-8164A>G	1	0	2.17	1.22	No alteration		0.002	NR
PTEN	NM_000314.4	c.80-14222G>A	1	0	2.67	0.10	No alteration		NR	NR
PTEN	NM_000314.4	c.165-9284G>A	2	0	1.85	0.78	No alteration		NR	NR
PTEN	NM_000314.4	c.209+2257del	3	0	3.76	0.59	No alteration		NR	NR
PTEN	NM_000314.4	c.802-51_802- 14del	7	AG 0.06 (95bp)	8.01	1.07	No alteration		0.017	Benign Benign Uncertain significance Benign

Gene	Transcript	Variant	Intron	Δ SpliceAI (4999) ^a	Δ ESRseq ^b	Abs. Difference ^c	Splicing outcome ^d	Comment	Population variant frequencies (gnomAD v2.1)	ClinVar review status (2021) ^e
RAD51C	NM_058216.1	c.572-2273G>T	3	AG 0.12 (-1bp) DG 0.09 (117bp)	0.59	1.92	No alteration		NR	NR
RBBP8	NM_002894.2	c.709+139T>C	9	0	1.90	0.53	No alteration		0.017	NR
TP53	NM_000546.5	c.97-53G>A	3	0	1.72	0.16	No alteration		NR	NR
TP53	NM_000546.5	c.993+88A>G	9	AG 0.2 (-1bp)/ DG 0.05 (- 224bp)	0.16	0.76	No alteration		NR	NR

^a For each variant, SpliceAI looks within a window (+/- 4999 in this case) to see how the variant affects the probabilities of different positions in the pre-mRNA being splice acceptors or donors. The values in brackets represent positions with the biggest probability of being used as splicing sites within the window. Negative values are upstream (5') of the variant, and positive are downstream (3') of the variant. In bold are the scores of ≥ 0.05 . AG: Acceptor Gain, DG: Donor Gain, AL: Acceptor Loss, DL: Donor Loss.

^b Δ ESRseq score was calculated as described by Ke et al. (2011)[134], and in bold are indicated values $\geq +0.63$.

^c Abs. difference: absolute value of the difference between the normalized SRE area of 100bp before and after a variant. In bold are indicated values ≥ 0.51 . ^d RNA splicing profiles from carriers were compared with non-carrier controls by capillary electrophoresis and Sanger sequencing. PSE: pseudoexon.

^e Variant classifications reported in ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>, accessed on 1st March 2021). NR: not reported; 1: criteria provided, one single submitter; classification without a number: no assertion criteria provided.

Yellow indicates that the variant is predicted as spliceogenic by our pipeline: Δ SpliceAI ≥ 0.05 or SpliceAI < 0.05 and Δ ESRq ≥ 0.63 and Abs. Dif. 0.51.

Supplementary Table 5. List of HBOC risk and Lynch genes and their respective NCBI reference transcripts used to annotate the ESRseq scores from canonical exons and adjacent intronic nucleotides

Gene	Transcript
<i>BARD1</i>	NM_000465.3
<i>BRCA1</i>	NM_007294.3
<i>BRCA2</i>	NM_000059.3
<i>BRIP1</i>	NM_032043.2
<i>CDH1</i>	NM_004360.3
<i>CHEK2</i>	NM_007194.3
<i>MLH1</i>	NM_000249.3
<i>MSH2</i>	NM_000251.2
<i>MSH6</i>	NM_000179.2
<i>PALB2</i>	NM_024675.3
<i>PTEN</i>	NM_000314.4
<i>RAD51C</i>	NM_058216.1
<i>RAD51D</i>	NM_002878.3
<i>STK11</i>	NM_000455.4
<i>TP53</i>	NM_000546.5

Supplementary Table 6. Primers and PCR conditions for characterizing the splicing effect of deep intronic variants using RNA from patients (whole blood).

Gene	Transcript	Name	Sequence (5' to 3')	T.A (°C)	FL size (bp)
ATM	NM_000051.3	R-ATM-1F*[6FAM]	GCCGCGGTGATACTACTTT	56.5	729
		R-ATM-6R	TCGCACACTGAATAGCCTTG		
		R-ATM-12F*[FAM]	TTTCCAAAGCGTGCCAGAA	57	203
		R-ATM-13R	ACAGCGATCCAGTGATTCCT		
		R-ATM-16F	TCCTGCGATTGTTAACATCAAAAG	56	212
		R-ATM-17R	TCCAGGTTGTTTGCATCAC		
		R-ATM-16F*[6FAM]	TCCTGCGATTGTTAACATCAAAAG	57	553
		R-ATM-20R	CAAAACATCTTGGTCACGACGA		
		R-ATM-43F*[6FAM]	AGGAATATGCAGTGGGACCA	57	409
		R-ATM-46R	TGTCCATTTCCTTTCCATCAGG		
		R-ATM-53.54*[6FAM]	CAGTGAAGACTCAGAGAAAAGG	57	897
		R-ATM-60.61R	CATGGTTTCTCACAGCATCTT		
		R-ATM-56F*[6FAM]	ACTGCAGAGAAACACGGAAC	57	495
		R-ATM-60R	GGAGTAGGAAGGATTTGCCCC		
		R-ATM-60F*[6FAM]	GGCATTACGGGTGTTGAAGG	56	274
		R-ATM-63R	AGACACGTTCAAGCTACTTTGT		
BARD1	NM_000465.3	R-ATM-61F	TGCTGTGAGAAAACCATGGA	57	165
		R-ATM-62R	CAGAGTAGGGTGAAGCTCAGT		
BRCA2	NM_000059.3	R-BARD1-4F*[6FAM]	GTGATGTCTAGTCCCTCAGCAATGA	62	877
		R-BARD1-11R	CACCTGCAGTGACGAGCTTA		
FAM175A	NM_139076.2	R-BRCA2-24F*[6FAM]	AGATTTACCAGCCACGGGAG	57	567
		R-BRCA2-27R	TGGCCATACAAAGTGATAAAGGA		
MLH1	NM_000249.3	R-FAM175A-5F*[6FAM]	AGAGAGAGGCTGCTTCACAA	57	243
		R-FAM175A-6R	TCGGCTAAACCAGTGGACA		
MSH2	NM_000251.2	R-MLH1-4F*[6FAM]	GCTTTGGCCAGCATAAGCCATG	61	653
		R-MLH1-11R	TCCTCGTGCAGGAAGTGAACCT		
MUTYH	NM_001128425.1	R-MLH1-11F*[6FAM]	AAGTTCACTTCCTGCACGAG	57	801
		R-MLH1-15R	CTGGCTCCGATAAACCTGAGA		
NF1	NM_001042492.2	R-MSH2-9F*[6FAM]	ACCATGAATTCCTGTAAACCT	56	717
		R-MSH2-13R	TGGACACTTCTGCTGACTCA		
		R-MUTYH-10F*[6FAM]	CTAGGGGCCACAGTGTGTAC	58	357
		R-MUTYH-13R	CTCCACAGTCCCTGCCAG		
		R-MUTYH-11F	CCTGACGTGGAGGAGTGTG	58.5	154
		R-MUTYH-12R	CCAGAACACAGGTGGCAGAG		
		R-NF1-15-F*[6FAM]	GGAATCCTGATGCTCCTCTAGTA	57.5	399

Gene	Transcript	Name	Sequence (5' to 3')	T.A (°C)	FL size (bp)
		R-NF1-18-R	ACATGTACAGGGCCACTTCT		
		R-NF1-29-F*[6FAM]	TGTTAGCTTTGAAATGGATCCTA	56	250
		R-NF1-32-R	CTGAGGGAAACGCTGGCT		
		R-NF1-38-F*[6FAM]	TTAGGCAGTTCTGACCCCGAG	57	493
		R-NF1-41-R	TTGATCCCAAGCCACCTGTT		
		R-NF1-54-F*[6FAM]	GCCAGTGTGTGTTTCCCAA	57.5	222
		R-NF1-56-R	TGAAAACGGTCTCTGCAAACC		
		R-PTEN-1F	CATTCCATCCTGCAGAAGA	58	795
PTEN	NM_000314.4	R-PTEN-7R*[6FAM]	TTGCTTCCCGTCGTGG		
		R-PTEN-1F (II)	TCGTTAGCAGAAACAAAAGGAGA	57	111
		R-PTEN-2R	CGCCTTCAAGTCTTCTGCA		
		R-PTEN-4F	ACGCCAAATTAATGCAG	56	479
		R-PTEN-7R*[6FAM]	TTGCTTCCCGTCGTGG		
		R-PTEN-6F*[6FAM]	ACTATTCCAGTCAGAGCG	57	567
		R-PTEN-9R	TGACGGCTCCTCTACTGTTT		
		R-RAD51C-3F*[6FAM]	CAGTAGATGTCAGATACCAGA	56	285
RAD51C	NM_058216.1	R-RAD51C-4R	AGTGTCTGAAAGGAAATCTGGA		
RBBP8	NM_002894.2	R-RBBP8-9F*[6FAM]	AGTAGCTGACACTTATGACCAAA	56	394
		R-RBBP8-12R	CCCAGGCTGTAAAAGAGAAGG		
TP53	NM_000546.5	R-TP53-2F*[6FAM]	CCCCTCTGAGTCAGGAAACA	57	574
		R-TP53-6R	ACACGCAAAATTCCTTCCAC		
		R-TP53-7F*[6FAM]	TGGCTCTGACTGTACCACCA	58	694
		R-TP53-11R	CCACAACAAAACACCAAGTGC		
		R-TP53-9F	CTCTCCCCAGCCCAAGAAGA	57.5	107
		R-TP53-10R	GTTCCAAAGGCCTCATTGAGC		

*PCR products amplified using one primer labelled with 6-Carboxyfluorescein (6-FAM) were analysed by high-resolution capillary electrophoresis

Supplementary Table 7. ΔESRseq threshold optimization (sensitivity + specificity).
This table is a summarized version of ST9 from Moles-Fernández et al. (2021)[287].

ΔESRseq Thresholds	Sensitivity + Specificity				Positive predictive Value (precision)				False Discovery Rate				Accuracy				MCC				False Negative Rate				False Omission Rate				Negative Predictive Value			
	Sensitivity	Specificity	+	Specificity	Sensitivity	False Positive Rate	False Positive Rate	Value	Rate	Rate	Rate	Rate	Accuracy	Accuracy	Accuracy	Accuracy	MCC	MCC	MCC	MCC	Rate	Rate	Rate	Rate	Rate	Rate	Rate	Rate	Value	Value	Value	Value
2.00	10.00	93.75	103.75	103.75	6.25	6.25	6.25	14.29	85.71	85.71	85.71	85.71	83.49	0.04	90.00	90.00	0.04	90.00	90.00	90.00	90.00	9.09	9.09	9.09	9.09	90.91	90.91	90.91	90.91	90.91	90.91	90.91
1.97	10.00	93.75	103.75	103.75	6.25	6.25	6.25	14.29	85.71	85.71	85.71	85.71	83.49	0.04	90.00	90.00	0.04	90.00	90.00	90.00	90.00	9.09	9.09	9.09	9.09	90.91	90.91	90.91	90.91	90.91	90.91	90.91
1.76	20.00	93.75	113.75	113.75	6.25	6.25	6.25	25.00	75.00	75.00	75.00	75.00	84.40	0.15	80.00	80.00	0.15	80.00	80.00	80.00	80.00	8.16	8.16	8.16	8.16	91.84	91.84	91.84	91.84	91.84	91.84	91.84
1.73	20.00	92.71	112.71	112.71	7.29	7.29	7.29	22.22	77.78	77.78	77.78	77.78	83.49	0.13	80.00	80.00	0.13	80.00	80.00	80.00	80.00	8.25	8.25	8.25	8.25	91.75	91.75	91.75	91.75	91.75	91.75	91.75
1.62	20.00	91.67	111.67	111.67	8.33	8.33	8.33	20.00	80.00	80.00	80.00	80.00	82.57	0.12	80.00	80.00	0.12	80.00	80.00	80.00	80.00	8.33	8.33	8.33	8.33	91.67	91.67	91.67	91.67	91.67	91.67	91.67
1.61	20.00	90.63	110.63	110.63	9.38	9.38	9.38	18.18	81.82	81.82	81.82	81.82	81.65	0.10	80.00	80.00	0.10	80.00	80.00	80.00	80.00	8.42	8.42	8.42	8.42	91.58	91.58	91.58	91.58	91.58	91.58	91.58
1.52	20.00	89.58	109.58	109.58	10.42	10.42	10.42	16.67	83.33	83.33	83.33	83.33	80.73	0.09	80.00	80.00	0.09	80.00	80.00	80.00	80.00	8.51	8.51	8.51	8.51	91.49	91.49	91.49	91.49	91.49	91.49	91.49
1.41	20.00	88.54	108.54	108.54	11.46	11.46	11.46	15.38	84.62	84.62	84.62	84.62	79.82	0.08	80.00	80.00	0.08	80.00	80.00	80.00	80.00	8.60	8.60	8.60	8.60	91.40	91.40	91.40	91.40	91.40	91.40	91.40
1.40	20.00	87.50	107.50	107.50	12.50	12.50	12.50	14.29	85.71	85.71	85.71	85.71	78.90	0.06	80.00	80.00	0.06	80.00	80.00	80.00	80.00	8.70	8.70	8.70	8.70	91.30	91.30	91.30	91.30	91.30	91.30	91.30
1.39	20.00	86.46	106.46	106.46	13.54	13.54	13.54	13.33	86.67	86.67	86.67	86.67	77.98	0.05	80.00	80.00	0.05	80.00	80.00	80.00	80.00	8.79	8.79	8.79	8.79	91.21	91.21	91.21	91.21	91.21	91.21	91.21
1.34	20.00	85.42	105.42	105.42	14.58	14.58	14.58	12.50	87.50	87.50	87.50	87.50	77.06	0.04	80.00	80.00	0.04	80.00	80.00	80.00	80.00	8.89	8.89	8.89	8.89	91.11	91.11	91.11	91.11	91.11	91.11	91.11
1.33	20.00	85.42	105.42	105.42	14.58	14.58	14.58	12.50	87.50	87.50	87.50	87.50	77.06	0.04	80.00	80.00	0.04	80.00	80.00	80.00	80.00	8.89	8.89	8.89	8.89	91.11	91.11	91.11	91.11	91.11	91.11	91.11
1.31	20.00	84.38	104.38	104.38	15.63	15.63	15.63	11.76	88.24	88.24	88.24	88.24	76.15	0.03	80.00	80.00	0.03	80.00	80.00	80.00	80.00	8.99	8.99	8.99	8.99	91.01	91.01	91.01	91.01	91.01	91.01	91.01
1.13	20.00	83.33	103.33	103.33	16.67	16.67	16.67	11.11	88.89	88.89	88.89	88.89	75.23	0.03	80.00	80.00	0.03	80.00	80.00	80.00	80.00	9.09	9.09	9.09	9.09	90.91	90.91	90.91	90.91	90.91	90.91	90.91
1.09	30.00	82.29	112.29	112.29	17.71	17.71	17.71	15.00	85.00	85.00	85.00	85.00	75.23	0.09	70.00	70.00	0.09	70.00	70.00	70.00	70.00	8.14	8.14	8.14	8.14	91.86	91.86	91.86	91.86	91.86	91.86	91.86
1.08	30.00	81.25	111.25	111.25	18.75	18.75	18.75	14.29	85.71	85.71	85.71	85.71	74.31	0.08	70.00	70.00	0.08	70.00	70.00	70.00	70.00	8.24	8.24	8.24	8.24	91.76	91.76	91.76	91.76	91.76	91.76	91.76
0.94	40.00	80.21	120.21	120.21	19.79	19.79	19.79	17.39	82.61	82.61	82.61	82.61	74.31	0.14	60.00	60.00	0.14	60.00	60.00	60.00	60.00	7.23	7.23	7.23	7.23	92.77	92.77	92.77	92.77	92.77	92.77	92.77
0.86	40.00	79.17	119.17	119.17	20.83	20.83	20.83	16.67	83.33	83.33	83.33	83.33	73.39	0.13	60.00	60.00	0.13	60.00	60.00	60.00	60.00	7.32	7.32	7.32	7.32	92.68	92.68	92.68	92.68	92.68	92.68	92.68
0.84	40.00	78.13	118.13	118.13	21.88	21.88	21.88	16.00	84.00	84.00	84.00	84.00	72.48	0.12	60.00	60.00	0.12	60.00	60.00	60.00	60.00	7.41	7.41	7.41	7.41	92.59	92.59	92.59	92.59	92.59	92.59	92.59
0.80	40.00	77.08	117.08	117.08	22.92	22.92	22.92	15.38	84.62	84.62	84.62	84.62	71.56	0.12	60.00	60.00	0.12	60.00	60.00	60.00	60.00	7.50	7.50	7.50	7.50	92.50	92.50	92.50	92.50	92.50	92.50	92.50
0.73	40.00	76.04	116.04	116.04	23.96	23.96	23.96	14.81	85.19	85.19	85.19	85.19	70.64	0.11	60.00	60.00	0.11	60.00	60.00	60.00	60.00	7.59	7.59	7.59	7.59	92.41	92.41	92.41	92.41	92.41	92.41	92.41
0.72	40.00	75.00	115.00	115.00	25.00	25.00	25.00	14.29	85.71	85.71	85.71	85.71	69.72	0.10	60.00	60.00	0.10	60.00	60.00	60.00	60.00	7.69	7.69	7.69	7.69	92.31	92.31	92.31	92.31	92.31	92.31	92.31
0.71	40.00	73.96	113.96	113.96	26.04	26.04	26.04	13.79	86.21	86.21	86.21	86.21	68.81	0.09	60.00	60.00	0.09	60.00	60.00	60.00	60.00	7.79	7.79	7.79	7.79	92.21	92.21	92.21	92.21	92.21	92.21	92.21
0.63	50.00	73.96	123.96	123.96	26.04	26.04	26.04	16.67	83.33	83.33	83.33	83.33	69.72	0.16	50.00	50.00	0.16	50.00	50.00	50.00	50.00	6.58	6.58	6.58	6.58	93.42	93.42	93.42	93.42	93.42	93.42	93.42
0.61	50.00	72.92	122.92	122.92	27.08	27.08	27.08	16.13	83.87	83.87	83.87	83.87	68.81	0.15	50.00	50.00	0.15	50.00	50.00	50.00	50.00	6.67	6.67	6.67	6.67	93.33	93.33	93.33	93.33	93.33	93.33	93.33
0.58	50.00	71.88	121.88	121.88	28.13	28.13	28.13	15.63	84.38	84.38	84.38	84.38	67.89	0.14	50.00	50.00	0.14	50.00	50.00	50.00	50.00	6.76	6.76	6.76	6.76	93.24	93.24	93.24	93.24	93.24	93.24	93.24
0.56	50.00	70.83	120.83	120.83	29.17	29.17	29.17	15.15	84.85	84.85	84.85	84.85	66.97	0.13	50.00	50.00	0.13	50.00	50.00	50.00	50.00	6.85	6.85	6.85	6.85	93.15	93.15	93.15	93.15	93.15	93.15	93.15

Δ ESRseq Thresholds	Sensitivity + Specificity				Positive predictive Value (precision)				False Discovery Rate				Accuracy				MCC				False Negative Rate				False Omission Rate				Negative Predictive Value			
	Sensitivity	Specificity	Sensitivity	Specificity	False Positive Rate	Positive predictive Value (precision)	False Discovery Rate	Accuracy	MCC	False Negative Rate	False Omission Rate	Negative Predictive Value	Accuracy	MCC	False Negative Rate	False Omission Rate	Negative Predictive Value	Accuracy	MCC	False Negative Rate	False Omission Rate	Negative Predictive Value	Accuracy	MCC	False Negative Rate	False Omission Rate	Negative Predictive Value					
0.55	50.00	69.79	119.79	30.21	14.71	85.29	66.06	0.12	50.00	6.94	93.06																					
0.43	50.00	68.75	118.75	31.25	14.29	85.71	65.14	0.12	50.00	7.04	92.96																					
0.37	50.00	67.71	117.71	32.29	13.89	86.11	64.22	0.11	50.00	7.14	92.86																					
0.33	50.00	66.67	116.67	33.33	13.51	86.49	63.30	0.10	50.00	7.25	92.75																					
0.32	50.00	65.63	115.63	34.38	13.16	86.84	62.39	0.10	50.00	7.35	92.65																					
0.29	50.00	64.58	114.58	35.42	12.82	87.18	61.47	0.09	50.00	7.46	92.54																					
0.28	50.00	63.54	113.54	36.46	12.50	87.50	60.55	0.08	50.00	7.58	92.42																					
0.26	50.00	62.50	112.50	37.50	12.20	87.80	59.63	0.08	50.00	7.69	92.31																					
0.19	50.00	61.46	111.46	38.54	11.90	88.10	58.72	0.07	50.00	7.81	92.19																					
0.12	50.00	60.42	110.42	39.58	11.63	88.37	57.80	0.06	50.00	7.94	92.06																					
0.11	50.00	59.38	109.38	40.63	11.36	88.64	56.88	0.06	50.00	8.06	91.94																					
0.09	50.00	58.33	108.33	41.67	11.11	88.89	55.96	0.05	50.00	8.20	91.80																					
0.07	60.00	57.29	117.29	42.71	12.77	87.23	55.96	0.10	40.00	6.78	93.22																					
0.00	60.00	56.25	116.25	43.75	12.50	87.50	55.05	0.10	40.00	6.90	93.10																					

In yellow indicated the cutoff with highest performance based on sum of sensitivity and specificity.

Supplementary Table 8. Performance of ΔSpliceAI alone and sequentially combined with ESRseq tool for 233 intronic variants.
Variants from the literature dataset (133 altering and 100 non altering splicing) (extended version of Table 11).

Pipeline	Positive									
	Sensitivity	Specificity	False Positive Rate	Positive predictive Value (precision)	False Discovery Rate	Accuracy	MCC	False Negative Rate	False Omission Rate	Negative Predictive Value
(1) Δ SpliceAI ≥ 0.05	93.98	92.00	8.00	93.98	6.02	93.13	0.86	6.02	7.55	86.79
(2) Δ SpliceAI ≥ 0.05 + SpliceAI < 0.05 and ΔESRseq ≥ 0.63	96.24	69.00	31.00	80.50	19.50	84.55	0.69	3.76	6.76	93.24
(3) Δ SpliceAI ≥ 0.05 + SpliceAI < 0.05 and ΔESRseq ≥ 0.63 and Abs. Dif. 0.51	95.49	86.00	14.00	90.07	9.93	91.42	0.83	4.51	6.52	93.48
In 1-3, ΔSpliceAI considering 4999bp window.										185.98
										165.24
										181.49

BRCA1 (NM_007294.3) Variant	ASpliceAI (4999) ^b				SPiP V2.1 ^c		ESRseq		Minigene		Patient carriers		mESC				Complementation ^d						
	Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutInPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	%FL Splicing fraction (Per- allele)		% c.5091_5092del Splicin g fraction (Per- allele)	% c.5078_5080del Splicin g fraction (Per- allele)	%Δ18 Splicing fraction (Per- allele)	%FL	%Δ18	HDR
c.5075-8T>G (p.?)	Pathogenic & Likely pathogenic	criteria provided,multiple	0.79	0.74	0.0	0	No	0.47	-0.08	-0.55	-	0.0	0.0	###	-	-	-	-	-	-	-	-	-
c.5075-7T>C (p.?)	Likely benign	criteria provided,multiple	0.06	0	0.0	0	No	0.80	2.14	1.34	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5075-6C>A (p.?)	Conflicting	criteria provided,conflicting interpretation	0.24	0.01	0.0	0	No	0.80	-1.39	-2.19	-	100.0	0.0	0.0	PAXgene	97.83(96.85)	-	-	0.00(-0.01)	-	-	-	-
c.5075-6C>T (p.?)	Likely benign	criteria provided,multiple	0	0.1	0.0	0	No	0.80	-1.32	-2.12	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5075-5T>C (p.?)	Likely benign	criteria provided,single submitter	0	0.07	0.0	0	No	0.80	0.55	-0.25	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5075-4G>A (p.?)	Likely benign	criteria provided,multiple	0.01	0.01	0.0	0	No	0.80	0.69	-0.11	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5075-3C>G (p.?)	Uncertain	criteria provided,single submitter	0.83	0	0.0	0	No	0.80	1.09	0.29	-	1.5	98.5	0.0	-	-	-	-	-	-	-	-	-
c.5075-3C>T (p.?)	Conflicting	criteria provided,conflicting interpretation	0.25	0.01	0.0	0	No	0.80	-1.64	-2.44	-	97.8	2.2	0.0	-	-	-	-	-	-	-	-	-
c.5075-2A>C (p.?)	Pathogenic	criteria provided,multiple	0.83	0.02	0.0	0	No	0.71	1.43	0.72	-	0.0	100.0	0.0	-	-	-	-	-	-	-	-	-
c.5075-2A>G (p.?)	Pathogenic	criteria provided,multiple	0.83	0.02	0.0	0	No	0.71	3.27	2.56	-	0.0	100.0	0.0	PAXgene	40.00(-18.83)	-	-	55.21(110.41)	The FL expression comes exclusively from the WT allele	-	-	-
c.5075-2A>T (p.?)	Pathogenic	no assertion criteria provided	0.83	0.01	0.0	0	No	0.71	0.94	0.24	-	0.0	100.0	0.0	-	-	-	-	-	-	-	-	-

BRCA1 (NM_007294.3) Variant	ASpliceAI (4999) ^b			SPiP V2.1 ^c			ESRseq			Minigene			Patient carriers			mESC			Complementation ^d	
	Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	FL allele-specific result	%FL	%Δ18		HDR
c.5075-2del (p.?)	Pathogenic & Likely pathogenic	criteria provided,multipe submitters	0.83	0	0.0	0	No					0.0	100.0	0.0	-	-	-	-	-	
c.5075-1G>A (p.?)	Pathogenic & Likely pathogenic	criteria provided,multipe submitters	0.83	0.01	0.0	0	No	0.64	-0.71	-1.35	-	0.0	100.0	0.0	-	-	-	-	-	
c.5075-1G>C (p.?)	Pathogenic	criteria provided,multipe submitters	0.83	0.01	0.0	0	No	0.64	-0.24	-0.89	-	0.0	100.0	0.0	-	-	-	-	-	
c.5075-1G>T (p.?)	Pathogenic & Likely pathogenic	criteria provided,multipe submitters	0.83	0.01	0.0	0	No	0.64	-1.24	-1.89	-	0.0	100.0	0.0	-	-	-	-	-	
c.5075_5078del (p.(Asp1692Valfs*9))	Pathogenic	criteria reviewed by expert panel	0.83	0.11	0.0	0	No	-	-	-	-	4.8	95.2	0.0	-	-	-	-	-	
c.5075A>C (p.(Asp1692Ala))	Likely pathogenic	criteria provided,single submitter	0.2	0	0.0	0	No	0.85	0.34	-0.51	0.51	96.8	3.2	0.0	-	-	-	-	-	
c.5075A>T (p.(Asp1692Val))	Conflicting	criteria provided,conflicting interpretations	0.32	0.01	0.0	0	No	0.85	0.13	-0.72	0.54	96.0	4.0	0.0	PAXgene	88.61(81.58)	-	11.39(22.16)	-	
c.5076del (p.(Asp1692Glnfs*10))	Pathogenic	reviewed by expert panel	0.34	0.01	0.0	0	No	-	-	-	-	78.8	21.2	0.0	-	-	-	-	-	
c.5077_5079delinsTTGATCTGCA (p.(Ala1693Phefs*12))	Not reported	-	0.06	0.02	0.0	0	No	-	-	-	-	3.6	96.4	0.0	-	-	-	-	-	
c.5077_5080delinsTTGATCTGCG (p.(Ala1693Glu1694delinsPheLeuGln))	not provided	-	0	0.03	0.0	0	No	-	-	-	-0.15	1.0	99.0	0.0	PAXgene	33.97(-27.69)	-	63.05(125.48)	The FL expression comes exclusively from the WT allele	
																	0.00	92.09	9.50	Good

BRCA1 (NM_007294.3) Variant		ΔSpliceAI (4999) ^b		SPiP V2.1 ^c		ESRseq		Minigene		Patient carriers		mESC		Complementation ^d							
Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	%FL Splicing fraction (Per- allele)	% c.5091_5092del Splicin g fractio n (Per- allele)	% c.5078_5080del Splicin g fractio n (Per- allele)	%Δ18 fraction (Per- allele)	%FL	%Δ18	HDR
c.5089_5092del (p.(Cys1697Asnfs*4))	Pathogenic & Likely pathogenic	0.07	0.02	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5090C>A (p.(Cys1697Tyr))	Likely pathogenic	0	0.03	0.0	0	No	0.56	-0.32	-0.88	0.39	99.9	0.1	0.0	-	-	-	-	-	-	-	-
c.5091_5092del (p.(Cys1697*))	Pathogenic	0.04	0	0.0	0	No	-	-	-	-	100.0	0.0	0.0	Whole-blood (trizol)	78.27(-)	17.81(3 5.62)	-	2.28(3.63)	WT allele contributes more than mutant allele to FL expression	-	-
c.5092_5093delinsAG (p.(Glu1698Arg))	Not reported	0.14	0	0.0	0	No	-	-	-	0.60	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5094A>G (p.(Glu1698=))	Likely benign	0.02	0	0.0	0	No	3.15	1.89	-1.27	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5095C>A (p.(Arg1699=))	Likely benign	0.13	0	0.0	0	No	2.77	1.08	-1.69	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5095C>T (p.(Arg1699Trp))	Pathogenic	0.14	0.01	0.0	0	No	2.77	1.79	-0.99	0.43	100.0	0.0	0.0	Whole-blood (MagNA pure))Whole- blood (Trizol))LCLs with purmoycin	94.97(94.31))8 9.39(81.38))93 .33(90.91)	-	1.24(1.85))10. 61(20.30))19(. 39)	98.89	0.00	16.30	Poor
c.5096G>A (p.(Arg1699Gln))	Pathogenic	0.15	0	0.0	0	No	2.93	1.20	-1.73	0.42	100.0	0.0	0.0	Whole-blood (trizol))PAXgen el	92.23(87.06))1 00.00(101.18))	-	6.05(11.17))0. 00(-0.01))	98.91	0.00	29.50	Poor
c.5096G>C (p.(Arg1699Pro))	Conflicting	0.02	0.01	0.0	0	No	2.93	0.84	-2.09	0.43	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5096G>T (p.(Arg1699Leu))	Likely pathogenic	0.19	0.12	0.0	0	No	2.93	1.15	-1.78	0.42	96.2	0.3	3.6	-	-	-	-	-	-	-	-
c.5097G>A (p.(Arg1699=))	Likely benign	0.07	0.01	0.0	0	No	2.69	1.79	-0.91	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5098A>G (p.(Thr1700Ala))	Conflicting	0.14	0	0.0	0	No	2.41	0.78	-1.63	0.41	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5098del (p.(Thr1700Hisfs*2))	Pathogenic	0.13	0.01	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5099C>T (p.(Thr1700Ile))	Likely pathogenic	0.03	0.01	0.0	0	No	1.63	1.24	-0.39	0.40	100.0	0.0	0.0	-	-	-	-	-	-	-	-

BRCA1 (NM_007294.3) Variant	ΔSpliceAI (4999) ^b	SPiP V2.1 ^c	ESRseq	Minigene			Patient carriers			mESC			Complementation ^d										
Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	%FLSplicing fraction (Per- allele)	%FLSplicing fraction (Per- allele)	Splicin g fraction (Per- allele)	%Δ18 Splicing fraction (Per- allele)	FL allele-specific result	%FL	%Δ18	HDR	
c.5100A>G (p.(Thr1700=))	Likely Benign	reviewed by expert panel	0	0.09	0.0	0	No	0.96	3.07	2.11	-	100.0	0.0	0.0	PBMCs PAXge nel	85.36(71.19) 9 9.23(102.83)	-	-	12.61(25.00) 0 .00(-0.63)	97.15	0.00	122.7 0	Good
c.5101C>A (p.(Leu1701Met))	Conflicting	criteria provided,con flicting interpretation s	0.01	0	0.0	0	No	0.96	0.73	-0.22	0.07	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5103G>A (p.(Leu1701=))	Likely benign	criteria provided,mul tiple submitters	0.13	0	0.0	0	No	0.35	-2.53	-2.88	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5102_5103del (p.(Leu1701Glnfs*14))	Pathogenic	reviewed by expert panel	0.12	0.01	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5104A>T (p.(Lys1702*))	not provided	-	0.18	0.07	0.0	0	No	-0.22	-1.47	-1.24	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5106del (p.(Lys1702Asnfs*4))	Pathogenic	reviewed by expert panel	0	0.04	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5106A>G (p.(Lys1702=))	Conflicting	criteria provided,con flicting interpretation s	0.01	0.05	0.0	0	No	-1.07	-0.21	0.86	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5107T>C (p.(Tyr1703His))	Uncertain	criteria provided,sing le submitter	0.01	0.05	0.0	0	No	-1.64	-0.22	1.41	0.37	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5107T>G (p.(Tyr1703Asp))	Uncertain	criteria provided,sing le submitter	0.01	0.07	0.0	0	No	-1.64	0.79	2.43	0.43	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5108A>C (p.(Tyr1703Ser))	not provided	-	0	0.06	0.0	0	No	-1.83	-0.32	1.51	0.42	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5108A>G (p.(Tyr1703Cys))	Conflicting	criteria provided,con flicting interpretation s	0	0.03	0.0	0	No	-1.83	-0.18	1.65	0.42	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5109T>C (p.(Tyr1703=))	Likely benign	reviewed by expert panel	0.01	0.06	0.0	0	No	-1.83	-0.61	1.22	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5109T>G (p.(Tyr1703*))	Pathogenic	reviewed by expert panel	0.13	0	0.0	0	No	-1.83	-1.96	-0.13	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5112del (p.(Leu1705*))	Pathogenic	reviewed by expert panel	0.01	0.02	0.0	0	No	-	-	-	-	97.7	0.8	1.5	-	-	-	-	-	-	-	-	-
c.5113C>G (p.(Leu1705Val))	Conflicting	criteria provided,con flicting interpretation s	0.05	0	0.0	0	No	-0.96	-2.40	-1.44	0.07	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-

BRCA1 (NM_007294.3) Variant		ΔSpliceAI (4999) ^a		SHIP V2.1 ^c		ESRseq		Minigene		Patient carriers		mESC		Complementation ^d								
Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	%FL Splicing fraction (Per- allele)	% c.5091 5092del (c.5078 5080del (Splicin g fractio n (Per- n (Per- allele) Splicin g fractio n (Per- n (Per- allele) %Δ18 Splicing fraction (Per- allele)	FL allele-specific result	%FL	%Δ18	HDR		
c.5113C>T (p.(Leu1705=))	Likely benign	reviewed by expert panel	0.12	0	0.0	0	No	-0.96	-3.48	-2.52	-	100.0	0.0	0.0	-	-	-	-	-	-		
c.5113_5114del (p.(Leu1705Argfs*10))	Not reported	-	0.18	0	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-		
c.5114_5121del (p.(Leu1705Argfs*8))	Pathogenic	reviewed by expert panel	0.01	0.14	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-		
c.5114T>C (p.(Leu1705Pro))	Pathogenic & Likely pathogenic	criteria provided,mul title submitters	0	0.15	0.0	0	No	-1.56	-0.45	1.11	0.34	100.0	0.0	0.0	-	-	-	-	-	-		
c.5114T>G (p.(Leu1705Arg))	Uncertain	criteria provided,sing le submitter	0.01	0.16	0.0	0	No	-1.56	2.23	3.78	0.31	100.0	0.0	0.0	-	-	-	-	-	-		
c.5115A>C (p.(Leu1705=))	Likely benign	reviewed by expert panel	0.01	0.15	0.0	0	No	-1.56	2.12	3.68	-	100.0	0.0	0.0	-	-	-	-	-	-		
c.5116G>A (p.(Gly1706Arg))	Pathogenic & Likely pathogenic	criteria provided,mul title submitters	0.01	0.06	0.0	0	No	-1.32	0.27	1.58	0.39	100.0	0.0	0.0	-	-	-	-	-	-		
c.5116G>C (p.(Gly1706Arg))	Pathogenic & Likely pathogenic	criteria provided,mul title submitters	0.01	0.16	0.0	0	No	-1.32	1.45	2.76	0.39	100.0	0.0	0.0	-	-	-	-	-	-		
c.5117G>C (p.(Gly1706Ala))	Benign	reviewed by expert panel	0.01	0.04	0.0	0	No	-0.78	-1.39	-0.62	0.23	100.0	0.0	0.0	PBMCsWhole -blood(Trizol)	100.00(100.46)85.37(73.34)]	-	0.00(- 0.22) 11.75(22 .57)]	96.22	0.00	74.00	Good
c.5117G>A (p.(Gly1706Glu))	Pathogenic	reviewed by expert panel	0.01	0.03	0.0	0	No	-0.78	-0.68	0.09	0.39	100.0	0.0	0.0	PBMCs)	100.00(100.46)	-	0.00(-0.22)	95.80	0.41	8.50	Poor
c.5120T>C (p.(Ile1707Thr))	Uncertain	criteria provided,mul title submitters	0	0.1	0.0	0	No	0.75	2.54	1.80	0.40	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5120T>G (p.(Ile1707Ser))	Uncertain	criteria provided,mul title submitters	0.01	0.03	0.0	0	No	0.75	0.50	-0.25	0.40	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5122G>A (p.(Ala1708Thr))	Uncertain	criteria provided,mul title submitters	0	0.01	0.0	0	No	1.26	0.65	-0.61	0.33	100.0	0.0	0.0	-	-	-	-	-	-	-	-
c.5122G>C (p.(Ala1708Pro))	not provided	-	0	0.08	0.0	0	No	1.26	2.18	0.91	0.42	100.0	0.0	0.0	-	-	-	-	-	-	-	-

BRCA1 (NM_007294.3) Variant	ΔSpliceAI (4999) ^b					SPiP V2.1 ^c		ESRseq		Minigene		Patient carriers				mESC				Complementation ^d			
Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	FL %	Δ18%	SS shift %	Sample type	%FLSplicing fraction (Per- allele)	% c.5091_5092del Splicin g fraction (Per- allele)	% c.5078_5080del Splicin g fraction (Per- allele)	%Δ18 Splicing fraction (Per- allele)	FL allele-specific result	%FL		%Δ18	HDR	
c.5123C>A (p.(Ala1708Glu))	Pathogenic	reviewed by expert panel	0.22	0	0.0	0	No	1.62	-0.96	-2.57	0.42	99.9	0.1	0.0	LCLs with purmoycin PBMCs Whole-blood(Trizol)	-	-	2.37(4.74) 19.26(38.31) 16.68(32.43)	Both alleles contribute equally to FL expression (94.3%FL in AE assay) WT allele contributes more than mutant allele to FL expression WT allele contributes more than mutant allele to FL expression	96.62	0.09	12.50	Poor
c.5123C>G (p.(Ala1708Gly))	Uncertain	criteria provided,single submitter	0.4	0	0.0	0	No	1.62	-1.04	-2.66	0.30	72.4	27.3	0.3	-	-	-	-	-	-	-	-	-
c.5123C>T (p.(Ala1708Val))	Uncertain	criteria provided,multiple submitters	0.13	0	0.0	0	No	1.62	-0.19	-1.81	0.43	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5123del (p.(Ala1708Glyfs*6))	Pathogenic	reviewed by expert panel	0.4	0	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5124G>A (p.(Ala1708=))	Likely Benign	reviewed by expert panel	0	0.1	0.0	0	No	1.36	-0.39	-1.76	-	100.0	0.0	0.0	PBMCs	100.00(100.46)	-	0.00(-0.22)	-	95.91	0.00	95.50	Good
c.5125G>A (p.(Gly1709Arg))	Uncertain	criteria provided,multiple submitters	0	0.09	0.0	0	No	1.48	1.69	0.21	0.34	97.2	2.6	0.2	-	-	-	-	-	-	-	-	-
c.5126del (p.(Gly1709Glufs*5))	Pathogenic	reviewed by expert panel	0	0.14	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5126G>A (p.(Gly1709Glu))	Uncertain	criteria provided,single submitter	0	0.13	0.0	0	No	1.61	2.56	0.94	0.34	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5128G>A (p.(Gly1710Arg))	Uncertain	criteria provided,multiple submitters	0	0.11	0.0	0	No	0.85	2.68	1.83	0.34	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5128G>T (p.(Gly1710*))	Pathogenic	reviewed by expert panel	0	0.01	0.0	0	No	0.85	1.32	0.47	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5129G>A (p.(Gly1710Glu))	Conflicting	criteria provided,conflicting interpretation	0	0.07	0.0	0	No	0.36	0.31	-0.05	0.40	99.7	0.0	0.3	-	-	-	-	-	-	-	-	-
c.5129del (p.(Gly1710Glufs*4))	Pathogenic	reviewed by expert panel	0	0.08	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-
c.5131A>C (p.(Lys1711Gln))	Conflicting	criteria provided,conflicting interpretation	0	0.06	0.0	0	No	0.19	1.09	0.90	0.05	100.0	0.0	0.0	-	-	-	-	-	-	-	-	-

BRCA1 (NM_007294.3) Variant		ASpliceAI (4999) ^b		SPiP V2.1 ^c		ESRseq		Minigene		Patient carriers		mESC				Complementation ^d				
Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	%FL Splicing fraction (Per- allele)	% c.5091_5092del Splicin g fractio n (Per- allele)	% c.5078_5080del Splicin g fractio n (Per- allele)	%FL Splicing fraction (Per- allele)	%Δ18	HDR
c.5131A>T (p.(Lys1711*))	Pathogenic	reviewed by expert panel	0.12	0	0.0	0	No	0.19	-0.10	-0.30	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5133dup (p.(Trp1712Metfs*4))	Pathogenic	reviewed by expert panel	0.03	0.02	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5133del (p.(Lys1711Asnfs*3))	Pathogenic	reviewed by expert panel	0.07	0	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5132A>G (p.(Lys1711Arg))	not provided	-	0.01	0.05	0.0	0	No	-0.48	2.24	2.72	-0.03	100.0	0.0	0.0	-	-	-	-	-	-
c.5135G>A (p.(Trp1712*))	Pathogenic	reviewed by expert panel	0.16	0.1	0.0	0	No	-2.38	-3.54	-1.16	-	85.5	0.0	14.5	-	-	-	-	-	-
c.5136G>A (p.(Trp1712*))	Pathogenic	reviewed by expert panel	0.02	0.03	0.0	0	No	-2.77	-0.72	2.06	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5137del (p.(Val1713*))	Pathogenic	reviewed by expert panel	0	0.04	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5137G>A (p.(Val1713Ile))	Conflicting	criteria provided,con flicting interpretation	0.02	0.15	0.0	0	No	-3.43	-0.36	3.08	-0.13	100.0	0.0	0.0	-	-	-	-	-	-
c.5138T>C (p.(Val1713Ala))	Conflicting	criteria provided,con flicting interpretation	0.02	0.14	0.0	0	No	-4.06	-1.03	3.03	0.40	100.0	0.0	0.0	-	-	-	-	-	-
c.5138T>G (p.(Val1713Gly))	Pathogenic	criteria provided,sing le submitter	0.02	0.13	0.0	0	No	-4.06	-0.70	3.36	0.40	100.0	0.0	0.0	-	-	-	-	-	-
c.5139_5141del (p.(Val1714del))	Not reported	-	0.01	0.14	0.0	0	No	-	-	-0.44	100.0	0.0	0.0	0.0	-	-	-	-	-	-
c.5141T>G (p.(Val1714Gly))	Pathogenic & Likely pathogenic	criteria provided,mul tiple submitters	0.16	0.95	0.0	0	No	-3.83	-4.40	-0.57	0.49	0.0	0.0	###	-	-	-	-	-	-
c.5142T>G (p.(Val1714=))	Likely benign	reviewed by expert panel	0.01	0.13	0.0	0	No	-4.03	-1.64	2.39	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5143A>C (p.(Ser1715Arg))	Pathogenic	reviewed by expert panel	0.01	0.15	0.0	0	No	-3.71	0.89	4.60	0.34	100.0	0.0	0.0	-	-	-	-	-	-
c.5143A>T (p.(Ser1715Cys))	Likely pathogenic	criteria provided,mul tiple submitters	0	0.11	0.0	0	No	-3.71	-1.22	2.49	0.35	100.0	0.0	0.0	-	-	-	-	-	-
c.5144G>A (p.(Ser1715Asn))	Pathogenic	reviewed by expert panel	0	0.11	0.0	0	No	-3.17	-1.12	2.04	-0.05	100.0	0.0	0.0	PBMCs	100.00(100.46)	0.00(-0.22)	-	-	-
c.5144G>T (p.(Ser1715Ile))	Uncertain	criteria provided,sing le submitter	0	0.11	0.0	0	No	-3.17	-0.68	2.49	0.36	96.3	0.0	3.7	-	-	-	-	-	-

BRCA1 (NM_007294.3) Variant	ΔSpliceAI (4999) ^b		SPiP V2.1 ^c		ESRseq		Minigene		Patient carriers		mESC		Complementation ^d											
	Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %		Δ18%	SS shift %	Sample type	%FLSplicing fraction (Per- allele)	%FLSplicing fraction (Per- allele)	%Δ18 Splicing fraction (Per- allele)	%FL	%Δ18	HDR		
c.5145C>A (p.(Ser1715Arg))	Uncertain	no assertion criteria provided	0.04	0	0.0	0	No	-3.09	-2.52	0.57	0.33	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5145C>G (p.(Ser1715Arg))	Pathogenic & Likely pathogenic	criteria provided,mul title	0.23	0.02	0.0	0	No	-3.09	-3.96	-0.87	0.33	100.0	0.0	0.0	PAXgene Whole e-blood (MagNA pure)	99.48(100.15) 94.90(94.16)	-	0.00(- 0.01) 1.18(1.7 3)	-	-	-			
c.5145del (p.(Tyr1716Ilefs*4))	Pathogenic	reviewed by expert panel	0.01	0.06	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5147A>C (p.(Tyr1716Ser))	Conflicting	criteria provided,con flicting interpretation	0	0.03	0.0	0	No	-2.48	-1.05	1.43	0.21	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5147A>G (p.(Tyr1716Cys))	Conflicting	criteria provided,con flicting interpretation	0	0.06	0.0	0	No	-2.48	-0.31	2.17	0.32	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5148T>A (p.(Tyr1716*))	Pathogenic	reviewed by expert panel	0.05	0	0.0	0	No	-1.66	-1.73	-0.07	-	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5148T>G (p.(Tyr1716*))	Pathogenic	reviewed by expert panel	0.2	0	0.0	0	No	-1.66	-1.99	-0.32	-	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5150del (p. phe1717Serfs*3))	Pathogenic	reviewed by expert panel	0.01	0.05	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5150T>A (p. phe1717Tyr)	not provided	-	0	0.04	0.0	0	No	-0.98	-0.99	-0.01	-0.15	100.0	0.0	0.0	-	-	-	-	-	-	-			
c.5152+1G>A (p.?)	Pathogenic	criteria provided,mul title	0.9	0.01	0.0	0	No	-0.93	-1.76	-0.83	-	0.0	99.3	0.7	-	-	-	-	-	-	-			
c.5152+1G>C (p.?)	Pathogenic	criteria provided,mul title	0.9	0.01	0.0	0	No	-0.93	-1.24	-0.31	-	0.0	98.5	1.5	Whole-blood (MagNA pure)	24.91(-45.82)	-	66.27(131.92)	0.00	82.85	8.00	The FL expression comes almost exclusively from the WT allele		Poor
c.5152+1G>T (p.?)	Pathogenic	reviewed by expert panel	0.9	0.03	0.0	0	No	-0.93	-2.89	-1.96	-	0.0	96.9	3.1	PAXgene PAXgene	36.18(- 26.45) 36.74(- 25.34)	-	60.73(121.45) 58.71(117.40)	2.32	81.89	12.00	Poor		Poor
c.5152+2T>A (p.?)	Pathogenic	criteria provided,single submitter	0.9	0	0.0	0	No	-1.60	0.21	1.81	-	0.0	100.0	0.0	-	-	-	-	-	-	-			-
c.5152+2T>G (p.?)	Pathogenic	criteria provided,single submitter	0.9	0	0.0	0	No	-1.60	0.64	2.24	-	0.0	100.0	0.0	-	-	-	-	-	-	-			-

BRCA1 (NM_007294.3) Variant	ASpliceAI (4999) ^a				SPiP V2.1 ^c		ESRseq		Minigene		Patient carriers		mESC				Complementation ^d								
	Clinvar Clinical Significance	Clinvar Review Status ^a	MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	%FLSplicing fraction (Per- allele)		% c.5091_5092del Splicin g	% c.5078_5080del Splicin g	%Δ18 Splicing fraction (Per- allele)	FL allele-specific result	%FL	%Δ18	HDR	
c.5152+2dup (p.?)	Pathogenic	criteria provided,single submitter	0.88	0	0.0	0	No	-	-	-	-	0.4	97.4	2.2	Short term lymphocytes treated with puromycin Whole-blood (MagNA pure) Whole-blood (MagNA pure)	29.45(-37.49) 35.71(-24.21) 35.12(-25.39)	-	-	43.43(86.87) 53.66(106.69) 54.75(108.88)	The FL expression comes almost exclusively from the WT allele The FL expression comes almost exclusively from the WT allele The FL expression comes almost exclusively from the WT allele	1.90	87.17	13.20	Poor	
c.5152+3A>C (p.?)	Conflicting	criteria provided,conflicting interpretations	0.88	0	0.0	0	No	-1.60	-1.51	0.09	-	0.0	99.7	0.4	PBMCs PAXgene Whole-blood (trizol)	36.41(-26.72) 29.74(-39.34) 34.39(-28.61)	-	-	63.59(126.97) 59.73(119.46) 58.33(115.73)	-	0.00	82.55	7.00	Poor	
c.5152+3_5152+4insG (p.?)	Not reported	-	0.9	0	-1.0	N	.	-	-	-	-	0.0	100.0	0.0	-	-	-	-	-	-	-	-	-	-	-
c.5152+4A>G (p.?)	Likely pathogenic	criteria provided,single submitter	0.8	0.01	0.0	0	No	-2.02	-3.89	-1.87	-	0.5	99.5	0.0	-	-	-	-	-	-	-	-	-	-	-
c.5152+4A>T (p.?)	not provided	-	0.75	0.02	0.0	0	No	-2.02	-1.56	0.46	-	2.4	91.9	5.7	Whole-blood (trizol) Whole-blood (trizol) Whole blood (trizol)	33.18(-31.03) 32.71(-31.99) 33.04(-31.31)	-	-	63.35(125.78) 64.42(127.92) 62.51(124.10)	-	1.52	88.02	10.50	Poor	
c.5152+4_5152+5insT (p.?)	Not reported	-	0.9	0.01	-1.0	N	.	-	-	-	-	0.0	98.6	1.4	-	-	-	-	-	-	-	-	-	-	-
c.5152+5G>A (p.?)	Pathogenic & Likely Pathogenic	criteria provided,multiple submitters	0.89	0.02	0.0	0	No	-2.67	-2.61	0.06	-	0.0	99.5	0.5	PBMCs	33.23(-33.08)	-	-	66.77(133.33)	The FL expression comes almost exclusively from the WT allele	0.00	85.49	10.30	Poor	
c.5152+5G>C (p.?)	Pathogenic	criteria provided,single submitter	0.9	0.03	0.0	0	No	-2.67	-1.72	0.94	-	0.0	97.1	2.9	PAXgene	31.75(-35.31)	-	-	60.85(121.69)	The FL expression comes almost exclusively from the WT allele	-	-	-	-	-
c.5152+6T>A (p.?)	Uncertain	criteria provided,single submitter	0.81	0	0.0	0	No	-2.90	-0.05	2.85	-	0.0	100.0	0.0	-	-	-	-	-	-	-	-	-	-	-
c.5152+6T>C (p.?)	Conflicting	criteria provided,conflicting interpretations	0.77	0	0.0	0	No	-2.90	-1.33	1.57	-	1.7	97.4	0.9	PAXgene	35.04(-28.75)	-	-	57.95(115.89)	The FL expression comes almost exclusively from the WT allele	-	-	-	-	-
c.5152+6T>G (p.?)	Conflicting	criteria provided,conflicting interpretations	0.84	0	0.0	0	No	-2.90	0.02	2.92	-	0.0	100.0	0.0	-	-	-	-	-	-	-	-	-	-	-
c.5152+7A>G (p.?)	Conflicting	criteria provided,conflicting interpretations	0.51	0	0.0	0	No	-2.83	-2.08	0.75	-	61.2	38.4	0.4	Whole-blood (trizol)	43.53(-10.34)	-	-	50.28(99.64)	-	64.50	30.18	75.80	Good	

BRCA1 (NM_007294.3) Variant	Clinvar Clinical Significance	Clinvar Review Status *	ASpliceAI (4999) ^a	SPiP V2.1 ^c				ESRseq				Minigene		Patient carriers		mESC				Complementation ^d
				MAX LOSS	MAX GAIN	deltaMES	BP	mutinPBarrea	WT	VAR	ΔESRseq	BayesDel (noAF)	FL %	Δ18%	SS shift %	Sample type	%FL Splicing fraction (Per-allele)	%FL Splicing fraction (Per-allele)	%FL Splicing fraction (Per-allele)	
Interpretation s																				
c.5152+9A>T (p.?)	Likely benign	criteria provided,multiple submitters	0.02	0	0.0	0	No	-2.46	-2.39	0.08	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+10A>G (p.?)	Benign & Likely Benign	criteria provided,multiple submitters	0.16	0.01	0.0	0	No	-2.04	-2.80	-0.77	-	100.0	0.0	0.0	PAXgene	100.00(101.18)	-	0.00(-0.01)	-	-
c.5152+11T>C (p.?)	Likely benign	criteria provided,single submitters	0.01	0	0.0	0	No	-1.64	-1.18	0.46	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+12A>G (p.?)	Likely benign	criteria provided,single submitters	0.02	0	0.0	0	No	-1.35	-1.41	-0.06	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+13C>A (p.?)	Uncertain	no assertion criteria provided	0.03	0	0.0	0	No	-1.41	-2.67	-1.26	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+15A>G (p.?)	Conflicting	criteria provided,conflicting interpretation	0.02	0	0.0	0	No	-1.52	0.00	1.52	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+19C>T (p.?)	Likely benign	criteria provided,single submitters	0	0.01	0.0	0	No	-0.50	-1.89	-1.39	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+20T>A (p.?)	Benign & Likely benign	criteria provided,multiple submitters	0.03	0	0.0	0	No	-0.09	-0.10	-0.01	-	100.0	0.0	0.0	Whole-blood (MagNA pure) PAXgene	96.48(97.33) 96.21(93.60)	-	0.70(0.77) 0.00(-0.01)	-	-
c.5152+23C>T (p.?)	Not reported	-	0	0	0.0	0	No	-0.15	0.72	0.87	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+28T>C (p.?)	Not reported	-	0	0.01	0.0	0	No	-0.28	-1.40	-1.12	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+31_5152+34del (p.?)	Not reported	-	0.05	0.01	0.0	0	No	-	-	-	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+39C>T (p.?)	Not reported	-	0	0.01	0.0	0	No	0.30	0.62	0.32	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+41T>C (p.?)	Not reported	-	0.01	0	0.0	0	No	0.30	-1.17	-1.47	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+44G>A (p.?)	Not reported	-	0	0.02	0.0	0	No	0.35	-1.40	-1.75	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+45A>C (p.?)	Not reported	-	0	0	0.0	0	No	0.14	-0.19	-0.34	-	100.0	0.0	0.0	-	-	-	-	-	-
c.5152+66G>A (p.?)	Benign	reviewed by expert panel	0	0	0.0	0	No	0.05	-0.89	-0.94	-	99.7	0.0	0.3	PBMCs	100.00(100.46)	-	0.00(-0.22)	-	-
c.5152+85del (p.?)	Benign	reviewed by expert panel	0.02	0.03	0.0	0	No	-	-	-	-	99.6	0.0	0.4	PAXgene	100.00(101.18)	-	0.00(-0.01)	-	-

- a. ClinVar review status “reviewed by expert panel”: refers to classifications provided by ENIGMA as external curator expert panel using old rules. The c.5090A>G is the unique variant reviewed externally under the new ENIGMA-VCEP rules (it was included in the pilot dataset from Parsons et al. (2024)[122])
- b. $\Delta\text{SpliceAI}$ raw, window of $\pm 4999\text{bp}$. MAX loss: maximum score between acceptor and donor loss prediction. MAX gain: maximum score between acceptor and donor gain prediction
- c. SPiP: $\text{MES} \leq 0$. -15% for variants in -12 to -18 intronic positions were predicted to affect polypyrimidine tract alterations. BP scoring 1 together with a “yes” prediction of variant being in BP area (mutInBParea) were used to identify variants potentially affecting branch point.
- d. Complementation phenotypes of BRCA1 exon 18 variants in mESC. poor: <20% viability (including no complementation), reduced: 20-50% viability and good: >50% viability

Supplementary Table 10. Detailed information from patients' samples.

ENIGMA collaborator group	Number of carriers	Number of controls (same sample type)	Type of sample	RNA isolation and clean-up	Variants
Institut Català d'Oncologia (Spain)	13 cancer	9 healthy	Frozen lymphocytes isolated from 8 mL of peripheral blood using Ficoll-Paque (GE Healthcare)	Trizol reagent (Invitrogen) and Ambion™ DNase I RNase-free (ThermoFisher)	c.5075-107A>G, c.5075-53C>T, c.5075-49del, c.5078_5080del, c.5100A>G, c.5117G>A, c.5117G>C, c.5123C>A, c.5124G>A, c.5144G>A, c.5152+3A>C, c.5152+5G>A, c.5152+66G>A
Centre François Baclesse (France)	9 cancer	9 healthy	PAXgene RNA tubes (QIAGEN)	PaxGene Blood miRNA Kit (QIAGEN)	c.5075-2A>G, c.5075A>T, c.5077_5080delinsTTCACTCTGC, c.5096G>A, c.5100A>G, c.5145C>G, c.5152+10A>G, c.5152+20T>A, c.5152+85del
Hospital Universitari Vall d'Hebron (Spain)	8 cancer	11 healthy	Whole blood-EDTA tubes	Trizol reagent (Invitrogen) and RNeasy Mini Kit (QIAGEN)	c.5075-53C>T, c.5091_5092del, c.5095C>T, c.5096G>A, c.5117G>C, c.5123C>A, c.5152+3A>C
Masaryk Memorial Cancer Institute (Czech Republic)	6 unknown status	5 unknown status	Whole blood-EDTA tubes	MagNA Pure Compact RNA Isolation Kit (Roche)	c.5095C>T, c.5145C>G, c.5152+1G>C, c.5152+2dup, c.5152+2dup, c.5152+20T>A
Hospital Universitario la Princesa (Spain)	3 unknown status	11 healthy	Whole blood-EDTA tubes	Trizol reagent (Invitrogen) and RNeasy Mini Kit (QIAGEN)	c.5152+4A>T
University Hospital Southampton (United Kingdom)	3 cancer	3 healthy	PAXgene blood RNA tubes (QIAGEN)	PaxGene Blood miRNA Kit (QIAGEN)	c.5075-6C>A, c.5152+5G>C, c.5152+6T>C
Manchester University NHS Foundation Trust (United Kingdom)	1 healthy 1 cancer	9 healthy	PAXgene blood RNA tubes (QIAGEN)	PaxGene Blood miRNA Kit (QIAGEN)	c.5152+1G>T
Ghent University Hospital (Belgium)	1 cancer	6 unknown status	Short-term lymphoblast culture using 1 µl/ml phytohemagglutinin-M (PHA-M; Gibco) and 50 µl/ml interleukin-2(IL-2; Sigma-Aldrich). Cells were treated with 0.2mg/mL Puromycin (Sigma-Aldrich), before total RNA extraction.	MagCore (Promega)	c.5152+2dup
Hospital Clínico San Carlos (Spain)	1 cancer	9 healthy	PAXgene blood RNA tubes (QIAGEN)	PaxGene Blood miRNA Kit (QIAGEN)	c.5152+3A>C

ENIGMA collaborator group	Number of carriers	Number of controls (same sample type)	Type of sample	RNA isolation and clean-up	Variants
Hospices Civils de Lyon (France) (cases) and Inserm U1245, Univ Rouen Normandie (France) (controls)	4 cancer	3 healthy	EBV-immortalized lymphoblastoid cell lines (LCLs) untreated/treated with puromycin. LCLs were cultivated at the Inserm U1245 laboratory (France) in RPMI medium (Life Technologies) supplemented with 2 mM L-glutamine and 10% fetal calf serum, at 37°C in a 5% CO2 atmosphere. Before RNA extraction, LCLs were transferred into 6-well plates, at 2.5x10 ⁶ cells/well, and incubated for 5.5 hours with/without 200 µg/ml Puromycin.	NucleoSpin RNA II kit (Macherey Nagel)	c.5078_5080del, c.5080G>T, c.5095C>T, c.5123C>A
Hospital Universitari Sant Joan de Reus (Spain)	1 cancer	11 healthy	Whole blood-EDTA tubes	Trizol reagent (Invitrogen) and RNeasy Mini Kit (QIAGEN)	c.5152+7A>G

Supplementary Table 11. Primers and PCR conditions.

Analyses of patients' RNA and DNA									
Primers features			PCR conditions						
Primers	Exon	Sequence	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature	Assay	
		(3'-->5')							
R-B1- 16F/ R-B1- 16F*	16	GGGAGAGCCAGAAATTGACA	498	58°C	15 sec	35	10min, 72°C	cDNA High sensitivity capillary-electrophoresis cDNA SANGER	
R-B1-22R	22	GGGTGAATGATGAAAGCTCCT							
BRCA1-16F	16	AATTCTTAACAGAGACCCAGAAC	450	58°C	15 sec	30	10min, 72°C	DNA SANGER	
BRCA1-16R	16	AAACTCTTTCCAGAAATGTTGT							
BRCA1-18F	18	GGACAGCACTTCCTGATTTTG	236	58°C	15 sec	30	10min, 72°C		
BRCA1-18R	18	CTCAGCATCAGCAAAAAACCTT							
R-BRCA1-15F.1	15	CTGGGCCACACGATTTGAC	475	57°C	15 sec	35	10min, 72°C	Allele-specific (cDNA SANGER)	
R-BRCA1-18R.1	18	GTCCGTTTCACACACAAACTCA							
R-B1-15F.2	15	GACGGAAACATCTTACTTTGCCA	497	57°C	15 sec	35	10min, 72°C		
R-B1-18R.2	18	CCCATTTTCTCCCGCAATT							
R-BRCA1-18F.1.	18	GAGTTTGTGTGTAACGGGACA	188	57°C	15 sec	35	10min, 72°C		
R-BRCA1-20R.1.	20	TGGGATTCTCTTGCTCGCTT							
B1ex18-19.InFus.BamHI-pSPL3m-F		GGCGCTGCAGGATCCCATACCTGCTGTATAATTAGTG	990=16+958+16	57	1 min	30	10 min, 72°C	PCR on genomic DNA	
B1ex18-19.InFus.MluI-pSPL3m-R		GTACGGGATCACGGGTTCACAGTGCAGGCCTGCAT							
RT_BRCA1_Ex16-F	16	GAATGTCCATGGTGGTGCTCTGG	387	55	30 sec	26	10 min, 72°C	RT-PCR on total RNA	
RT_BRCA1_Ex21-R	21	CTGTGGGCATGTTGGTGAAGG							

Analyses of patients' RNA and DNA									
Primers features				PCR conditions					
Primers	Exon	Sequence		Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature	Assay
		(3' -->5')							
SnapShot_B1Ex18-5078_5080del-R	18	TGTCGGTTCACACAAACTCA		-	-	-	-	-	Primer extension reactions on PCR (gDNA) and RT-PCR (RNA) products
SnapShot_BRCA1Ex18_c.5080GT -R	18	GTGTCCGTTCAACACAAACT		-	-	-	-	-	
SnapShot_BRCA1Ex18_c.5123CA-F	18	CGGACACTGAAATATTTTCTAGGAATTG		-	-	-	-	-	
SnapShot_BRCA1Ex18_c.5123CA-R	18	ATAGCTAACTACCCATTTTCTCTCC		-	-	-	-	-	

*Primer labelled with 6-Carboxyfluorescein (6-FAM)

B)

Minigene splicing assay								
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature	Purpose	
B1ex18-19.InFus.BamHI-pSPL3m-F	GGCGCTGCAGGATCCCATACCTGC	990	57	1 min	30	10min, 72°C	Minigene preparation: cloning, mutagenesis, Sanger sequencing	
	GTACGGGATCAGCGGCTCACAGTGCAG							
B1ex18-19.InFus.MluI-pSPL3m-R	GCCTGCAT						Minigene preparation: cloning, mutagenesis, Sanger sequencing	
pspl3-937F	CCTTGGGATGTTGATGATCTG	-	-	-	-	-	Minigene preparation: Sanger sequencing	
pspl3-1191R	ACTTCTGTGGGTGGGGTC	-	-	-	-	-	Minigene preparation: Sanger sequencing	
BRCA1 c.5075-82_SeqF	CTGAGGCTCTTTAGCTTCTTAGGAC	-	-	-	-	-	Minigene preparation: Sanger sequencing	
BRCA1 c.5152+108_SeqR	GGTAACTCAGACTCAGCATCAGC	-	-	-	-	-	Minigene preparation: Sanger sequencing	
BRCA1_Ex18_c.5075-107AG.F	GAA TCA TAA ATC CGG ATT GAT CTT GGG	-	-	-	-	-	Minigene preparation: mutagenesis	
BRCA1_Ex18_c.5075-53CT.F	CTT AGG ACA GCA TTT CCT GAT TTT G	-	-	-	-	-	Minigene preparation: mutagenesis	
BRCA1_Ex18_c.5075-49del.F	CTT AGG ACA GCA CTT CTG ATT TTG	-	-	-	-	-	Minigene preparation: mutagenesis	
BRCA1_Ex18_c.5075-37del-F	GATTTGTTTCAACTTCTAATC	-	-	-	-	-	Minigene preparation: mutagenesis	
BRCA1_Ex18_c.5075-35AC-F	GATTTGTTTTCCACTTCTAATCC	-	-	-	-	-	Minigene preparation: mutagenesis	

Minigene splicing assay						
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature
BRCA1_Ex18_c.5075-33CT-F	TGTTTTCAATTTCTAAATCCCTTG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-32TC-F	TGTTTTCAACCTCTAATCCCTTG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-30CT-F	GTTTTCAACTTTTAAATCCTTTGAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-27AC-F	TTCAACTTCTACTCCTTTGAGTG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-24CT-F	CAACTTCTAATCTTTTGAGTGTTT	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-23TG-F	CAACTTCTAATCCGTTGAGTGTTT	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-18GA-F	AATCCTTTGAATGTTTTTCATTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-13TG-F	CTTTGAGTGTTGTTCAATCTGCAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-11TC-F	TGAGTGTTTCCATCTGCAGATG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-9AG-F	GAGTGTTTTTCGTTCTGCAGATG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-9AT-F	GAGTGTTTTTCTTTCTGCAGATG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-8TG-F	GTGTTTTTCAGTCTGCAGATG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-7TC-F	GTGTTTTTCATCCTGCAGATGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-6CA-F	GTTTTTCATTATGCAGATGCTG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-6CT-F	GTTTTTCATTTTGCAGATGCTG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-5TC-F	GTTTTTCATTCGCGCAGATGCTG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-4GA-F	TTTTTCATTCTACAGATGCTGAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-3CG-F	TTTCATTCTGGAGATGCTGAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-3CT-F	TTTCATTCTGTAGATGCTGAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-2AC-F	TTTCATTCTGCCGATGCTGAGTT	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-2AG-F	TTTCATTCTGGGATGCTGAGTT	-	-	-	-	Minigene preparation: mutagenesis

Minigene splicing assay							
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature	Purpose
BRCA1_Ex18_c.5075-2AT-F	TTCATTCTGCTGATGCTGAGTT	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-2del-F	TTCATTCTGCGATGCTGAGTT	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-1GA-F	TCATTCTGCAAAATGCTGAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-1GC-F	TCATTCTGCACATGCTGAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075-1GT-F	TCATTCTGCATATGCTGAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075AG-F	CATTCTGCAGCTGCTGAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075AT-F	CATTCTGCAGTTGCTGAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5075_5078del-F	CATTCTGCAGTGAGTTTGTGTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5076del-F	CATTCTGCAGAGCTGAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5077_5079delins-F	TCATTCTGCAGATTTCATTCTGCAGAGTTTGTGTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18c.5077_5080delinsC-F	TCATTCTGCAGATTTCATTCTGCAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18c.5077_5080delinsG-F	TCATTCTGCAGATTTCATTCTGCAGTTTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5078_5080del	CATTCTGCAGATGAGTTTGTGTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5080GA-R	CACACACAAACCTAGCATCTGCAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5080GT-F	CTGCAGATGCTTAGTTTGTGTGTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5084_5085del-F	CAGATGCTGAGTGTGTGAACGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5085TA-F	GATGCTGAGTTAGTGTGAAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5085dup-F	GATGCTGAGTTTGTGTGAAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5086GA-F	GATGCTGAGTTTATGTGTGAACG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5086GC-F	GATGCTGAGTTTCTGTGAACGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5086GT-F	GATGCTGAGTTTGTGTGAACG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5087TA-F	GCTGAGTTTGAGTGTGAACGGAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5088GA-F	CTGAGTTTGTATGTGAACGGAC	-	-	-	-	-	Minigene preparation: mutagenesis

Minigene splicing assay							
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature	Purpose
BRCA1_Ex18_c.5089TA-F	TCAGTTTGTGAGTGAACGGACAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5089_5092del-F	CTGAGTTTGTGAACGGACACTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5090GA-F	GAGTTTGTGTATGAACGGACAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5091_5092del-F	CTGAGTTTGTGTGAACGGACACTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5092_5093delinsAG-F	GTTTGTGTGTAGACGGACACTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5094AG-F	TTGTGTGTGAGCGGACACTGAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5095CA-F	TTGTGTGTGAAGGACACTGAAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5095CT-R	TTTCAGTGTCCATTTCACACACAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5096GA-F	GTGTGTGAACAGACACTGAAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5096GC-F	GTGTGTGAACCGACACTGAAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5096GT-F	GTGTGTGAACITGACACTGAAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5097GA-R	ATTTCAGTGTTCGTTTCACACAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5098AG-R	ATATTCAGTGCCCGTTCACAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5098del-F	GTGTGAACGGCACTGAAATAT	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5099CT-R	ATATTCAGTATCCGTTTCACAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5100AG-F	GTGAACGGACGCTGAAATATTTTC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5101CA-F	GTGAACGGACAATGAAATATTTTC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5102_5103del-R	AGAAAAATATTGTGTCGGTTCAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5103GA-F	GAACGGACACTAAAAATATTTTC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5104AT-R	GAAATATTACAGTGTCGGTTC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5106AG-F	CGGACACTGAAGTATTTTCTAGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5106del-F	CGGACACTGAATATTTTCTAGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5107TC-F	GGACACTGAAACATTTTCTAGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5107TG-F	GGACACTGAAAGATTTTCTAGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5108AC-F	GACACTGAAATCTTTTCTAGGAATTG	-	-	-	-	-	Minigene preparation: mutagenesis

Minigene splicing assay							
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature	Purpose
BRCA1_Ex18_c.5108AG-F	GACACTGAAATGTTTCTAGGAATTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5109TC-F	GACACTGAAATACCTTCTAGGAATTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5109TG-F	GACACTGAAATAGTTTCTAGGAATTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5112del-F	CTGAAATATTCTAGGAATTGCG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5113_5114del-R	CGCAATTCTCTAAAAATATTTCAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5113CG-R	GCAATTCTCTACAAAATATTTCAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5113CT-R	GCAATTCTCTATAAAATATTTCAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5114TC-F	GAAATATTTTCCAGGAATTGCGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5114TG-F	GAAATATTTTTCGAGGAATTGCGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5114_5121del-F	CTGAAATATTTTCGCGGGAGGAAAATG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5115AC-F	GAAATATTTTCTCGGAATTGCG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5116GA-F	ATATTTCTAAGAAATTGCGGGAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5116GC-F	ATATTTCTACGAATTGCGGGAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5117GA.F	GAAATATTTTCTAGAAAATGCGGGAGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5117GC-R	CTCCCGCAATTGCTAGAAAATATTTC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5120TC-F	TTTCTAGGAACTGCGGGAGGAAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5120TG-F	TTTCTAGGAAGTGC GGAGGAAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5122GA-R	CATTTTCCTCCCGTAATTCCTAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5122GC-F	TCTAGGAATTCGGGAGGAAAATG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5123CA-R	CATTTTCCTCCCTCAATTCCTAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5123CG-R	ATTTTCCTCCCGCAATTCCTAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5123CT-R	ATTTTCCTCCCGCAATTCCTAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5123del-R	CATTTTCCTCCCGCAATTCCTAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5124GA-R	CATTTTCCTCTGCAATTCCTAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5125GA-F	TCTAGGAATTGCGAGAGGAAAATG	-	-	-	-	-	Minigene preparation: mutagenesis

Minigene splicing assay						
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature
BRCA1_Ex18_c.5126GA-R	CCCATTTCCTCCGGCAATTCC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5126del-R	CCATTTTCCTCCGGCAATTCTAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5128GA-R	CTACCCATTTCCTCCGGCAATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5128GT-R	CTACCCATTTCCTCCGGCAATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5129GA-R	CTACCCATTTCCTCCGGCAATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5129del-R	CTACCCATTTCCTCCGGCAATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5131AC-R	TAACTACCCATTTCCTCCGGCAA	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5131AT-R	AACCTACCCATTTCCTCCGGCAA	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5132AG-R	TAACTACCCATTTCCTCCGGCAA	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5133del-R	GCTAACTACCCATTTCCTCCG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5133dup-R	GCTAACTACCCATTTCCTCCG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5135GA-R	AGCTAACTACCTATTTCCTCC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5136GA-R	TAGCTAACTACTCATTTTCCTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5137GA-R	ATAGCTAACTATCCATTTCCTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5137del-R	ATAGCTAACTACCATTTTCCTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5138TC-R	AATAGCTAACTGCCCCATTTCCTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5138TG-R	AATAGCTAACTGCCCCATTTCCTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5139_5141del-R	CAGAAATAGCTAACCCATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5141TG-R	GAAATAGCTACCTACCCATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5142TG-R	CAGAAATAGCTCACTACCCATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5143AC-R	CAGAAATAGCGAACTACCCATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5143AT-R	CAGAAATAGCAAACTACCCATTTC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5144GA-R	CTTACAGAAATAGTTAACTACCC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5144GT-R	TACAGAAATAGATACTACCCAT	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5145CA-R	CTTACAGAAATATCTAACTACCC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5145CG-R	CTTACAGAAATACCTAACTACCCA	-	-	-	-	Minigene preparation: mutagenesis

Minigene splicing assay						
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature
BRCA1_Ex18_c.5145del-R	CTTACAGAAATACTAACTACCCA	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5147AC-R	TACTTACAGAAAAGAGCTAACTACC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5147AG-R	TACTTACAGAAAACAGCTAACTACC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5148TA-R	TACTTACAGAAATTAGCTAACTACC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5148TG-R	TACTTACAGAACTAGCTAACTACC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5150TA-R	TACTTACAGTAATAGCTAACTAC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5150del-R	TACTTACAGAACTAGCTAACTAC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+1G	GTATTATACCTTAGAAATAGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+1G C-R	GTATTATACCTTAGAGAAATAGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+1GT-R	GTATTATACCTTAAAGAAATAGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+2TA-R	GTATTATACCTTCAGAAATAGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+2TG-R	GTATTATACCTCCAGAAATAGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+2dup-R	GTATTATACCTTAACAGAAATAGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+3A C-R	GTATTATACCTGACAGAAATAGC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+3_+4insG-R	AGTATTATACCTTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+4A G-R	AGTATTATACCTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+4AT .R	GTATTATACATACAGAAATAGCTAACTAC	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+4_+5insT-R	ATAGTATTATACATTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+5G A-R	ATAGTATTATTTTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+5G C-R	ATAGTATTATAGTTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+6TA-R	AATAGTATTATCTTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+6TG-R	AATAGTATTATCCTTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+6TC-R	AATAGTATTATGCTTACAGAAATAG	-	-	-	-	Minigene preparation: mutagenesis

Minigene splicing assay							
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature	Purpose
BRCA1_Ex18_c.5152+7A G-R	GAAATAGTATTACACTTACAGA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+9AT -R	GAGAAATAGTATAACTTACAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+10A G-R	GAGAAATAGTACTATATCTTACAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+11T C-R	GAGAAATAGTGTATTACTTACAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+12A G-R	GGAGAAATAGCATTATACCTTAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+13C A-R	GGAGAAATATTATTATCTTAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+15A G-R	GAGGGGAGAAACAGTATTATAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+19C T-R	GAGGAGGGGAAAAATAGTATTATAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+20T A-R	GAGGAGGGGTGAAATAGTATTATAC	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+23C T-R	GTAAAGGGGAGGAGGAGGAGAAATAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+28T C-R	GTGTTAAAGGGGGGAGGGGAGAGAAA	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+31_+34del-R	TCTGAGGTGTTGGAGGAGGGGGAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+39C T-R	GCAATTCTGAGATGTTAAAGGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+41T C-R	ATGCAATTCTGGGTGTTAAAGG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+44 G-R	AAATGCAATTTTGGGTGTTAAAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+45A C-R	AAATGCAATGCTGAGGTGTTAAAG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+66 G-R	AAACCTTAGGTGTTAAATGTTAGGTG	-	-	-	-	-	Minigene preparation: mutagenesis
BRCA1_Ex18_c.5152+85del-R	GACTCAGCATCAGCAAAACCTTAG	-	-	-	-	-	Minigene preparation: mutagenesis
6FAM-pSPL3-SD6 (5'-fluor)*	TCTGAGTCACCTGGACAACC	381	57	30 sec	30 (200 ng RNA)	10 min, 72°C	Analysis of minigene transcripts: RT-PCR
pSPL3-SD6	TCTGAGTCACCTGGACAACC	-	-	-	-	-	Analysis of minigene transcripts: sequencing of RT-PCR products

Minigene splicing assay						
Primer name	Primer sequence (5'-3')	Amplicon size (bp)	Annealing temperature	Elongation time	Cycles	Final elongation time and temperature
pSPL3-SA2	ATCTCAGTGGTATTGTGAGC	-	-	-	-	-

* Primer labelled with 6-Carboxyfluorescein (6-FAM)

Supplementary Table 12. Splicing events detected by semi-quantitative analysis in patients and mESC RNA samples.

Splicing event ^a	RNA ^b	Inferred / Sanger Confirmed	Functional annotation ^c	Coding sequence	Previously described	Amplicon size (RT-PCR 16F and 22R)
Δ17	r.4987_5074del	Inferred by size	PTC-NMD		PMID: 24569164, PMID: 29774201, PMID: 32133419, PMID: 28905878, PMID: 28919163	410
Δ18	r.5075_5152del	Confirmed by Sanger	No FS	p.(Asp1692_Trp1718delinsGly)	PMID: 24569164, PMID: 29774201, PMID: 28919163	420
Δ18q	r.5137_5152del	Inferred by size	PTC-NMD		PMID: 30101128(identified in RNA sample from one carrier of c.5152+5G>T variant)	482
Δ19	r.5153_5193del	Inferred by size	PTC-NMD		PMID: 24569164 and PMID: 29774201 (mentioned, not validated)	457
Δ20	r.5194_5277del	Inferred by size	No FS	p.(His1732_Lys1759del)	PMID: 24569164, PMID: 29774201, PMID: 28919163	414
Δ21	r.5278_5332del	Inferred by size	PTC-NMD		PMID: 24569164, PMID: 29774201, PMID: 32133419, PMID: 28905878, PMID: 28919163	443
Δ17_18	r.4987_5152del	Confirmed by Sanger	PTC-NMD		PMID: 28919163	332
Δ17_19	r.4987_5193del	Inferred by size	No FS	p.(Met1663_Glu1731del)	PMID: 24569164, PMID: 29774201 (mentioned, not validated), PMID: 28919163	291
Δ17_20	r.4987_5277del	Inferred by size	No FS	p.(Met1663_Lys1759del)	PMID: 28919163	206
Δ18_19	r.5075_5193del	Inferred by size	PTC-NMD		PMID: 29774201, PMID: 28919163	379
Δ18_20	r.5075_5277del	Inferred by size	PTC-NMD		PMID: 24569164 and PMID: 29774201 (mentioned, not validated)	295
Δ18_21	r.5075_5332del	Inferred by size	No FS	p.(Asp1692_Thr1777del)	Not found in literature, but inferred by size (detected in some carriers of exon 18 variants)	240
Δ18+Δ20	r.5075_5152del + r.5194_5277del	Inferred by size	No FS	p.(Asp1692_Trp1718delinsGly) + p.(His1732_Lys1759del)	Not found in literature, but inferred by size (detected in some carriers of exon 18 variants)	336
Δ18+Δ21	r.5075_5152del + r.5278_5332del	Inferred by size	PTC-NMD		Not found in literature, but inferred by size (detected in some carriers of exon 18 variants)	365
▼ 135bp ins?	Unknown	Inferred by size	Unknown		633 bp peak detected which could not be inferred as any transcript described before and too low expressed to be sequenced	633

Splicing event ^a	RNA ^b	Inferred / Sanger Confirmed	Functional annotation ^c	Coding sequence	Previously described	Amplicon size (RT-PCR 16F and 22R)
▼ 146bp int19	r.5193_-1234_5194-1089ins146	Inferred by size	PTC-NMD		PMID: 28919163	644
▼ 16q	r.4986_4987ins4986+1_4_986+65	Inferred by size	PTC-NMD		PMID: 32133419, PMID: 28905878, PMID: 28919163	563
▼ 2bp ins ?	Unknown	Inferred by size	Unknown		500 bp peak detected in mESCs that could not be inferred as any transcript described before and too low expressed to be sequenced	500
▼ 52bp ins ?	Unknown	Inferred by size	Unknown		550 bp peak detected in patients that could not be inferred as any transcript described before and too low expressed to be sequenced	550
Full-length		Confirmed by Sanger				498

^a. Exon legacy numbers: 22 coding exons of the reference full-length transcript are numbered from 2 to 24 with no exon 4 defined.

^b. We have designated splicing events as in Colombo et al. (2014)[176] : Δ=skipping, ▼ = retention, and q= donor shift.

^c. Coding sequence annotation according <https://varnomen.hgvs.org/> last accessed November 2023. Ensemble reference transcript ENST00000357654 and ensemble reference protein ENSP00000350283. ^eFunctional annotation was obtained through Variant Effect Predictor web tool from Ensemble (<https://www.ensembl.org/Tools/VEP>). PTC-NMD: splicing event that introduces premature termination codons that are predicted to induce the nonsense-mediated mRNA decay, no FS: splicing events not altering the reading frame.

Supplementary Table 13. Complete semi-quantitative mRNA splicing profiling of variant carriers and controls by capillary electrophoresis of fluorescent amplicons, resulting from RT-PCR using primers in exon 16 and exon 22 of *BRCA1*.

Due to the large amount of data that this table contains, the information is available here:
https://drive.google.com/drive/folders/1T9aCggWsrMPzhU_31aseVRcnn9vEHkT?usp=drive_link

Supplementary Table 14. Δ SpliceAI(4999bp) performance in *BRCA1* exon 18 variants analysed in patients (A) and minigenes (B).

	Patients	Minigenes
Sensitivity	76.47%	87.80%
Specificity	80.00%	94.19%
Positive predictive value (PPV)	86.67%	87.80%
Negative predictive value (NPV)	66.67%	94.19%
Accuracy	77.78%	92.13%

A)Patients	Number of unique variants*
True positives (Δ SpliceAI max ≥ 0.2 and $\Delta 18 > 3.84\%$)	13
False positives (Δ SpliceAI max ≥ 0.2 and $\Delta 18 \leq 3.84\%$)	2
True negatives (Δ SpliceAI max ≤ 0.1 and $\Delta 18 \leq 3.84\%$)	8
False negatives (Δ SpliceAI max ≤ 0.1 and $\Delta 18 > 3.84\%$)	4

*For a variant present in more than one carrier we considered the mean of $\Delta 18\%$ splicing fraction in all carriers

Max: maximum raw SpliceAI Δ score, defined as the maximum probability of altered splicing across 4 output probabilities (designated by the tool as loss of acceptor or donor sites, gain of acceptor or donor sites) and the maximum distance of 10,000 nucleotides ($\pm 4,999$ nucleotides from the variant of interest).

B)Minigene	Number of unique variants
True positives (Δ SpliceAI max ≥ 0.2 and FL $< 99\%$)	36
False positives (Δ SpliceAI max ≥ 0.2 and FL $\geq 99\%$)	5
True negatives (Δ SpliceAI max ≤ 0.1 and $\Delta 18$ FL $\geq 99\%$)	81
False negatives (Δ SpliceAI max ≤ 0.1 and FL $< 99\%$)	5

Max: maximum raw SpliceAI Δ score, defined as the maximum probability of altered splicing across 4 output probabilities (designated by the tool as loss of acceptor or donor sites, gain of acceptor or donor sites) and the maximum distance of 10,000 nucleotides ($\pm 4,999$ nucleotides from the variant of interest).

Supplementary Table 15. Semi-quantitative mRNA splicing profiling of minigene constructs.

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
WT	-	100.0	0.1	0.0	0.0	0.0	0.1	12
c.5075-107 A>G	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	99.7	0.1	0.0	0.0	0.3	0.1	2
c.5075-53 C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-49del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-37del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-35A>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-33C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	3
c.5075-32T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-30C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-27A>C	-	98.1	0.6	1.9	0.6	0.0	0.0	2
c.5075-24C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-23T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-18G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	4
c.5153-13T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-11T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-9A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-9A>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-8T>G	Confirmed by sequencing: Creation of a new 3'ss (at c.5075-7) leading to retention of the last 7 nt of intron 17, PTC-NMD	0.0	0.0	0.0	0.0	100.0	0.0	2
c.5075-7T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-6C>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-6C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-5T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
c.5075-4G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5075-3C>G	-	1.5	0.4	98.5	0.4	0.0	0.0	2
c.5075-3C>T	-	97.8	0.2	2.2	0.2	0.0	0.0	2
c.5075-2A>C	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5075-2A>G	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5075-2A>T	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5075-2del	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5075-1G>A	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5075-1G>C	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5075-1G>T	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5075A>C	-	96.8	0.6	3.2	0.6	0.0	0.0	2
c.5075A>T	-	96.0	0.7	4.0	0.7	0.0	0.0	2
c.5075_5078del	-	4.8	0.6	95.2	0.6	0.0	0.0	2
c.5076del	-	78.8	0.7	21.2	0.7	0.0	0.0	2
c.5077_5079delinsTTCAATTCTGCA	-	3.6	0.0	96.4	0.0	0.0	0.0	2
c.5077_5080delinsTTCAATTCTGC	-	1.0	0.2	99.0	0.2	0.0	0.0	2
c.5077_5080delinsTTGATTCTGC	-	3.4	0.2	96.6	0.2	0.0	0.0	2
c.5078_5080del	-	86.3	0.9	13.7	0.9	0.0	0.0	4
c.5080G>A	-	78.9	1.9	21.1	1.9	0.0	0.0	4
c.5080G>T	-	29.5	2.0	70.5	2.0	0.0	0.0	4
c.5084_5085del	Inferred by size: Slight activation of an exonic cryptic 3' ss , Δ18p (r-5075_5116del)	99.0	1.4	0.5	0.7	0.5	0.7	2
c.5085T>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5085dup	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5086G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
c.5086G>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5086G>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5087T>A	-	100.0	0.0	0.0	0.0	0.0	0.0	4
c.5088G>A	-	99.7	0.4	0.3	0.4	0.0	0.0	2
c.5089T>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5089_5092del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5090G>A	-	99.9	0.2	0.1	0.2	0.0	0.0	4
c.5091_5092del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5092_c.5093delinsAG	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5094A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5095C>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5095C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5096G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5096G>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5096G>T	Inferred by size: Slight activation of an exonic cryptic 3' ss, Δ18p (r.5075_5116del)	96.2	0.4	0.3	0.4	3.6	0.0	2
c.5097G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5098A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5098del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5099C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5100A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5101C>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5103G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5102_5103del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5104A>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
c.5106del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5106A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5107T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5107T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5108A>C	-	100.0	0.0	0.0	0.0	0.0	0.0	4
c.5108A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5109T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5109T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5112del	Inferred by size: Slight activation of an exonic cryptic 3' ss : Δ18p (r.5075_5116del)	97.7	3.2	0.8	1.1	1.5	2.1	2
c.5113C>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5113C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5113_5114del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5114T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5114T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5115A>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5116G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5116G>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5117G>C	-	100.0	0.0	0.0	0.0	0.0	0.0	4
c.5117 G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5120T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5120T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	4
c.5114_5121del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5122G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5122G>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
c.5123C>A	-	99.9	0.2	0.1	0.2	0.0	0.0	6
c.5123C>G	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	72.4	5.5	27.3	5.6	0.3	0.3	7
c.5123C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	3
c.5123del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5124G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5125G>A	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	97.2	3.2	2.6	3.0	0.2	0.2	4
c.5126del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5126G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5128G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5128G>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5129G>A	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	99.7	0.5	0.0	0.0	0.3	0.5	2
c.5129del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5131A>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5131A>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5133dup	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5133del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5132A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5135G>A	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	85.5	0.2	0.0	0.0	14.5	0.2	2
c.5136G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5137del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5137G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5138T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5138T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
c.5139_5141del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5141T>G	Confirmed by sequencing: Strong activation/creation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)-PTC-NMD	0.0	0.0	0.0	0.0	100.0	0.0	2
c.5142T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5143A>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5143A>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5144G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5144G>T	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	96.3	0.3	0.0	0.0	3.7	0.3	2
c.5145C>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5145C>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5145del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5147A>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5147A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5148T>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5148T>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5150del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5150T>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+1G>A	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.0	0.0	99.3	1.0	0.7	1.0	2
c.5152+1G>C	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.0	0.0	98.5	0.1	1.5	0.1	2
c.5152+1G>T	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.0	0.0	96.9	0.2	3.1	0.2	2
c.5152+2T>A	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5152+2T>G	-	0.0	0.0	100.0	0.0	0.0	0.0	2

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
c.5152+2dup	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.4	0.5	97.4	0.8	2.2	0.3	2
c.5152+3A>C	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.0	0.0	99.7	0.5	0.4	0.5	2
c.5152+4A>G	-	0.5	0.7	99.5	0.7	0.0	0.0	2
c.5152+4A>T	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	2.4	0.0	91.9	0.2	5.7	0.2	2
c.5152+3_5152+4insG	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5152+4_5152+5insT	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.0	0.0	98.6	0.1	1.4	0.1	2
c.5152+5G>A	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.0	0.0	99.5	0.7	0.5	0.7	2
c.5152+5G>C	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	0.0	0.0	97.1	0.2	2.9	0.2	2
c.5152+6T>A	-	0.0	0.0	100.0	0.0	0.0	0.0	2
c.5152+6T>C	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	1.7	0.7	97.4	1.0	0.9	0.4	2
c.5152+6T>G	-	0.0	0.0	100.0	0.0	0.0	0.0	3
c.5152+7A>G	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	61.2	13.2	38.4	13.5	0.4	0.4	5
c.5152+9A>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+10A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+11T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+12A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+13C>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+15A>G	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+19C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+20T>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+23C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2

Status of pSPL3mBK2-BRCA1. ex18-19 minigene (WT or variants)	Splice site shift (ss shift) description	Full-length %		Δ18%		% ss shift		n
		Mean	SD	Mean	SD	Mean	SD	
c.5152+28T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+31_5152+34del	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+39C>T	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+41T>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+44G>A	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+45A>C	-	100.0	0.0	0.0	0.0	0.0	0.0	2
c.5152+66G>A	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	99.7	0.0	0.0	0.0	0.3	0.0	2
c.5152+85del	Inferred by size: Slight activation of an exonic cryptic 5' ss, Δ18q (r.5137_5152del)- PTC-NMD	99.6	0.1	0.0	0.0	0.4	0.1	2

SS: splice site, PTC-NMD: splicing event that introduces premature termination codons that are predicted to induce the nonsense-mediated mRNA decay, SD: standard deviation, n: number of replicates.

Supplementary Table 16. Complete semi-quantitative mRNA splicing profiling of mESC by capillary electrophoresis of fluorescent amplicons, resulting from RT-PCR using primers in exon 16 and exon 22 of BRCA1.

Due to the large amount of data that this table contains, the information is available here:
https://drive.google.com/drive/folders/1T9aCggWsrMPzhU_31aseVRcn9vEHkT?usp=drive_link

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PV51+PM5_PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classification: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN															
c.5075-8T>G (p.?)	0.79	0.74	PP3_splicing	-	-	-	NA no reference variant already classified by ENGMA at acceptor motif/ same position	-	-	PP3	VUS	-	PVS1_S(RNA)	PVS1_S (RNA)	Only minigene data showing complete aberrant splicing	PVS1_S (RNA)	VUS
c.5075-7T>C (p.?)	0.06	0	BP4_splicing	-	-	-	-	-	-	BP4	VUS	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5075-6C>A (p.?)	0.24	0.01	PP3_splicing	-	-	-	NA no reference variant already classified by ENGMA at acceptor motif/ same position	-	-	PP3	VUS	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (Patient and minigene)	Contradictory: PP3 (1)+ BP7_S(RNA)(-4)=-3	LB
c.5075-4C>T (p.?)	0	0.1	BP4_splicing	-	-	-	-	-	-	BP4	VUS	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5075-5T>C (p.?)	0	0.07	BP4_splicing	-	-	-	-	-	-	BP4	VUS	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5075-4G>A (p.?)	0.01	0.01	BP4_splicing	-	-	-	-	-	-	BP4	VUS	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5075-3C>G (p.?)	0.83	0	PP3_splicing	-	-	-	NA no reference variant already classified by ENGMA at acceptor motif/ same position	-	-	PP3	VUS	-	PVS1_S(RNA)	PVS1_S (RNA)	Only minigene data showing complete aberrant splicing	PVS1_S (RNA)	VUS
c.5075-3C>T (p.?)	0.25	0.01	PP3_splicing	-	-	-	NA no reference variant already classified by ENGMA at acceptor motif/ same position	-	-	PP3	VUS	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	Contradictory: PP3 (1)+ BP7_S(RNA)(-4)=-3	LB
c.5075-2A>C (p.?)	0.83	0.02	-	-	-	-	NA because any reference variant already classified by ENGMA at acceptor motif	-	-	PVS1	LP	-	PVS1(RNA)	PVS1 (RNA)	Canonical splicing site	PVS1(RNA)	LP
c.5075-2A>G (p.?)	0.83	0.02	-	-	-	-	NA because any reference variant already classified by ENGMA at	-	-	PVS1	LP	PVS1(RNA)	PVS1(RNA)	PVS1(RN A)	Variant at canonical splicing site leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1(RNA)	LP

Variant	SpliceAI	Splicing in silico evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	In silico missense e/in-frame evidence	In silico final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PV51+PM5_PTC coding codes	Combined evidence: in silico + coding effect + position	Classification: in silico + coding effect + variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: in silico + coding effect + position + splicing observed impact	Classification considering in silico + coding effect + variant position + splicing observed impact
acceptor motif																	
c.5075-2A>T (p.?)	0.83	0.01	-	-	-	-	NA because any reference variant already classified by ENIGMA at acceptor motif	-	-	PVS1	LP	-	PVS1(RNA)	PVS1(RNA)	Canonical splicing site	PVS1(RNA)	LP
NA because any reference variant already classified by ENIGMA at acceptor motif																	
c.5075-2del (p.?)	0.83	0	-	-	-	-	NA because any reference variant already classified by ENIGMA at acceptor motif	-	-	PVS1	LP	-	PVS1(RNA)	PVS1(RNA)	Canonical splicing site	PVS1(RNA)	LP
NA because any reference variant already classified by ENIGMA at acceptor motif																	
c.5075-1G>A (p.?)	0.83	0.01	-	-	-	-	NA because any reference variant already classified by ENIGMA at acceptor motif	-	-	PVS1	LP	-	PVS1(RNA)	PVS1(RNA)	Canonical splicing site	PVS1(RNA)	LP
NA because any reference variant already classified by ENIGMA at acceptor motif																	
c.5075-1G>C (p.?)	0.83	0.01	-	-	-	-	NA because any reference variant already classified by ENIGMA at acceptor motif	-	-	PVS1	LP	-	PVS1(RNA)	PVS1(RNA)	Canonical splicing site	PVS1(RNA)	LP
NA because any reference variant already classified by ENIGMA at acceptor motif																	
c.5075-1G>T (p.?)	0.83	0.01	-	-	-	-	NA because any reference variant already classified by ENIGMA at acceptor motif	-	-	PVS1	LP	-	PVS1(RNA)	PVS1(RNA)	Canonical splicing site	PVS1(RNA)	LP
NA because any reference variant already classified by ENIGMA at acceptor motif																	
c.5075A>C (p.Asp1692Ala)	0.2	0	PP3_splicing	-	0.51	PP3_prot ein	NA no reference variant already classified by ENIGMA at acceptor motif/ same position	-	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
NA no reference variant already classified by ENIGMA at acceptor motif/ same position																	

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICING	PS1_MISSENSE	Premature stop codon: PVS1+PMS_PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classification: <i>in silico</i> + coding effect + variant position	Splicing observed Impact- Patient	Splicing observed Impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
c.5075A>T (p.(Asp1692Val))	0.32	0.01	PP3_splicing	-	0.54	PP3_prot ein	PP3 both	NA no reference variant already classified by ENIGMA at acceptor motif/ same position	-	PP3	VUS	NA	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5075_5078del (p.(Asp1692Valfs*9))	0.83	0.11	-	-	-	-	-	-	PVS1+PMS_PTC_S	PVS1+PMS_PTC_C_S	P	-	PVS1_S(RNA)	PVS1_S (RNA)/PVS1 frameshift +PMS_PTC_S	Only minigene data showing complete aberrant splicing. We considered in a conservative manner, that PVS1 (frameshift) should be retained for variant interpretation because splicing defect have not yet been confirmed in patients, and it might be possible that alleles carrying this variant express transcripts with a premature stop codon, thus meeting PVS1 and PMS_PTC codes.	PVS1+PMS_PTC_S+PVS1_S(RNA)	P
c.5076del (p.(Asp1692Glufs*10))	0.34	0.01	-	-	-	-	-	-	PVS1+PMS_PTC_S	PVS1+PMS_PTC_C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PMS_PTC_S	P
c.5077_5078delinsTTTCATTGCA (p.(Ala1693Phefs*12))	0.06	0.02	-	-	-	-	-	-	PVS1+PMS_PTC_S	PVS1+PMS_PTC_C_S	P	-	PVS1_S(RNA)	PVS1_S (RNA)/PVS1 frameshift +PMS_PTC_S	Only minigene data showing complete aberrant splicing. We considered in a conservative manner, that PVS1 (frameshift) should be retained for variant interpretation because splicing defect have not yet been confirmed in patients, and it might be possible that alleles carrying this variant express transcripts with a premature stop codon, thus meeting PVS1 and PMS_PTC codes.	PVS1+PMS_PTC_S+PVS1_S(RNA)	P
c.5077_5080delinsTTTCATTCTGC (p.(Ala1693_Glu1694delinsPheLeuGln))	0	0.03	BP4_splicing	-	-0.15	BP4_prot ein	BP4	-	-	BP4	VUS	PVS1(RNA)	PVS1_S(RNA)	PVS1(RNA)	Variant leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1(RNA)	LP

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PV51+PM5_PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classification: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
c.5077_5080delinsTTGATT CTGC (p.(Ala1693_Glu1694delinsLeuLeuGln))	0	0.03	BP4_splicing	-	BP4_prot ein	BP4	-	-	-	BP4	VUS	-	PV51_S(RNA)	PV51_S (RNA)	Only minigene data showing complete aberrant splicing	PV51_S (RNA)	VUS
c.5078_5080del (p.(Ala1693del))	0.18	0	NA	-	BP4_prot ein	NA	-	-	-	-	VUS	Unclear	Unclear	NA	PVS1 could not be applied since proportion exon 18 skipping was non-concordant between minigenes and also because patients showed heterogeneous results regarding PVS1 criteria application.	-	VUS
c.5080G>A (p.(Glu1694Lys))	0.21	0	PP3_splicing	-	BP4_prot ein	PP3_splicing	NA, pathogenic variant in the same nucleotide is a nonsense variant previously classified by its coding effect and Variant under assessment presents lower spliceAI prediction (0.21 vs 0.41)	-	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5080G>T (p.(Glu1694*))	0.41	0	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PTC_S	P	PVS1 (RNA)	PVS1 for coding effect	PVS1 (RNA)	Variant leading to complete exon skipping in carriers (quantitative AE result in a sample from LCL treated with puromycin), consequently only PVS1 (RNA) can be used for variant classification despite no PTC transcripts would be generated (PM5_PTC not met)—only patient data considered because minigene did not show a concordant result.	PVS1 (RNA)	LP
c.5084_5085del (p.(Phe1695Cysfs*3))	0.03	0.01	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PTC_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5085T>A (p.(Phe1695Leu))	0.15	0	NA	-	BP4_prot ein	NA	-	-	-	-	VUS	-	-	NA	Misense variant with FL% >30%	-	VUS

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PVS1+PM5 _PTC_S	Combined evidence: <i>in silico</i> + coding effect + position	Classificat ion: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN															
c.5085dup (p.(Val1696Cysfs*3))	0.05	0.02	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5086G>A (p.(Val1696Met))	0.04	0	BP4_splicing	-	0.14	BP4_prot ein	BP4	-	-	BP4	VUS	-	NA	NA	Misense variant with FL% >30%	BP4	VUS
c.5086G>C (p.(Val1696Leu))	0	0.04	BP4_splicing	-	0.15	NA	NA	-	-	-	VUS	-	NA	NA	Misense variant with FL% >30%	-	VUS
c.5086G>T (p.(Val1696Leu))	0.05	0	BP4_splicing	-	0.15	NA	NA	-	-	-	VUS	-	NA	NA	Misense variant with FL% >30%	-	VUS
c.5087T>A (p.(Val1696Glu))	0	0.03	BP4_splicing	-	0.42	PP3_prot ein	PP3_prot in	-	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5088G>A (p.(Val1696=))	0.04	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA A)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA A)	LB
c.5089T>A (p.(Cys1697Ser))	0.01	0.02	BP4_splicing	-	0.40	PP3_prot ein	PP3_prot in	-	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5089_5092del (p.(Cys1697Asnfs*4))	0.07	0.02	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	-	NA	PVS1	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5090G>A (p.(Cys1697Trp))	0	0.03	BP4_splicing	-	0.39	PP3_prot ein	PP3_prot in	-	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5091_5092del (p.(Cys1697=))	0.04	0	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	NA	NA	-	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5092_5093delinsAG (p.(Glu1698Arg))	0.14	0	NA	-	0.60	PP3_prot ein	PP3_prot in	-	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5094A>G (p.(Glu1698=))	0.02	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA A)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA A)	LB
c.5095C>A (p.(Arg1699=))	0.13	0	NA	NA	-	-	-	-	-	-	VUS	-	BP7_S(RNA A)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP7_S(RNA A)	LB
c.5095C>T (p.(Arg1699Trp))	0.14	0.01	NA	-	0.43	PP3_prot ein	PP3_prot in	-	-	PP3	VUS	NA	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5095G>A (p.(Arg1699Gln))	0.15	0	NA	-	0.42	PP3_prot ein	PP3_prot in	-	-	PP3	VUS	NA	NA	NA	Misense variant with FL% >30%	PP3	VUS

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PVS1+PM5_PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classificat ion: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN															
c.5096G>C (p.(Arg1699Pro))	0.02	0.01	BP4_splicing	-	0.43	PP3_prot ein	PP3_prot ein	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5096G>T (p.(Arg1699Leu))	0.19	0.12	NA	-	0.42	PP3_prot ein	PP3_prot ein	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5097G>A (p.(Arg1699=))	0.07	0.01	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5098A>G (p.(Thr1700Ala))	0.14	0	NA	-	0.41	PP3_prot ein	PP3_prot ein	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5098del (p.(Thr1700Hisfs*2))	0.13	0.01	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PTC_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5099C>T (p.(Thr1700Ile))	0.03	0.01	BP4_splicing	-	0.40	PP3_prot ein	PP3_prot ein	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5100A>G (p.(Thr1700=))	0	0.09	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RN A)	Synonymous variant with no impact (Patient and minigene)	BP4 + BP7_S(RNA)	LB
c.5101C>A (p.(Leu1701Met))	0.01	0	BP4_splicing	-	0.07	BP4_prot ein	BP4	-	-	BP4	VUS	-	-	NA	Misense variant with FL% >30%	BP4	VUS
c.5103G>A (p.(Leu1701=))	0.13	0	NA	NA	-	-	-	-	-	-	VUS	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP7_S(RNA)	LB
c.5102_5103del (p.(Leu1701Glnfs*14))	0.12	0.01	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PTC_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5104A>T (p.(Lys1702*))	0.18	0.07	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PTC_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5106del (p.(Lys1702Asnfs*4))	0	0.04	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PTC_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5106A>G (p.(Lys1702=))	0.01	0.05	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB

Variant	SpliceAI	Splicing/ <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense/ <i>e/in-frame</i> evidence	<i>In silico</i> final evidence for missense/ <i>/in-frame</i> variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PV51+PM5_PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classificat ion: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN															
c.5107T>C (p.(Tyr1703His))	0.01	0.05	BP4_splicing	0.37	PP3_prot ein	PP3_prot ein	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5107T>G (p.(Tyr1703Asp))	0.01	0.07	BP4_splicing	0.43	PP3_prot ein	PP3_prot ein	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5108A>C (p.(Tyr1703Ser))	0	0.06	BP4_splicing	0.42	PP3_prot ein	PP3_prot ein	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5108A>G (p.(Tyr1703Cys))	0	0.03	BP4_splicing	0.42	PP3_prot ein	PP3_prot ein	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5109T>C (p.(Tyr1703=))	0.01	0.06	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5109T>G (p.(Tyr1703*))	0.13	0	-	-	-	-	-	-	PV51+PM5_PTC_S	PV51+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PV51+PM5_PTC_S	P
c.5112del (p.(Leu1705*))	0.01	0.02	-	-	-	-	-	-	PV51+PM5_PTC_S	PV51+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PV51+PM5_PTC_S	P
c.5113C>G (p.(Leu1705Val))	0.05	0	BP4_splicing	0.07	BP4_prot ein	BP4	-	-	-	BP4	VUS	-	-	NA	Missense variant with FL% >30%	BP4	VUS
c.5113C>T (p.(Leu1705=))	0.12	0	NA	-	-	-	-	-	-	-	VUS	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP7_S(RNA)	LB
c.5113_5114del (p.(Leu1705Arg*10))	0.18	0	-	-	-	-	-	-	PV51+PM5_PTC_S	PV51+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PV51+PM5_PTC_S	P
c.5114T>C (p.(Leu1705Pro))	0	0.15	NA	0.34	PP3_prot ein	PP3_prot ein	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5114T>G (p.(Leu1705Arg))	0.01	0.16	NA	0.31	PP3_prot ein	PP3_prot ein	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5115A>C (p.(Leu1705=))	0.01	0.15	NA	-	-	-	-	-	-	-	VUS	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP7_S(RNA)	LB

Variant	SpliceAI	Splicing/ in silico evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	In silico missense e/in- frame evidence	In silico final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PV51+PM5 _PTC coding codes	Combined evidence: In silico + coding effect + position	Classificat ion: In silico+ coding effect+ variant position	Splicing observed impact- Patent	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: In silico+ coding effect+ position+ splicing observed impact	Classification considering In silico+ coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN															
c.5116G>A (p.(Gly1706Arg))	0.01	0.06	BP4_splicing	-	0.39	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5116G>C (p.(Gly1706Arg))	0.01	0.16	NA	-	0.39	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5117G>C (p.(Gly1706Ala))	0.01	0.04	BP4_splicing	-	0.23	NA	NA	-	-	-	VUS	NA	NA	NA	Misense variant with FL% >30%	-	VUS
c.5117G>A (p.(Gly1706Glu))	0.01	0.03	BP4_splicing	-	0.39	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	NA	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5120T>C (p.(Ile1707Thr))	0	0.1	BP4_splicing	-	0.40	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5120T>G (p.(Ile1707Ser))	0.01	0.03	BP4_splicing	-	0.40	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5114_5121del (p.(Leu1705Argfs*9))	0.01	0.14	-	-	-	-	-	-	PV51+PM5 _PTC_S	PV51+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PV51+PM5_PTC_S	P
c.5122G>A (p.(Ala1708Thr))	0	0.01	BP4_splicing	-	0.33	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5122G>C (p.(Ala1708Pro))	0	0.08	BP4_splicing	-	0.42	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5123C>A (p.(Ala1708Glu))	0.22	0	PP3_splicing	-	0.42	PP3_prot ein	PP3 both	NA, cannot be used as reference variant for PS1_splicing due to leaky effect.	-	PP3	VUS	NA	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5123C>G (p.(Ala1708Gly))	0.4	0	PP3_splicing	-	0.30	PP3_prot ein	PP3 both	NA, outside splice site no spliceogenic reference variant at same position	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS
c.5123C>T (p.(Ala1708Val))	0.13	0	NA	-	0.43	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	-	NA	NA	Misense variant with FL% >30%	PP3	VUS

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PVS1+PM5 _PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classification: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN															
c.5123del (p.(Ala1708Glyfs*6))	0.4	0	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5124G>A (p.(Ala1708-I))	0	0.1	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RNA)	Synonymous variant with no impact (Patient and minigene)	BP4 + BP7_S(RNA)	LB
c.5125G>A (p.(Gly1709Arg))	0	0.09	BP4_splicing	-	PP3_prot ein	PP3_prot in	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5126del (p.(Gly1709Glyfs*5))	0	0.14	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5126G>A (p.(Gly1709Glu))	0	0.13	NA	-	PP3_prot ein	PP3_prot in	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5128G>A (p.(Gly1710Arg))	0	0.11	NA	-	PP3_prot ein	PP3_prot in	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5128G>T (p.(Gly1710*))	0	0.01	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5129G>A (p.(Gly1710Glu))	0	0.07	BP4_splicing	-	PP3_prot ein	PP3_prot in	-	NA, no reference variant	-	PP3	VUS	-	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5128del (p.(Gly1710Glyfs*4))	0	0.08	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5131A>C (p.(Lys1711Gln))	0	0.06	BP4_splicing	-	BP4_prot ein	BP4	-	-	-	BP4	VUS	-	NA	NA	Missense variant with FL% >30%	BP4	VUS
c.5131A>T (p.(Lys1711*))	0.12	0	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	-	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P

Variant	SpliceAI	Splicing/ <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PVS1+PM5 _PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classificat ion: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN														
c.5133dup (p.(Trp1712Metfs*4))	0.03	0.02	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5133del (p.(Lys1711Asnfs*3))	0.07	0	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5132A>G (p.(Lys1711Arg))	0.01	0.05	BP4_splicing	-	-0.03	BP4_prot ein	BP4	-	-	BP4	VUS	NA	NA	Missense variant with FL% >30%	BP4	VUS
c.5135G>A (p.(Trp1712*))	0.16	0.1	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5136G>A (p.(Trp1712*))	0.02	0.03	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5137del (p.(Val1713*))	0	0.04	-	-	-	-	-	-	PVS1+PM5 _PTC_S	PVS1+PM5_PT C_S	P	NA	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5137G>A (p.(Val1713Ile))	0.02	0.15	NA	-	-0.13	BP4_prot ein	NA	-	-	-	VUS	NA	NA	Missense variant with FL% >30%	-	VUS
c.5138T>C (p.(Val1713Ala))	0.02	0.14	NA	-	0.40	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5138T>G (p.(Val1713Gly))	0.02	0.13	NA	-	0.40	PP3_prot ein	PP3_prot in	NA, no reference variant	-	PP3	VUS	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5139 5'41del (p.(Val1714del))	0.01	0.14	NA	-	-0.44	BP4_prot ein	NA	-	-	-	VUS	NA	NA	In-frame deletion with no splicing impact	-	VUS
c.5141T>G (p.(Val1714Gly))	0.16	0.95	PP3_splicing	-	0.49	PP3_prot ein	PP3 both	NA, outside splice site motifs and no spliceogenic reference	-	PP3	VUS	PVS1_S(RNA)	PVS1_S (RNA)	Only minigene data showing complete aberrant splicing	PVS1_S (RNA)	VUS

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PVS1+PM5_PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classification: <i>in silico</i> + coding effect+ variant position	Splicing observed impact- Minigene	Splicing observed impact- Patient	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
variant at same position																	
c.5142T>G (p.(Val1714=))	0.01	0.13	NA	-	-	-	-	-	-	-	VUS	BP7_S(RNA)	-	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP7_S(RNA)	LB
c.5143A>C (p.(Ser1715Arg))	0.01	0.15	NA	0.3415 52	PP3_prot ein	PP3_prot ein	-	Reference variant: Well established pathogenic variant that acts via protein change used for PS1 criteria	-	Reference variant: Well established pathogenic variant that acts via protein change used for PS1 criteria	Reference variant: Well established Pvariant that acts via protein change used for PS1 criteria	NA	-	NA	Missense variant with FL% >30%	Reference missense variant	Reference missense variant
c.5143A>T (p.(Ser1715Oys))	0	0.11	NA	0.35	PP3_prot ein	PP3_prot in	-	NA, no reference variant	-	PP3	VUS	NA	-	NA	Missense variant with FL% >30%	PP3	VUS
c.5144G>A (p.(Ser1715Asn))	0	0.11	NA	-0.05	BP4_prot ein	NA	-	-	-	-	VUS	NA	NA	NA	Missense variant with FL% >30%	-	VUS
c.5144G>T (p.(Ser1715Ile))	0	0.11	NA	0.36	PP3_prot ein	PP3_prot in	-	NA, no reference variant	-	PP3	VUS	NA	-	NA	Missense variant with FL% >30%	PP3	VUS
c.5145C>A (p.(Ser1715Arg))	0.04	0	BP4_splicing	0.33	PP3_prot ein	PP3_prot in	-	PS1	-	PP3+PS1	VUS	NA	-	NA	Missense variant with FL% >30%	PP3+PS1	VUS
c.5145C>G (p.(Ser1715Arg))	0.23	0.02	PP3_splicing	0.33	PP3_prot ein	PP3 both	NA, outside splice site motifs and no spliceogenic reference variant at same position	-	-	PP3	VUS	NA	NA	NA	Missense variant with FL% >30%	PP3	VUS
c.5145del (p.(Tyr1716Ilefs*4))	0.01	0.06	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PT C_S	P	NA	-	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P
c.5147A>C (p.(Tyr1716Ser))	0	0.03	BP4_splicing	0.21	NA	NA	-	-	-	-	VUS	-	-	NA	Missense variant with FL% >30%	-	VUS
c.5147A>G (p.(Tyr1716Oys))	0	0.06	BP4_splicing	0.32	PP3_prot ein	PP3_prot in	-	NA, no reference variant	-	PP3	VUS	-	-	NA	Missense variant with FL% >30%	PP3	VUS
c.5148T>A (p.(Tyr1716=))	0.05	0	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PT C_S	P	NA	-	NA	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	PVS1+PM5_PTC_S	P

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense e/in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICING	PS1_MISSENSE	Premature stop codon: PVS1+PM5_PTC_S coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classification: <i>in silico</i> + coding effect+ variant position	Comment	Splicing observed impact- Minigene	Splicing observed impact- Patient	Final splicing observed impact code	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
	MAX LOSS	MAX GAIN															
c.5148T>G (p.(Tyr171Gln))	0.2	0	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PT_C_S	P	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	NA	-	NA	PVS1+PM5_PTC_S	P
c.5150del (p.(Phe171Serfs*3))	0.01	0.05	-	-	-	-	-	-	PVS1+PM5_PTC_S	PVS1+PM5_PT_C_S	P	Nonsense or frameshift variant that generates a PTC predicted to undergo NMD degradation. Variant is present in a biologically-relevant transcript	NA	-	NA	PVS1+PM5_PTC_S	P
c.5150T>A (p.(Phe171Tyr))	0	0.04	BP4_splicing	-	-0.15	BP4_prot ein	BP4	-	-	BP4	VUS	Misense variant with FL% >30%	NA	-	NA	BP4	VUS
c.5152+1G>A (p.?)	0.9	0.01	-	-	-	-	PS1_P	-	-	PVS1+PS1_P	LP	Canonical splicing site	PVS1(RNA)	-	PVS1(RNA)	PVS1(RNA)	LP
c.5152+1G>C (p.?)	0.9	0.01	-	-	-	-	PS1_P	-	-	PVS1+PS1_P	LP	Variant at canonical splicing site leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1(RNA)	-	PVS1(RNA)	PVS1(RNA)	LP
c.5152+1G>T (p.?)	0.9	0.03	-	-	-	-	Reference variant: well-established pathogenic variant used for PS1(splicing)	-	-	Reference variant: well-established pathogenic variant	Reference variant: well-established P variant	Variant at canonical splicing site leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1(RNA)	-	PVS1(RNA)	Reference variant: well-established pathogenic variant	Reference variant: well-established P variant
c.5152+2T>A (p.?)	0.9	0	-	-	-	-	PS1_P	-	-	PVS1+PS1_P	LP	Canonical splicing site	PVS1(RNA)	-	PVS1(RNA)	PVS1(RNA)	LP
c.5152+2T>G (p.?)	0.9	0	-	-	-	-	PS1_P	-	-	PVS1+PS1_P	LP	Canonical splicing site	PVS1(RNA)	-	PVS1(RNA)	PVS1(RNA)	LP
c.5152+2dup (p.?)	0.88	0	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	Variant leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1(RNA)	-	PVS1(RNA)	PVS1(RNA)	LP
c.5152+3A>C (p.?)	0.88	0	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	Variant leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1(RNA)	-	PVS1(RNA)	PVS1(RNA)	LP
c.5152+4A>G (p.?)	0.8	0.01	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	Only minigene data showing complete aberrant splicing	PVS1_S (RNA)	-	PVS1_S (RNA)	PVS1_S (RNA)	VUS
c.5152+4A>T (p.?)	0.75	0.02	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	Variant leading to complete exon skipping in carriers and minigenes	PVS1_S (RNA)	-	PVS1_S (RNA)	PVS1_S (RNA)	LP

Variant	SpliceAI	Splicing <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense/In-frame evidence	<i>In silico</i> final evidence for missense/In-frame variants	PS1_SPLICI NG	PS1_MISSEN SE	Premature stop codon: PV51+PM5_PTC coding codes	Combined evidence: <i>In silico</i> + coding effect + position	Classification: <i>In silico</i> + coding effect+ variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>In silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>In silico</i> + coding effect+ variant position+ splicing observed impact	
																		MAX LOSS
(quantitative concordant result)																		
c.5152+3_5152+4insG (p.?)	0.9	0	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	-	PVS1_S(RNA)	PVS1_S (RNA)	PVS1_S (RNA)	Only minigene data showing complete aberrant splicing	VUS	
c.5152+4_5152+5insT (p.?)	0.9	0.01	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	-	PVS1_S(RNA)	PVS1_S (RNA)	PVS1_S (RNA)	Only minigene data showing complete aberrant splicing	VUS	
c.5152+5G>A (p.?)	0.89	0.02	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	PVS1 (RNA)	PVS1_S(RNA)	PVS1 (RNA)	Variant leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1 (RNA)	LP	
c.5152+5G>C (p.?)	0.9	0.03	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	PVS1 (RNA)	PVS1_S(RNA)	PVS1 (RNA)	Variant leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1 (RNA)	LP	
c.5152+6T>A (p.?)	0.81	0	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	-	PVS1_S(RNA)	PVS1_S (RNA)	PVS1_S (RNA)	Only minigene data showing complete aberrant splicing	VUS	
c.5152+6T>C (p.?)	0.77	0	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	PVS1 (RNA)	PVS1_S(RNA)	PVS1 (RNA)	Variant leading to complete exon skipping in carriers and minigenes (quantitative concordant result)	PVS1 (RNA)	LP	
c.5152+6T>G (p.?)	0.84	0	PP3_splicing	-	-	-	PS1_M	-	-	PP3+PS1_M	VUS	-	PVS1_S(RNA)	PVS1_S (RNA)	PVS1_S (RNA)	Only minigene data showing complete aberrant splicing	VUS	
c.5152+7A>G (p.?)	0.51	0	PP3_splicing	NA	-	-	NA, outside splice site motifs and no spliceogenic reference variant at same position	-	-	PP3	VUS	PVS1_S (RNA)	BP7_S(RNA)	PVS1_S (RNA)* (Based on one carrier RT-PCR result)	PVS1 unclear. Discordant result between minigene and carriers. This variants lead to near complete skipping in one carrier analysed by RT-PCR, and minigene result was not considered for classification. Strong is the maximum strenght applicable for assays using patient's mRNA without allele-specific quantification. However, results for additional two carriers analysed by RT-PCRseq and RNAseq showed lower levels of exon skipping	PVS1_S (RNA)?	VUS	
c.5152+9A>T (p.?)	0.02	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+10A>G (p.?)	0.16	0.01	NA	NA	-	-	-	-	-	-	VUS	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RN A)	Intronic variant with no impact (Patient and minigene)	BP7_S(RNA)	LB	

Variant	SpliceAI	Splicing/ <i>in silico</i> evidence	BP7 for silent and intronic variants outside splicing motifs (at or beyond +7/-21) if BP4 met	Bayes Del (NoAF)	<i>In silico</i> missense /in-frame evidence	<i>In silico</i> final evidence for missense /in-frame variants	PS1_SPLICING	PS1_MISSENSE	Premature stop codon: PV51+PM5_PTC coding codes	Combined evidence: <i>in silico</i> + coding effect + position	Classification: <i>in silico</i> + coding effect + variant position	Splicing observed impact- Patient	Splicing observed impact- Minigene	Final splicing observed impact code	Comment	Combined evidences: <i>in silico</i> + coding effect+ position+ splicing observed impact	Classification considering <i>in silico</i> + coding effect+ variant position+ splicing observed impact
c.5152+11T>C (p.?)	0.01	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+12A>G (p.?)	0.02	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+13C>A (p.?)	0.03	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+15A>G (p.?)	0.02	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+19C>T (p.?)	0	0.01	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+20T>A (p.?)	0.03	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (Patient and minigene)	BP4 + BP7_S(RNA)	LB
c.5152+23C>T (p.?)	0	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+28T>C (p.?)	0	0.01	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+31_5152+34del (p.?)	0.05	0.01	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+39C>T (p.?)	0	0.01	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+41T>C (p.?)	0.01	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+44G>A (p.?)	0	0.02	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+45A>C (p.?)	0	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	-	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (minigene)	BP4 + BP7_S(RNA)	LB
c.5152+46G>A (p.?)	0	0	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (Patient and minigene)	BP4 + BP7_S(RNA)	LB
c.5152+85del (p.?)	0.02	0.03	BP4_splicing	BP7	-	-	-	-	-	BP4+BP7	LB	BP7_S(RNA)	BP7_S(RNA)	BP7_S(RNA A)	Intronic variant with no impact (Patient and minigene)	BP4 + BP7_S(RNA)	LB

Supplementary Table 18. Functional data for *BRCA1* exon 18 variants' classification.

Due to the large amount of data that this table contains, the information is available here:
https://drive.google.com/drive/folders/1T9aCgg/WsrMPzhU_31aseVRcn9vEHkT?usp=drive_link

Supplementary Table 19. Variant allele frequency data for *BRCA1* exon 18 variants' classification.

Due to the large amount of data that this table contains, the information is available here:
https://drive.google.com/drive/folders/1T9aCgg/WsrMPzhU_31aseVRcn9vEHkT?usp=drive_link

Supplementary Table 20. Case-control data from BRIDGES study [37] for BRCA1 exon 18 variants' classification.

cNomen (BRCA1, NM_007294.3)	pNomen	Cases	Controls	Total cases	Total controls	Dataset	Odds Ratio (OR) OR≥4	C.I.95%	p-value	z-statistic	Evidence code NA: non-applicable
c.5075-1G>T	p.?	1	0	60466	53461	truncating_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5075-1G>T	p.?	1	0	48826	50703	truncating_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5095C>T	p.(Arg1699Trp)	2	0	60466	53461	missense_all_samples	4.42	0.2122 to 92.0886	0.34	0.96	NA
c.5095C>T	p.(Arg1699Trp)	1	0	48826	50703	missense_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5096G>A	p.(Arg1699Gln)	15	8	60466	53461	missense_all_samples	1.66	0.7028 to 3.9104	0.25	1.15	NA
c.5096G>A	p.(Arg1699Gln)	14	8	48826	50703	missense_population	1.82	0.7623 to 4.3323	0.18	1.35	NA
c.5099C>T	p.(Thr1700Ile)	1	0	60466	53461	missense_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5108A>G	p.(Tyr1703Cys)	1	0	60466	53461	missense_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5108A>G	p.(Tyr1703Cys)	1	0	48826	50703	missense_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5117G>C	p.(Gly1706Ala)	2	4	60466	53461	missense_all_samples	0.44	0.0810 to 2.4137	0.35	0.94	NA
c.5117G>C	p.(Gly1706Ala)	1	4	48826	50703	missense_population	0.26	0.0290 to 2.3229	0.23	1.21	NA
c.5123C>A	p.(Ala1708Glu)	7	1	60466	53461	missense_all_samples	6.19	0.7614 to 50.3069	0.09	1.71	NA
c.5123C>A	p.(Ala1708Glu)	7	0	48826	50703	missense_population	15.58	0.8896 to 272.7518	0.06	1.88	NA
c.5125G>A	p.(Gly1709Arg)	1	0	60466	53461	missense_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5125G>A	p.(Gly1709Arg)	1	0	48826	50703	missense_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5145C>G	p.(Ser1715Arg)	2	0	48826	50703	missense_population	5.19	0.2493 to 108.1592	0.29	1.06	NA
c.5080G>T	p.(Glu1694*)	8	1	60466	53461	truncating_all_samples	7.07	0.8846 to 56.5564	0.07	1.84	NA
c.5080G>T	p.(Glu1694*)	3	0	48826	50703	truncating_population	7.27	0.3754 to 140.7379	0.19	1.31	NA
c.5109T>G	p.(Tyr1703*)	1	0	60466	53461	truncating_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5109T>G	p.(Tyr1703*)	1	0	48826	50703	truncating_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5136G>A	p.(Trp1712*)	1	0	60466	53461	truncating_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5136G>A	p.(Trp1712*)	1	0	48826	50703	truncating_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5137del	p.(Val1713*)	1	0	60466	53461	truncating_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5137del	p.(Val1713*)	1	0	48826	50703	truncating_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5152+1G>C	p.?	2	0	60466	53461	truncating_all_samples	4.42	0.2122 to 92.0886	0.34	0.96	NA
c.5152+1G>C	p.?	1	0	48826	50703	truncating_population	3.12	0.1269 to 76.4797	0.49	0.70	NA
c.5152+1G>T	p.?	1	0	60466	53461	truncating_all_samples	2.65	0.1080 to 65.1161	0.55	0.60	NA
c.5152+1G>T	p.?	1	0	48826	50703	truncating_population	3.12	0.1269 to 76.4797	0.49	0.70	NA

Data from: <https://bcac.ccge.medschl.cam.ac.uk/bcacdata/bridges-summary-results-breast-cancer-risk-genes-2021/> (accessed in February 2024)

Supplementary Table 21. Evidence code assignment based on multifactorial likelihood clinical data.

BRCA1 exon 18 variant	Source(s)	Co-Segregation LR	Tumour Pathology LR	Co-occurrence LR	Family History LR	Combined Clinical LR for PP4 or BP5	Code/s to apply
c.5075-9A>T	Easton et al 2007, PMID:17924331	1	-	0.00	1.23	0.0046	BP5_S
c.5075A>T	Li et al 2020, PMID: 31853058	-	-	-	-	15.20	PP4_M
c.5078_5080del	Parsons et al 2019, PMID:31131967	3.64	9.99	1.03	0.27	10.12	PP4_M
	Li et al 2020, PMID: 31853058	-	-	-	-	1.13	
	Combined LR	-	-	-	-	11.39	
c.5084_5085del	Li et al 2020, PMID: 31853058	-	-	-	-	0.47	BP5
c.5085T>A	Parsons et al 2019, PMID:31131967	-	-	1.03	0.78	0.81	PP4/BP5 not met
	Li et al 2020, PMID: 31853058	-	-	-	-	0.89	
	Combined LR	-	-	-	-	0.72	
	Li et al 2020, PMID: 31853058	-	-	-	-	1.25	PP4/BP5 not met
c.5089T>A	Li et al 2020, PMID: 31853058	-	-	-	-	1.00	PP4/BP5 not met
c.5090G>A	Li et al 2020, PMID: 31853058	-	-	-	-	0.48	BP5
c.5095C>T	Easton et al 2007, PMID:17924331	1	-	1.82	21877.62	39810.72	PP4_VS
	Li et al 2020, PMID: 31853058	-	-	-	-	1047.21	
	Combined LR	-	-	-	-	41690323.67	
	Goldgar et al 2004, PMID:15290653	2.84	-	1.4	-	3.98	PP4_S
c.5096G>A	Li et al 2020, PMID: 31853058	-	-	-	-	25.49	
	Combined LR	-	-	-	-	101.37	
	Parsons et al 2019, PMID:31131967	-	-	1.03	0.73	0.76	PP4/BP5 not met
c.5099C>T	Parsons et al 2019, PMID:31131967	1.32	0.49	-	-	0.65	PP4/BP5 not met
c.5106del	Li et al 2020, PMID: 31853058	-	-	-	-	2.21	PP4
c.5108A>G	Li et al 2020, PMID: 31853058	-	-	-	-	1.64	PP4/BP5 not met

BRCA1 exon 18 variant	Source(s)	Co-Segregation LR	Tumour Pathology LR	Co-occurrence LR	Family History LR	Combined Clinical LR for PP4 or BP5	Code/s to apply
c.5109T>G	Li et al 2020, PMID: 31853058	-	-	-	-	9.74	PP4_M
c.5112del	Li et al 2020, PMID: 31853058	-	-	-	-	2.93	PP4_P
c.5113C>G	Li et al 2020, PMID: 31853058	-	-	-	-	0.02	BP5_S
c.5114T>C	Parsons et al 2019, PMID: 31131967	3.82	4.41	-	-	16.83	PP4_M
c.5116G>A	Caputo et al 2021, PMID: 34597585	61.93	11.79	-	3.17	2312.62	PP4_VS
c.5117G>A	Easton et al 2007, PMID: 17924331	1.95	-	1.35	223.87	588.84	PP4_VS
	Li et al 2020, PMID: 31853058	-	-	-	-	0.96	
	Caputo et al 2021, PMID: 34597585	7.37	3.12	-	2.61	59.97	
	Combined LR	-	-	-	-	33858.72	
c.5117G>C	Caputo et al 2021, PMID: 34597585	0.00000147	0.0000119	-	0.35	6.09E-12	BP5_VS
c.5120T>C	Li et al 2020, PMID: 31853058	-	-	-	-	0.55	PP4/BP5 not met
c.5122G>A	Li et al 2020, PMID: 31853058	-	-	-	-	0.50	PP4/BP5 not met
c.5123C>A	Easton et al 2007, PMID: 17924331	1	-	12.59	1.65959E+11	2.0893E+12	PP4_VS
	Li et al 2020, PMID: 31853058	-	-	-	-	498698.23	
	Combined LR	-	-	-	-	1.04E+18	
c.5123C>T	Li et al 2020, PMID: 31853058	-	-	-	-	0.003	BP5_M
	ENIGMA member unpublished	0.19	819.17	-	0.29	43.40	
	Combined LR	-	-	-	-	0.13	
c.5125G>A	Li et al 2020, PMID: 31853058	-	-	-	-	0.79	PP4/BP5 not met
c.5129G>A	Parsons et al 2019, PMID: 31131967	-	0.64	-	-	0.64	PP4/BP5 not met
c.5135G>A	Li et al 2020, PMID: 31853058	-	-	-	-	6.22	PP4_M
c.5136G>A	Li et al 2020, PMID: 31853058	-	-	-	-	263.46	PP4_S

<i>BRCA1</i> exon 18 variant	Source(s)	Co-Segregation LR	Tumour Pathology LR	Co-occurrence LR	Family History LR	Combined Clinical LR for PP4 or BP5	Code/s to apply
c.5137delG	Li et al 2020, PMID: 31853058	-	-	-	-	0.64	PP4/BP5 not met
c.5143A>C	Easton et al 2007, PMID:17924331	1	-	1.23	36.31	44.67	PP4_S
c.5143A>T	Parsons et al 2019, PMID:31131967	-	-	1.03	2.96	3.06	PP4/BP5 not met
	Li et al 2020, PMID: 31853058	-	-	-	-	0.62	
	Combined LR	-	-	-	-	1.91	
c.5144G>A	Parsons et al 2019, PMID:31131967	57.86	3.73	1.30	3.63	1016.90	PP4_VS
	Caputo et al 2021, PMID: 34597585	167.97	-	-	6.42	1078.99	
	Combined LR	-	-	-	-	1097228.21	
c.5145C>G	Caputo et al 2021, PMID: 34597585	28.8	13.91	-	6.78	2715.44	PP4_VS
c.5147A>C	Li et al 2020, PMID: 31853058	-	-	-	-	0.45	BP5_P
c.5147A>G	Li et al 2020, PMID: 31853058	-	-	-	-	0.26	BP5_P
c.5152+1G>A	Li et al 2020, PMID: 31853058	-	-	-	-	3.46	PP4_P
c.5152+2dup	Parsons et al 2019, PMID:31131967	-	-	1.03	1.30	1.35	PP4/BP5 not met
c.5152+7A>G	Parsons et al 2019, PMID:31131967	-	-	1.03	0.89	0.92	PP4/BP5 not met

Supplementary Table 22. Variant classification based on evidence available for the 166 variants included in the Part 2.

Due to the large amount of data that this table contains, the information is available here:

https://drive.google.com/drive/folders/1T9aCggWsrMPzHU_31aseVRcnn9vEHkT?usp=drive_link

Supplementary Table 23. Clinical, tumoral, familial and multifactorial data obtained from carriers of *BRCA1* c.5152+7A>G.

Source(s)	Family number	Carrier ID (F: female, M: male)	Primary Tumour Pathology (age onset) ^a	Family history	Co-occurrence with other <i>BRCA1</i> variants	Co-occurrence LR	Co-Segregation LR (COOL software)	Family History (CanRisk) ^b	Tumour Pathology LR (PMID: 25857409)	Combined LR
Hospital Universitari Sant Joan de Reus (Spain)	1	Carrier 1 (F)	Breast cancer, grade 3, ER-negative (<50)	Paternal grandfather lung cancer (56 years old, variant status unknown)	No	-	-	0.027	3.160	0.0059
Amby Genetics	2	Carrier 2(F)	Breast cancer, triple negative (≥ 50)	Maternal aunt and maternal cousin breast cancer (unknown age of onset and unknown variant status)	No	-	-	0.011	4.410	
Amby Genetics	3	Carrier 3(F)	Breast cancer (50), and second breast cancer (70)	Mother breast cancer (90 years old, variant status unknown), brother prostate cancer (70 years old, variant status unknown) and maternal aunt breast cancer (50 years old, variant status unknown)	No	-	0.661	0.010	-	
Amby Genetics	3	Carrier 4(M)	Unaffected (≥ 50)	Son of carrier 3	No	-	-		-	
Amby Genetics	4	Carrier 5(F)	Unaffected (≥ 50)	Brother esophageal cancer (unknown age and unknown variant status), paternal aunt breast cancer (30 years old, variant status unknown), paternal aunt 2 breast cancer (60 years old, unknown variant status), paternal aunt 3 Adenocarcinoma not otherwise specified (unknown age of onset and unknown variants status), 2 maternal cousins with breast cancer (both at 50 years old, variant status unknown).	No, but also carries a likely pathogenic variant in BRIP1	-	-	Not considered because the presence of a likely pathogenic variant in BRIP1	-	
Amby Genetics	5	Carrier 6(F)	Breast cancer, grade 3, ER-positive (≥ 50)	Father prostate cancer (50 years old, variant status unknown), mother-melanoma (80 years old, variant status unknown), maternal aunt breast cancer (70 years old, variant status unknown), maternal uncle prostate cancer (60 years old, variant status unknown), maternal-maternal great aunt breast cancer (unknown age and unknown variant status)	No	-	-	0.003	0.900	

Source(s)	Family number	Carrier ID (F: female, M: male)	Primary Tumour Pathology (age onset) ^a	Family history	Co-occurrence with other <i>BRCA1</i> variants	Co-occurrence LR	Co-Segregation LR (COOL software)	Family History (CanRisk) ^b	Tumour Pathology LR (PMID: 25857409)	Combined LR
Amby Genetics	6	Carrier 7 (M)	Prostate cancer (>60) and oncocytoma (50)	Mother ovarian cancer (Early 70s, variant status unknown), maternal aunt ovarian cancer (early 50s, variant status unknown), maternal grandmother ovarian cancer (Early 70s, variant status unknown), father prostate cancer (early 70s, variant status unknown)	No	-	-	0.064	-	
Amby Genetics	7	Carrier 8 (F)	Unaffected (<50)	Daughter of carrier 9	No	-	0.938	0.002	-	
Amby Genetics	7	Carrier 9 (F)	NA	NA	No	-			-	
Amby Genetics	8	Carrier 10 (F)	Unaffected (≥ 50)	Brother prostate cancer (>60 years old, variant status unknown), mother ovarian cancer (<70 years old, variant status unknown), maternal aunt ovarian cancer (40 years old) and breast cancer (50 years old, variant status unknown), maternal grandmother ovarian cancer (70 years old, variant status unknown)	No	-	-	0.308	-	
Amby Genetics	9	Carrier 11 (F)	Colorectal cancer (50)	Colorectal cancer in father and paternal grandparents, maternal grandmother ovarian cancer (50 years old, variant status unknown)	No	-	-	0.002	-	
Myriad Genetics	10	Carrier 12 (F)	Breast cancer (40)	Paternal grandmother breast cancer (40 years old, variant status unknown), mother lung cancer (60 years old, variant status unknown)	c.5266dupC (p.Gln1756Profs*74) in <i>trans</i> , without Anemia de Fanconi phenotype	0.001	-	0.008	-	
Total	10	12				0.001	0.619	0.570 (LR)	12.542	0.0059

^a. ER: oestrogen receptor

^b. BOADICEA risk-prediction scores indicated for each family. Total likelihood ratio (LR) was obtained from the mean of all family scores divided by *BRCA1/2* negative families score (0.085) from Spurdte et al. (2012)[277].

Supplementary Table 24. Published splicing results from variants included in Part 2.

Variant	Source of RNA ^a	Splicing result	Reference	Minigene (Δ18%)	Minigene (SS shifts%)	Carrier 1 Δ18% (SF) ^b	Sample Carrier 1 ^a	Carrier 2 Δ18% (SF) ^b	Sample Carrier 2 ^a	Carrier 3 Δ18% (SF) ^b	Sample Carrier 3 ^a	Clinvar classification	Review status	Final classification
c.5075-107A>G	Short-term cultured lymphocytes treated with puromycin	no aberration	PMID: 217696 58	0.0	0.3	0.00	PBMCs	-	-	-	-	Likely Benign	no assertion criteria provided	Benign
c.5075-49del	Short-term cultured lymphocytes treated with puromycin	no aberration	PMID: 267805 56	0.0	0.0	0.00	PBMCs	-	-	-	-	-	-	Likely Benign
c.5075-9A>T	Patient plus allele-specific (PAXGENE)	no aberration	PMID: 225050 45	0.0	0.0	0.0	-	-	-	-	-	Benign & Likely Benign	criteria provided, multiple submitters, no conflicts	Benign
c.5075-8T>G	Patient (LCL with puromycin)	Use of a de novo site 7 nt upstream from 3' ss	PMID: 225050 45	0.0	100.0	-	-	-	-	-	-	Pathogenic & Likely Pathogenic	criteria provided, multiple submitters, no conflicts	Likely pathogenic
c.5075-6C>A	Patient (PAXGENE)	no aberration	PMID: 297502 58, PMID: 32123317	0.0	0.0	0.00	PAXgene	-	-	-	-	Conflicting interpretations of Pathogenicity	criteria provided, conflicting interpretations	Benign
c.5075-2A>G	Patient not allele-specific (PBMCs and LCL without Puromycin)	exon 18 deletion	PMID: 232399 86	100.0	0.0	55.21	PAXgene	-	-	-	-	Pathogenic	criteria provided, multiple submitters, no conflicts	Pathogenic
c.5075-1G>C	Minigene	exon 18 deletion	PMID: 246677 79	100.0	0.0	-	-	-	-	-	-	Pathogenic	criteria provided, multiple submitters, no conflicts	Pathogenic
c.5077-5080delin sTTGATTCTGC	LCL without puromycin	exon 18 deletion	PMID: 225050 45 *	99.0	0.0	63.05	Whole blood-Magna Pure	-	-	-	-	not provided	no assertion provided	Likely Pathogenic
c.5077-5080delin sTTGATTCTGC	LCL without puromycin	exon 18 deletion	PMID: 225050 45 *	96.6	0.0	-	-	-	-	-	-	Pathogenic	criteria provided, single submitter	Uncertain

Variant	Source of RNA ^a	Splicing result	Reference	Minigene (Δ18%)	Minigene (SS shifts%)	Carrier 1 Δ18% (SF) ^b	Sample Carrier 1 ^a	Carrier 2 Δ18% (SF) ^b	Sample Carrier 2 ^a	Carrier 3 Δ18% (SF) ^b	Sample Carrier 3 ^a	Clinvar classification	Review status	Final classification
c.5078_5080del	Patients (PBMCs), PAXgene, LCL with puromycin	exon 18 deletion	PMID:12955719, PMID:22505045	13.7	0.0	56.18	PBMCs	32.40	LCLs treated with puromycin	-	-	Pathogenic & Likely Pathogenic	criteria provided, multiple submitters, no conflicts	Likely pathogenic
c.5080G>A	LCL with puromycin	Slight exon 18 skipping	PMID:22505045	21.1	0.0	-	-	-	-	-	-	Uncertain significance	criteria provided, single submitter	Likely Pathogenic
c.5080G>T	Patient with allele-specific result / Minigene. No NMD inhibition indicated in any of the studies.	exon 18 deletion	PMID:9497265 / PMID:18391021	70.5	0.0	57.06	LCLs treated with puromycin	-	-	-	-	Pathogenic	reviewed by expert panel	Pathogenic
c.5096G>A	Patient (PAXGENE)	no aberration	PMID:32123317, PMID:22505045	0.0	0.0	6.05	Whole blood- Trizol	0.00	PAXgene	-	-	Pathogenic	reviewed by expert panel	Likely Pathogenic
c.5100A>G	Short-term cultured lymphocytes treated with puromycin	no aberration	PMID:26780556	0.0	0.0	12.61	PBMCs	0.00	Whole blood- Magna Pure	-	-	Likely Benign	reviewed by expert panel	Benign
c.5116G>A	stimulated T lymphocytes with puromycin	no aberration	PMID:22505045	0.0	0.0	-	-	-	-	-	-	Pathogenic & Likely Pathogenic	criteria provided, multiple submitters, no conflicts	Pathogenic
c.5117G>C	Patients (PBMCs)	no aberration	PMID:12955719	0.0	0.0	0.00	PBMCs	11.75	Whole blood- Trizol	-	-	Benign	reviewed by expert panel	Benign
c.5117G>A	Short-term cultured lymphocytes treated with puromycin and whole blood (PAXGENE)	no aberration	PMID:26780556, PMID:32123317	0.0	0.0	0.00	PBMCs	-	-	-	-	Pathogenic	reviewed by expert panel	Pathogenic

Variant	Source of RNA ^a	Splicing result	Reference	Minigene (Δ18%)	Minigene (SS shifts%)	Carrier 1 Δ18% (SF) ^b	Sample Carrier 1 ^a	Carrier 2 Δ18% (SF) ^b	Sample Carrier 2 ^a	Carrier 3 Δ18% (SF) ^b	Sample Carrier 3 ^a	Clinvar classification	Review status	Final classification
c.5123C>A	Short-term cultured lymphocytes treated with puromycin/ Peripheal blood lymphocytes /peripheral-blood leukocytes and minigenes / LCLs with cycloheximide	exon 18 deletion/partial exon 18 skipping / partial exon 18 skipping/ no aberration	PMID: 26780556/ PMID: 20215541/PMID: 19404736/ PMID: 15923272	0.1	0.0	2.37	LCLs treated with puromycin	19.26	PBMCs	16.68	Whole blood- Trizol	Pathogenic	reviewed by expert panel	Pathogenic
c.5124G>A	Minigene	no aberration	PMID: 21673748	0.0	0.0	0.00	PBMCs	-	-	-	-	Likely Benign	reviewed by expert panel	Benign
c.5144G>A	Patient not allele-specific (PBMCs) and minigene	no aberration	PMID: 21673748 and PMID: 12955719	0.0	0.0	0.00	PBMCs	-	-	-	-	Pathogenic	reviewed by expert panel	Pathogenic
c.5152+1G>C	Patient not allele-specific; aberrant transcripts consistent with loss of function	exon 18 deletion	PMID: 23239986	98.5	1.5	66.27	Whole blood- Magna Pure	-	-	-	-	Pathogenic	criteria provided, multiple submitters, no conflicts	Pathogenic
c.5152+1G>T	LCLs (treated with puromycin or cycloheximide)	exon 18 deletion	PMID: 30101128	96.9	3.1	60.73	PAXgene	58.71	PAXgene	-	-	Pathogenic	reviewed by expert panel	Pathogenic
c.5152+2dup	phytohemagglutinin stimulated short-term lymphocyte cultures with puromycin	exon 18 deletion	PMID: 29280214	97.4	2.2	43.43	Short term lymphocyte cultures treated with puromycin	53.66	Whole blood- Magna Pure	54.75	Whole blood- Magna Pure	Pathogenic	criteria provided, single submitter	Likely Pathogenic
c.5152+5G>A	Patient (PBMCs)	exon 18 deletion	PMID: 12955719	99.5	0.5	66.77	PBMCs	-	-	-	-	Pathogenic & Likely Pathogenic	criteria provided, multiple submitters, no conflicts	Pathogenic
c.5152+5G>C	Patient (PAXGENE)	exon 18 deletion	PMID: 32123317	97.1	2.9	60.85	PAXgene	-	-	-	-	Pathogenic	criteria provided, single submitter	Pathogenic

Variant	Source of RNA ^a	Splicing result	Reference	Minigene (Δ18%)	Minigene (SS shifts%)	Carrier 1 Δ18% (SF) ^b	Sample Carrier 1 ^a	Carrier 2 Δ18% (SF) ^b	Sample Carrier 2 ^a	Carrier 3 Δ18% (SF) ^b	Sample Carrier 3 ^a	Clinvar classification	Review status	Final classification
c.5152+6T>C	Patient (PAXGENE)	exon 18 deletion	PMID: 32123317	97.4	0.9	57.95	PAXgene	-	-	-	-	Conflicting interpretations of Pathogenicity	criteria provided, conflicting interpretations	Pathogenic
c.5152+10A>G	Cyclohexamide-treated and untreated LCLs	no aberration	PMID: 21394826	0.0	0.0	0.00	PAXgene	-	-	-	-	Benign & Likely Benign	criteria provided, multiple submitters, no conflicts	Benign
c.5152+85del	Patients (PAXGENE) and minigene	no aberration	PMID: 22505045	0.0	0.4	0.00	PAXgene	-	-	-	-	Benign	reviewed by expert panel	Benign

^a. PBMCs: peripheral blood mononuclear cell, LCLs: Lymphoblastoid Cell lines

^b. SF: splicing fraction levels (%)

* In PMID: 22505045 variant nomenclature is "c.5077_5080del4ins10", hence it is unclear which exact indel variant was analysed.

Supplementary Table 25. Clinical and molecular characteristics of patients and tumours evaluated in Part 3.

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germine variant	Exon	Protein clinically relevant ^d	Coding effect	Cohort	Sample ID	Sample type ^e	Tumour ^f	Tumour type ^g	Breast tumour molecular subtype ^h	Age at diagnosis	Grade	Y-H2AX%	RAD51%	BRCA1%
HUVH	1	F	BRCA1	c.1121_1123delinsT	11	-	Frameshift	Pathogenic	T1	Bx	BC	IDC	TN	35	G3	90.0	1.1	42.0
HUVH	2	F	BRCA1	c.1912del	11	-	Frameshift	Pathogenic	T2	Bx	BC	IDC	TN	52	G3	90.9	7.0	1.6
HUVH	3	F	BRCA1	c.211A>G	5	RING	Missense (splicing)	Pathogenic	T3	Bx	BC	IDC	TN	47	G3	99.0	8.5	4.8
HUVH	4	F	BRCA1	c.211A>G	5	RING	Missense (splicing)	Pathogenic	T4	Bx	BC	IDC	TN	32	G3	68.0	2.7	2.0
HUVH	5	F	BRCA1	c.211A>G	5	RING	Missense (splicing)	Pathogenic	T5	Bx	BC	IDC	ER+	40	G2	62.0	8.9	50.0
HUVH	6	F	BRCA1	c.211A>G	5	RING	Missense (splicing)	Pathogenic	T6	Bx	BC	IDC	TN	38	G3	98.0	0.0	0.0
HUVH	7	F	BRCA1	c.2722G>T	11	-	Nonsense	Pathogenic	T7	Bx	BC	IDC	TN	48	G3	96.0	6.9	70.0
HUVH	8	F	BRCA1	c.2921T>A	11	-	Nonsense	Pathogenic	T8	Bx	BC	IDC	ER+	33	G2	70.0	3.5	19.0
HUVH	9	F	BRCA1	c.3331_3334del	11	-	Frameshift	Pathogenic	T9	Bx	BC	IDC	TN	48	G3	90.0	21.0	85.0
HUVH	10	F	BRCA1	c.3331_3334del	11	-	Frameshift	Pathogenic	T10	Sx	BC	IDC	TN	46	NA	58.3	0.0	0.0
HUVH	11	F	BRCA1	c.3481_3491del	11	-	Frameshift	Pathogenic	T11.1	Bx	BC	IDC	ER+	60	G2	88.0	1.9	34.0
HUVH	11	F	BRCA1	c.3481_3491del	11	-	Frameshift	Pathogenic	T11.2	Bx	BC	IDC	ER+	60	G2	57.0	1.5	43.0
HUVH	12	F	BRCA1	c.3583del	11	-	Frameshift	Pathogenic	T12	Bx	BC	IDC	TN	42	G2	87.0	6.9	40.0
HUVH	13	F	BRCA1	c.3756_3759del	11	-	Frameshift	Pathogenic	T13	Bx	BC	IDC	TN	34	G3	82.0	2.2	25.0
HUVH	14	F	BRCA1	c.3839_3843delinsAGGC	11	-	Nonsense	Pathogenic	T14	Bx	BC	IDC	TN	34	G3	90.0	1.5	20.0
HUVH	15	F	BRCA1	c.4843dup	16	-	Frameshift	Pathogenic	T15.1 (BX)	Bx	BC	IDC	TN	53	G3	72.0	0.0	0.0
HUVH	15	F	BRCA1	c.4843dup	16	-	Frameshift	Pathogenic	T15.2 (SX)	Sx	BC	IDC	TN	53	G3	36.0	0.0	2.0
HUVH	16	F	BRCA1	c.494dupT	8	-	Frameshift	Pathogenic	T16	Bx	BC	IDC	TN	52	G3	96.0	0.0	3.0
HUVH	17	F	BRCA1	c.5091_5092del	18	BRTC	Nonsense	Pathogenic	T17	Bx	BC	IDC	TN	53	NA	50.0	1.4	0.0
HUVH	18	F	BRCA1	c.5095C>T	18	BRTC	Missense	Pathogenic	T18	Bx	BC	IDC	TN	60	G3	92.0	2.8	40.0

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germine variant	Exon	Protein clinically relevant ^d	Coding effect	Cohort	Sample ID	Sample type ^e	Tumour type ^f	Breast tumour molecular subtype ^g	Age at diagnosis	Grade	Y-H2AX%	RAD51%	BRCA1%	
HUVH	19	F	BRCA1	c.5123C>A	18	BRTC	Missense	Pathogenic	T19	Bx	BC	IDC	TN	28	G2	72.0	0.0	0.0
HUVH	20	F	BRCA1	c.5123C>A	18	BRTC	Missense	Pathogenic	T20	Bx	BC	IDC	TN	47	G3	96.0	3.7	1.5
HUVH	21	F	BRCA1	c.5154G>A	19	BRTC	Nonsense	Pathogenic	T21	Bx	BC	IDC	ER+	38	G3	50.0	1.9	7.0
HUVH	22	F	BRCA1	c.5324T>G	21	BRTC	Missense	Pathogenic	T22	Bx	BC	IDC	TN	49	G3	96.0	3.3	1.0
HUVH	23	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T23	Bx	BC	IDC	TN	43	G3	70.0	1.6	7.0
HUVH	24	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T24	Bx	BC	IDC	TN	29	G3	68.0	2.5	0.0
HUVH	25	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T25L	Bx	BC	IDC	TN	54	G2	64.0	0.0	3.0
HUVH	25	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T26R	Bx	BC	IDC	ER+	54	G2	75.0	2.3	60.0
HUVH	26	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T27	Bx	BC	IDC	ER+	42	G3	63.0	1.0	1.0
HUVH	27	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T28	Bx	BC	IDC	TN	73	G3	72.0	0.0	0.0
HUVH	28	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T29	Bx	BC	IDC	ER+	45	G3	92.0	3.6	4.0
HUVH	29	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T30	Bx	BC	IDC	ER+	29	G3	51.0	5.6	3.0
HUVH	30	F	BRCA1	c.68_69del	2	RING	Frameshift	Pathogenic	T31	Bx	BC	IDC	ER+	36	G2	50.0	1.2	4.0
HUVH	31	F	BRCA1	c.815_824dup	11	-	Frameshift	Pathogenic	T32	Bx	BC	IDC	TN	67	G3	84.0	0.4	24.5
HUVH	32	F	BRCA1	c.845C>A	11	-	Nonsense	Pathogenic	T33	Sx	BC	IDC	TN	60	G3	78.0	0.0	5.0
NCCS	33	F	BRCA1	c.181T>G	5	RING	Missense	Pathogenic	T34	Sx	OC	NA	-	NA	NA	64.3	8.0	1.4
HSJR	34	F	BRCA1	c.5123C>A	18	BRTC	Missense	Pathogenic	T35	Sx	OC	High-grade serous ovarian carcinoma	-	46	NA	92.5	59.3	NE
HSJR	35	U	BRCA1	c.5123C>A	18	BRTC	Missense	Pathogenic	T36	Bx	BC	IDC	TN	47	NA	81.3	1.2	NE
HUVH	36	M	BRCA2	c.1389_1390del	10	-	Frameshift	Pathogenic	T37	Bx	BC	IDC	ER+	54	G3	86.0	0.0	NA
HUVH	37	F	BRCA2	c.1597del	10	-	Frameshift	Pathogenic	T38	Bx	BC	IDC	ER+	35	G2	34.0	1.0	83.0
HUVH	38	F	BRCA2	c.1597del	10	-	Frameshift	Pathogenic	T39	Bx	BC	IDC	ER+	48	G3	76.0	1.6	90.0
HUVH	39	F	BRCA2	c.1842dupT	10	-	Nonsense	Pathogenic	T40	Bx	BC	IDC	ER+	38	G3	92.0	1.6	68.0

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germine Variant	Exon	Protein clinically relevant domain ^d	Coding effect	Cohort	Sample ID	Sample type ^e	Tumour type ^f	Breast tumour molecular subtype ^g	Age at diagnosis	Grade	Y-H2AX%	RAD51%	BRCA1%
HUVH	40	F	BRCA2	c.1842dupT	10	-	Nonsense	Pathogenic	T41	Sx	BC	IDC	TN	G3	54.0	0.9	NA
HUVH	41	F	BRCA2	c.2367del	11	-	Frameshift	Pathogenic	T42	Bx	BC	IDC	ER+	G3	42.0	1.9	NA
HUVH	42	F	BRCA2	c.2808_2811del	11	-	Frameshift	Pathogenic	T43	Bx	BC	IDC	ER+	G3	55.0	2.0	97.5
HUVH	43	F	BRCA2	c.3059_3060del	11	-	Nonsense	Pathogenic	T44	Bx	BC	IDC	ER+	G3	80.0	30.0	83.0
HUVH	44	F	BRCA2	c.3264dup	11	-	Frameshift	Pathogenic	T45	Bx	BC	IDC	ER+	G3	90.0	0.0	46.0
HUVH	45	F	BRCA2	c.3264dup	11	-	Frameshift	Pathogenic	T46	Bx	BC	IDC	TN	G3	98.0	0.7	86.7
HUVH	46	F	BRCA2	c.3264dup	11	-	Frameshift	Pathogenic	T47	Sx	BC	IDC	ER+	G3	47.5	2.0	80.0
HUVH	47	F	BRCA2	c.3922G>T	11	-	Nonsense	Pathogenic	T48	Bx	BC	IDC	ER+	NA	85.0	0.5	80.0
HUVH	48	F	BRCA2	c.541del	7	-	Frameshift	Pathogenic	T49	Bx	BC	IDC	ER+	G2	65.0	23.4	76.0
HUVH	49	F	BRCA2	c.5576_5579del	11	-	Frameshift	Pathogenic	T50	Bx	BC	ILC	ER+	G2	76.0	0.0	92.5
HUVH	50	F	BRCA2	c.5681dup	11	-	Nonsense	Pathogenic	T51	Bx	BC	ILC	ER+	G2	76.0	3.3	100.0
HUVH	51	F	BRCA2	c.5720_5723del	11	-	Nonsense	Pathogenic	T52	Bx	BC	ILC	ER+	G2	70.0	0.0	93.3
HUVH	52	F	BRCA2	c.63del	2	PALB2 INTERACTING DOMAIN	Frameshift	Pathogenic	T53	Bx	BC	IDC	ER+	G3	76.0	0.5	96.7
HUVH	53	F	BRCA2	c.7617+1G>A	i15	DBD	Splicing	Pathogenic	T54	Bx	BC	IDC	ER+	G3	76.0	4.0	78.0
HUVH	54	F	BRCA2	c.7617+1G>A	i15	DBD	Splicing	Pathogenic	T55	Bx	BC	IDC	ER+	G2	52.0	0.8	NA
HUVH	55	F	BRCA2	c.8067T>A	18	DBD	Nonsense	Pathogenic	T56	Bx	BC	IDC	ER+	G2	26.0	0.0	32.5
HUVH	56	F	BRCA2	c.8168A>G	18	DBD	Missense	Pathogenic	T57	Bx	BC	IDC	ER+	G2	70.0	1.9	37.0
HUVH	57	F	BRCA2	c.9026_9030del	23	DBD	Frameshift	Pathogenic	T58	Bx	BC	IDC	ER+	G3	92.0	1.6	80.0
HUVH	58	F	BRCA2	c.9097dup	23	DBD	Frameshift	Pathogenic	T59	Bx	BC	IDC	ER+	G2	74.0	0.7	75.0
HUVH	59	F	BRCA2	c.9100C>T	23	DBD	Nonsense	Pathogenic	T60	Bx	BC	IDC	ER+	G2	68.0	1.0	90.0
HUVH	60	F	BRCA2	c.9285C>G	25	DBD	Missense	Pathogenic	T61	Bx	BC	IDC	TN	G3	94.0	7.7	86.0
HUVH	61	F	BRCA2	c.9382C>T	25	DBD	Nonsense	Pathogenic	T62	Bx	BC	IDC	ER+	G2	97.3	9.3	82.0

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germine Variant	Exon	Protein clinically relevant domain ^d	Coding effect	Cohort	Sample ID	Sample type ^e	Tumour ^f	Tumour type ^g	Breast tumour molecular subtype ^h	Age at diagnosis	Grade	γ-H2AX%	RAD51%	BRCA1%
HUVH	62	F	BRCA2	c.9382C>T	25	DBD	Nonsense	Pathogenic	T63	Sx	BC	IDC	ER+	32	G3	50.0	11.2	53.3
HUVH	63	F	BRCA2	c.9382C>T	25	DBD	Nonsense	Pathogenic	T64	Bx	BC	IDC	ER+	46	G2	58.0	0.0	NA
UHP	64	F	BRCA2	c.7975A>G	17	DBD	Missense	Pathogenic	T65	Sx	BC	IDC	NA	48	G3	44.2	2.3	7.6
HUVH	65	F	PALB2	c.3113G>A	10	WD40	Nonsense (Splicing?)	Pathogenic	T66	Bx	BC	IDC	ER+	43	G3	60.0	0.0	67.0
HUVH	66	F	PALB2	c.3256C>T	12	WD40	Nonsense	Pathogenic	T67	Sx	BC	Adenosquamous carcinoma	TN	44	G3	42.0	2.8	22.5
HUVH	67	F	PALB2	c.1050_1053del	4	-	Frameshift	Pathogenic	T68	Bx	BC	IDC	ER+	36	G2	86.0	0.8	76.0
HUVH	68	F	PALB2	c.3256C>T	12	WD40	Nonsense	Pathogenic	T69	Bx	BC	IDC	ER+	41	G3	74.0	0.0	87.5
HUVH	69	M	PALB2	c.3201+5G>T	i10	WD40	Splicing	Pathogenic	T70	Sx	BC	IDC	ER+	54	G2	63.0	3.0	NA
HUVH	70	F	PALB2	c.1111G>T	4	-	Nonsense	Pathogenic	T71	Bx	BC	ILC	ER+	45	G2	74.0	0.0	70.0
HUVH	71	F	PALB2	c.3256C>T	12	WD40	Nonsense	Pathogenic	T72R	Bx	BC	IDC	TN	40	G2	96.0	1.3	95.0
HUVH	71	F	PALB2	c.3256C>T	12	WD40	Nonsense	Pathogenic	T73L	Bx	BC	IDC	ER+	40	G2	84.5	0.5	67.5
HUVH	72	F	PALB2	c.1240C>T	4	-	Nonsense	Pathogenic	T74	Bx	BC	IDC	TN	38	G3	76.0	1.0	68.3
HUVH	73	F	PALB2	c.3362del	13	WD40	Frameshift	Pathogenic	T75	Sx	BC	IDC	ER+	41	NA	64.0	0.5	41.0
HUVH	74	F	PALB2	c.3362del	13	WD40	Frameshift	Pathogenic	T76	Sx	BC	IDC	ER+	43	NA	41.0	2.0	NA
HUVH	75	F	PALB2	c.1675C>T	4	-	Nonsense	Pathogenic	T77	Bx	BC	IDC	ER+	51	G3	91.5	0.5	90.0
HUVH	76	F	PALB2	c.1675C>T	4	-	Nonsense	Pathogenic	T78	Bx	BC	IDC	ER+	58	G1	70.0	0.0	70.0
HUVH	77	F	PALB2	c.2257C>T	5	-	Nonsense	Pathogenic	T79	Bx	BC	IDC	ER+	46	G3	64.0	5.0	65.0
HUVH	78	F	PALB2	c.3247_3248insT	12	WD40	Frameshift	Pathogenic	T80	Bx	BC	IDC	ER+	28	G2	63.3	1.4	90.0
HUVH	79	F	PALB2	c.695del	4	-	Frameshift	Pathogenic	T81	Bx	BC	IDC	TN	56	G3	100.0	1.0	95.0
HUVH	80	F	PALB2	Del exons 4-7	4 to 7	WD40	In-frame	Pathogenic	T82	Bx	BC	IDC	ER+	37	G2	80.0	0.0	86.7
HUVH	81	F	PALB2	c.1240C>T	4	-	Nonsense	Pathogenic	T83	Sx	BC	IDC	ER+	44	G1	30.0	0.0	40.0
HUVH	82	F	PALB2	c.50T>G	2	CC	Nonsense	Pathogenic	T84R	Bx	BC	IDC	ER+	47	NA	87.5	18.0	87.5

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germine Variant	Exon	Protein clinically relevant domain ^d	Coding effect	Cohort	Sample ID	Sample type ^e	Tumour type ^f	Breast tumour molecular subtype ^h	Age at diagnosis	Grade	Y-H2AX%	RAD51%	BRCA1%	
HUVH	97	F	BRCA1/BRCA2	c.442-34C>T, c.1067A>G/c.-26G>A, c.67+82C>G, c.1114A>C, c.3396A>G, c.4563A>G, c.6513G>C, c.6841+80_6841+83del, c.6938-120T>C, c.7242A>G, c.7806-14T>C, c.8755-66T>C	17, 11/5'UTR, i2, 10, 11, 11, 11, 11, i12, 14, i16, i21	-/-, -/5'UTR, -/-, -/-, -/-, -/-, DBD, DBD'	missense, intronic/UTR, intronic, missense, synonymous, synonymous, synonymous, intronic, intronic, synonymous, intronic, intronic	Benign	T100	Bx	BC	ER+	41	G3	55.6	51.7	80.6	
UHP	98	F	BRCA2	c.3516G>A	11	-	Synonymous	Benign	T101	Sx	BC	IDC	NA	34	G3	87.5	9.0	50.0
UHP	99	F	BRCA2	c.10110G>A	10	-	Synonymous	Benign	T102	Sx	BC	IDC	NA	43	na	81.5	38.5	66.7
MyBrCa	100	U	BRCA1	c.2566T>C	11	-	Missense	Benign	T103	Sx	BC	IDC	ER+	44	G2	52.4	31.7	60.5
MyBrCa	101	U	BRCA1	c.2566T>C	11	-	Missense	Benign	T104	Sx	BC	MP	NA	44	NA	71.2	27.3	78.0
MyBrCa	102	U	BRCA2	c.10234A>G	10	-	Missense	Benign	T105	Sx	BC	IDC	ER+	43	G2	33.3	27.5	53.5
MyBrCa	103	U	BRCA2	c.10234A>G	10	-	Missense	Benign	T106	Sx	BC	IDC	ER+	69	G2	51.3	5.8	35.0
MyBrCa	104	U	BRCA2	c.5785A>G	11	-	Missense	Benign	T107	Sx	BC	IDC	ER+	41	G3	84.4	22.3	46.0
MyBrCa	105	U	BRCA2	c.5785A>G	11	-	Missense	Benign	T108	Sx	BC	IDC	ER+	38	G3	65.7	20.4	33.3
MyBrCa	106	U	BRCA2	c.5785A>G	11	-	Missense	Benign	T109	Sx	BC	IDC	ER+	58	G3	76.1	20.9	17.4
MyBrCa	107	U	BRCA2	c.8187G>T	18	DBD	Missense	Benign	T110	Sx	BC	IDC	HER2+	74	G3	73.7	7.2	6.5
MyBrCa	108	U	BRCA2	c.8187G>T	18	DBD	Missense	Benign	T111	Sx	BC	IDC	HER2+	44	G3	87.1	11.9	50.8
MyBrCa	109	U	BRCA2	c.8851G>A	22	DBD	Missense	Benign	T112	Sx	BC	IDC	ER+	44	G2	85.5	59.8	84.8
MyBrCa	110	U	BRCA1	c.427G>A	7	-	Missense	Benign	T113	Sx	BC	IDC	HER2+	57	G3	71.6	30.8	34.6
MyBrCa	111	U	BRCA2	c.8525G>A	20	DBD	Missense	Benign	T114	Sx	BC	IDC	TN	66	G3	47.6	60.6	51.2
HUVH	112	F	BRCA1	c.1655G>A	11	-	Missense	Unreviewed →Benign	T115	Bx	BC	IDC	ER+	52	G2	49.1	48.2	61.1

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germine Variant	Exon	Protein clinically relevant domain ^d	Coding effect	Cohort	Sample ID	Sample type ^e	Tumour ^f	Tumour type ^g	Breast tumour molecular subtype ^h	Age at diagnosis	Grade	Y-H2AX%	RAD51%	BRCA1%
HUVH	113	F	BRCA1	c.398G>A	7	-	Missense	Unreviewed →Benign	T116	Sx	OC	High-grade serous ovarian carcinoma	-	61	G3	66.0	6.9	0.0
HUVH	114	F	BRCA1	c.932C>G	11	-	Missense	Unreviewed →Benign	T117	Bx	BC	IDC	ER+	53	G3	77.4	22.5	71.1
HUVH	115	F	BRCA1	c.3247A>G	11	-	Missense	Unreviewed →Benign	T118	Bx	BC	IDC	ER+	38	G1	67.9	18.1	35.7
HUVH	116	F	BRCA2	c.7522G>A	15	DBD	Missense	Unreviewed →Benign	T119	Bx	BC	IDC	ER+	38	G2	58.0	19.0	23.7
HUVH	117	F	BRCA2	c.3398C>T	11	-	Missense	Unreviewed →Benign	T120	Sx	OC	High-grade serous ovarian carcinoma	-	48	G3	85.0	16.2	47.2
HUVH	118	F	BRCA2	c.1645A>G	10	-	Missense	Unreviewed →Benign	T121	Bx	BC	IDC	TN	40	G3	58.7	47.4	80.0
HUVH	119	M	BRCA2	c.3216A>T	11	-	Missense	Unreviewed →Benign	T122	Sx	BC	PAP	ER+	55	G1	75.0	6.8	60.2
HUVH	120	F	BRCA2	c.6382A>G and c.6700T>C	11	-	Missense	Unreviewed →Benign	T123	Sx	OC	Clear cell	-	59	G3	67.4	4.8	7.9
HUVH	121	F	BRCA2	c.3032C>G	11	-	Missense	Unreviewed →Benign	T124	Bx	BC	IDC	TN	40	NA	66.0	16.3	48.9
HUVH	122	F	BRCA2	c.6288C>T	11	-	Missense	Unreviewed →Benign	T125	Sx	BC	IDC	ER+	36	NA	65.3	3.0	23.1
HUVH	123	F	PALB2	c.925A>G	4	-	Missense	Unreviewed →Benign	T126	Bx	BC	IDC	TN	41	NA	82.5	28.3	88.5
HUVH	124	F	PALB2	c.925A>G	4	-	Missense	Unreviewed →Benign	T127	Bx	BC	IDC	TN	58	G3	69.2	3.8	5.6
HUVH	125	F	PALB2	c.629C>T	4	-	Missense	Unreviewed →Benign	T128	Sx	BC	DCIS	ER+	50	G2	54.8	6.7	17.6
HUVH	126	F	PALB2	c.1861C>A	5	-	Missense	Unreviewed →Benign	T129	Sx	BC	IDC	ER+	60	G2	48.6	1.9	41.3

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germine Variant	Exon	Protein clinically relevant ^d	Coding effect	Cohort	Sample ID	Sample type ^e	Tumour ^f	Tumour type ^g	Breast tumour molecular subtype ^h	Age at diagnosis	Grade	γ-H2AX%	RAD51%	BRCA1%
HUVH	127	F	BRCA1/2	c.2123C>A (BRCA1) and c.6560C>T (BRCA2)	11 and 11	-	Missense	Unreviewed →Benign	T130	Bx	BC	IDC	TN	37	NA	92.6	2.8	5.4
UHP	128	F	BRCA2	c.3083A>C	11	-	Missense	Unreviewed →Benign	T131	Bx	BC	IDC	NA	56	G1	82.6	24.1	75.6
UHP	129	F	BRCA2	c.4316C>A	11	-	Missense	Unreviewed →Benign	T132	Bx	BC	IDC	NA	34	G2	94.0	18.4	54.1
UHP	130	F	BRCA2	c.800G>A	10	-	Missense	Unreviewed →Benign	T133	Bx	BC	IDC	NA	22	G3	92.9	19.2	82.5
UHP	131	F	BRCA2	c.8117A>G	18	DBD	Missense	Unreviewed →Benign	T134	Sx	BC	IDC	NA	63	G2	30.3	59.4	NE
UHP	132	F	BRCA2	c.3310A>C and c.3503T>A	11	-	Missense	Unreviewed →Benign	T135	Bx	BC	IDC	NA	70	G1	58.8	14.3	63.0
UHP	133	F	BRCA2	c.7022G>A	14	-	Missense	Unreviewed →Benign	T136	Sx	BC	IDC	ER+	68	G2	61.5	22.4	31.4
UHP	134	F	BRCA2	c.2799A>G	11	-	Synonymous	Unreviewed →Benign	T137	Sx	BC	IDC	NA	55	G3	59.6	9.1	2.2
UHP	135	F	BRCA2	c.4928T>C	11	-	Missense	Unreviewed →Benign	T138	Bx	BC	IDC	NA	43	G3	63.4	26.1	86.0
HUVH	136	F	BRCA2	c.9410C>T	25	DBD	Missense	Unreviewed →VUS	T139	Sx	OC	High-grade serous ovarian carcinoma	-	62	G3	72.9	15.1	62.7
HUVH	137	F	PALB2	c.2093T>C	5	-	Missense	Unreviewed →VUS	T140	Bx	BC	IDC	ER+	57	G2	61.4	31.0	78.8
HUVH	138	F	PALB2	c.101G>A	2	CC	Missense	Unreviewed →VUS	T141	Bx	BC	IDC	ER+	30	NA	87.3	2.8	11.5
HUVH	139	F	PALB2	c.2612A>G	7	WD40	Missense	Unreviewed →VUS	T142	Sx	BC	IDC	ER+	26	G3	67.3	3.4	47.1
HUVH	140	F	PALB2	c.2401G>A	5	-	Missense	Unreviewed →VUS	T143	Sx	BC	IDC	ER+	41	NA	43.5	22.6	25.5
HUVH	141	F	PALB2	c.2425A>G	5	-	Missense	Unreviewed →VUS	T144	Sx	BC	DCIS	ER+	35	G3	63.0	13.7	64.7
HUVH	142	F	BRCA1	Deletion exons 10-14	10 to 14	-	In-frame	Unreviewed →VUS	T145	Bx	BC	IDC	TN	45	G2	81.0	1.0	79.0

Centre ^a	Patient ID	Sex ^b	Gene ^c	Germ-line variant	Exon	Protein clinically relevant domain ^d	Coding effect	Cohort	Sample ID	Sample type	Tumour ^e	Tumour type ^f	Breast tumour molecular subtype ^h	Age at diagnosis	Grade	Y-H2AX%	RAD51%	BRCA1%
HUVH	143	F	BRCA1	c.791_794delGTTTC	11	-	Frameshift	Unreviewed →VUS	T146	Bx	BC	IDC	TN	38	G3	95.0	58.0	89.0

^a. HUVH: Hospital Universitari Vall d'Hebron, MyBrCA: Cancer Research Malaysia, HSR: Hospital Universitari Sant Joan de Reus, UHP: University Hospital of Parma
^b. F: female, M: male, U: unknown
^c. Reference transcripts: BRCA1(NM_007294.3), BRCA2 (NM_000059.3), PALB2 (NM_024675.3)
^d. Only clinically relevant domains indicated. BRCA1: RING (from residue 2 to 101), CC (from residue 1391-1424) and BRCT (from residue 1650 to 1587). BRCA2: PALB2-interacting domain (from residues 10 to 40) and DBD (from residue 2481 to 3186). PALB2: CC (from residues 9 to 44), and WD40 (from residue 854 to 1186)
^e. Bx: biopsy, Sx: surgical specimen
^f. BC: breast cancer, OC: ovarian cancer
^g. DCIS: ductal carcinoma in situ, IDC: invasive ductal carcinoma, ILC: invasive lobular carcinoma, MUCI: Mucinous, PAP: papillary, MP: metaplastic
^h. TN: triple negative breast cancer, ER+: oestrogen receptor positive, HER2+ (ER-, PR- and HER2+).

Supplementary Table 26. Non-evaluable and excluded tumour samples.

Centre ^a	Patient ID	Cohort	Sample ID	Sample type ^b	Tumour type	Reasoning ^c
HUVH	144	Pathogenic	T148	Sx	BC	Not-available y-H2AX staining
HUVH	145	Pathogenic	T149	Bx	BC	<25% y-H2AX (low)
HUVH	32	Pathogenic	T33. sx2	Sx	BC	Insufficient (<40 GMN+)/absence tumour cells
HUVH	146	Pathogenic	T150	Sx	OC	Not-available y-H2AX staining
HUVH	147	Pathogenic	T151.1	Sx	BC	<25% y-H2AX (low)
HUVH	147	Pathogenic	T151.2	Bx	BC	<25% y-H2AX (low)
HUVH	148	Pathogenic	T152	Bx	BC	Not-available y-H2AX staining
HUVH	149	Pathogenic	T153	Bx	BC	<25% y-H2AX (low)
HUVH	150	Pathogenic	T154	Bx	BC	Not-available y-H2AX staining
HUVH	82	Pathogenic	T84	Sx	BC	Unclear staining
HUVH	82	Pathogenic	T85	Sx	BC	Unclear staining
HSJR	151	Pathogenic	T155	Bx	BC	Unclear staining (weak GMN)
MyBrCa	152	Bening	T156	Sx	BC	<25% y-H2AX (low)
MyBrCa	153	Bening	T157	Sx	BC	Unclear staining (weak GMN)
MyBrCa	154	Bening	T158	Sx	BC	Unclear staining (weak GMN)
MyBrCa	155	Bening	T159	Sx	BC	Unclear staining (weak GMN)
MyBrCa	156	Bening	T160	Sx	BC	Unclear staining (weak GMN)
MyBrCa	157	Bening	T161	Sx	BC	<25% y-H2AX (low)
MyBrCa	158	Bening	T162	Sx	BC	Unclear staining (weak GMN)
MyBrCa	159	Bening	T163	Sx	BC	Insufficient (<40 GMN+)/absence tumour cells
MyBrCa	160	Bening	T164	Sx	BC	Unclear staining (weak GMN)
MyBrCa	161	Bening	T165	Sx	BC	Unclear staining (weak GMN)
MyBrCa	162	Bening	T166	Sx	BC	Unclear staining (background)
MyBrCa	163	Bening	T167	Sx	BC	Unclear staining
MyBrCa	164	Bening	T168	Sx	BC	Unclear staining (background)
MyBrCa	165	Bening	T169	Sx	BC	Unclear staining (weak GMN)
MyBrCa	166	Bening	T170	Sx	BC	Unclear staining (weak GMN)
MyBrCa	167	Bening	T171	Sx	BC	Unclear staining (weak GMN and background)

Centre ^a	Patient ID	Cohort	Sample ID	Sample type ^b	Tumour type	Reasoning ^c
MyBrCa	168	Bening	T172	Sx	BC	Unclear staining (background)
MyBrCa	169	Bening	T173	Sx	BC	Unclear staining (weak GMN)
MyBrCa	170	Bening	T174	Sx	BC	Unclear staining (weak GMN)
MyBrCa	171	Bening	T175	Sx	BC	<25% y-H2AX (low)
MyBrCa	172	Bening	T176	Sx	BC	Unclear staining (background)
MyBrCa	173	Bening	T177	Sx	BC	Unclear staining (weak GMN)
MyBrCa	174	Bening	T178	Sx	BC	Unclear staining (background)
MyBrCa	175	Bening	T179	Sx	BC	Unclear staining (weak GMN and DAPI)
MyBrCa	176	Bening	T180	Sx	BC	Unclear staining (weak GMN)
MyBrCa	177	Bening	T181	Sx	BC	Unclear staining (weak GMN and background)
MyBrCa	178	Bening	T182	Sx	BC	<25% y-H2AX (low)
MyBrCa	179	Bening	T183	Sx	BC	Unclear staining (weak GMN)
MyBrCa	180	Bening	T184	Sx	BC	Insufficient (<40 GMN+)/absence tumour cells
HUVH	181	Bening	T185.1	Bx	BC	Unclear staining
HUVH	181	Bening	T185.2	Sx	BC	Unclear staining
HUVH	182	Bening	T186	Sx	BC	Unclear staining
HUVH	183	Bening	T187	Sx	BC	Unclear staining (background and weak geminin)
UHP	184	Bening	T188	Sx	BC	Insufficient (<40 GMN+)/absence tumour cells
HUVH	185	Unreviewed	T189	Sx	BC	<25% y-H2AX (low)
HUVH	186	Unreviewed	T190	Sx	BC	Unclear staining (high autofluorescence)
HUVH	187	Unreviewed	T191	Sx	BC	<25% y-H2AX (low)
HUVH	188	Unreviewed	T192	Sx	BC	Unclear staining
HUVH	189	Unreviewed	T193	Sx	BC	<25% y-H2AX (low)
HUVH	190	Unreviewed	T194	Bx	BC	Insufficient (<40 GMN+)/absence tumour cells
HUVH	191	Unreviewed	T195	Sx	BC	Insufficient (<40 GMN+)/absence tumour cells
HUVH	192	Unreviewed	T196	Sx	BC	<25% y-H2AX (low)
HUVH	193	Unreviewed	T197	Bx	BC	Insufficient (<40 GMN+)/absence tumour cells and unclear staining
UHP	194	Unreviewed	T198	Bx	BC	Insufficient (<40 GMN+)/absence tumour cells
UHP	195	Unreviewed	T199	Sx	BC	Unclear staining
UHP	196	Unreviewed	T200	Sx	BC	<25% y-H2AX (low)
UHP	128	Unreviewed	T201	Bx	BC	Insufficient (<40 GMN+)/absence tumour cells

Centre ^a	Patient ID	Cohort	Sample ID	Sample type ^b	Tumour type	Reasoning ^c
UHP	197	Unreviewed	T202	Sx	BC	Unclear staining
UHP	198	Unreviewed	T203.1	Sx	BC	<25% y-H2AX (low)
UHP	198	Unreviewed	T203.2	Sx	BC	Unclear staining
UHP	133	Unreviewed	T204	Sx	BC	Unclear staining
UHP	199	Unreviewed	T205	Sx	BC	<25% y-H2AX (low)
UHP	200	Unreviewed	T206	Sx	BC	<25% y-H2AX (low)

^a. HUVH: Hospital Universitari Vall d'Hebron, MyBrCA: Cancer Research Malaysia , HSJR: Hospital Universitari Sant Joan de Reus, UHP: University Hospital of Parma

^b. Bx: biopsy, Sx: surgical specimen

^c. GMN: geminin

Supplementary Table 27. *In silico* evidence for unreviewed variants' classification.

Gene ^a	Variant	Coding effect	ΔSpliceAI Raw (4999bp) ^b	ΔSpliceAI Masked (4999bp) ^b	BayesDel (noAF) ^c Only for missense BRCA1/2	Splicing code "NA": not applicable for scores between 0.1 and 0.2 "-": code cannot be considered: nonsense, frameshift and variants at dinucleotide positions	PS1 splicing Only for variants meeting PP3, splicing, or PVS1 because they are affecting dinucleotide positions	Protein domain ^d	Missense in silico code	PS1 missense Only for missense BRCA1/2 meeting PP3, missense	Final <i>In silico</i> codes
BRCA1	c.398G>A	Missense	0.01	0.01	0.34	BP4	-	no	BP1_S	-	BP1_S
BRCA1	Del ex10-14	In-frame	-	-	-	NA (too large for prediction)	-	no	-	-	-
BRCA1	c.791_794del	Frameshift	0.66	0.66	-	-	-	no	-	-	-
BRCA1	c.932C>G	Missense	0.01	0	0.07	BP4	-	no	BP1_S	-	BP1_S
BRCA1	c.1655G>A	Missense	0.01	0	0.14	BP4	-	no	BP1_S	-	BP1_S
BRCA1	c.2123C>A	Missense	0	0	-0.26	BP4	-	no	BP1_S	-	BP1_S
BRCA1	c.3247A>G	Missense	0.01	0	-0.38	BP4	-	no	BP1_S	-	BP1_S
BRCA1	c.5152+3A>C	Intronic	0.88	0.88	-	PP3	PS1_M (reference pathogenic variant: c.5152+1G>T)	BRTC	-	-	PP3+PS1_M
BRCA1	c.5311_5332+1del	Frameshift	0.99	0.6	-	-	-	BRTC	-	-	-
BRCA1	c.5467G>A	Missense	0.73	0.73	0.12	PP3	NA, variant at donor motif but no reference variant	BRTC	BP4	-	PP3
BRCA2	c.800G>A	Missense	0	0	-0.47	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.1645A>G	Missense	0	0	-0.31	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.2004del	Frameshift	0.01	0.01	-	-	-	no	-	-	-
BRCA2	c.2799A>G	Synonymous	0.01	0.01	-0.9	BP4	-	no	NA	-	BP1_S
BRCA2	c.3032C>G	Missense	0	0	0.43	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.3083A>C	Missense	0	0	0.07	BP4	-	no	BP1_S	-	BP1_S

Gene ^a	Variant	Coding effect	Δ SpliceAI Raw (4999bp) ^b	Δ SpliceAI Masked (4999bp) ^b	BayesDel (noAF) ^c Only for missense BRCA1/2	Splicing code "NA": not applicable for scores between 0.1 and 0.2 "-": code cannot be considered: nonsense, frameshift and variants at dinucleotide positions	PS1 splicing Only for variants meeting PP3_splicing, or PVS1 because they are affecting dinucleotide positions	Protein domain ^d	Missense in silico code	PS1 missense Only for missense BRCA1/2 meeting PP3_missense	Final in silico codes
BRCA2	c.3216A>T	Missense	0	0	-0.61	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.3310A>C	Missense	0	0	-0.09	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.3398C>T	Missense	0	0	-0.25	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.3503T>A	Missense	0	0	-0.49	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.4316C>A	Missense	0	0	-0.19	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.4928T>C	Missense	0.01	0	-0.46	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.6268C>T	Missense	0	0	-0.54	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.6382A>G	Missense	0.02	0	-0.6	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.6560C>T	Missense	0.02	0.02	-0.22	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.6700T>C	Missense	0.01	0.01	0.09	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.7022G>A	Missense	0.01	0.01	-0.08	BP4	-	no	BP1_S	-	BP1_S
BRCA2	c.7522G>A	Missense	0.11	0.11	0.35	NA	-	DBD	PP3	NA: no other (likely) pathogenic reference variants	PP3
BRCA2	c.7985C>G	Missense	0.23	0.23	-0.15	PP3	NA, variant is outside acceptor motif (is 8th position inside exon)/ no reference variant at the same position)	DBD	BP4	-	PP3
BRCA2	c.8117A>G	Missense	0	0	-0.36	BP4	-	DBD	BP4	-	BP4
BRCA2	c.8847del	Frameshift	0.02	0.02	-	-	-	DBD	-	-	-
BRCA2	c.9410C>T	Missense	0	0	0.07	BP4	-	DBD	BP4	-	BP4
PALB2	c.101G>A	Missense	0	0	-	NA	-	CC	-	-	BP1
PALB2	c.629C>T	Missense	0.01	0.01	-	NA	-	no	-	-	BP1

Gene ^a	Variant	Coding effect	Δ SpliceAI Raw (499bp) ^b	Δ SpliceAI Masked (499bp) ^b	BayesDel (noAF) ^c Only for missense BRCA1/2	Splicing code "NA": not applicable for scores between 0.1 and 0.2 "-": code cannot be considered: nonsense, frameshift and variants at dinucleotide positions	PS1 splicing Only for variants meeting PP3 splicing, or PVS1 because they are affecting dinucleotide positions	Protein domain ^d	Missense in silico code	PS1 missense Only for missense BRCA1/2 meeting PP3 missense	Final <i>in silico</i> codes
PALB2	c.925A>G	Missense	0.03	0.03	-	NA	-	no	-	-	BP1
PALB2	c.1861C>A	Missense	0.03	0	-	NA	-	MBD	-	-	BP1
PALB2	c.2093T>C	Missense	0	0	-	NA	-	no	-	-	BP1
PALB2	c.2401G>A	Missense	0.01	0.01	-	NA	-	no	-	-	BP1
PALB2	c.2425A>G	Missense	0	0	-	NA	-	no	-	-	BP1
PALB2	c.2612A>G	Missense	0.28	0.28	-	PP3	NA: Activation of a cryptic and loss of natural site and not at splicing acceptor motif	WD40	-	-	PP3 + BP1

^a. Reference transcripts: BRCA1 (NM_007294.3), BRCA2 (NM_000059.3), PALB2 (NM_024675.3)

^b. Maximum raw or masked Δ SpliceAI score, defined as the maximum probability of altered splicing across 4 output probabilities (designated by the tool as loss of acceptor or donor sites, gain of acceptor or donor sites) and the maximum distance of 10,000 nucleotides ($\pm 4,999$ nucleotides from the variant of interest).

^c. BayesDel (noAF): model that does not incorporate allele frequencies

^d. Only clinically relevant domains indicated. BRCA1: RING (from residue 2 to 101), CC (from residue 1391-1424) and BRCT (from residue 1650 to 1587). BRCA2: PALB2-interacting domain (from residue 10 to 40) and DBD (from residue 2481 to 3186). PALB2: CC (from residue 9 to 44), and WD40 (from residue 854 to 1186)

In all codes, NA: not-applicable because variant does not meet criteria specification, "-": code cannot be considered for this variant, S: strong, M: moderate

Supplementary Table 28. Splicing observed impact evidence for unreviewed variants' classification.

Gene ^a	Variant	Coding effect	Variant location	Protein domain ^b	Reference	Splicing result	Methodology	Evidence
BRCA1	c.791_794del	Frameshift	Exon 11	-	Internal and collaborators mRNA results published in a poster communication (PMID: 37165069)	Quantitative results from carriers demonstrated that the variant allele generates almost exclusively Δ11q. This in-frame event removes >10% of coding sequence but is not affecting RING or BRTC domain, hence it meets PVS1_MOD.	RT-PCR allele-specific assay using mRNA from patient's cultured PMBCs (NMD inhibited)	PVS1_M(RNA)
BRCA1	c.5152+3A>C	Intronic	Intron 18	BRTC	Unpublished internal data (part 2 of this thesis)	Quantitative results from patients and minigenes showed generates exclusively Δ18 (≥90%, i.e. normal transcript <10%). This is an in-frame deletion affecting BRTC domain, hence PVS1(RNA) must be considered.	RT-PCR using mRNA from patient's whole blood and minigenes	PVS1(RNA)
BRCA1	c.5467G>A	Missense	Exon 23	BRTC	PMID: 12915465	Results from an humanised mouse model showed this variant produced Δ23. ▼ 23q(insertion of 5 intronic bp) and full-length, this suggesting partial effect which did not allow to apply PVS1 (RNA) code.	ES stem cell/RT-PCR	PVS1(RNA)_NA. Bioinformatic codes kept as splicing results are unclear.
BRCA2	c.3032C>G	Missense	Exon 11	-	PMID: 32123317	No effect	RT-PCR using patients' mRNA	NA
BRCA2	c.7985C>G	Missense	Exon 18	DBD	PMID: 28339459	Minigene results showed this variant generated Δ18 (92%) Δ18q (del 191 bp) (5%), others (4.8%). Both Δ18 and Δ18q events are predicted to introduce a premature termination codon and both transcripts are susceptible to be NMD, hence this results allot to apply PVS1(RNA)_strong	Minigenes	PVS1_S(RNA)

^a. Reference transcripts: BRCA1(NM_007294.3), BRCA2 (NM_000059.3)

^b. Only clinical important domains considered. BRCA1: RING (from residue 2 to 101), CC (from residue 1391-1424) and BRCT (from residue 1650 to 1587). BRCA2: PALB2-interacting domain (from residue 10 to 40) and DBD (from residue 2481 to 3186).

NA: not-applicable, S: strong, M: moderate

Supplementary Table 29. Allele frequency evidence for unreviewed variants' classification.

Due to the large amount of data that this table contains, the information is available here:
https://drive.google.com/drive/folders/1T9aCggWsrMPzhU_31aseVRcn9vEHkT?usp=drive_link

Supplementary Table 30. Functional evidence for unreviewed variants' classification.

Gene ^a	cNomen	pNomen	Variant location	Protein domain ^b	Functional results and references	Functional evidence
<i>BRCA1</i>	c.398G>A	p.(Arg133His)	Exon 7	-	Reported by one calibrated study to exhibit protein function similar to benign control variants (PMID:30219179)	BS3
<i>BRCA1</i>	c.2123C>A	p.(Ser708Tyr)	Exon 11	-	Reported by one calibrated study to exhibit protein function similar to benign control variants (PMID:32546644)	BS3
<i>BRCA1</i>	c.5152+3A>C	p.?	Intron 18	BRTC	Reported by two calibrated study incorporating mRNA splicing effects to exhibit protein function similar to pathogenic control variants (PMID:30209399 and PART2 of this thesis)	PS3
<i>BRCA1</i>	c.5467G>A	p.(Ala1823Thr)	Exon 23	BRTC	Missense variant shown to alter splicing. Reported by one calibrated study incorporating mRNA splicing effect to exhibit function similar to pathogenic control variants (PMID:30209399)	PS3
<i>BRCA2</i>	c.7522G>A	p.(Gly2508Ser)	Exon 15	DBD	Reported by two or more calibrated studies with discordant results. Functional effect similar to benign control variants (PMIDs: 33609447) and between what was observed for benign and control variants (PMID: 29988080, 32444794)	NA
<i>BRCA2</i>	c.7559G>T	p.(Arg2520Leu)	Exon 15	DBD	Reported by one calibrated study to exhibit protein function similar to benign control variants (PMID:33609447)	BS3
<i>BRCA2</i>	c.9086C>T	p.(Ala3029Val)	Exon 23	DBD	Missense variant predicted to alter splicing. Results from one calibrated study with cDNA based design not considered for code application (PMID:33609447)	NA
<i>BRCA2</i>	c.9383G>A	p.(Arg3128Gln)	Exon 25	DBD	Reported by one calibrated study to exhibit protein function similar to benign control variants (PMID:33609447)	BS3
<i>BRCA2</i>	c.4928T>C	p.(Val1643Ala)	Exon 11	-	Reported by one calibrated study to exhibit protein function similar to benign control variants (PMID:33293522)	BS3

^a. Reference transcripts: *BRCA1*(NM_007294.3), *BRCA2*(NM_000059.3), *PALB2*(NM_024675.3)

^b. Only clinically relevant domains considered. *BRCA1*: RING (from residue 2 to 101), CC (from residue 1391-1424) and BRCT (from residue 1650 to 1587). *BRCA2*: PALB2-interacting domain (from residue 10 to 40) and DBD (from residue 2481 to 3186).
 NA: not applicable.

Supplementary Table 31. Multifactorial likelihood clinical data for unreviewed variants' classification.

Gene	Variant	Source(s)	Co-Segregation LR	Tumour Pathology LR ^a	Co-occurrence LR	Family History LR	Combined LR	Combined Clinical LR for PP4 or BP5 ^b	Code/s to apply ^c
BRCA1	c.398G>A	Parsons et al 2019, PMID:31131967	-	-	1.14	0.12	0.13	0.15	BP5_M
		Tumour pathology LR	-	1.13	-	-	-	-	-
BRCA1	Del exons 10-14	Li et al 2020, PMID: 31853058	-	-	-	-	0.06	0.07	BP5_M
		Tumour pathology LR	-	1.36	-	-	-	-	-
BRCA1	c.791_794del	Li et al 2020, PMID: 31853058	-	-	-	-	2.94	9.29	PP4_M
		Tumour pathology LR	-	3.16	-	-	-	-	-
BRCA1	c.2123C>A	Parsons et al 2019, PMID:31131967	-	3.73	-	-	3.73	3.73	PP4_P
		Tumour pathology LR	-	Patient included in Parsons et al 2019, PMID:31131967	-	-	-	-	-
BRCA1	c.5467G>A	Parsons et al 2019, PMID:31131967	0.96	3.16	-	-	3.04	12.57	PP4_M
BRCA2	c.800G>A	Parsons et al 2019, PMID:31131967	1.17	1.58	1.41	0.35	0.90	0.0007	BP5_VS
		Caputo et al 2021, PMID: 34597585	5.05	0.05	0.01	0.53	0.00	-	-
BRCA2	c.1645A>G	Li et al 2020, PMID: 31853058	-	-	-	-	0.52	0.36	BP5_P
		Tumour pathology LR	-	0.69	-	-	0.69	-	-
BRCA2	c.3032C>G	Parsons et al 2019, PMID:31131967	3.55	0.27	1.28	2.11	2.61	2.61	PP4_P
		Tumour pathology LR	-	Patient included in Parsons et al 2019, PMID:31131967	-	-	-	-	-
BRCA2	c.3216A>T	Parsons et al 2019, PMID:31131967	-	0.88	-	-	0.88	0.88	UNINFORMATIVE

Gene	Variant	Source(s)	Co-Segregation LR	Tumour Pathology LR ^a	Co-occurrence LR	Family History LR	Combined LR	Combined Clinical LR for PP4 or BP5 ^b	Code/s to apply ^c
BRCA2	c.3310A>C	Tumour pathology LR	-	Patient included in Parsons et al 2019, PMID:31131967	-	-	-	-	
		Li et al 2020, PMID: 31853058	-	-	-	-	0.11	0.04	BP5_S
		Tumour pathology LR	-	0.35	-	-	0.35	-	
BRCA2	c.3398C>T	Li et al 2020, PMID: 31853058	-	-	-	-	0.41	0.47	BP5_P
		Tumour pathology LR	-	1.15	-	-	-	-	
BRCA2	c.4316C>A	Li et al 2020, PMID: 31853058	-	-	-	-	1.05	0.92	UNINFORMATIVE
		Tumour pathology LR	-	0.88	-	-	0.88	-	
BRCA2	c.4928T>C	Parsons et al 2019, PMID:31131967	8.57	0.74	1.24	0.19	1.52	1.64	UNINFORMATIVE
		Tumour pathology LR	-	1.08	-	-	1.08	-	
BRCA2	c.7022G>A	Parsons et al 2019, PMID:31131967	-	-	1.02	0.49	0.50	0.30	BP5_P
		Li et al 2020, PMID: 31853058	-	-	-	-	0.68	-	
		Tumour pathology LR	-	0.89	-	-	0.89	-	
BRCA2	c.7522G>A	Li et al 2020, PMID: 31853058	-	-	-	-	0.44	0.47	BP5_P
		Tumour pathology LR	-	1.07	-	-	1.07	-	
BRCA2	c.8117A>G	Parsons et al 2019, PMID:31131967	-	0.69	1.41	1.23	1.19	1.49	UNINFORMATIVE
		Tumour pathology LR	-	1.25	-	-	1.25	-	

^a. Tumour pathology LR were calculated according to Spurdle et al. (2014)[198] for breast cancer and to O'Mahony et al. (2023)[199] for ovarian tumours. Several patients from Vall d'Hebron Hospital had previously been included in the tumour pathology LR calculations by Parsons et al., (2019). Consequently, we did not recalculate the LR for these variants.

^b. Combined clinical likelihood ratios (LR) were calculated by multiplying individual LR.

^c. BP5 and PP4 weights were applied according to Tavtigian et al. (2018)[118]. VS: strong, S: strong, M: moderate, P: supporting.

Supplementary Table 32. Case-control data from BRIDGES study for unreviewed variants' classification.

Gene ^a	Variant	gNomen ^b	Cases	Total cases	Controls	Total controls	Database	OR ^c	CI 95%	p-value	z-statistic	Evidence code ^c
BRCA1	c.398G>A p.(Arg133His)	g.41256182C>T	3	60466	0	53461	ALL_population (BRIDGES)	6.19	0.3197 to 119.8267	0.23	1.21	NA
BRCA1	c.2123C>A p.(Ser708Tyr)	g.41245425G>T	2	60466	6	53461	ALL_population (BRIDGES)	0.29	0.0595 to 1.4603	0.13	1.50	NA
BRCA1	c.3247A>G p.(Met1083Val)	g.41244301T>C	1	60466	0	53461	ALL_population (BRIDGES)	2.65	0.1080 to 65.1161	0.55	0.60	NA
BRCA1	c.5467G>A p.(Ala1823Thr)	g.41199660C>T	2	60466	0	53461	ALL_population (BRIDGES)	4.42	0.2122 to 92.0886	0.34	0.96	NA
BRCA2	c.800G>A p.(Gly267Glu)	g.32906415G>A	4	60466	5	53461	ALL_population (BRIDGES)	0.71	0.1899 to 2.6342	0.61	0.52	NA
BRCA2	c.1645A>G p.(Lys549Glu)	g.32907260A>G	0	60466	2	53461	ALL_population (BRIDGES)	0.18	0.0085 to 3.6835	0.26	1.12	NA
BRCA2	c.3032C>G p.(Thr1011Arg)	g.32911524C>G	5	60466	2	53461	ALL_population (BRIDGES)	2.21	0.4288 to 11.3937	0.34	0.95	NA
BRCA2	c.3310A>C p.(Thr1104Pro)	g.32911802A>C	0	60466	2	53461	ALL_population (BRIDGES)	0.18	0.0085 to 3.6835	0.26	1.12	NA
BRCA2	c.3503T>A p.(Met1168Lys)	g.32911995T>A	0	60466	2	53461	ALL_population (BRIDGES)	0.18	0.0085 to 3.6835	0.26	1.12	NA
BRCA2	c.4928T>C p.(Val1643Ala)	g.32913420T>C	11	60466	5	53461	ALL_population (BRIDGES)	1.95	0.6758 to 5.5987	0.22	1.23	NA
BRCA2	c.6560C>T p.(Pro2187Leu)	g.32915052C>T	0	60466	1	53461	ALL_population (BRIDGES)	0.29	0.0120 to 7.2351	0.45	0.75	NA
BRCA2	c.7022G>A p.(Arg2341His)	g.32929012G>A	1	60466	0	53461	ALL_population (BRIDGES)	2.65	0.1080 to 65.1161	0.55	0.60	NA
BRCA2	c.7522G>A p.(Gly2508Ser)	g.32930651G>A	22	60466	25	53461	ALL_population (BRIDGES)	0.78	0.4387 to 1.3800	0.39	0.86	NA
BRCA2	c.8117A>G p.(Asn2706Ser)	g.32937456A>G	6	60466	4	53461	ALL_population (BRIDGES)	1.33	0.3742 to 4.7000	0.66	0.44	NA
PALB2	c.101G>A p.(Arg34His)	g.23649398C>T	4	60466	1	53461	ALL_population (BRIDGES)	3.54	0.3953 to 31.6438	0.29	1.13	NA
PALB2	c.629C>T p.(Pro210Leu)	g.23647238G>A	16	60466	15	53461	ALL_population (BRIDGES)	0.94	0.4662 to 1.9078	0.87	0.16	NA
PALB2	c.1861C>A p.(Pro621Thr)	g.23641614G>T	0	60466	1	53461	ALL_population (BRIDGES)	0.29	0.0120 to 7.2351	0.45	0.75	NA
PALB2	c.2093T>C p.(Leu698Pro)	g.23641382A>G	2	60466	0	53461	ALL_population (BRIDGES)	4.42	0.2122 to 92.0886	0.34	0.96	NA
PALB2	c.2401G>A p.(Asp801Asn)	g.23641074C>T	1	60466	0	53461	ALL_population (BRIDGES)	2.65	0.1080 to 65.1161	0.55	0.60	NA
PALB2	c.2612A>G p.(Asp871Gly)	g.23637693T>C	2	60466	3	53461	ALL_population (BRIDGES)	0.59	0.0985 to 3.5278	0.56	0.58	NA

Data source: <https://bcac.ccg.medschl.cam.ac.uk/bcacdata/bridges-summary-results-breast-cancer-risk-genes-2021/> (accessed in February 2024)

^a. Reference transcripts: BRCA1 (NM_007294.3), BRCA2 (NM_000059.3), PALB2 (NM_024675.3)

^b. GRCh37 / hg19

^c. Evidence PS4 for BRCA1/2: p-value ≤0.05 and Odds Ratio (OR) ≥4 with lower 95%CI excludes 2. For PALB2: p-value ≤0.05 and Odds Ratio (OR) ≥3 or lower 95%CI ≥1.5. In column "OR", values in red are those greater than the cutoff defined for each gene.

NA: not applicable

Supplementary Table 33. Likelihood ratio association analyses of RAD51% HRR status based on BC molecular subtype, tumour grade and age at diagnosis.

All genes		All ages					<50 years old				≥50 years old			
		Grade	HRD by RAD51 test	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	LR (95%CI)
All subtypes	I	Yes (HRD)	2 (100.0)	1 (25.0)	4.00 (0.73-21.84)	Non-informative	1 (100.0)	0 (0.0)	-	-	1 (100.0)	1 (33.3)	3.00 (0.61-14.86)	Non-informative
		No (HRP)	0 (0.0)	3 (75.0)	-	-	0 (0.0)	1 (100.0)	-	-	0 (0.0)	2 (66.7)	-	-
	II	Yes (HRD)	27 (96.4)	3 (21.4)	4.50 (1.65-12.30)	Pathogenic moderate	20 (95.2)	0 (0.0)	-	-	7 (100.0)	3 (50.0)	2.00 (0.90-4.45)	Non-informative
		No (HRP)	1 (3.6)	11 (78.6)	0.045 (0.01-0.32)	Benign strong	1 (4.8)	8 (100.0)	0.05 (0.01-0.32)	Benign strong	0 (0.0)	3 (50.0)	-	-
	III	Yes (HRD)	45 (91.8)	6 (35.3)	2.60 (1.36-4.98)	Pathogenic supporting	30 (90.1)	2 (22.2)	4.09 (1.20-13.95)	Pathogenic supporting	15 (93.7)	4 (50.0)	1.88 (0.93-3.79)	Non-informative
		No (HRP)	4 (8.2)	11 (64.7)	0.13 (0.05-0.34)	Benign moderate	3 (9.1)	7 (77.8)	0.12 (0.04-0.36)	Benign moderate	1 (6.3)	4 (50.0)	0.13 (0.02-0.94)	Benign moderate
TN	ALL	Yes (HRD)	32 (94.1)	3 (42.9)	2.20 (0.93-5.19)	Non-informative	21 (95.5)	2 (40.0)	2.39 (0.81-7.01)	Non-informative	12 (92.3)	1 (50.0)	1.85 (0.46-7.45)	Non-informative
		No (HRP)	2 (5.9)	4 (57.1)	0.10 (0.02-0.46)	Benign moderate	1 (4.5)	3 (60.0)	0.08 (0.01-0.59)	Benign moderate	1 (7.7)	1 (50.0)	0.15 (0.01-1.59)	Non-informative
	I	Yes (HRD)	0	0	-	-	0	0	-	-	0	0	-	-
		No (HRP)	0	0	-	-	0	0	-	-	0	0	-	-
	II	Yes (HRD)	4 (100.0)	0	-	-	3 (100.0)	0	-	-	1 (100.0)	0	-	-
		No (HRP)	0 (0.0)	0	-	-	0 (0.0)	0	-	-	0 (0.0)	0	-	-
ER+	III	Yes (HRD)	25 (92.6)	2 (50.0)	1.85 (0.69-4.96)	Non-informative	16 (94.1)	1 (50.0)	1.88 (0.47-7.57)	Non-informative	9 (90.0)	1 (50.0)	1.80 (0.44-7.31)	Non-informative
		No (HRP)	2 (7.4)	2 (50.0)	0.15 (0.03-0.78)	Benign moderate	1 (5.9)	1 (50.0)	0.12 (0.01-1.24)	Non-informative	1 (10.0)	1 (50.0)	0.20 (0.02-2.03)	Non-informative
	ALL	Yes (HRD)	49 (90.7)	6 (28.6)	3.18 (1.61-6.28)	Pathogenic supporting	35 (87.5)	1 (8.3)	10.5 (1.60-68.83)	Pathogenic moderate	15 (100.0)	5 (55.6)	1.80 (1.00-3.23)	Non-informative
		No (HRP)	5 (9.3)	15 (71.4)	0.13 (0.05-0.31)	Benign moderate	5 (12.5)	11 (91.7)	0.14 (0.06-0.32)	Benign moderate	0 (0.0)	4 (44.4)	-	-
	I	Yes (HRD)	2 (100.0)	1 (50.0)	2.00 (0.50-8.00)	Non-informative	1 (100.0)	0 (0.0)	-	-	1 (100.0)	1 (100.0)	1.00 (1.00-1.00)	Non-informative
		No (HRP)	0 (0.0)	1 (50.0)	-	-	0 (0.0)	1 (100.0)	-	-	0 (0.0)	0 (0.0)	-	-
II	Yes (HRD)	23 (95.8)	3 (27.3)	3.51 (1.33-9.26)	Pathogenic supporting	17 (94.4)	0 (0.0)	-	-	6 (100.0)	3 (60.0)	1.67 (0.81-3.41)	Non-informative	
	No (HRP)	1 (4.2)	8 (72.7)	0.06 (0.01-0.40)	Benign moderate	1 (5.56)	6 (100.0)	0.06 (0.01-0.37)	Benign moderate	0 (0.0)	2 (40.0)	-	-	

All genes		All ages						<50 years old				≥50 years old			
		Grade	HRD by RAD51 test	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength
		III	Yes (HRD)	19 (90.5)	1 (16.7)	5.43 (0.90-32.66)	Non-informative	13 (86.7)	0 (0.0)	-	-	6 (100.0)	1 (33.3)	3.00 (0.61-14.86)	Non-informative
			No (HRP)	2 (9.5)	5 (83.3)	0.11 (0.03-0.45)	Benign moderate	2 (13.3)	3 (100.0)	0.13 (0.04-0.48)	Benign moderate	0 (0.0)	2 (66.7)	-	-
HER2+		All	Yes (HRD)	0	1 (25.0)	-	-	0	0 (0.0)	-	-	0	1 (50.0)	-	-
			No (HRP)	0	3 (75.0)	-	-	0	2 (100.0)	-	-	0	1 (50.0)	-	-

LR>1: association with pathogenic variant, pathogenic evidence; LR<1: no association with pathogenic variant, benign evidence. Evidence strength was measured based on Bayesian modelling of ACMG/AMP rules (**Table 8**); Benign supporting ($0.23 \leq LR < 0.48$), Benign moderate ($0.05 \leq LR < 0.23$), Pathogenic supporting ($2.08 \leq LR < 4.30$), non-informative ($0.48 \leq LR < 2.08$). LR estimates reaching informative ACMG/AMP strengths at a statistically significant CI (i.e., not spanning 1), are highlighted in bold. n: sample size, LR: Likelihood ratio; CI: Confidence Interval, ACMG/AMP: American College of Medical Genetics/Association for Molecular Pathology. TN: triple negative, ER: oestrogen receptor, PR: progesterone receptor.

Supplementary Table 34. Likelihood ratio association analyses of RAD51% HRR status in tumours from germline *BRCA1* pathogenic variant carriers based on BC molecular subtype, tumour grade and age at diagnosis.

<i>BRCA1</i>	All ages				<50 years old				≥50 years old			
	Grade	HRD by RAD51 test	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)
All subtypes	ALL	Yes (HRD)	35 (94.6)	12	3.39 (2.08-5.51)	Pathogenic supporting	22 (95.7)	4 (15.4)	6.22 (2.51-15.38)	Pathogenic moderate	13 (93.8)	8 (47.1)
		No (HRP)	2 (5.4)	31 (72.1)	0.07 (0.02-0.29)	Benign moderate	1 (4.3)	22 (84.6)	0.051 (0.01-0.35)	Benign moderate	1 (6.2)	9 (52.9)
	I	Yes (HRD)	0	1 (25.0)	-	-	0	0 (0.0)	-	-	0	1 (33.3)
		No (HRP)	0	3 (75.0)	-	-	0	1 (100.0)	-	-	0	2 (66.7)
	II	Yes (HRD)	9 (100.0)	3 (21.4)	4.67 (1.71-12.72)	Pathogenic moderate	5 (100.0)	0 (0.0)	-	-	4 (100.0)	3 (50.0)
		No (HRP)	0 (0.0)	11 (78.6)	-	-	0 (0.0)	8 (100.0)	-	-	0 (0.0)	3 (50.0)
TN	III	Yes (HRD)	23 (92.0)	6 (35.3)	2.61 (1.36-5.01)	Pathogenic supporting	15 (93.8)	2 (22.2)	4.22 (1.23-14.42)	Pathogenic supporting	8 (88.9)	4 (50.0)
		No (HRP)	2 (8.0)	11 (64.7)	0.12 (0.03-0.49)	Benign moderate	1 (6.3)	7 (77.8)	0.08 (0.01-0.55)	Benign moderate	1 (11.1)	4 (50.0)
	ALL	Yes (HRD)	25 (92.6)	3 (42.9)	2.16 (0.91-5.12)	Non-informative	15 (93.8)	2 (40.0)	2.34 (0.80-6.91)	Non-informative	10 (90.9)	1 (50.0)
		No (HRP)	2 (7.4)	4 (57.1)	0.13 (0.03-0.57)	Benign moderate	1 (6.3)	3 (60.0)	0.10 (0.01-0.79)	Benign moderate	1 (9.1)	1 (50.0)
	I	Yes (HRD)	0	0	-	-	0	0	-	-	0	0
		No (HRP)	0	0	-	-	0	0	-	-	0	0
ER+	II	Yes (HRD)	0	0	-	-	0	0	-	-	0	0
		No (HRP)	0	0	-	-	0	0	-	-	0	0
	III	Yes (HRD)	19 (90.5)	2 (50.0)	1.81 (0.67-4.87)	Non-informative	11 (91.7)	1 (50.0)	1.83 (0.45-7.41)	Non-informative	8 (88.9)	1 (50.0)
		No (HRP)	2 (9.5)	2 (50.0)	0.19 (0.04-0.98)	Benign moderate	1 (8.3)	1 (50.0)	0.17 (0.02-1.72)	Non-informative	1 (11.1)	1 (50.0)
	ALL	Yes (HRD)	10 (100.0)	6 (28.6)	3.50 (1.78-6.88)	Pathogenic supporting	7 (100.0)	1 (8.3)	12.00 (1.84-78.37)	Pathogenic moderate	3 (100.0)	5 (55.6)
		No (HRP)	0 (0.0)	15 (71.4)	-	-	0 (0.0)	11 (91.7)	-	-	0 (0.0)	4 (44.4)
ER+	I	Yes (HRD)	0	1 (50.0)	-	-	0	0 (0.0)	-	-	0	1 (100.0)
		No (HRP)	0	0	-	-	0	0	-	-	0	0

All ages									
BRCA1	Grade	HRD by RAD51 test		Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	<50 years old	
		No (HRP)	Yes (HRD)					Pathogenic n(%)	Benign n(%)
	II	No (HRP)	Yes (HRD)	0	1 (50.0)	-	-	0	0 (0.0)
		6 (100.0)	3 (27.3)	3 (100.0)	0 (0.0)	3.67 (1.40-9.62)	-	3 (100.0)	3 (60.0)
	III	No (HRP)	Yes (HRD)	0 (0.0)	8 (72.7)	-	-	0 (0.0)	2 (40.0)
		4 (100.0)	1 (16.7)	4 (100.0)	0 (0.0)	6.00 (1.00-35.91)	-	0	1 (33.3)
HER2+	All	No (HRP)	Yes (HRD)	0 (0.0)	5 (83.3)	-	-	0 (0.0)	2 (66.7)
		0	1 (25.0)	0	0 (0.0)	-	-	0	1 (50.0)
		No (HRP)	Yes (HRD)	0	3 (75.0)	-	-	0	1 (50.0)
		0	3 (75.0)	0	2 (100.0)	-	-	0	1 (50.0)

LR>1: association with pathogenic variant, pathogenic evidence; LR<1: no association with pathogenic variant, benign evidence. Evidence strength was measured based on Bayesian modelling of ACMG/AMP rules (**Table 8**); Benign supporting (0.05 ≤ LR < 0.23), Pathogenic supporting (2.08 ≤ LR < 4.30), non-informative (0.48 ≤ LR < 2.08). LR estimates reaching informative ACMG/AMP strengths at a statistically significant CI (i.e., not spanning '1'), are highlighted in bold. n: sample size, LR: Likelihood ratio; CI: Confidence Interval, ACMG/AMP: American College of Medical Genetics/Association for Molecular Pathology. TN: triple negative, ER: oestrogen receptor, PR: progesterone receptor.

Supplementary Table 35. Likelihood ratio association analyses of RAD51% HRR status in tumours from germline *BRCA2* pathogenic variant carriers based on BC molecular subtype, tumour grade and age at diagnosis.

<i>BRCA2</i>	All ages				<50 years old				≥50 years old			
	Grade	HRD by RAD51 test	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)
All subtypes	All	Yes (HRD)	29 (90.6)	12 (27.9)	3.25 (1.98-5.32)	Pathogenic supporting	21 (87.5)	4 (15.4)	5.69 (2.28-14.19)	Pathogenic moderate	8 (100.0)	8 (47.1)
		No (HRP)	3 (9.4)	31 (72.1)	0.13 (0.04-0.39)	Benign moderate	3 (12.5)	22 (84.6)	0.15 (0.05-0.43)	Benign moderate	0 (0.0)	9 (52.9)
	I	Yes (HRD)	0	1 (25.0)	-	-	0	0 (0.0)	-	-	0	1 (33.3)
		No (HRP)	0	3 (75.0)	-	-	0	1 (100.0)	-	-	0	2 (66.7)
	II	Yes (HRD)	11 (91.7)	3 (21.4)	4.28 (1.55-11.83)	Pathogenic supporting	9 (90.0)	0 (0.0)	-	-	2 (100.0)	3 (50.0)
		No (HRP)	1 (8.3)	11 (78.6)	0.11 (0.02-0.71)	Benign moderate	1 (10)	8 (100.0)	0.10 (0.02-0.64)	Benign moderate	0 (0.0)	3 (50.0)
TN	III	Yes (HRD)	15 (88.2)	6 (35.3)	2.50 (1.28-4.87)	Pathogenic supporting	10 (83.3)	2 (22.2)	3.75 (1.08-13.07)	Pathogenic supporting	5 (100.0)	4 (50.0)
		No (HRP)	2 (11.8)	11 (64.7)	0.18 (0.05-0.70)	Benign moderate	2 (16.7)	7 (77.8)	0.21 (0.06-0.80)	Benign moderate	0 (0.0)	4 (50.0)
	ALL	Yes (HRD)	3 (100.0)	3 (42.9)	2.33 (0.99-5.49)	Non-informative	3 (100.0)	2 (40.0)	2.50 (0.85-7.31)	Non-informative	0	1 (50.0)
		No (HRP)	0 (0.0)	4 (57.1)	-	-	0 (0.0)	3 (60.0)	-	-	0	1 (50.0)
	I	Yes (HRD)	0	0	-	-	0	0	-	-	0	0
		No (HRP)	0	0	-	-	0	0	-	-	0	0
ER+	II	Yes (HRD)	0	0	-	-	0	0	-	-	0	0
		No (HRP)	0	0	-	-	0	0	-	-	0	0
	III	Yes (HRD)	3 (100.0)	2 (50.0)	2.00 (0.75-5.33)	Non-informative	3 (100.0)	1 (50.0)	2.00 (0.50-8.00)	Non-informative	0	1 (50.0)
		No (HRP)	0 (0.0)	2 (50.0)	-	-	0 (0.0)	1 (50.0)	-	-	0	1 (50.0)
	ALL	Yes (HRD)	25 (89.3)	6 (28.6)	3.13 (1.57-6.22)	Pathogenic supporting	17 (85.0)	1 (8.3)	10.20 (1.55-67.22)	Pathogenic moderate	8 (100.0)	5 (55.6)
		No (HRP)	3 (10.7)	15 (71.4)	0.15 (0.05-0.45)	Benign moderate	3 (15.0)	11 (91.7)	0.16 (0.06-0.47)	Benign moderate	0 (0.0)	4 (44.4)
ER+	I	Yes (HRD)	0	1 (50.0)	-	-	0	0 (0.0)	-	-	0	1 (100.0)
		No (HRP)	0	0	-	-	0	0	-	-	0	0

All ages											
BRCA2	Grade	HRD by RAD51 test	Pathogenic			Benign			ACMG/AMP strength		
			n(%)	n(%)	LR (95%CI)	n(%)	n(%)	LR (95%CI)	n(%)	n(%)	ACMG/AMP strength
	II	No (HRP)	0	1 (50.0)	-	1 (100.0)	0	-	0	0 (0.0)	-
		Yes (HRD)	11 (91.7)	3 (27.3)	3.36 (1.26- 8.96)	9 (90.0)	0 (0.0)	-	2 (100.0)	3 (60.0)	1.67 (0.81- 3.41)
		No (HRP)	1 (8.3)	8 (72.7)	0.11 (0.02- 0.77)	1 (10.0)	6 (100.0)	0.10 (0.02- 0.64)	0 (0.0)	2 (40.0)	-
	III	Yes (HRD)	11 (84.6)	1 (16.7)	5.08 (0.84- 30.84)	6 (75.0)	0 (0.0)	-	5 (100.0)	1 (33.3)	3.00 (0.61- 14.86)
		No (HRP)	2 (15.4)	5 (83.3)	0.18 (0.05- 0.69)	2 (25.0)	3 (100.0)	0.25 (0.08- 0.83)	0 (0.0)	2 (66.7)	-
		Yes (HRD)	0	1 (25.0)	-	0	0 (0.0)	-	0	1 (50.0)	-
HER2+	All	No (HRP)	0	3 (75.0)	-	0	2 (100.0)	-	0	1 (50.0)	-

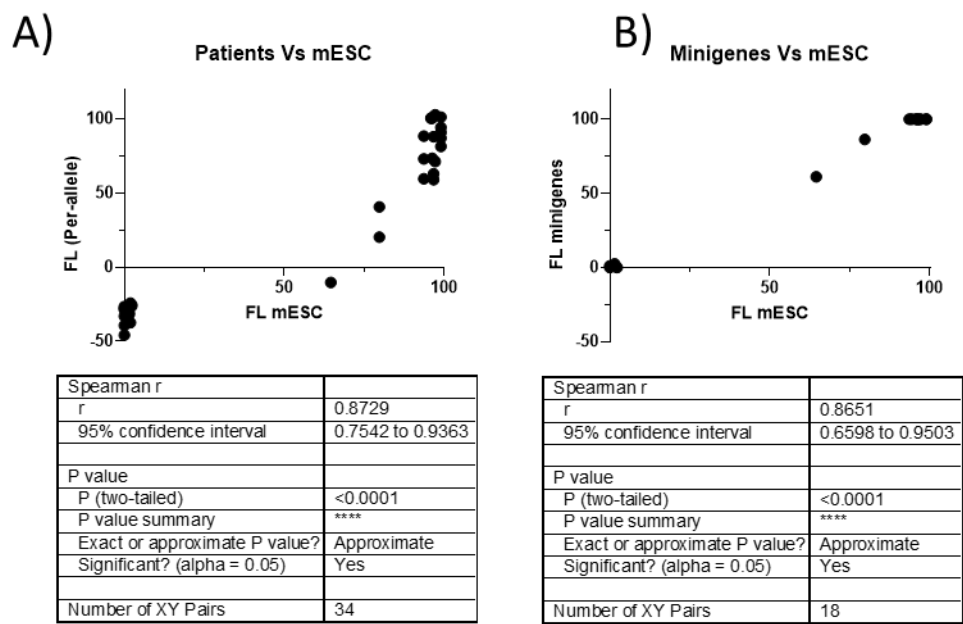
LR>1: association with pathogenic variant, pathogenic evidence; LR<1: no association with pathogenic variant, benign evidence. Evidence strength was measured based on Bayesian modelling of ACMG/AMP rules (**Table 8**); Benign supporting (0.23 ≤ LR < 0.48), Benign moderate (0.05 ≤ LR < 0.23), Pathogenic supporting (2.08 ≤ LR < 4.30), non-informative (0.48 ≤ LR < 2.08), LR estimates reaching informative ACMG/AMP strengths at a statistically significant CI (i.e., not spanning '1'), are highlighted in bold. n: sample size, LR: Likelihood ratio; CI: Confidence Interval, ACMG/AMP: American College of Medical Genetics/Association for Molecular Pathology. TN: triple negative, ER: oestrogen receptor, PR: progesterone receptor.

Supplementary Table 36. Likelihood ratio association analyses of RAD51% HRR status in tumours from germline *PALB2* pathogenic variant carriers based on BC molecular subtype, tumour grade and age at diagnosis.

<i>PALB2</i>	Grade	All ages				<50				≥50			
		HRD by RAD51 test	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength	Pathogenic n(%)	Benign n(%)	ACMG/AMP strength
All subtypes	All	Yes (HRD)	18 (90.0)	12 (27.9)	3.23 (1.95-5.33)	Pathogenic supporting	14 (87.5)	4 (15.4)	5.69 (2.27-14.28)	Pathogenic moderate	4 (100.0)	8 (47.1)	2.13 (1.28-3.52)
		No (HRP)	2 (10.0)	31 (72.1)	0.14 (0.04-0.52)	Benign moderate	2 (12.5)	22 (84.6)	0.15 (0.04-0.55)	Benign moderate	0 (0.0)	9 (52.9)	-
	I	Yes (HRD)	2 (100.0)	1 (25.0)	4.00 (0.73-21.84)	Non-informative	1 (100.0)	0 (0.0)	-	-	1 (100.0)	1 (33.3)	3.00 (0.61-14.86)
		No (HRP)	0 (0.0)	3 (75.0)	-	-	0 (0.0)	1 (100.0)	-	-	0 (0.0)	2 (66.7)	-
	II	Yes (HRD)	7 (100.0)	3 (21.4)	4.67 (1.71-12.72)	Pathogenic moderate	6 (100.0)	0 (0.0)	-	-	1 (100.0)	3 (50.0)	2.00 (0.9-4.45)
		No (HRP)	0 (0.0)	11 (78.6)	-	-	0 (0.0)	8 (100.0)	-	-	0 (0.0)	3 (50.0)	-
	III	Yes (HRD)	7 (100.0)	6 (35.3)	2.83 (1.49-5.39)	Pathogenic supporting	5 (100.0)	2 (22.2)	4.50 (1.33-15.28)	Pathogenic moderate	2 (100.0)	4 (50.0)	2.00 (1.00-4.00)
		No (HRP)	0 (0.0)	11 (64.7)	-	-	0 (0.0)	7 (77.8)	-	-	0 (0.0)	4 (50.0)	-
	ALL	Yes (HRD)	4 (100.0)	3 (42.9)	2.33 (0.99-5.49)	Non-informative	3 (100.0)	2 (40.0)	2.50 (0.85-7.31)	Non-informative	1 (100.0)	1 (50.0)	2.00 (0.50-8.00)
		No (HRP)	0 (0.0)	4 (57.1)	-	-	0 (0.0)	3 (60.0)	-	-	0 (0.0)	1 (50.0)	-
TN	I	Yes (HRD)	0	0	-	-	0	0	-	-	0	0	-
		No (HRP)	0	0	-	-	0	0	-	-	0	0	-
	II	Yes (HRD)	0	0	-	-	0	0	-	-	0	0	-
		No (HRP)	0	0	-	-	0	0	-	-	0	0	-
	III	Yes (HRD)	3 (100.0)	2 (50.0)	2.00 (0.75-5.33)	Non-informative	2 (100.0)	1 (50.0)	2.00 (0.50-8.00)	Non-informative	1 (100.0)	1 (50.0)	2.00 (0.50-8.00)
		No (HRP)	0 (0.0)	2 (50.0)	-	-	0 (0.0)	1 (50.0)	-	-	0 (0.0)	1 (50.0)	-
	ALL	Yes (HRD)	14 (87.5)	6 (28.6)	3.06 (1.52-6.17)	Pathogenic supporting	11 (84.6)	1 (8.3)	10.15 (1.53-67.27)	Pathogenic moderate	3 (100.0)	5 (55.6)	1.80 (1.00-3.23)
		No (HRP)	2 (12.5)	15 (71.4)	0.18 (0.05-0.66)	Benign moderate	2 (15.4)	11 (91.7)	0.17 (0.05-0.61)	Benign moderate	0 (0.0)	4 (44.4)	-
	ER+	Yes (HRD)	14 (87.5)	6 (28.6)	3.06 (1.52-6.17)	Pathogenic supporting	11 (84.6)	1 (8.3)	10.15 (1.53-67.27)	Pathogenic moderate	3 (100.0)	5 (55.6)	1.80 (1.00-3.23)
		No (HRP)	2 (12.5)	15 (71.4)	0.18 (0.05-0.66)	Benign moderate	2 (15.4)	11 (91.7)	0.17 (0.05-0.61)	Benign moderate	0 (0.0)	4 (44.4)	-

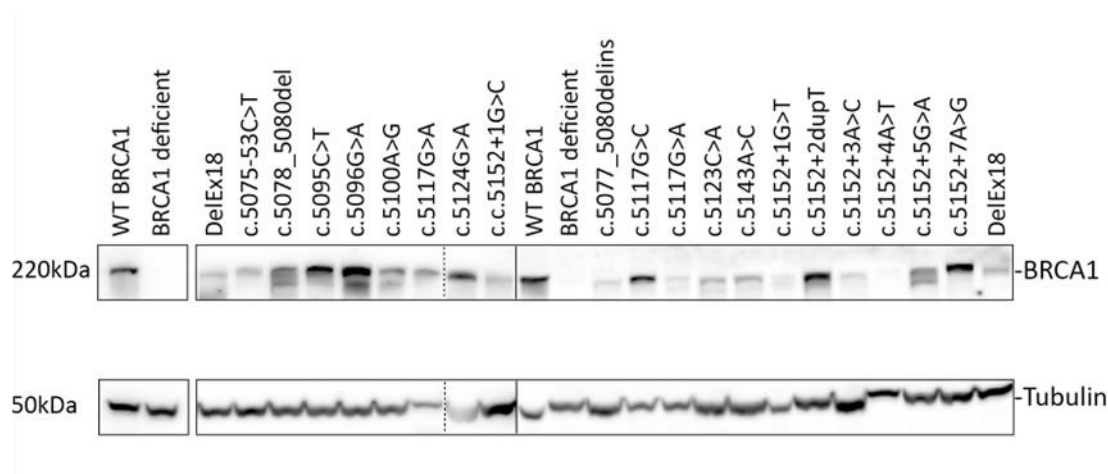
All ages															
PALB2	Grade	HRD by RAD51 test	Pathogenic		Benign n(%)	LR (95%CI)	ACMG/AMP strength	<50		≥50		Pathogenic n(%)	Benign n(%)	LR (95%CI)	ACMG/AMP strength
			n(%)	n(%)				n(%)	n(%)						
	I	Yes (HRD)	2 (100.0)	1 (50.0)	2.00 (0.50-8.00)	Non-informative		1 (100.0)	0 (0.0)	-	-	1 (100.0)	1 (100.0)	1.00 (1.00-1.00)	Non-informative
		No (HRP)	0 (0.0)	1 (50.0)	-	-		0 (0.0)	1 (100.0)	-	-	0 (0.0)	0 (0.0)	-	-
	II	Yes (HRD)	6 (100.0)	3 (27.3)	3.67 (1.40-9.62)	Pathogenic supporting		5 (100.0)	0 (0.0)	-	-	1 (100.0)	3 (60.0)	1.67 (0.81-3.41)	Non-informative
		No (HRP)	0 (0.0)	8 (72.7)	-	-		0 (0.0)	6 (100.0)	-	-	0 (0.0)	2 (40.0)	-	-
	III	Yes (HRD)	4 (100.0)	1 (16.7)	6.00 (1.00-35.91)	Non-informative		3 (100.0)	0	-	-	1 (100.0)	1 (33.3)	3.00 (0.61-14.86)	Non-informative
		No (HRP)	0 (0.0)	5 (83.3)	-	-		0 (0.0)	3 (100.0)	-	-	0 (0.0)	2 (66.7)	-	-
	All	Yes (HRD)	0	1 (25.0)	-	-		0	0 (0.0)	-	-	0	1 (50.0)	-	-
		No (HRP)	0	3 (75.0)	-	-		0	2 (100.0)	-	-	0	1 (50.0)	-	-

LR>1: association with pathogenic variant, pathogenic evidence; LR<1: no association with pathogenic variant, benign evidence. Evidence strength was measured based on Bayesian modelling of ACMG/AMP rules (**Table 8**); Benign supporting (0.23 ≤ LR < 0.48), Benign moderate (0.05 ≤ LR < 0.23), Pathogenic supporting (2.08 ≤ LR < 4.30), non-informative (0.48 ≤ LR < 2.08). LR estimates reaching informative ACMG/AMP strengths at a statistically significant CI (i.e., not spanning 1), are highlighted in bold. n: sample size, LR: Likelihood ratio; CI: Confidence Interval, ACMG/AMP: American College of Medical Genetics/Association for Molecular Pathology. TN: triple negative, ER: oestrogen receptor, PR: progesterone receptor.

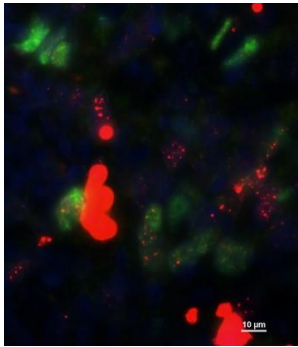
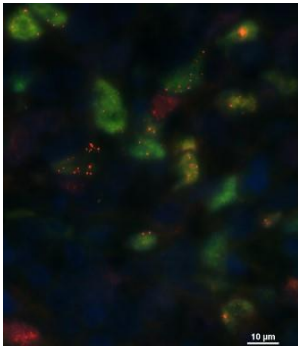
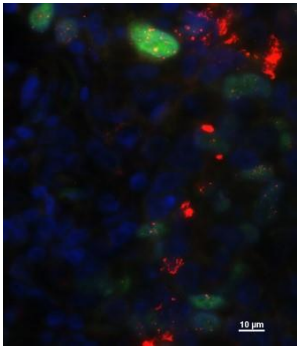
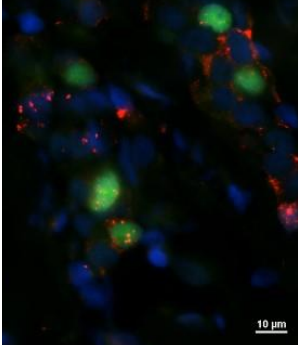
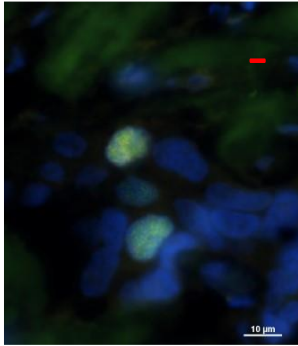
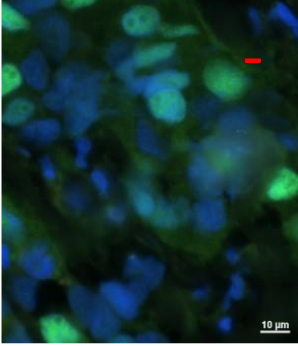


Supplementary figure 1. Correlation analysis between FL levels observed in mESC and Patients' FL per-allele fraction (A) and FL in minigenes (B).

A) Patients' per-allele FL levels (Y-axis) were calculated from FL splicing fraction obtained by RT-PCR semiquantitative analysis derived from at least two PCRs using primers amplifying the cDNA region between exon 16 to exon 22. Specifically, FL splicing fraction was calculated by dividing the peak are of each isoform by the total peak areas detected in the sample. Then per-allele FL levels were obtained as two times the level FL observed in the variant carrier minus the average level in non-carrier control sample. Fl per-allele result from carriers of the same variant were treated individually since interindividual differences were observed and different sources of RNA were used for semiquantitative splicing profiling. FL levels from BRCA1 variant-expressing mESCs (monoallelic model) (X-axis) were obtained by RT-PCR semiquantitative analysis using RNA mESC after mBrca1 depletion, by amplifying the cDNA region between exon 16 to exon 22. At least three PCRs experiments were performed. **B)** FL in minigenes came from semi-quantitative mRNA splicing by capillary electrophoresis of fluorescent amplicons. Relative expression of each isoform was estimated as explained for splicing fractions in patients. Spearman's rank correlation factor showed a positive significant correlation between mESC with both patients and minigenes.



Supplementary figure 2. Western blot analysis of protein samples from mESC expressing WT BRCA1 or BRCA1 exon 18 variants. Tubulin was used as the loading control. Dashed lines represent a marking of a different set of samples on the same blot, whereas continuous lines mark samples from distinct and separately exposed blots.

	γ H2AX/GMN/DAPI	RAD51/GMN/DAPI	BRCA1/GMN/DAPI
T126 (BC, _____) γ H2AX-82.5% RAD51-28.3% BRCA1-88.5%			
T127 (BC, _____) γ H2AX-69.2% RAD51-3.8% BRCA1-5.6%			

Supplementary figure 3. Representative immunofluorescence images of two breast tumours.

Tumours were from two different germline carriers of the c.925A>G *PALB2* variant. All pictures were taken at 600x magnification. Scale bar = 10 µm. *BC*: breast cancer, *BX*: biopsy.

APPENDIX II

ENIGMA PROJECT PROPOSAL: “Analysis of potentially spliceogenic variants affecting BRCA1 exon 18”

Project Proposal to ENIGMA Steering Committee

Proposals will be reviewed by the SC to facilitate collaboration and avoid overlap between different projects. Approved proposals will be posted on ENIGMA site for the information of all members. Information from the ENIGMA variant master list can be used to design projects, but detailed information (including variant submitters) can only be made available after a project has been submitted and approved. A report on progress is to be provided at each ENIGMA meeting (these can be presented on behalf of the Study Leader). Projects that are inactive for lengthy periods of time (> 1 year), or have very poor productivity over 3 years, may be terminated after review by the SC unless sufficient cause is demonstrated. A written report is to be provided for projects that are terminated with no resulting publication(s).

Study Leader: Orland Diez

Study Collaborators:

Sara Gutierrez-Enriquez

Miguel de la Hoya

All Enigma studies having cases with potentially spliceogenic exonic/intronic variants in or close to BRCA1 exon 18.

Submission Date: July 2015

Proposal Title: Analysis of potentially spliceogenic variants affecting BRCA1 exon 18.

Background: There are many potentially spliceogenic variants in exon 18 of the BRCA1 gene or in adjacent intronic regions. Most of them have not been thoroughly studied and remain unclassified despite appearing in databases or publications, or have been classified differently in different laboratories, as they have done or not in vitro studies, leading to conflicting lab reports.

According to different sources (databases, publications, personal communications, etc., there are many variants that could be included in the study: c.5075 A>T, -1G>A, -1G>C, -2A>C, -2A>T, -9A>T, c.5077_5079del3, c.5150T>A, c.5152+1G>A, +1G>C, +1G>T, +3A>C, +3insT, +4A>T, +5G>A, +6T>A, +6T>C, +6T>G. Some of these variants are recurrent in some countries (for example Spain: c.5152+3A> C, c.5152+5G>A) and it would be possible to collect data from different families to better establish the associated risk.

Aims: To perform a comprehensive study of potentially spliceogenic variants affecting BRCA1 exon 18.

Approach/Statistical Considerations: We propose to study the variants located in exon 18 of BRCA1 and adjacent introns by different methodologies: analysis of mRNA, specific expression

of the carrier allele, potential presence of isoforms, analysis by minigenes when necessary, tumour analysis, cosegregation analysis, co-occurrence of pathogenic variants in trans, multifactorial study, and any other kind of study considered appropriate

Funding: no

Timeline: 2015-2016

Data Fields Required: The PI will request to collaborators information of *in vitro* results, clinical /pathological data of carrier individuals, pedigrees, etc.

Outcome: data collection and required laboratory analysis to be carried out during 2015, to aim for publication soon after.

ENIGMA PROJECT PROPOSAL: “The use of a RAD51-based functional test in FFPE tumour tissue as a predictor of *BRCA1*, *BRCA2*, and *PALB2* variant pathogenicity”

Project Proposal to ENIGMA Steering Committee

Proposals will be reviewed by the SC to facilitate collaboration and avoid overlap between different projects. Approved proposals will be posted on ENIGMA site for the information of all members. Information from the ENIGMA variant master list can be used to design projects, but detailed information (including variant submitters) can only be made available after a project has been submitted and approved. A report on progress is to be provided at each ENIGMA meeting (these can be presented on behalf of the Study Leader). Projects that are inactive for lengthy periods of time (> 1 year), or have very poor productivity over 3 years, may be terminated after review by the SC unless sufficient cause is demonstrated. A written report is to be provided for projects that are terminated with no resulting publication(s).

Study Leader: Sara Gutiérrez-Enríquez and Maaïke Vreeswijk

Study Collaborators: Members of ENIGMA providing tumour samples and participating in the analysis of the tumour RAD51 predictive value for variant interpretation. ENIGMA collaborators so far (in alphabetical order) include: Judith Balmaña; Kathleen Claes, Fergus Couch, Miguel de la Hoya, Orland Diez, Joanna Domènech, Joanne Ngeow Yuen Yie; Fabienne Lesueur, Jean-Yves Masson; Alejandro Moles-Fernández; Sandro Orrù; Jacques Simard; Inge Søkilde Pedersen; Teo Soo Hwang; Melissa Southey; Mads Thomassen

Submission Date: December 2020-January 2021

Proposal Title: The use of a RAD51-based functional test in FFPE tumour tissue as a predictor of *BRCA1*, *BRCA2*, and *PALB2* variant pathogenicity.

Background: RAD51 protein plays a key role in the repair of DNA double strand breaks (DSBs) via homologous recombination (HR) and acts downstream of *BRCA1*, *PALB2* and *BRCA2*[353]. Accumulation of RAD51 at DSBs is a functional biomarker for DNA repair via homologous recombination: the presence of RAD51 foci in the nucleus is an indicator for HR proficiency whereas the absence of foci is indicative of a HR deficiency[354]. It has been demonstrated that the levels of RAD51 foci in tumour cells, without challenging cells with DNA damaging agents, predicts PARP inhibitor response in breast tumours from *BRCA1*, *BRCA2* (*BRCA1/2*) and *PALB2* patients [247,248]. Importantly, these two publications also demonstrate that the RAD51 tumour assay has high sensitivity in identification of *BRCA1/2* and *PALB2* deficient tumours and hence this assay might be useful in the classification of variants of unknown significance (VUS). Interestingly, Duran-Lozano et al.(2019) [183] recently reported the use of RAD51 tumour assay as an additional tool in the classification of a variant of unknown significance in *PALB2*. This proposal is presented to the ENIGMA members in the context of the RAD51Predict project, a

European ERAPerMed funded project headed by Violeta Serra (VHIO, Barcelona, Spain). The overarching goal of this project is to establish the prevalence of functional HR deficiency (HRD) and its predictive value for personalized treatment with platinum salts and PARPi in breast, ovarian, prostate and endometrial cancer, using the RAD51 immuno-assay on formalin-fixed paraffin-embedded (FFPE) tumour tissue and genomic assays. One of the work packages assesses the utility of the RAD51 tumour assay as a predictor of *BRCA1*, *BRCA2*, and *PALB2* variant pathogenicity. To assess the performance of the RAD51 tumour assay in variant classification we need sufficient numbers of tumours with “classic” (i.e. truncating) pathogenic variants which we have already secured in the RAD51Predict project. Additionally, we like to include tumours with *BRCA1*, *BRCA2* and *PALB2* missense variants classified as pathogenic or benign. Given the rarity of occurrence of pathogenic missense variants, a collaborative international consortium project (i.e. amid ENIGMA members) is needed to get access to sufficient numbers of relevant tumour samples.

Aims: To determine the predictive value of the RAD51 tumour assay to classify variants in *BRCA1*, *BRCA2* and *PALB2*. We specifically want to:

1. Establish the performance of the RAD51 tumour assay in primary, chemotherapy naïve tumours bearing either pathogenic or benign missense variants in *BRCA1*, *BRCA2* and *PALB2*.
2. To assess RAD51 score in primary, chemotherapy naïve tumours with VUS in *BRCA1*, *BRCA2* and *PALB2*

Approach/Statistical Considerations: Confusion matrix, Matthews correlation coefficient (MCC) and receiver operating characteristic (ROC) curve analyses.

Funding: The RAD51 tumour foci assay will be conducted either in Vall d’Hebron Institute of Oncology by Sara Gutiérrez-Enríquez’s team, supported by EraPerMed RAD51predict project and a project financed by Instituto Salud Carlos III, or by Maaïke Vreeswijk’s team supported by funding from the Dutch Cancer Society.

Timeline:

Year 1: Collect tumour samples from ENIGMA members

Year 2: Perform the RAD51 assay and foci scoring

Year 3: Data analysis and writing manuscript

Material and Data Fields Required:

- Five slides of 3 µm of formalin-fixed paraffin-embedded (FFPE) or tissue microarrays (TMA) of primary, chemotherapy naïve tumour obtained at the time of the biopsy or surgery.
- A file with key genetic variant information and data on homologous recombination testing based on DNA if available (e.g. LOH, genomic scars and mutational signatures).
- The RAD51 foci will be determined as described in Castroviejo-Bermejo et al. (2018); Cruz et al. (2018) [247,248].

Outcome:

- At this moment only very few tumour characteristics are robust enough to be included in current ACMG rules for germline variant classification in cancer predisposition genes[250,355]. Our project might provide a functional marker for the HR status in the tumour which might be integrated into the existing criteria for functional evidence (PS3 criterion) or patient phenotype (PP4 criterion) in the classification of variants following ENIGMA or ACMG rules[99].
- At least one paper which describes the performance of the RAD51 tumour assay with respect to proven pathogenic and benign missense variants, VUS evaluation and (if data permit) comparison of the RAD51 functional marker with DNA based HRD tests.

APPENDIX III

Publications resulting from collaboration in other projects during this thesis:

Herencia-Roper, A., Llop-Guevara, A., Staniszewska, A. D., **Domènech-Vivó, J.**, García-Galea, E., Moles-Fernández, A., Pedretti, F., Domènech, H., Rodríguez, O., Guzmán, M., Arenas, E. J., Verdaguer, H., Calero-Nieto, F. J., Talbot, S., Tobalina, L., Leo, E., Lau, A., Nuciforo, P., Dienstmann, R., Macarulla, T., ... Serra, V. (2024). The PARP1 selective inhibitor saruparib (AZD5305) elicits potent and durable antitumor activity in patient-derived BRCA1/2-associated cancer models. *Genome medicine*, 16(1), 107. <https://doi.org/10.1186/s13073-024-01370-z>

ACKNOWLEDGEMENTS

Després de més de 400 pàgines dedicades a classificar variants puc dir que l'etapa "Tesi" està acabant, i tot i que ha sigut un camí llarg, l'he fet molt ben acompanyada. Aquesta tesi és meua i de tots els que d'una manera o altra n'heu format part.

En primer lloc he d'agrair als meus directors, la **Sara** i l'**Orland**, l'oportunitat de poder fer la tesi sota la seva supervisió, gràcies per tot el que m'heu ensenyat i tot el suport que m'heu brindat durant aquests anys. **Sara**, gràcies per transmetre'm la teua passió per la ciència des del primer dia, per estar sempre disposada a escoltar cada una de les meues idees, per la confiança depositada en tot el que he fet i per la teua comprensió en els moments difícils. **Orland**, gràcies per tot el suport i confiança, per continuar treballant en aquesta tesi fins al final i per tots els consells al llarg dels anys. No hauria pogut tenir millors referents en aquesta etapa, i soc la científica que soc gràcies a vosaltres.

El meu pas pel laboratori d'Oncogenètica (sempre es dirà així al meu cor tot i que ara siguem el "*Hereditary Cancer Genetics Group*") es remunta al desembre del 2018. El que va començar com unes pràctiques de màster ha acabat sent una de les aventures més grans de la meua vida acadèmica. Durant aquests quasi 6 anys he compartit moltíssimes reunions, una pandèmia, quatre canvis de laboratori, congressos, sopars, dinars, scape rooms, partides de bolos, i moltes experiències més amb un equip fantàstic. Gràcies per haver fet que aquests anys siguin increïbles, els recordaré sempre!

Ester, com estic sobrevivint a arribar al laboratori i no trobar-te allí? No ho sé. Vam començar juntes aquesta etapa, i he tingut una sort enorme de poder fer tot aquest camí al teu costat. Gràcies pel teu suport incondicional, per estar sempre allí, per no saltar-te ninguna reunió de splicing i focis, per dir sempre que si a un cafè, per tots els moments compartits entre les parets de VHIO i voltant per món. Puc dir amb total certesa que ets la millor companya que tindrè mai, i tot i que ja no et vegi cada matí, la nostra amistat és per sempre. T'estimo molt Estersita!

Ale, mi maestro y mentor, ¿qué habría hecho sin ti? Gracias por enseñarme todo lo que se de *splicing*, por cuidarme tanto, por todas nuestras charlas tanto en VHIO como en los bares, por tus consejos y por estar siempre ahí. Ha sido un placer trabajar contigo, y llegar a publicar juntos el pequeño "*cancers*". Tu partida también se me hizo difícil, pero he de reconocer que el hecho de que te quedaras bien cerquita y pudiéndote llamar cada día han hecho que lo lleve mejor. Se que tengo un maestro y amigo para siempre. Te quiero mucho!

Continuant amb els *former members* del lab no puc deixar de mencionar a la **Sandra**. Moltes gràcies pels teus consells, per ensenyar-me tot el que se de cultius i CRISPR, vaig aprendre molt al teu costat! **Gemma**, tot i que vam coincidir només una setmana dins del laboratori, vas deixar un *super legado d'splicing* i sempre t'estaré agraïda. He après molt més del que creus llegint les teves llibretes i power points. M'hagués agradat compartir més ciència amb tu, però ha sigut un plaer conèixer-te fora del lab!

I com no he de mencionar a les noves incorporacions. **Laia**, la meva petita gran *student*, gràcies per posar mans a la *locura* del funcional, per ser tan perseverant, estar sempre disposada a anar més enllà i no deixar que res et pari. És difícil trobar a algú amb una vocació científica com la teva, no ho perdís per molt difícil que es posin les coses (que s'hi posaran perquè la tesi és així). Tens un futur brillant. **Sety**, thanks for being such a sunshine, for always proposing nice plans and supporting me in all my tempts to make team building! Your happiness is contagious, and all your efforts will also bring you wonderful opportunities. Em sento molt afortunada per acabar aquesta etapa treballant amb vosaltres, i també ben tranquil·la per que sé que el futur del grup està més que assegurat.

I per últim a la resta del grup, l'equip de la UARPC i del mòdul blau, **Judith, Mara, Adrià, Estela, Edu, Alejandra, Laura, Anna, Adriana, Júlia, Sara T., Maite i Berta**, gràcies per fer-me plenament conscient de que darrera de cada número de mostra hi ha una pacient i una família. Sense la vostra feina res d'aquesta tesi tindria sentit. Em sento molt afortunada per haver pogut compartir molta ciència i des de diferents punts de vista amb vosaltres. **Judith**, gràcies per la teva implicació en cada un dels projectes d'aquesta tesi, pels consells i per estar sempre disponible per qualsevol consulta. **Edu**, també mereixes una menció especial per tota la feina que t'he donat amb les sangs, i per tot el que m'has ensenyat a les consultes, gràcies per tot!

Més enllà del grup he d'agrair a la **Mercè** la seva simpatia i capacitat resolutiva, estaríem perduts sense tu! I també al **Víctor**, per estar sempre disposat a resoldre qualsevol dubte estadístic amb un somriure.

Pel nom del grup on he realitzat la tesi s'intueix que la nostra especialitat és la genètica, i per això un terç d'aquesta feina tampoc hauria sigut possible sense l'estreta col·laboració amb el grup de Teràpies Experimentals del VHIO. **Violeta i Alba**, gràcies per introduir-me al món dels *foci*, per creure en el potencial de RAD51 per classificar variants i estar sempre pendents de les

VUS, he après moltíssim treballant amb vosaltres. **Olga** i **Marta** gràcies per totes les IF i protocols. **Sara S**, **Heura** i **Andrea**, gràcies per estar sempre disposades a ajudar-me amb els casos dubtosos i les fotos, no me n'hauria sortit sense vosaltres! I deixant un moment de banda la ciència... **Andrea**, mereces una mención especial por estar siempre ahí, dispuesta a escuchar y ayudar en todo. Gracias por ser una amiga increíble, es imposible estar mal a tu lado. **Heura**, la meva “*tocaya*” de cognom, quin gran descobriment, la teva passió per la ciència es admirable i no em puc imaginar una festa sense tu donant-ho tot!

Obviously, this thesis wouldn't have been possible either without the collaborative efforts from the **ENIGMA** consortium. I am deeply grateful for the opportunity to learn from all its members and to continue growing as part of the VCEP team. Thank you for your unwavering dedication and efforts in conducting science, always with the ultimate goal of helping people.

I anant una mica més enllà del recolzament científic, durant aquests anys al VHIO he conegut persones increïbles que s'han convertit en amics. Sense vosaltres aquest camí hauria sigut tres cops més complicat i la meitat de divertit.

A la meva segona casa al VHIO, el *Upper GI cancer translational research group*, o més coneguts com Chromatins o Tian's. Gracias, esta tesis no hubiera sido posible sin vuestro apoyo. **Carmen**, la vida nos ha presentado hasta tres veces, y estoy convencida de que es porque nuestro destino era ser amigas y acompañarnos siempre. No sé cómo agradecerte todo, has sido mi referente, confidente y casa en todo momento. Gracias por todo lo que me cuidas, por tus consejos, por las risas, las cervecitas, las lloraditas y por todo lo que nos queda por pasar juntas. ¡Te quiero mucho, hermana! Y mis chiquitos... **Mariana** gracias por ser la dulzura personificada. Eres una gran investigadora y sé que llegarás tan lejos como te propongas, mucho ánimo con lo que queda de tesis, esto casi está hecho! **Joel**, **Sofia** y **Cheska**, el trio perfecto. ¿Hay alguien que se apunte más rápido a un plan, sea el que sea? No. Gracias por hacernos reír, estar siempre dispuestos a escuchar un problema o un cotilleo y por celebrar todo lo que es celebrable y más. Diversión aparte, también sois de lo más trabajador que conozco. Estoy más que convencida que haréis unas tesis increíbles, y sabéis que seguiré vuestros pasos bien de cerca. **Natalia**, ha sido un placer trabajar contigo estos últimos días, gracias por hacerme volver a pipetear! Viendo tu pasión y dedicación en cada proyecto estoy convencida de que tu tesis será todo un éxito. **Jess**, gracias per salvar-me amb els teus coneixements, he

sigut una lab manager decent gràcies als teus consells! En resum, gràcies per l'adopció, heu fet que aquesta etapa sigui meravellosa!

Iris (la meva eterna veïna), **Iñigo, Hugo, Fabio, Ari, Pablo, Andreu, María, Paula, Patri, Sara A., Almudena, Caye, Laura, Santi, Enrique, Agus, Claudia, Jon Ander, Mariona...** Heu fet que el VHIO sigui casa durant aquest casi 6 anys. M'enduc molts julitos, Beer sessions, viatges, sopars de nadal, cumple, VHIO findes, vermouths, incomptables nits al almodobar i plataforma, paddle surf traumatic experiencies, defenses de tesis amb les seves respectives comentadetes ... Us estaré sempre agraïda per haver-me acompanyat durant aquesta etapa fent que el camí sigui tremendo! Y aquí no puedo olvidarme de **Sara B. y Marc**, gracias por hacer que los primeros años fueran mucho más llevaderos, y por seguir ahí después de tomar caminos alejados de VHIO. Y **Javi**, mi motopapi de confianza que ha llegado hace poquito, ¡pero también ha contribuido en hacer que todo sea más divertido!

Tot això tampoc hauria sigut possible sense la gent que tinc més enllà del VHIO i de Barcelona.

Al 2018 vaig deixar la meva estimada Lleida, però no vaig deixar els meus amics de sempre. **Bassi i Marina**, gràcies per aquesta amistat de més de 20 anys. He crescut al vostre costat i espero envellir tenint-vos ben a prop. Sou la viva imatge de la diversió, crec que si *carpe diem* fos una cara seria la vostra. No m'imagino ni Lleida ni cap racó del món sense vosaltres! Us estimo molt.

Evi i Neus, la vida ens va presentar pels carrerons d'un poble petit de la Noguera i aquí estem més de dos dècades després! Se que durant aquest "any tesis" ens hem vist poquet, però us vull agrair tots els moments que hem compartit. Espero celebrar aquesta tesi i moltes coses més al vostre costat.

Toni, qui em diria que tindria tant a agrair a una persona que vaig conèixer fa una dècada a Mallorca? Gràcies per tots els moments compartits des de llavors. **Teresa**, el meu gran descobriment. Les nostres quedades, sopars, *apperols* i *audios* em donen la vida. T'has convertit en un suport indispensable. **Pol**, crec que mai coneixeré ningú amb una curiositat tan aguditzada com la teva. Gràcies estar sempre allí, es impossible no riure al teu costat! **Laura, Manu i Mireia**, el subgrup Lleidatà! Gràcies per tot el que hem viscut aquests anys, aquesta tesi no s'hagués fet igual sense tots els vermouths, esmorzars, bbqs i festes al vostre costat.

Joan, Èlia, Aitor, Pablo i Júlia, portem caminant junts des de batxillerat i sé que ho continuarem fent molts anys mes. Tinc molta sort de tenir amics amb els que puc comptar per sempre!

I a la meva família, gràcies per tot.

Padrina, amb tot el que l'ha vida t'ha posat per davant tu sempre saps com sortir-te'n i això es el que he après de tu, a riure de tot el que passa per que sempre passa de tot. Gràcies per ensenyar-me a ser així, per cuidar-me tant i per trucar-me cada dimarts des de que vaig marxar de Lleida. **Padrí**, estiguis on estiguis sé que estàs veient fins on he arribat i que estàs molt orgullós. Gràcies per ensenyar-me a ser pacient, i forta. Mai oblidaré tot el que has fet per nosaltres, ni les cares que posaves i com reies quan et preguntava que et semblaven les figures de la tesi mentre anava escrivint al teu costat. Ets un referent de com es pot lluitar fins al final i aquí et trobem molt a faltar. Us estimo molt.

A totes les meves **tietes, tiets i cosins**, soc molt afortunada per tenir una família com vosaltres! En especial a la **Nati**, el **Beni**, el **Sergi**, l'**Imanol** i l'**Alba**, moltes gràcies per ser casa des de que em vaig mudar a Barcelona, per acollir-me i estar sempre pendents de mi. Tinc molta sort de tenir-vos.

Mari i Eliseus, aquesta tesi no s'hagués arribat a escriure sense la vostra ajuda i suport incondicional. Gràcies per cuidar-me sempre, i en especial en aquesta etapa tan dura. No sé com podré arribar a agrair-vos tot!

Als meus pares, res d'això hauria sigut possible sense vosaltres. **Mama** ets tota una referent de bondat i amabilitat. **Papa**, el teu instint aventurer és tota una inspiració. Als dos, gràcies per creure en mi, per animar-me i fer-me costat sempre, per donar-me forces quan no les tenia, per donar-ho tot i més per a que tant la Maria com jo poguéssim arribar tant lluny com volguéssim. Us admiro i estimo moltíssim.

Maria, sempre hem estat juntes (literalment), i tu sempre m'has cuidat i protegit, fent de germana gran sense ser-ho. Tot i viure separades per 158km, et noto sempre al meu costat preparada per continuar cuidant-me i recolzant-me. Com he dit sempre, no hi hauria una Joanna sense una Maria. Gràcies per ser la millor germana, t'estimo. ¡Y **Yeray**, no me olvido de ti! Gracias por acogerme siempre con una sonrisa.

I a l'**Oscar**. Ets qui més ha patit les conseqüències d'aquesta tesi, i per això vull expressar-te el meu més sincer agraïment. Gràcies per la teva paciència i comprensió infinita, per la teva capacitat de fer-me riure en qualsevol circumstància, per ser la calma que necessitava en els moments de més pressió, i pel teu suport incondicional. Ha sigut un camí difícil, però tu no m'has deixat sola en cap moment. No sé què depara el futur, però estic segura que, amb tu al meu costat, tindrè el millor company *d'aventuritas* per viure el que sigui. T'estimo molt.

Per últim, vull agrair a totes les **pacients** que, de manera totalment altruista, han col·laborat en aquesta tesi. La vostra generositat fa possible que continuem avançant.

