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.

Miguel de la Hoya[1,*], Omar Soukarieh[2], Irene López-Perolio[1], Ana Vega[3], Logan C. Walker[4], Yvette van Ierland[5], Diana Baralle[6], Marta Santamariña[7], Vanessa Lattimore[4], Juul Wijnen[5, 8], Philip Whiley[9], Ana Blanco[3], Michela Raponi[6], Jan Hauke[10], Barbara Wappenschmidt[10], Alexandra Becker [10], Thomas v. O. Hansen[11], Raquel Behar[1], KConFaB investigators[12], Diether Niederacher[13], Norbert Arnold[14], Bernd Dworniczak[15], Doris Steinemann[16], Ulrike Faust[17], Wendy Rubinstein[18], Peter J. Hulick[19], Claude Houdayer[20], Sandrine M. Caputo[21], Laurent Castera[22], Tina Pesaran[23], Elizabeth Chao[23], Carole Brewer[24], Melissa C. Southey[25], Christi J. van Asperen[5], Christian F. Singer[26], Jan Sullivan[27], Nicola Poplawski[28, 29], Phuong Mai[30], Julian Peto[31], Nichola Johnson[32, 33], Barbara Burwinkel[34, 35], Harald Surowy[34, 35], Stig E. Bojesen[36-38], Henrik Flyger[39], Annika Lindblom[40], Sara Margolin[41], Jenny Chang-Claude[42, 43], Anja Rudolph[42], Paolo Radice[44], Laura Galastri[45], Janet E. Olson[46], Emily Hallberg[46], Graham G. Giles[47, 48], Roger L. Milne[47, 48], Irene L. Andrulis[49, 50], Gord Glendon[49], Per Hall[51], Kamila Czene[51], Fiona Blows[52], Mitul Shah[52], Qin Wang[53], Joe Dennis[53], Kyriaki Michailidou[53, 54], Lesley McGuffog[55], Manjeet K. Bolla[53], Antonis C. Antoniou[55], Douglas F. Easton[52, 53], Fergus J. Couch[46, 56], Sean Tavtigian[57], Maaike Vreeswijk[5], Michael Parsons[9], Huong Meeks[57], Alexandra Martins[2], David E. Goldgar [58, 59], Amanda B. Spurdle[9, 59, *]

[1] Molecular Oncology Laboratory, Instituto de Investigacion Sanitaria San Carlos (IdISSC), Hospital Clinico San Carlos, Madrid, Spain.

[2] Inserm U1079-IRIB, University of Rouen, Normandy Centre for Genomic and Personalized Medicine, Rouen, France.

[3] Fundacion Publica Galega de Medicina Xenómica-SERGAS Grupo de Medicina Xenómica-USC, IDIS, CIBERER, Santiago de Compostela, 15706, Spain.

[4] Department of Pathology, University of Otago, Christchurch, New Zealand.

[5] Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.

[6] Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

[7] CIBERER, Grupo de Medicina Xenómica-USC, Universidade de Santiago de Compostela, Fundacion Galega de Medicina Xenómica (SERGAS), Santiago de Compostela, Spain.

[8] Department of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands.

[9] Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, 4006, Australia.

[10] Center for Hereditary Breast and Ovarian Cancer, Center for Integrated Oncology (CIO), Medical Faculty, University Hosptial Cologne, Cologne, Germany.

[11] Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, DK-2100, Denmark.

[12] Peter MacCallum Cancer Center, The University of Melbourne, Melbourne, 3002, Australia.

[13] Department of Gynaecology and Obstetrics, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Germany.

[14] Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Campus Kiel, Christian-Albrechts University Kiel, Germany.

[15] Institute of Human Genetics, University of Münster, Münster, Germany.

[16] Institute of Human Genetics, Hannover Medical School, Hannover, Germany.

[17] Institute of Medical Genetics and Applied Genomics, Univeristy Hosptial Tuebingen, Tuebingen, Germany.

[18] National Institutes of Health, Bethesda, MD, USA.

[19] Center for Medical Genetics, NorthShore University Health System, University of Chicago Pritzker School of Medicine, 1000 Central Street, Suite 620, Evanston, IL, USA.

[20] Service de Génétique, Department de Biologie des Tumeurs, Institut Curie and INSERM U830, centre de recherche de l'Institut Curie, Paris, and Universite Paris Descartes, Sorbonne Paris Cite, Paris, France.

[21] Service de Génétique, Department de Biologie des Tumeurs, Institut Curie, Paris, France.

[22] Centre Francois Baclesse, Laboratoire de Biologie et de Genetique du Cancer, 14076, Caen, France.

[23] Ambry Genetics, 15 Argonaut, Aliso Viejo, CA 92656, USA.

[24] Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK.

[25] Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Vic, 3010, Australia.

[26] Dept of Obstetrics and Gynaecology, Medical University of Vienna, Vienna, Austria, Waehringer Guertel 18-20, A 1090 Vienna, Austria.

[27] Genetic Health Service NZ, South Island Hub, Christchurch Hospital, New Zealand.

[28] Adult Genetics Unit, South Australian Clinical Genetics Service, SA Pathology at the Women's and Children's Hospital, 72 King William Road, North Adelaide, SA 5067, Australia.

[29] University Department of Paediatrics, University of Adelaide, North Terrace, Adelaide, SA 5000, Australia.

[30] Clinical Genetics Branch,DCEG, NCI, NIH, 9609 Medical Center Drive, Room 6E-454, Bethesda, MD, USA.

[31] London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK.

[32] Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, SW3 6JB, UK.

[33] Division of Breast Cancer Research, The Institute of Cancer Research, London, SW3 6JB, UK.

[34] Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, University of Heidelberg, Heidelberg, 69120, Germany.

[35] Molecular Epidemiology Group, German Cancer Research Center, DKFZ, Heidelberg, 69120, Germany.

[36] Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen, Denmark.

[37] Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, 2730, Denmark.

[38] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 1165, Denmark.

[39] Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Denmark.

[40] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, SE-171 77, Sweden.

[41] Department of Oncology Pathology, Karolinska Institutet, Sweden.

[42] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany.

[43] University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

[44] Unit of "Molecular bases of genetic risk and genetic testing", Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy.

[45] Associazione Volontari Italiani Sangue (AVIS) comunale di Milano, Milano, Italy.

[46] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905, USA.

[47] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Vic, 3010, Australia.

[48] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Vic, 3004, Australia.

[49] Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada.

[50] Department of Molecular Genetics, University of Toronto, Canada.

[51] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE-171 77, Sweden.

[52] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK.

[53] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK.

[54] Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Cyprus.

[55] Strangeways Research Laboratory, Worts Causeway,Cambridge CB1 8RN, UK.

[56] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, 55905, USA.

[57] Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine.

[58] Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah, USA.

[59] These authors contributed equally.

[*] Corresponding authors.

Amanda Spurdle: QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, 4006, Australia; Tel +617 3362 0371; Email, amanda.spurdle@qimr.edu.au

Miguel de la Hoya: Laboratorio de Oncología Molecular, Instituto de Investigación Sanitaria San Carlos (IdISSC), Hospital Clínico San Carlos, c/Martín Lagos s/n, Madrid 28040, Spain; Tel +34 913303348; Fax +34 913303544; Email, mhoya@hotmail.com

Abstract

A recent analysis using family history weighting and co-observation classification modeling indicated that *BRCA1* c.594-2A>C (IVS9-2A>C), previously described to cause exon 10 skipping (a truncating alteration), displays characteristics inconsistent with those of a high risk pathogenic *BRCA1* variant. We used large-scale genetic and clinical resources from the ENIGMA, CIMBA and BCAC consortia to assess pathogenicity of c.594-2A>C. The combined odds for causality considering case-control, segregation, and breast tumor pathology information was 3.23×10^{-8}. Our data indicate that c.594-2A>C is always in *cis* with c.641A>G.

The spliceogenic effect of c.[594-2A>C;641A>G] was characterized using RNA analysis of human samples and splicing minigenes. As expected, c.[594-2A>C; 641A>G] caused exon 10 skipping, albeit not due to c.594-2A>C impairing the acceptor site but rather by c.641A>G modifying exon 10 splicing regulatory element(s). Multiple blood-based RNA assays indicated that the variant allele did not produce detectable levels of full-length transcripts, with a *per allele BRCA1* expression profile comprised of ≈70-80% truncating transcripts, and ≈20-30% of in-frame Δ9,10 transcripts predicted to encode a BRCA1 protein with tumor suppression function.

We confirm that *BRCA1*c.[594-2A>C;641A>G] should not be considered a high-risk pathogenic variant. Importantly, results from our detailed mRNA analysis suggest that BRCA-associated cancer risk is likely not markedly increased for individuals who carry a truncating variant in *BRCA1* exons 9 or 10, or any other *BRCA1* allele that permits 20-30% of tumor suppressor function. More generally, our findings highlight the importance of assessing naturally occurring alternative splicing for clinical evaluation of variants in disease-causing genes.

Introduction

Sequence variants that alter the highly conserved intronic dinucleotides at splice donor and acceptor sites of high-risk disease predisposition genes are often assumed to be pathogenic, due to their high likelihood to alter RNA splicing. Although such variants will almost certainly lead to disruption of normal splicing patterns, the exact nature of the resulting alternate splicing patterns cannot be reliably predicted. Indeed, a standardized classification scheme recently developed for mismatch repair gene variants through consensus across multiple international sites (1) proposes that mRNA assay and/or clinical data are necessary to upgrade dinucleotide donor and acceptor variant classification from "likely pathogenic" to "pathogenic".

The dinucleotide acceptor site variant *BRCA1* c.594-2A>C (also known as IVS9-2A>C) has recently been reported associated with clinical characteristics inconsistent with a high risk of cancer expected for a pathogenic *BRCA1* variant (2). Previous RNA analyses of carriers of *BRCA1* c.594-2A>C indicate that this variant is associated with an aberrant mRNA profile (3, 4), including production of exon 10 deleted out-of-frame transcripts. These observations indicate that the relationship between splicing aberrations and increased risk is not straightforward, and pose the question of which measures of mRNA transcript dysregulation best reflect variant pathogenicity, considering recommendations already published by the ENIGMA Splicing Working Group (5). We undertook a study to assess level of risk associated with *BRCA1* c.594-2A>C using segregation and large-scale case-control analysis, and detailed mRNA analyses correlating genotype with aberrant mRNA profiles.

Results

Genetic studies

Characteristics of *BRCA1* c.594-2A>C variant carriers identified in BCAC, CIMBA, and ENIGMA are detailed in Supplementary Tables 1 and 2.

BRCA1 c.594-2A>C (rs80358033) was identified in 7/24,605 invasive breast cancer cases and 9/25,836 controls, when including only the 11 studies with at least one observation (Supplementary Table 1). Standard case-control analysis yielded an odds ratio (OR) of 0.82 (95% CI 0.26-2.47), which was little different after adjustment for principle components (OR 0.83, 95% CI 0.41-2.24). However, some studies indicated that they had performed *BRCA1/2* mutation screening of cases and may have excluded cases with pathogenic variants. Since *BRCA1* c.594-2A>C has generally been assumed to be pathogenic on the basis of its location at a splice acceptor site, this could create a bias due to preferential exclusion of c.594-2A>C carriers cases but not controls. After exclusion of four studies that did such genetic testing, we were left with 5/20,992 cases and 6/22,332 controls that carried the c.594-2A>C variant (See Supplementary Table 1), yielding a revised OR of 0.87 (95% CI 0.26-2.86) after adjustment for principle components. The odds for causality based on carrier frequency and ages at diagnosis/interview in these cases and controls was 7.3×10^{-5} (equating to an odds against pathogenicity of 13770:1). The case-control findings demonstrate that the *BRCA1* c.594-2A>C variant is clearly not associated with a high risk of breast cancer, and is unlikely to be associated with even a moderate (~3-5-fold) risk of breast cancer. There were 15 *BRCA1* c.594-2A>C carrier individuals from 13 families identified in the CIMBA dataset through genotyping with the iCOGS array. It was confirmed with the submitting sites that none of these individuals carried another pathogenic variant in *BRCA1*, and that 8 of these families overlapped with those identified via ENIGMA while the proband for another family was also recruited into BCAC. Overall, information for segregation analysis was available for 14 probands from ENIGMA/CIMBA (Supplementary Table 1), and breast tumor pathology information for 32 cases from ENIGMA, CIMBA

or BCAC (Supplementary Tables 1 and 2). The combined odds for causality based on segregation analysis, assuming *BRCA1* age-specific risks of breast and ovarian cancer as estimated in the large study of Antoniou et al (6), was 0.10 (ranging from 0.02 to 6.85 for individual families). The breast tumor pathology features of variant carriers were not consistent with those found commonly for high-risk *BRCA1* pathogenic variant carriers. The majority of tumors were ER positive (25/32), and the odds for causality based on pathology information was 4.98×10^{-6} (200994:1 against causality).

After contacting the submitting centres and through re-investigation of original genetic test results, the *BRCA1* exonic variant rs55680408 (c.641A>G, p.Asp214Gly) was confirmed to be present in all ENIGMA/CIMBA c.594-2A>C families included in the final analysis, and another 13 c.594-2A>C carriers identified by Ambry Genetics that were excluded from analysis due to lack of relevant clinical information. Specifically, review of genetic testing data by Ambry Genetics identified a total of 20 carriers of *BRCA1* c.[594-2A>C; 641A>G] from >213,000 tests, including two siblings and a mother-daughter pair; there was clear evidence that the alleles were *in cis* from next generation sequencing reads, and neither allele was observed alone in 2636 unrelated parent exomes. Further, all carrier individuals from BCAC were shown to share the same *BRCA1* haplotype (data not shown). Based on the haplotype and genotype information, it was assumed that all *BRCA1* c.594-2A>C carriers in the BCAC dataset were also carriers of c.641A>G (p.Asp214Gly). Considering case-control, segregation and pathology information, the combined odds for causality was 3.61×10^{-11} (i.e. 2.77×10^{10}:1 *against* causality). These results indicate that individuals carrying *BRCA1*c.[594-2A>C; 641A>G] (Supplemental Figure 1A) should be counselled as *not* having a high risk of *BRCA1*-associated disease.

Splicing studies

Comprehensive characterization of BRCA1 alternative splicing landscape in c.[594-2A>C; 641 A>G] carriers by capillary electrophoresis and high throughput RNA sequencing (RNAseq).

To search for a plausible biological mechanism explaining the lack of evidence for an increased cancer risk in *BRCA1* c.[594-2A>C; 641A>G] carriers, we first performed a comprehensive characterization of the *BRCA1* alternative splicing landscape in the vicinity of exon 10.With this aim, we performed a series of capillary electrophoresis analyses on RNAs obtained from lymphoblastoid cell lines (LCLs) (see methods). We have shown previously that this approach is highly sensitive, allowing comprehensive identification, characterization, and semi-quantification of alternative splicing (4, 7). Experiments performed with two combinations of forward and reverse primers located in exons 8 and 11detected up to five different alternative splicing events both in LCLs from one c.[594-2A>C; 641 A>G] carrier (Carrier 1) and healthy controls (Figure 1), including 3 in-frame (full-length (FL), Δ9,10, and ▼10p), and two out-of-frame (Δ9 and Δ10). All but ▼10p (r.594-21_594-1*ins*) have been described previously as naturally occurring *BRCA1* alternative splicing events in control samples (7). No c.[594-2A>C; 641A>G] specific events were identified. Overall, experiments conducted in the presence of puromycin (Puro+ experiments in Figure 1 and Supplemental Figure 1) indicated that Δ9,10 splicing fraction ($\Delta 9,10^{SF}$) is similar in Carrier 1 and Controls (≈29%), $\Delta 10^{SF}$ is considerably higher (≈38% vs. ≈1%), and FL^{SF} much lower (≈31% vs. ≈66%). $\Delta 9^{SF}$ (<3%) and $▼10p^{SF}$ (<1%)were rather minor alternative splicing events in all tested samples. As expected, Puro- experiments measured higher $\Delta 9,10^{SF}$ in Carrier 1 than in Controls (Figure 1A and Supplemental Figure 1) due to a drop in $\Delta 10^{SF}$, probably reflecting nonsense mediated decay (NMD) degradation of out-of-frame Δ10 transcripts.

Overall, findings were confirmed by comparable experiments performed by other contributing centers (Supplemental Figure 2), assaying up to eight individual *variant allele* carriers identified in four unrelated families and 3 different types of samples (LCLs, Leukocytes (LEUs), and fresh whole blood); there was similar $\Delta 9,10^{SF}$ in Carriers and Controls (range 20-30% depending on specific protocols and/or sample used for experiments), and a significant increase of $\Delta 10^{SF}$ (with corresponding decrease of FL^{SF}) in Carriers. Complementary analyses performed in the subpopulation of *BRCA1*Δ11q transcripts were coincident, with similar $(\Delta 9,10+\Delta 11q)^{SF}$ in Carriers and Controls, and a significant increase of $(\Delta 10+\Delta 11q)^{SF}$ (with corresponding decrease of $\Delta 11q^{SF}$) in Carriers (Supplemental Figure 2C). Incidentally, our data supports ▼10p as a naturally occurring *BRCA1* alternative splicing event not previously reported, probably due to its very low SF. Capillary electrophoresis findings (in particular the lack of *variant allele* specific transcripts, and the detection of ▼10p in Controls) were confirmed by RNA-seq experiments (Supplemental Figure 3).

Quantitative analyses combined with alternative splicing event specific biallelic expression analyses confirms that c.[594-2A>C; 641A>G] modifies the BRCA1 alternative splicing landscape, but not the overall BRCA1 expression level.

The comprehensive analysis of the *BRCA1* alternative splicing landscape described above did not provide an obvious explanation for why c.[594-2A>C; 641A>G] carriers do not display features of a standard pathogenic *BRCA1* variant. Yet, the absence of carrier-specific transcripts prompted us to speculate that it is perhaps the actual level of naturally occurring in-frame transcripts in *variant allele* carriers that may explain the genetic findings, in particular levels of FL and Δ9,10 transcripts given that▼10p transcript levels were very low. Since capillary electrophoresis is a semi-quantitative approach, we decided to perform further analyses with quantitative PCR (qPCR) and digital PCR (dPCR) that, overall, confirmed

capillary electrophoresis findings (Figure 2). qPCR absolute quantification of individual alternative splicing events in Carrier 1 estimated for $\Delta 9,10^{SF}$ a value of 19%±0.9, in the upper-limit of Controls (ranging from 4% to 17%)(Figure 2A, left), together with an obvious reduction of FL transcripts (Figure 2A, right). In addition to Δ9,10, three other naturally occurring in-frame alternative splicing events involving exon 10 and/or nearby exon 11 have been described, namely Δ9_11, Δ11, and Δ11q (7). We used qPCR absolute quantification to estimate the SF of these alternative splicing events, detecting an increase of $(\Delta 9,10+\Delta 11q)^{SF}$ in Carrier 1 (9%±0.8) if compared with Controls (average of 7%). No differences were observed with regard to $\Delta 11^{SF}$ and $\Delta 9\_11^{SF}$ (Supplemental Figure 4). Similarly, dPCR analyses (Figure 2B) revealed a modest increase of $\Delta 9,10^{SF}$ in Carrier 1 (24%±0.9) if compared with Controls (average of 17%), together with a 50% reduction of FL^{SF} that is fully compatible with lack of FL transcripts arising from the *variant allele*.

Alternative splicing event specific reverse transcription and PCR amplification (RT-PCR) sequencing experiments (Supplemental Figure 5) performed in carriers 3 to 5 (from one Dutch family) confirmed that Δ9,10 expression is biallelic, whereas Δ10 expression is essentially monoallelic (arising from the *variant allele*). Neither qPCR absolute quantification (Supplemental Figure 4A), nor biallelic expression analysis (Supplemental Figure 5B) suggested higher overall *BRCA1*expression level in c.[594-2A>C; 641A>G] carriers. Yet, to further exclude this possibility we performed dPCR analyses of *BRCA1*Δ9,10 and FL expression relative to *BRCA2* (Supplemental Figure 6). The data indicated that Δ9,10 relative expression level is similar in LCLs from Carrier 1 and Controls, while FL expression level shows a 50% reduction, again supporting that the *variant allele* is not producing FL transcripts.

Taken together, capillary electrophoresis analyses of RT-PCR products, RNAseq, qPCR, dPCR and alternative splicing event specific sequencing experiments supported a model in which the *variant allele* does not produce novel *BRCA1* transcripts, nor increases overall *BRCA1*expression level, but rather

substitutes FL transcripts (containing exons 9 and 10) with out-of-frame Δ10 transcripts, such that the contribution of in-frame Δ9,10 transcripts to the overall expression level is *similar* or *slightly higher* (see Figure 2, Supplemental Figure 2) to that observed in wild-type (WT) alleles. Of note, according to our data the overall model is also probably true in the subset of *BRCA1*Δ11q transcripts (see Supplemental Figures 2C and 4B). According to this model, *BRCA1*Δ9 (out-of-frame) and *BRCA1* ▼10p (in-frame) contribution to the overall expression level are very low both in *variant* and *WT* alleles (see Figure 2B and 2C), and hence irrelevant to explain the lack of risk observed in *variant allele* carriers.

Splicing reporter minigene analyses reveal that c.641A>G is causing exon 10 skipping in c.[594-2A>C; 641 A>G] carriers.

We also performed minigene assay experiments to dissect the contribution of the individual variants c.594-2A>C and c.641A>G to the splicing pattern observed in *variant allele* carriers. Experiments were performed with two minigene assays (pCAS2-BRCA1-Exon10, and pB1). A schematic representation of these reporter minigenes is shown in Figure 3. pCAS2-BRCA1-Exon10 and pB1 experiments performed in HeLa cells, as well as pB1 experiments performed in breast (MCF7 and HBL100) and ovarian (IGROV-1) cell lines, revealed that both c.594-2A>C and c.641A>G impair normal exon 10 splicing, albeit with different outcomes (Figure 3). pCAS2-BRCA1-Exon10 c.594-2A>C and pB1 c.594-2A>C predominantly produced▼10p transcripts, but also a minor amount of Δ10 transcripts (Figure 3A), a finding confirming previous pSPL3-BRCA1-Exon10 experiments performed in COS-7 cells (8). By contrast, pCAS2-BRCA1-Exon10 c.641A>G and pB1 c.641A>G mostly produced Δ10 but no detectable▼10p. The finding that c.641A>G causes exon 10 skipping albeit being located outside the splice site, suggests that this variant disturbs the regulation of exon 10 splicing, probably by destroying splicing enhancer elements and/or by creating splicing silencer elements, a hypothesis supported by an *in silico* analysis based on ESRseq scores (Supplemental Figure 7A). The presence of regulatory

mechanisms underlying *BRCA1* exon 10 splicing was further supported by small interfering RNA (siRNA) experiments performed in MDA-MB231 cells showing that endogenous *BRCA1* depends on Tra2-β for exon 10 inclusion (Supplemental Figure 7B). Double mutant pCAS2-BRCA1-Exon10 c.[594-2A>C; 641 A>G] and pB1 c.[594-2A>C; 641 A>G] experiments mimicking the *variant* allele observed *in vivo* produced detectable levels of both Δ10 and ▼10p, with Δ10 being the predominant outcome in all cell lines tested (Figure 3).

Discussion

In the present study we have demonstrated that c.[594-2A>C; 641A>G] carriers (but not necessarily carriers of a potential *BRCA1* allele in which c.594-2A>C is not linked to c.641A>G) should not be considered at high-risk of developing *BRCA1*-associated cancers. The finding is remarkable, since the variant allele causes exon 10 skipping, a frame-shift alteration. In addition, we propose a plausible biological mechanism underlying the finding, the so-called *BRCA1* Δ9,10 *rescue model*, and we show the relevance of the findings for developing disease gene variant classification algorithms.

The first study addressing the spliceogenic impact of *BRCA1*c.594-2A>C demonstrated an association with exon 10 skipping (3), supporting the initial pathogenic classification by Myriad Genetics (2). Here we confirm exon 10 skipping in c.594-2A>C carriers, and we show that contrary to expectations this splicing alteration is not driven by c.594-2A>C, but rather by the linked variant c.641A>G. Further, we show that the *variant allele* does not produce full-length (FL) transcripts, nor other in-frame transcripts apart from *normal* levels of Δ9,10 and residual levels of ▼10p transcripts. These findings lead us to conclude that Δ9,10 transcripts arising from the *variant allele* confer sufficient tumor suppressor activity *in vivo* to compensate for the lack of FL transcripts. To be more precise, the combined genetic and splicing data lead us to formulate a Δ9,10 *rescue model* in which *BRCA1* alleles with an associated $\Delta 9,10^{SF}$ of ≈20%-30% (as measured in blood related samples) confer tumor suppressor haplosufficiency (Figure 4). The actual value is probably closer to 20% than to 30% (according both to qPCR and dPCR estimations in Carrier 1, and to capillary electrophoresis estimations in Carriers 2 to 8), but at any rate is very similar to that observed in control samples. The finding that Δ9,10 is a predominant alternative splicing event not only in blood derived samples but also in clinically relevant tissues such as breast and ovary (Supplemental Figure 8) is critical to support our *rescue model* for both breast and ovarian cancer. Indeed, family history of breast and/or ovarian cancer is a key criterion for genetic testing for most

participating ENIGMA and CIMBA sites, and segregation analysis modelled both breast and ovarian cancer risk, providing no indication that *BRCA1* c.594-2A>C (IVS9-2A>C) could be associated with increased ovarian cancer risk only. Further, similar to our findings reported for breast cancer, case-control data from a parallel study by the Ovarian Cancer Association Consortium does not support an association with ovarian cancer risk, with *BRCA1* c.594-2A>C identified in 2/16,121 cases and 4/26,167 controls (OCAC, unpublished data). Note that the *BRCA1* Δ9,10 *rescue model* predicts lack of breast and ovarian cancer risk not only for *BRCA1* variants causing exon 10 skipping (or exon 9 skipping), *but* to any loss-of-function mutation in exons 9 or 10 (nonsense or frame shift mutations), provided that the mutant allele produces *normal* levels of Δ9,10 transcripts (Figure 4).

Evidently, the *BRCA1* Δ9,10 *rescue model* presumes that Δ9,10 transcripts encode a protein isoform ($BRCA1^{p.Gly183\_Lys223del}$) that has tumor suppressor activity. To our knowledge, this BRCA1 isoform (lacking only 41 out of 1863 amino acid residues) has not been detected *in vivo*, nor functionally characterized *in vitro*, but tumor suppressor activity is fully compatible with structural considerations: 1) the 41 missing residues are unlikely to affect protein folding, since they are embedded in an intrinsically disordered protein region spanning amino acids 170-1649 (9); 2) $BRCA1^{Gly183\_Lys223del}$ includes all known functional domains/residues critical for tumor suppression, including the RING domain (spanning amino acids 2-103) that mediates binding to BARD1, an obligated heterodimer partner *in vivo* (10). Interestingly, $BRCA1^{p.Gly183\_Lys223del}$ lacks some residues critical for E3 ligase activity (11), a *BRCA1* function that appears to be dispensable for tumor suppression (12, 13). Yet, the most compelling argument supporting $BRCA1^{p.Gly183\_Lys223del}$ tumor suppressor activity stems from combined genetic and splicing analyses of *BRCA1* c.591C>T (rs1799965). This variant, *also* not associated with the high risk of cancer expected for a pathogenic *BRCA1* variant (current odds for causality of 8.50×10^{-16} based on segregation and pathology information, ENIGMA unpublished data), expresses mostly Δ9,10 transcripts, a significant proportion of out-of-frame Δ9 transcripts, and very few FL transcripts (14), strongly pointing to $BRCA1^{p.Gly183\_Lys223del}$ as a protein with tumor suppressor function. As far as we know, the only cancer

predisposition gene for which a similar alternative splicing rescue model has been proposed is the tumor suppressor *adenomatous poliposis coli* (*APC*) gene, albeit in this case loss of function variants in the alternatively spliced region of *APC* exon 9 are not associated with lack of risk, but with a milder phenotype, termed attenuated familial adenomatous polyposis(15).

The *BRCA1*Δ9,10 *rescue model* highlights the often neglected relevance of naturally occurring alternative splicing in the clinical arena, and has obvious implications for variant classification algorithms. The ENIGMA consortium has developed and documented criteria for the 5-tier classification of *BRCA1/2* genetic variants based on qualitative and quantitative information (http://www.enigmaconsortium.org/). According to these rules, and consistent with those proposed by InSiGHT for Mismatch Repair gene variants (1), *BRCA1/2* variants considered extremely likely to alter splicing based on position (typically IVS±1 or IVS±2) were initially all considered Class-4 (likely pathogenic) if untested for splicing alterations. However, the findings presented in this study have been pivotal to support amendment to these classification criteria, specifying need for particular caution in interpreting variants in instances where Δ9,10 (or other known naturally occurring in-frame alternative splicing events) might rescue gene functionality (see Supplemental Table 3). Hence, we also recommend caution in interpreting coding sequence variants that lead to premature termination codons in *BRCA1* exons 9 and 10. This conservative stance is consistent with recent American College of Medical Genetics (ACMG) guidelines (16), which recommend considering the presence of alternative gene transcripts, understanding which are biologically relevant, and in which tissues the products are expressed. Thus, caution should be exercised when interpreting the impact of truncating variants confined to only a subset of transcripts, given the presence of other protein isoforms.

Of note, our results have additional implications unrelated to alternative splicing. More precisely, our study suggests that *BRCA1* tumor suppressor activity tolerates a substantial reduction in expression level *in vivo*. Indeed, results shown in Figure 4 indicate that a *BRCA1* allele producing as much as ≈70-80% of

transcript encoding tumor suppressor deficient protein (as measured in blood-related samples) may not necessarily confer high-risk of developing cancer. This observation supports the conservative viewpoint of the ENIGMA consortium that, in the absence of other information, a variant can be considered pathogenic due to an effect on mRNA integrity if it only produces transcripts carrying a premature stop codon or an in-frame deletion disrupting known functional domain(s), as determined by semi-quantitative or quantitative methods.

In brief, there are several broad messages arising from the present study. Our results confirm that mRNA and genetic studies are warranted to inform the clinical significance of sequence alterations at the highly conserved intronic dinucleotides of splice donor and acceptor sites, and highlight the need to consider both variant haplotype and alternative splicing events in the design and interpretation of assays assessing the functional consequences of variants of uncertain clinical significance. We have also shown that comprehensive understanding of alternative splicing, paired with clinical genetic studies, is critical to understand the clinical consequences of complex splicing profiles observed for certain spliceogenic variants. Lastly, we provide a baseline hypothesis for future investigation and interpretation of other likely spliceogenic *BRCA1/2* variants, a hypothesis that has implications for informing standards for generic variant classification guidelines.

Materials and Methods

1. *Genotyping and Sample Sets*

We undertook screening of *BRCA1* c.594-2A>C by direct genotyping, as part of the iCOGS experiment detailed elsewhere (17, 18). This study included genotype and pathology results from breast cancer cases and controls participating in the Breast Cancer Association Consortium (BCAC; http://apps.ccge.medschl.cam.ac.uk/consortia/bcac//), and from carriers of *BRCA1* assumed pathogenic

variants participating in the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA; http://apps.ccge.medschl.cam.ac.uk/consortia/cimba//). In addition, via the Evidence-based Network for Investigating Germline Mutant Alleles (ENIGMA, http://enigmaconsortium.org/, (19)), we identified probands recruited through familial cancer clinics who were found to be positive for *BRCA1* c.594-2A>C via clinical genetic testing. All study participants were enrolled into national or regional studies under ethically-approved protocols.

Information was recorded for all variant carriers regarding cancer status, age at diagnosis/interview, breast tumor pathology (grade, and Estrogen Receptor (ER), Progesterone Receptor (PR), and Herceptin-2 (HER2) status), and also pedigree and segregation information where available. For carriers identified though CIMBA and ENIGMA, the genotype for exonic variant c.641A>G (p.Asp214Gly) was sought from the original clinical testing report.

The BCAC dataset included 53,354 breast cancer cases and 49,720 controls and documented age at diagnosis/interview from 45 studies, detailed in (17). The denominator reduced to 24,605 cases and 25,836 controls when including only invasive breast cancer cases and controls from the 11 studies with at least one observation (Supplementary Table 1). These 11 studies included only individuals of European ancestry, and four (MCBS, MBCCSG, KARBAC, OFBCR) had undergone testing for germline *BRCA1/2* pathogenic variants (4% - 100% of samples, depending on the BCAC study), including two of four studies which sampled cases on the basis of reported family history or presence of bilateral disease.

The CIMBA dataset included 11,105 female *BRCA1* pathogenic variant carriers aged ≥ 18 y from 46 studies in CIMBA recruited through cancer genetics clinics. There were 4,845 females without report of cancer, 4,713 breast cancer cases, 933 ovarian cancer cases, and 614 individuals reporting both breast and ovarian cancer.

By contact with submitters and examination of clinical information, it was established that 11 of the 15 CIMBA probands overlapped with individuals included in the ENIGMA dataset, and one of proband was

also a participant in BCAC site (Supplementary Table 2). Only non-overlapping data was included in multifactorial likelihood analysis.

2. *Statistical methods:*

We evaluated the effect of the *BRCA1* c.594-2A>C variant on breast cancer risk in BCAC, using logistic regression models with adjustment for censoring age and population structure, based on six principal components which defined any residual population sub-structure. Censoring age was defined as age at breast cancer diagnosis, or age at last interview/follow-up. Only case-control studies in which the variant was observed at least once were included in the analysis.

In order to place case-control data into the same likelihood ratio (LR) framework as the other lines of evidence used for multifactorial likelihood analysis (20, 21), we compared the likelihood of the distribution of *BRCA1* c.594-2A>C variant carriers among cases and controls under the hypothesis that the variant has the same age specific relative risks as the "average" pathogenic *BRCA1* mutation compared to that under the hypothesis that it is not associated with any increased breast cancer risk. Specifically, we used the age at diagnosis of cases and age at interview for controls together with the relative risks of breast cancer estimated from case series unselected for family history (6) to calculate the probability that each individual carrying *BRCA1* c.594-2A>C in the sample is a *BRCA1* pathogenic variant carrier given their affected status and age. Under the hypothesis that *BRCA1* c.594-2A>C is a benign variant and does not confer increased breast cancer risk, we calculated the probability of the distribution of cases and controls among *BRCA1* c.594-2A>C carriers as a simple binomial probability with p=proportion of cases in the sample. These two likelihoods were then compared to derive the appropriate LR.

Bayes scores for segregation were derived as described previously (22), and pathology LRs were applied as indicated in Spurdle et al (23). The segregation scores, pathology LRs and case-control LRs are

mutually independent and were combined to derive a combined odds for causality as described previously (20).

We used the program Phase 2.0 (24) to estimate the most likely haplotypes of the BCAC cases and controls based on 29 variants in the region within and surrounding the *BRCA1* locus, in order to examine if all c.594-2A>C variant carriers were observed on the same haplotypic background. Variants used for phasing were those submitted by ENIGMA for inclusion on the iCOGS chip design, the most common of which were rs8176258, rs1799967, rs1799950, rs4986852, and rs1799966.

3. *mRNA Analysis methods:*

3.1. Nomenclature. We use as reference sequences to describe *BRCA1* genetic variants the GenBank reference sequences U14680.1 (cDNA) and NC_000017.11 (genomic). When referring to *BRCA1* exons, we use exon numbering according to U14680.1. To characterize the *BRCA1* alternative splicing landscape in c.[594-2A>C; 641A>G] carriers (sometimes referred throughout the text as *variant allele* carriers), we performed different RNA splicing analyses at the immediate vicinity of *BRCA1* exon 10 (defined as the gene region spanning exons 8 to 11). Since our methodology do not allow analysis of complete transcripts (from 5'-end to poly(A) tail), we refer throughout the text to alternative splicing event containing transcripts, or alternative splicing events, rather than to alternative splicing transcripts or RNA isoforms (7) . For the very same reason, full-length (FL) refers throughout the text to *BRCA1* exons 9- and 10-containing transcripts (transcripts containing the exons9/10 junction defined in the GenBank reference sequenceU14680.1), and not necessarily to the complete 5711nt mRNA described in U14680.1. We have designated alternative splicing events by combining U14680.1 exon numbering with the following symbols: Δ (exon skipping), ▼ (intron retention), p (proximal, or 5'), and q (distal, or 3'). 3.2 *RNA analysis of human samples.* Up to seven contributing laboratories (sites 1 to 7) performed RNA splicing analyses with various methodologies, including fluorescent RT-PCR followed by capillary

electrophoresis, real-time quantitative PCR (qPCR), digital PCR (dPCR), Sanger sequencing, and RNAseq (see Supplemental Methods for further details). Experiments were performed in RNAs extracted from lymphoblastoid cell lines (LCLs), short-term (3-6 days) cultured Leukocytes (LEU), or fresh peripheral blood. RNAs were derived from 8 individual c.[594-2A>C; 641A>G] carriers (hereafter referred as Carriers 1 to 8) identified in four unrelated families from Australia (Carrier 1, LCL), Germany (Carrier 2, LCL), The Netherlands (Carriers 3 to 7, LEUs), and France (Carrier 8, peripheral blood), and healthy controls. We conducted several experiments designed to characterize the *BRCA1* alternative splicing landscape observed in *variant carriers*. We used as quantitative description the splicing fraction (*SF*), defined here as the contribution of individual alternative splicing events to the overall *BRCA1* expression level (expressed as a percentage). As proxies for overall expression level, we used the Σ of all peak areas detected (capillary electrophoresis), or the signal obtained with a TaqMan assay recognizing the *BRCA1* exons 23-24 junction (dPCR). The latter was selected since both *BRCA1* exons 23 and 24 are likely constitutive exons (7). Note that *SF* is a relative measure between signals arising from the same locus (in this case *BRCA1*), so that it is neither directly related to the actual expression level on individual splicing events, nor with the overall expression level from that locus. It is formally possible that increments in the *SF* of one particular alternative splicing event correlate with actual reductions in the expression level of that splicing event. For that reason, we determine the absolute expression level of individual alternative splicing events by qPCR with standard curves (see supplemental methods for further details), and we performed relative expression analyses by dPCR, using as a reference a TaqMan assay recognizing the *BRCA2* exons 26-27 junction. When indicated, we used as a positive control RNA extracted from LCLs carrying the *BRCA1* variant c.591C>T [p.= (Cys197Cys)], known to increase $\Delta 9^{SF}$ and $\Delta 9{,}10^{SF}$(14). Many experiments were performed in parallel with cultured cells treated/untreated with a nonsense mediated mRNA decay pathway (NMD) inhibitor, either Puromycin (Puro+/- experiments), or Cycloheximide (Cyclo+/- experiments). RNA from Carrier 8 was directly extracted from fresh peripheral blood. Biallelic expression was assessed by alternative splicing eventspecific RT-PCR followed by

Sanger sequencing through rs1060915 (an informative exonic SNP located at *BRCA1* exon 13), using primers and protocols previously described (14). In addition, we searched for *BRCA1* tissue specific alternative splicing landscape in clinically relevant samples by comparing RNAs extracted from healthy control fresh peripheral blood, a pool of 10 healthy breast tissues (enriched normal epithelial areas selected by a pathologist) adjacent to breast tumor samples, and commercial RNAs from healthy breast and ovarian human tissues. Experiments were performed by capillary electrophoresis of RT-PCR products, and by dPCR. Depending on the contributing laboratories, different RNA isolation and cDNA synthesis approaches were used (see Supplemental Methods for further details).

3.3. *Minigene Splicing Assays.* To dissect the contribution of the individual BRCA1 variants c.594-2A>C and c.641A>G to the splicing alteration observed in c.[594-2A>C; 641A>G] carriers, we performed splicing assays with 2 different types of reported minigenes: pCAS2-BRCA1-Exon10 and pB1 (a minigene spanning BRCA1 exons 8 to 12). See Supplemental Methods and Figure 3 for further details.

3.4. *RNA interference experiments*. To identify splicing regulatory proteins involved in *BRCA1* exon 10 splicing, we performed a series of RNA interference experiments knocking down diverse splicing regulatory factors (hnRNPA1, Tra2β, SF2/ASF, and SC35). Experiments were performed in the breast cancer cell line MDAMD231 (see Supplemental Methods for further details).

Acknowledgements

Funding

The research described was supported by 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 [PI12/00539 and PI15/00059 to MdH, PI13/02030 to AV]; the French Ministry of Higher Education and Research [to OS]; the University of Otago, Mackenzie Charitable Foundation, Maria Lupton, and .Health Research Council of New Zealand [to LW]; UK Higher Education Funding Council Senior Fellowship Scheme, the University of Southampton [to DB]; Cancer research UK [to DB, MR]; FamilienHede Nielsen Foundation fund [to TvOH]; Cancer Research-UK Senior Cancer Research Fellowship [to ACA]; National Institute of Health [CA128978 and CA11616 to FJC]; an NIH specialized program of research excellence in breast cancer to the Mayo Clinic [P50 CA116201 to FJC]; and the US Breast Cancer Research Foundation [to FJC]; translational grant from the French National Cancer Institute and Direction Générale de l'Offre des Soins (INCa-DGOS AAP/CFB/CI) and a grant from the French North-West Canceropole (CNO) [to AM]; The Cancer Council Queensland [APP1086286 to ABS]; the NHMRC Senior Research Fellowship Scheme [ID 1061779 to ABS]; NHMRC Project grant scheme [ID #1010719 to ABS].

Additional infrastructure support to consortium members is as follows:
kConFab infrastructure has been supported by funding from the National Breast Cancer Foundation, National Health and Medical Research Council, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia [to kConFab, and the kConFab Clinical Follow-up study].
The German Consortium of Hereditary Breast and Ovarian Cancer (GC-HBOC) is supported by the German Cancer Aid (grant no 109076, Rita K. Schmutzler) and by the Center for Molecular Medicine

Cologne (CMMC)

The French consortium is supported by the French National Cancer Institute.

EMBRACE is supported by Cancer Research UK Grants C1287/A10118 and C1287/A11990.

BCFR was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR.

The BBCS is funded by Cancer Research UK and Breakthrough Breast Cancer (recently merged with Breast Cancer Campaign forming Breast Cancer Now) and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN).

The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ).

The CGPS was supported by the Chief Physician Johan Boserup and LiseBoserup Fund, the Danish Medical Research Council and Herlev Hospital

KARBAC was supported financially through the regional agreement on medical training and clinical research (ALF) between Stockholm City Council and KarolinskaInstitutet, and from the Stockholm Cancer Foundation and the Swedish Cancer Society.

KARBAC was supported financially through the regional agreement on medical training and clinical research (ALF) between Stockholm City Council and KarolinskaInstitutet, and from the Stockholm Cancer Foundation and the Swedish Cancer Society.

The MARIE study was supported by the Deutsche Krebshilfee.V. [70-2892-BR I], the Hamburg Cancer Society, the German Cancer Research Center and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402].

The MARIE study was supported by the Deutsche Krebshilfee.V. [70-2892-BR I], the Hamburg Cancer Society, the German Cancer Research Center and the Federal Ministry of Education and Research

(BMBF) Germany [01KH0402].

MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment according to Italian laws in support of the Fondazione IRCCS IstitutoNazionaleTumori.

The MCBCS was supported by the NIH grant CA128978 and a Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation.

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria.

OFBCR was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR.

The pKARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer, and the Swedish Medical Research Counsel.

SEARCH was supported by grants CRUK A490/A11021, C490/A16561.

Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research

Foundation, and the Ovarian Cancer Research Fund.

CIMBA data management was supported by Cancer Research-UK grant C12292/A11174 and C1287/A10118.

BCAC is funded by Cancer Research UK [C1287/A10118, C1287/A12014] and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS).

We thank all the families and individuals that participated in this research. We thank Paul Pharoah and the Ovarian Cancer Association Consortium for providing summary information on the frequency of the *BRCA1* c.594-2A>C variant in ovarian cancer cases and controls.

We acknowledge the contributions of Georgia Chenevix-Trench to CIMBA and the kConFaB resource, and additional study-specific acknowledgements as noted below.

NZBCS: Anne Smith, Bridget Robinson, Caroline Lintott, John Pearson, Yen Phung George Wiggins and the family members for their valuable contributions.

*The Netherlands Consortium:*FransHogervorst for assistance in data collation, Dr. M. Olderode-Berends from the University Medical Centre Groningen and Dr. E. van Riel from University Medical Centre Utrecht for providing blood samples for RNA studies, and Elsa Bik from the Leiden University Medical Centre, The Netherlands for excellent technical assistance.

kConFab: Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study for their contributions to this resource.

GC-HBOC: Dieter Schäfer, Center Frankfurt, for providing DNA samples and JulianeKöhler for excellent technical assistance.

Northshore: Tina Selkirk for assistance in data collation.

French Consortium: the French oncogeneticists, and the UNICANCER Genetic Group (UGG) lead by

Dr Catherine Noguès.

BCFR-AU: Maggie Angelakos, Judi Maskiell, Gillian Dite, Helen Tsimiklis.

MUV: Daniela Muhr for assistance in data collation.

BBCS: Eileen Williams, Elaine Ryder-Mills, Kara Sargus

BSUCH: Peter Bugert, Medical Faculty Mannheim

CGPS: Staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: DortheUldall Andersen, Maria BirnaArnadottir, Anne Bank, DortheKjeldgård Hansen. The Danish Breast Cancer Group (DBCG) is acknowledged for the tumor information. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases.

MARIE: Alina Vrieling, Katharina Buck, MuhabbetCelik, Ursula Eilber and Sabine Behrens.

MBCSG: SiranoushManoukian, Bernard Peissel, Jacopo Azzollini and Fernando Ravagnani of the Fondazione IRCCS IstitutoNazionaleTumori (INT), Milan, Italy; BernandoBonanni, Monica Barile and Irene Feroce of the IstitutoEuropeo di Oncologia (IEO), Milan, Italy; and the personnel of the Cogentech Cancer Genetic Test Laboratory, Milan, Italy.

OFBCR: Teresa Selander, NayanaWeerasooriya

SEARCH: Marie Mack

COGS: Paul Pharoah, Andrew Berchuck (OCAC), Georgia Chenevix-Trench, Ken Offit (CIMBA), Alison M. Dunning, Andrew Lee,Ed Dicks, Craig Luccarini, the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira, the staff of the CNIO genotyping unit, Jacques Simard, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière, Frederic Robidoux, the staff of the McGill University and Génome Québec Innovation Centre, Sune F. Nielsen, Borge G. Nordestgaard, the staff of the Copenhagen DNA laboratory, Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer, the staff of Mayo Clinic Genotyping Core Facility.

AM and INSERM: Dr Sophie Krieger for contributing with DNA samples, Prof Thierry Frébourg for providing patient's samples for RNA analysis, and AurélieDrouet and Gaia Castelain for technical

assistance.

Conflicts of Interest

Tina Pesaran and Elizabeth Chao are paid employees of Ambry Genetics.

References

1 Thompson, B.A., Spurdle, A.B., Plazzer, J.P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capella, G., den Dunnen, J.T. *et al.* (2014) Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.*, **46**, 107-115.

2 Rosenthal, E.T., Bowles, K.R., Pruss, D., van Kan, A., Vail, P.J., McElroy, H. and Wenstrup, R.J. (2015) Exceptions to the rule: Case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes. *Clin. Genet.*, **88**, 533-541.

3 Tesoriero, A.A., Wong, E.M., Jenkins, M.A., Hopper, J.L., Brown, M.A., Chenevix-Trench, G., Spurdle, A.B., Southey, M.C. and kConFab. (2005) Molecular characterization and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple-case breast cancer families. *Hum. Mutat.*, **26**, 495.

4 Whiley, 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.* (2014) Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin. Chem.*, **60**, 341-352.

5 Walker, L.C., Whiley, P.J., Houdayer, C., Hansen, T.V., Vega, A., Santamarina, M., Blanco, A., Fachal, L., Southey, M.C., Lafferty, A. *et al.* (2013) 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.*, **34**, 1424-1431.

6 Antoniou, A., Pharoah, P.D., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A. *et al.* (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.*, **72**, 1117-1130.

7 Colombo, M., Blok, M.J., Whiley, P., Santamarina, M., Gutierrez-Enriquez, S., Romero, A., Garre, P., Becker, A., Smith, L.D., De Vecchi, G. *et al.* (2014) Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum. Mol. Genet.*, **23**, 3666-3680.

8 Steffensen, A.Y., Dandanell, M., Jonson, L., Ejlertsen, B., Gerdes, A.M., Nielsen, F.C. and Hansen, T. (2014) Functional characterization of BRCA1 gene variants by mini-gene splicing assay. *Eur. J. Hum. Genet.*, **22**, 1362-1368.

9 Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabartty, A. and Arrowsmith, C.H. (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? *J. Mol. Biol.*, **345**, 275-287.

10 Clark, S.L., Rodriguez, A.M., Snyder, R.R., Hankins, G.D. and Boehning, D. (2012) Structure-Function Of The Tumor Suppressor BRCA1. *Comput. Struct. Biotechnol. J.*, **1**, pii: e201204005.

11 Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J. and Fields, S. (2015) Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics*, **200**, 413-422.

12 Shakya, R., Reid, L.J., Reczek, C.R., Cole, F., Egli, D., Lin, C.S., deRooij, D.G., Hirsch, S., Ravi, K., Hicks, J.B. *et al.* (2011) BRCA1 tumor suppression depends on BRCT phosphoprotein binding, but not its E3 ligase activity. *Science*, **334**, 525-528.

13 Savage, K.I. and Harkin, D.P. (2015) BRCA1, a 'complex' protein involved in the maintenance of genomic stability. *FEBS J.*, **282**, 630-646.

14 Dosil, V., Tosar, A., Canadas, C., Perez-Segura, P., Diaz-Rubio, E., Caldes, T. and de la Hoya, M. (2010) Alternative splicing and molecular characterization of splice site variants: BRCA1 c.591C>T as a case study. *Clin. Chem.*, **56**, 53-61.
15 Nieuwenhuis, M.H. and Vasen, H.F. (2007) Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. *Critical reviews in oncology/hematology*, **61**, 153-161.
16 Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.*, **17**, 405-424.
17 Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K. *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, **45**, 353-361, 361e351-352.
18 Couch, F.J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K.B., Soucy, P., Fredericksen, Z., Barrowdale, D., Dennis, J. *et al.* (2013) Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet.*, **9**, e1003212.
19 Spurdle, A.B., Healey, S., Devereau, A., Hogervorst, F.B., Monteiro, A.N., Nathanson, K.L., Radice, P., Stoppa-Lyonnet, D., Tavtigian, S., Wappenschmidt, B. *et al.* (2012) 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.*, **33**, 2-7.
20 Goldgar, D.E., Easton, D.F., Deffenbaugh, A.M., Monteiro, A.N., Tavtigian, S.V., Couch, F.J. and Breast Cancer Information Core Steering, C. (2004) Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am. J. Hum. Genet.*, **75**, 535-544.
21 Goldgar, D.E., Easton, D.F., Byrnes, G.B., Spurdle, A.B., Iversen, E.S., Greenblatt, M.S. and Group, I.U.G.V.W. (2008) Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum. Mutat.*, **29**, 1265-1272.
22 Thompson, D., Easton, D.F. and Goldgar, D.E. (2003) A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am. J. Hum. Genet.*, **73**, 652-655.
23 Spurdle, A.B., Couch, F.J., Parsons, M.T., McGuffog, L., Barrowdale, D., Bolla, M.K., Wang, Q., Healey, S., Schmutzler, R., Wappenschmidt, B. *et al.* (2014) 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.*, **16**, 3419.
24 Delaneau, O., Marchini, J. and Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179-181.

Legends to Figures

Figure 1. Capillary Electrophoresis analyses of *BRCA1*alternative splicing landscape in LCLs from one *BRCA1*c.[594-2A>C; 641A>G] carrier (Carrier 1) and 10 Controls. Panel A shows representative examples of capillary electrophoresis analysis of RT-PCR products generated with the E8.1-E11p assay in LCLs treated (Puro+) or untreated (Puro-) with the nonsense mediated decay inhibitor puromycin. The fluorescence intensity of each peak (Y-axis) is expressed in arbritary units (AU). The analyses detected the full-length transcript (FL), and up to four alternative splicing events, two in-frame (Δ9,10 and ▼10p) and two out-of-frame (Δ9, and Δ10). In these particular examples, ▼10p transcripts are detected only in Carrier 1, but we have detected ▼10p transcripts in Controls, as summarized in panel B. The presence of ▼10p in Controls has been further confirmedby RNAseq (see Supp. Figure 3).The boxplots in **Panel B** (displaying low, Q1, median, Q3, and high values) show the splicing fraction (SF) of in-frame transcripts (Δ9,10, FL and ▼10p) observed in Carrier 1 (3 technical replicas) and 10 Controls. SF expressed as the % of the corresponding peak area to the Σ of all five peak areas detected by capillary electrophoresis. This particular experiment was performed with the E8.2-E11q.2 assay. Note that the $▼10p^{SF}$ is rather minor (<1%) regardless of the LCL tested. The FL^{SF} was much lower in Carrier 1 than in Control samples. The boxplots in **Panel C** (displaying low, Q1, median, Q3, and high values) show the SF of out-of-frame transcripts (Δ9 and Δ10) observed in Carrier 1 (3 technical replicates) and 10 Controls. The relative contribution of Δ10 to the overall signal was much higher in Carrier 1 than in Control samples. Normal outliers (>1.5 interquartile range, IQR) display small circles. (** represents P≤0.01) (*** represents P≤0.001) (ns=non-significant).

Figure 2. Quantification of major in-frame transcripts Δ9,10 and full-length (FL) in LCLs from one *BRCA1*c.[594-2A>C; 641A>G] carrier (Carrier 1) and Controls. Experiments were performed in LCLs treated with Puromycin (Puro+). **Panel A** displays $\Delta 9,10^{SF}$ and FL^{SF}, estimated as the ratio between the GADPH normalized absolute numbers of Δ9,10 (or FL) molecules and absolute number of *all BRCA1* transcripts, as determined by qPCR analysis performed with standard curves (see Supplemental Methods and Supplemental Figure 4). Standard deviation of 3 independent measures is shown. **Panel B** displays dPCR data measuring $\Delta 9,10^{SF}$ and FL (inclusion of exons 9 and 10)^{SF}, using exon23-24 junction as a proxy for overall *BRCA1* expression level. The precision of each measure (as determined by the QuantStudio 3D Analysis Cloud Software) is indicated. Two technical replicates of Carrier 1 are shown. We included as positive control a LCL carrying the *BRCA1* c.591C>T variant, known to increase $\Delta 9,10^{SF}$. The $\Delta 9,10^{SF}$ in Carrier 1 was higher than in Controls (24% in two technical replicates of Carrier 1 vs. an average of 17% in 7 control samples, Mann-Whitney U test; p=0.028 for difference between groups), but a 50% reduction of FL^{SF} (50% in two technical replicas of Carrier 1 vs. an average of 94% in 6 control samples, Mann-Whitney U test; p=0.036 for difference between groups).

Figure 3. Analysis of *BRCA1* c.594-2A>C and c.641A>G variants with splicing reporter minigene assays. The figure shows schematic non-scale representations of the splicing reporter minigenes pCAS2-BRCA1-exon10 (panel A) and pB1 (panel B) used for splicing assays. Minigenes were constructed as described under Supplemental Methods. PCMV indicates the cytomegalovirus promoter, boxes represent exons and lines in between indicate introns. *BRCA1* sequences are highlighted in black. Arrows represent primers used in RT-PCR reactions. With the exception of pB1 *BRCA1* intron 11 (402 nt-long full-length IVS11), minigenes harbor partial segments of *BRCA1* introns. For comparative purposes, the size in nucleotides of each segment is shown together with the size corresponding to the endogenous full-length *BRCA1* introns shown in brackets. As indicated, pB1 carries an additional cytosine (+3insC) in exon 8 to keep the ORF with α-globin exon 1 (Raponi et al., 2012). Splicing assays were performed by analyzing the splicing pattern of WT and mutant minigenes (c.594-2A>C, c.641A>G, and c.[594-2A>C; 641A>G]) transiently expressed in human cells (HeLa, COS-7, MCF7, HBL100 or IGROV-1) as described under Supplemental Methods. The images show RT-PCR products separated in ethidium bromide-stained agarose gels. FL, full-length; Δ9, exon 9 skipping; Δ10, exon 10 skipping; Δ9,10, skipping of both exons 9 and 10; *, retention of 21 intronic nucleotides immediately upstream exon 10 (▼10p). One can note that: (i) the relative level of alternatively spliced pB1(WT) transcripts is higher in IGROV-1 than in HeLa, MCF-7, or HBL100 cells, and (ii) the predominant alternative splicing event of pB1(WT) in these cell lines is Δ10, whereas that of endogenous wild-type *BRCA1* in blood related samples is Δ9,10 (Figure 4 and Supplemental Figures 1 and 2).

Figure 4. Combined genetic and splicing analyses of *BRCA1* c.[594-2A>C; 641A>G] and *BRCA1* c.591C>T supports a *BRCA1*Δ9,10 rescue model with far-reaching clinical implications. Panel A (top) shows the splicing fraction (SF) of five alternative splicing events detected by capillary electrophoresis analysis of RT-PCR products generated with the E8.2-E11q.2 assay (Puro+ experiments, 36 cycle PCRs, see Figure 1 and Supplemental Figure 1 for further details). As shown, this description of the *BRCA1* alternative splicing landscape in the vicinity of exon 10 is different in healthy control samples, c.[594-2A>C; 641A>G] carriers, and c.591C>T carriers. Yet, we show in the present study that none of these 3 *BRCA1* splicing landscapes is associated with high risk of developing *BRCA1* related cancers. The chart displays SFs that, in carriers, represent a combined signal from the variant allele and the accompanying WT allele. **Panel A (bottom).** Deduced *per allele* SFs are shown. Assuming that SFs arising from the accompanying WT allele equal to the average SFs observed in 10 Control samples (as shown in the central chart bar), we deconvoluted the SFs corresponding to c.[594-2G; c.641G] (left chart bar) and c.591T (right chart bar) alleles. **Panel B.** The cartoon represents the relative *per allele* (100% equals to the overall expression level arising from one individual allele) and *per cell* (100% equals to the overall expression arising from a diploid genome) expression (BRCA1 exons 7 to 11) in a c.[594-2G; c.641G] carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL and Δ9,10 transcripts are shown, albeit Δ9 and ▼10p transcripts account for ≈5% of the *per cell* expression. Truncating (out-of-frame) events are highlighted with a red cross. The analysis suggests that expressing up to ≈35% of *BRCA1* PTC-NMD transcripts (*per* diploid genome) is not associated with high-risk of developing cancer. The analysis suggests as well that a *BRCA1* allele expressing up to ≈70% (*per allele*) *BRCA1* PTC-NMD transcripts is not associated with high-risk of developing cancer (a relevant finding in the context of the two-hit model). **Panel C.** The cartoon represents the relative *per allele* (100% equals to the overall expression level arising from one individual allele) and *per cell* (100% equals to the overall expression arising from a diploid genome) expression (*BRCA1* exons 7 to 11) in a c.591C>T carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL, Δ9,10 and Δ9 (variant allele) are shown, albeit Δ9 (wt allele), Δ10 (wt and variant allele), and ▼10p (wt and variant allele) transcripts account for ≈5% of the *per cell* expression. The data strongly suggests that *BRCA1*Δ9,10 transcripts, representing up to 51% (per diploid genome) and up to 71% (*per* allele) of the overall *BRCA1* expression code for a BRCA1 protein with tumor suppressor activity. The model displayed in this figure is intended to illustrate the most relevant findings of our study. Yet, some limitations should be highlighted. First, the model assumes (based on 36-cycle PCR capillary EP data) that $\Delta 9,10^{SF}$ in Controls and c.[594-2A>C; 641A>G] carriers is ≈29%, while other experiments suggests that the actual value is probably lower in both instances (Figure 2, Supplemental Figure 2), albeit slightly increased in Carriers

vs. Controls. The model has been elaborated with data obtained in LCLs, not in clinically relevant tissues such as breast or ovarian.

Abbreviations

Cyclo- Cycloheximide absent

Cyclo+ Cycoeximide present

dPCR digital PCR

qPCR quantitative PCR

FL full-length

LCL lymphoblastoid cell line

LEU leukocyte

NMD nonsense mediated decay

PTC premature termination codon

Puro- Puromycin absent

Puro+ Puromycin present

qPCR quantitative PCR

RNAseq high-throughput RNA sequencing

RT reverse transcription

SF splicing fraction

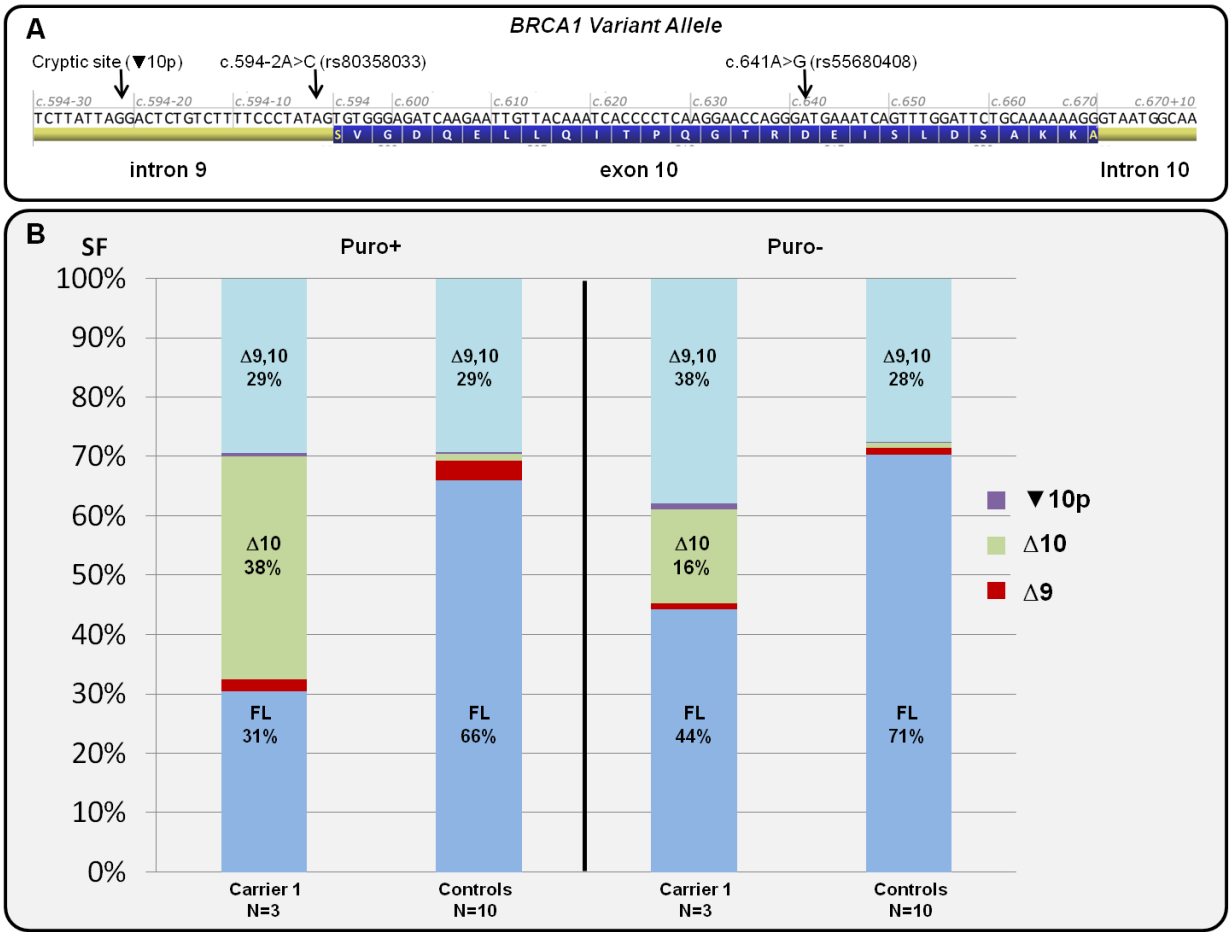
siRNA small interference RNA

WT wildtype

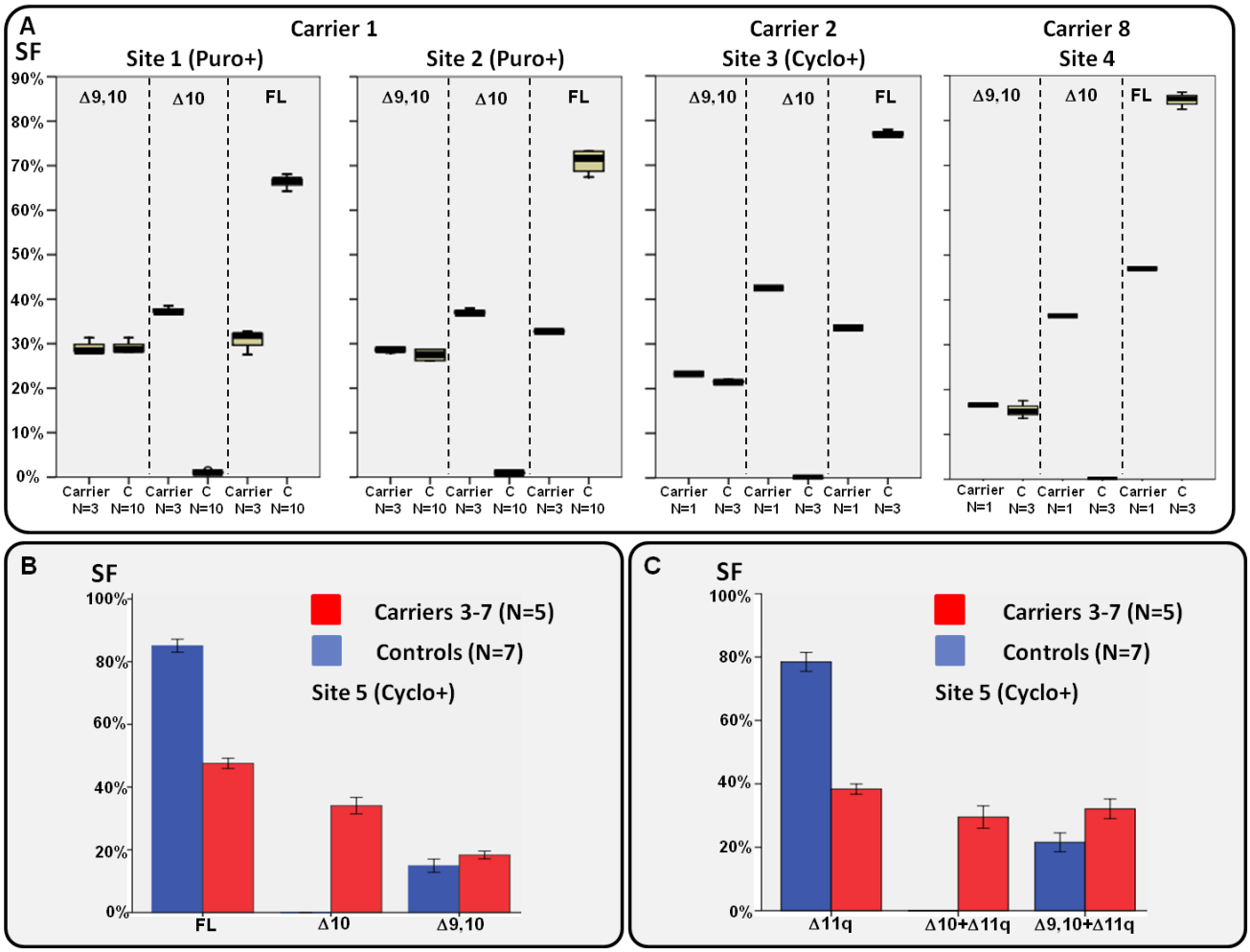
Supplemental Data

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.

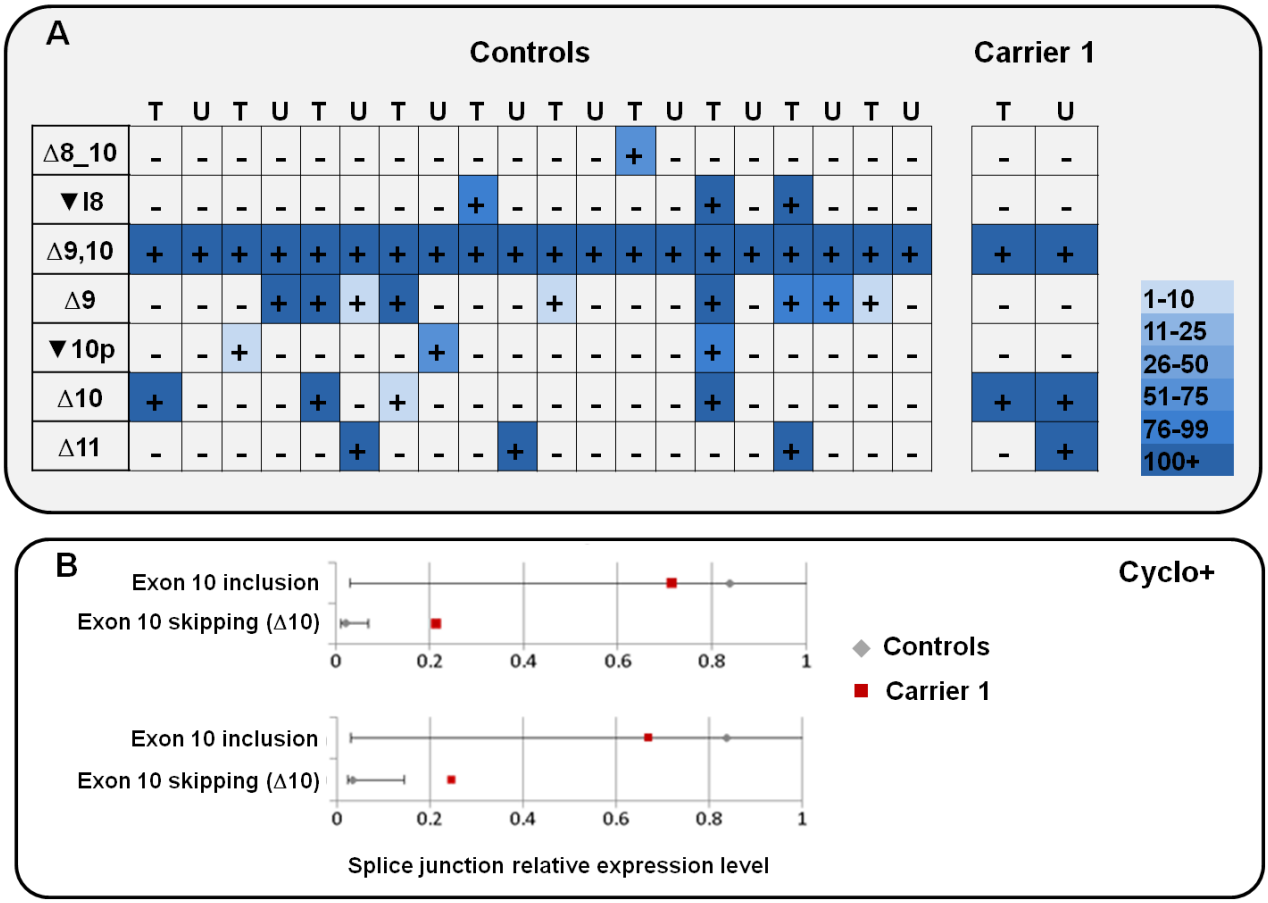
de la Hoya et al.



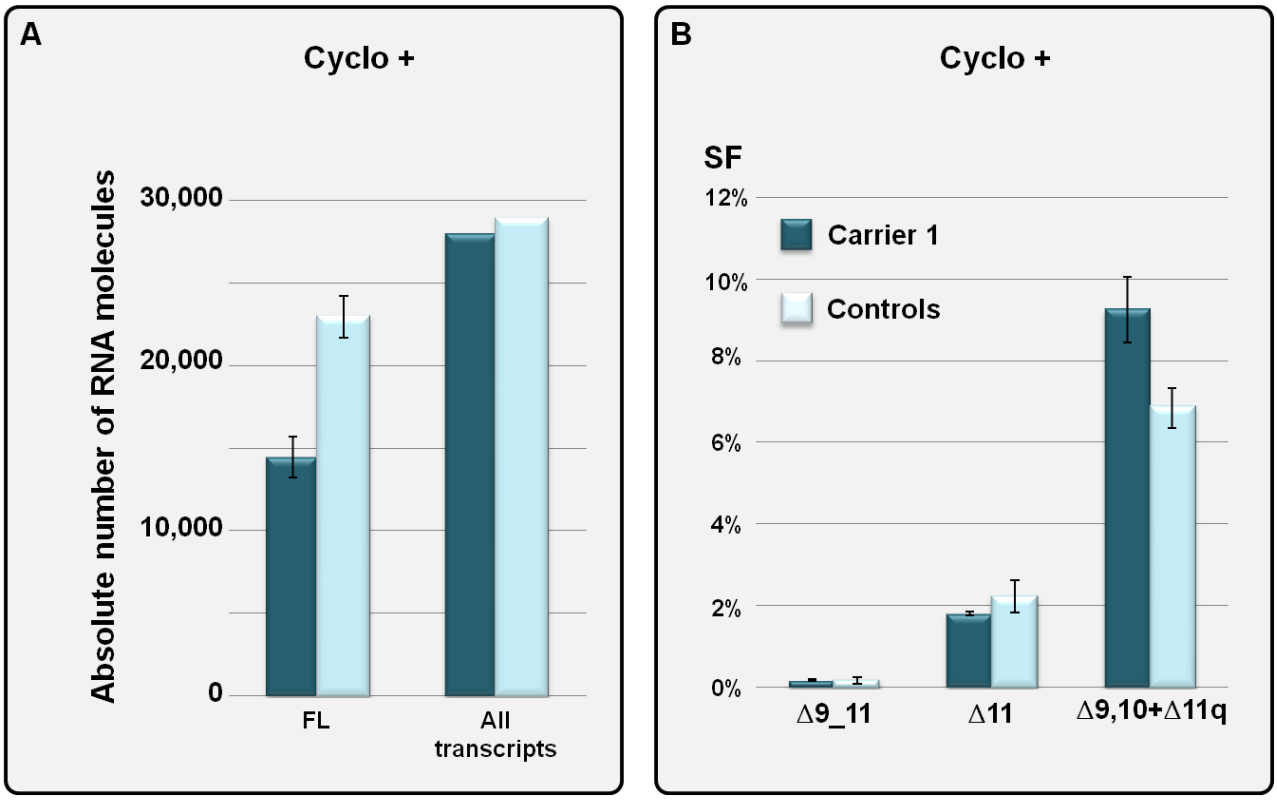
Supplemental Figure 1. *BRCA1* alternative splicing landscape in LCLs from one *BRCA1* c.[594-2A>C; 641A>G] carrier (Carrier 1) and 10 Controls. Panel A shows a schematic representation of the c.[594-2A>C; 641A>G] *variant allele* under investigation. **Panel B** Capillary electrophoresis analysis of RT-PCR products generated with the E8.2-E11q.2 assay (36 PCR cycles) detects up to five different alternative splicing events both in c.[594-2A>C; 641A>G] carriers and Controls, including 3 in-frame (Δ9,10, Full-length (FL) and ▼10p), and 2 out-of-frame (Δ9 and Δ10) events (see Figure 1). The charts represent the splicing fraction (SF) of these five splicing events (expressed as the % of the corresponding peak area to the Σ of all five peak areas) observed in Carrier 1 (average of 3 technical replicas) and Controls (average of 10 different samples). The analysis shows that increased $\Delta 10^{SF}$ (and corresponding decrease of FL^{SF}) is the predominant effect observed in carriers. The data is compatible as well with ▼10p being up-regulated in carriers (as predicted by splicing reporter minigene experiments, Figure 3), but splicing fractions are too low both in Carrier 1 (0.6% in Puro+ and 1% in Puro- experiments) and controls (0.3% in Puro+ and 0,2% in Puro- experiments) to be reliable measured by capillary electrophoresis. Note that SFs observed in experiments performed with or without Puromycin are different, suggesting nonsense-mediate decay (NMD) of out-of-frame Δ9 and Δ10 transcripts in Puro- experiments. In the case of Controls, differences are subtle, since $\Delta 9^{SF}$ (3,4% in Puro+ and 1,2% in Puro- experiments) and $\Delta 10^{SF}$ (1,1% in Puro+ and 0,8% in Puro- experiments) are rather minor contributors to the overall expression level. However, differences in Carrier 1 are significant, due to the fact that Δ10 is a predominant contributor to the overall expression level in carriers ($\Delta 10^{SF}$ of 38% in Puro+ experiments and 16% in Puro- experiments). The experiments summarized in this figure, as well as most experiments reported in the present study, restrict the characterization of *BRCA1* alternative splicing landscape to the vicinity of *BRCA1* exon 10 i.e. we have not characterized/isolated complete transcripts. For this reason, we refer throughout the text to alternative splicing events rather than to alternative splicing transcripts.



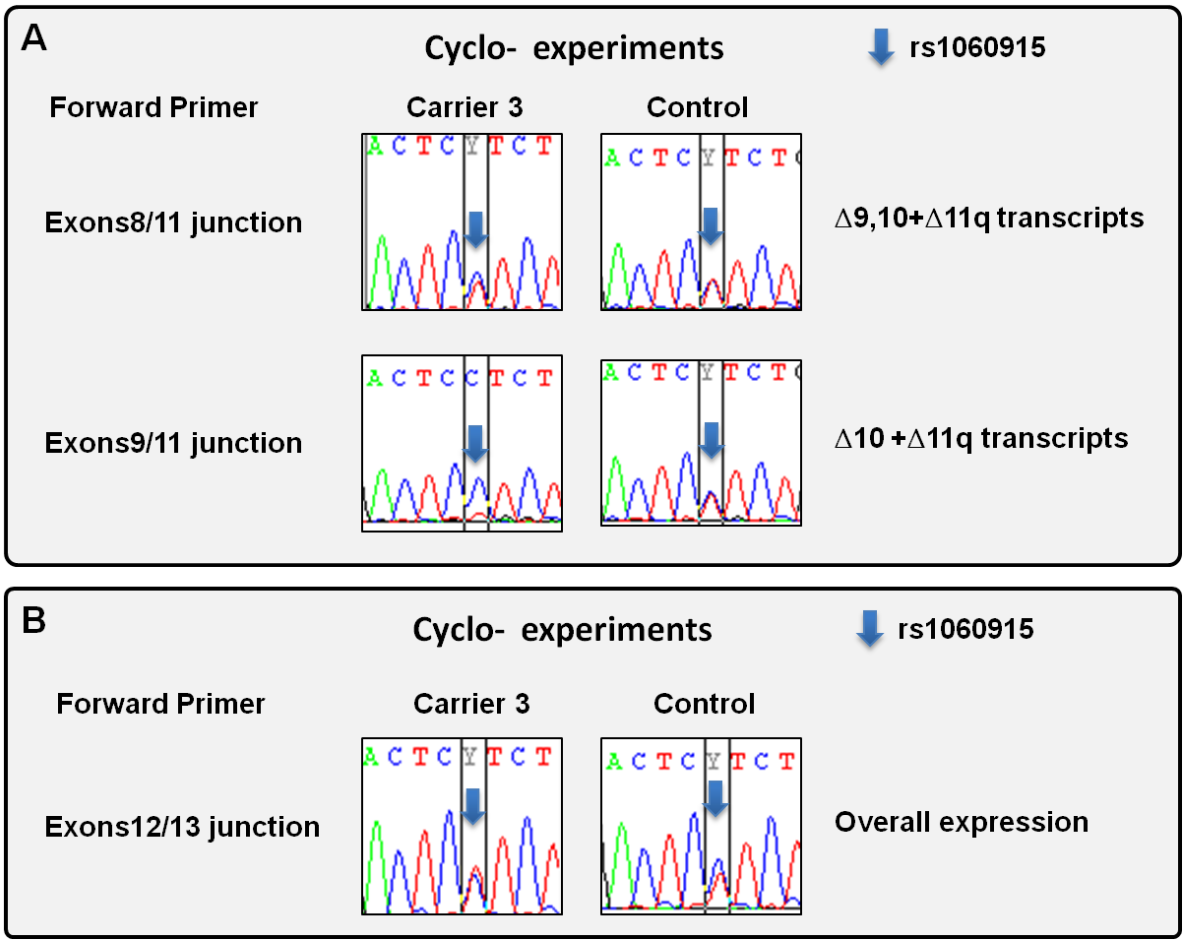
Supplemental Figure 2. Δ9,10 (and Δ9,10+Δ11q), Δ10 (and Δ10+Δ11q), FL and Δ11q *SF* determined by capillary electrophoresis in up to 8 independent c.[594-2A>C;641A>G] carriers. The boxplots in **Panel A** (displaying low, Q1, median, Q3, and high values) summarize experiments performed by four contributing laboratories (sites 1 to 4) in two independent LCL carriers named Carrier 1, and Carrier 2 , one fresh blood sample (PAXgene system for blood sampling, no puromycin/cycloheximide treatment) named Carrier 8, and its corresponding control. Note the high level of Δ10 transcripts in Carrier 8, despite the fact that NMD is not inhibited by the PAXgene system (C) samples. In the case of carriers, N represents the number of technical replicates. Site 1 data (the same data summarized in Supplemental Figure 1) was generated with the E8.2-E11q.2 RT-PCR assay (36 PCR cycles). Site 2 data was generated with the E8.1-E11q1 RT-PCR assay (36 PCR cycles). Site 3 data was generated with the E7-E11q.1 RT-PCR assay (40 PCR cycles). Finally, Site 4 data was generated with the E8.2-E11q.2 RT-PCR assay (30 PCR cycles). All RT-PCR assays are described in Supplemental Methods Table 1. **Panel B** summarizes experiments performed by site 5 in five independent leukocytes (LEU) carriers (Carriers 3 to 7) and seven controls. Data was generated with the E8.1-E11q1 RT-PCR assay (28 PCR cycles). **Panel C** summarizes experiments identical to those described in Panel B, except that the reverse primer mapped to exon 12 (E8.1-E12 RT-PCR assay). Note that in this case, the *SF* is not calculated relative to the "overall *BRCA1* expression level", but to the "overall *BRCA1* Δ11q expression level". Since *BRCA1* exon11q (3309nt) is too long to be PCR amplified with standard protocols for short amplicons, this primer combination allowed us to analyse alternative splicing in the subpopulation of *BRCA1* transcripts lacking exon 11q. Standard deviations are shown. When indicated, LCLs were treated with Puromycin (Puro+), or Cycloheximide (Cyclo+), prior to RNA extraction.



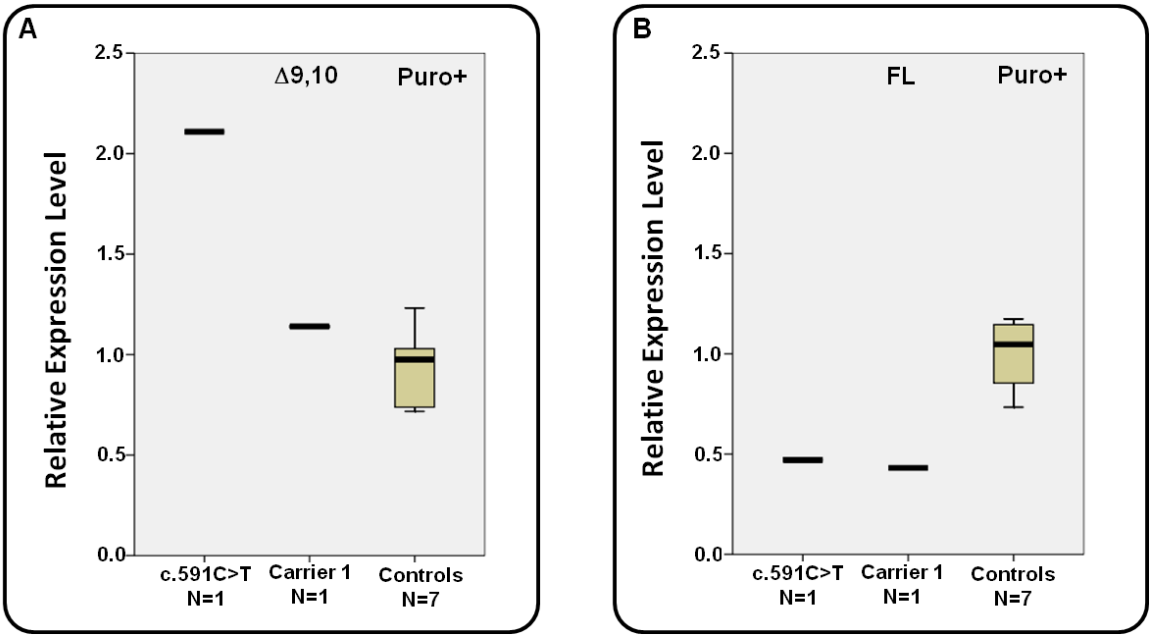
Supplemental Figure 3. RNAseq data in one *BRCA1* c.[594-2A>C; 641A>G] carrier (Carrier 1) and Controls. Expression data were generated using the TruSeq Targeted RNA Expression platform (Illumina) and RNA from 11 LCLs (10 Controls and Carrier 1). Splice junction reads defining 47 different BRCA1 alternative splicing events were identified in at least one sample of the analyzed cohort. For simplicity, only *BRCA1* splicing events involving exons 7 to 12 are shown in the heatmap **(Panel A)**. The color key indicates the number of reads at exon-exon junctions defining each splicing event. Cyclo+ (T) and Cyclo – (U) experiments were performed. Experiments show the presence of ▼10p transcripts in Control samples, as well as absence of c.[594-2A>C; 641A>G] specific splice junctions, thus confirming capillary electrophoresis findings. **Panel B** shows the relative proportion of sequence reads for *BRCA1* Δ10 (exons 9/11 junction) and exon 10 inclusion (exons 10/11 junction) relative to *BRCA1* exons2/3 junction (upper chart), or *BRCA2* exons 22/23 junction (bottom chart). Mean value and standard deviation for 10 Controls is indicated in grey. Only Cyclo+ experiments are shown.



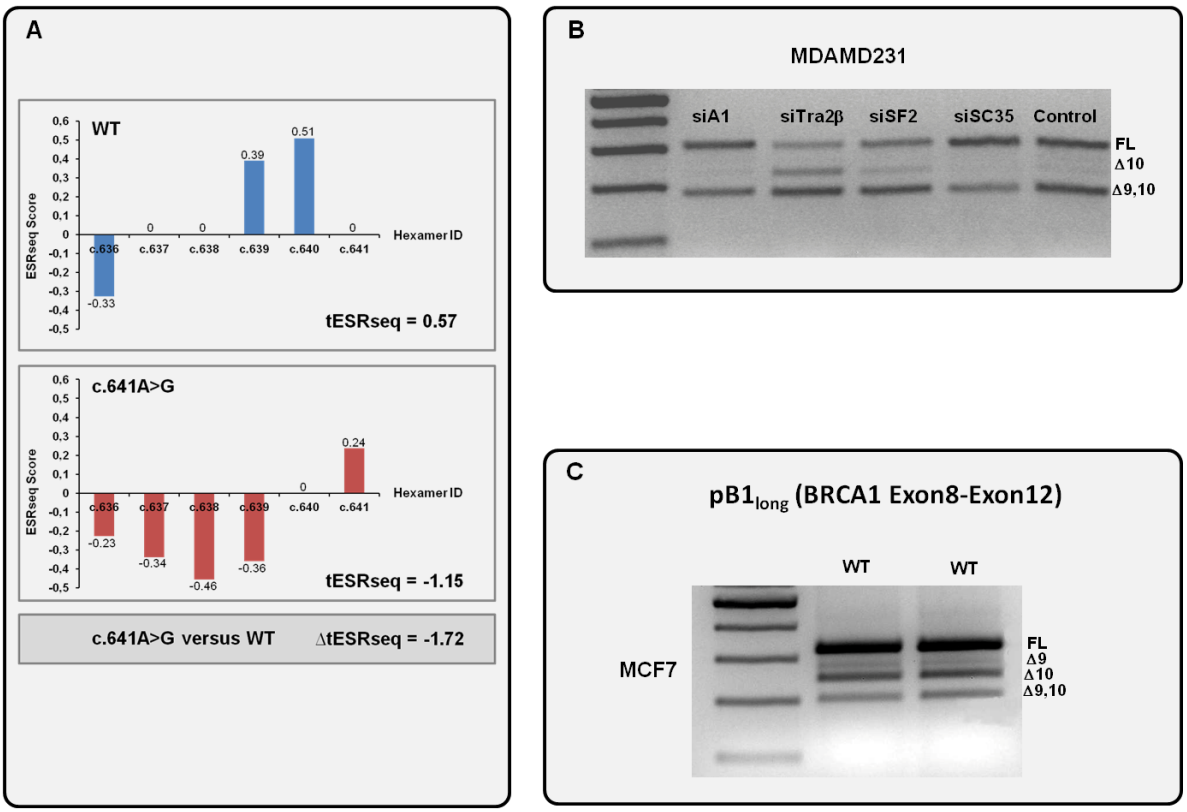
Supplemental Figure 4. qPCR analysis of BRCA1 alternative splicing landscape in LCLs from one *BRCA1* c.[594-2A>C; 641A>G] carrier (Carrier 1) and 11 Controls. We performed several quantitative real-time PCR (qPCR) experiments with standard curves to determine the GADPH normalized absolute number of *BRCA1* FL transcripts (exon 9 and exon 10 containing transcripts), as well as *BRCA1* transcripts containing the following alternative splicing events : Δ9,10, Δ9, Δ10, Δ9_11, Δ11q, Δ11, and Δ9,10+Δ11q. The chart in **Panel A** displays the normalized absolute number of FL transcripts in Carrier 1 and 11 Controls (standard error of 3 experiments is shown), as well as the normalized absolute number of all *BRCA1* transcripts (see supplemental methods for further details). The data indicate that the level of exons 9 and 10 containing transcripts drops in Carrier 1 if compared with healthy controls, but the overall *BRCA1* expression level remains roughly constant. The Chart in **Panel B** shows $\Delta 9\_11^{SF}$, $\Delta 11^{SF}$, and $(\Delta 9,10+\Delta 11q)^{SF}$ estimated as the ratio between the GADPH normalized absolute number of transcripts containing the indicated alternative splicing event, and the GADPH normalized absolute number of all *BRCA1* transcripts shown in Panel A (standard error of 3 experiments in shown). As already observed for $\Delta 9,10^{SF}$ (see Figure 2), the $(\Delta 9,10+\Delta 11q)^{SF}$ appears to be slightly increased in Carrier 1 if compared with healthy control samples.



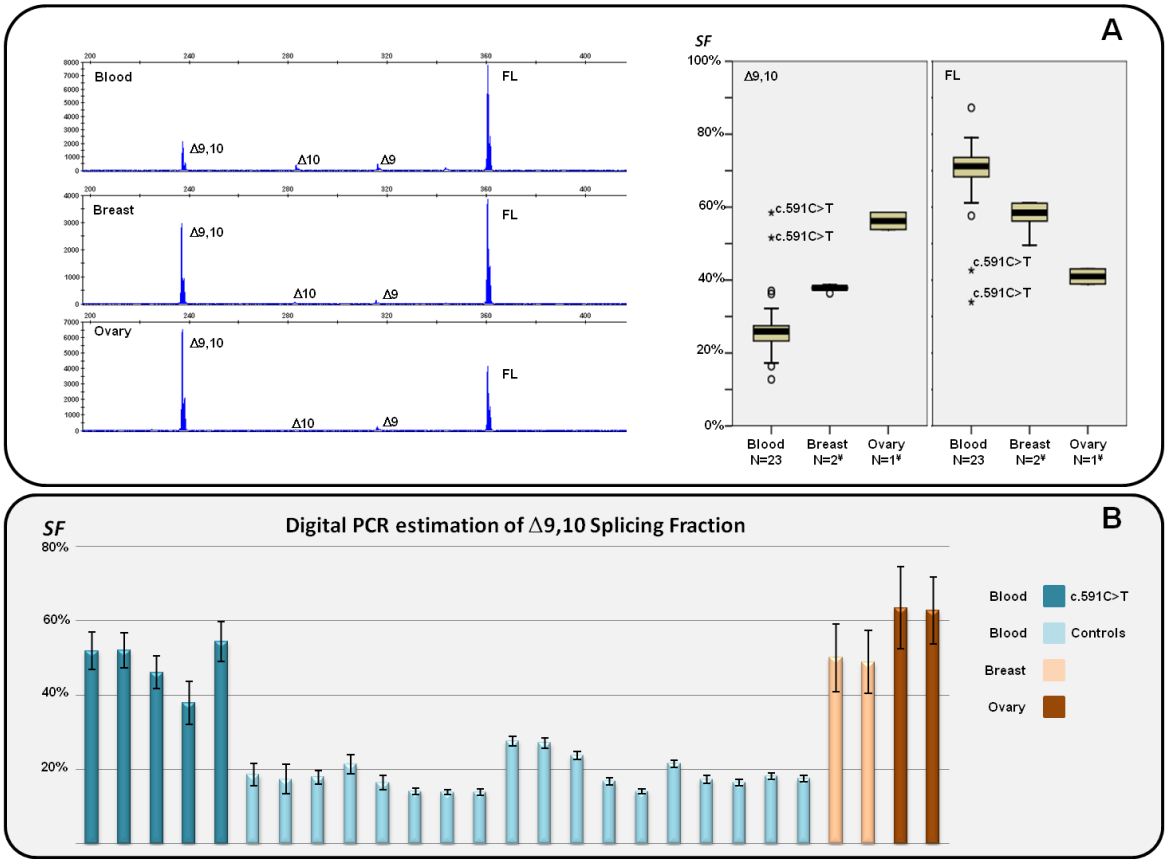
Supplemental Figure 5. Biallelic expression analyses in LEUs from one BRCA1 c.[594-2A>C; 641A>G] carrier (Carrier 3) and one Control. The figure shows representative examples of RT-PCR sequencing at the informative SNP rs1060915 (c.4308C>T, p.=), located in *BRCA1* exon 13. RT-PCR amplifications were performed with a reverse primer located in *BRCA1* exon 15, and different forward primers (left) designed to be specific for certain alternative splicing events. Since *BRCA1* exon 11 is too long to be PCR amplified with our protocol, experiments shown in panel A are restricted to the subset of *BRCA1* Δ11q transcripts, as indicated (right). A forward primer located in exons12/13 junction **(panel B)** is considered a proxy for overall expression. In Carrier 3, the rs1060615 C allele was demonstrated to be *in cis* with c.[594-2A>C;641A>G] (data not shown). Overall, data obtained with exons8/11 junction forward primer **(panel A, top)** shows that in carriers, both alleles contribute similarly to Δ9,10 overall expression level. A slight imbalance in favor of the variant allele (c.4308C) observed in Carrier 3 is compatible with other results suggesting that levels of Δ9,10 transcripts in c.[594-2A>C; 641A>G] carriers is in the upper limit of Control samples (Figure 2, Supplemental Figure 6). Data obtained with exons 9/11 junction forward primer **(panel A, bottom)** shows that Δ10 transcripts arise mostly from the variant allele. **Panel B** shows similar experiments performed with exons 12/13 forward primers, a proxy for overall expression level. The data suggests a slight imbalance in favor of the WT allele (c.4308T) in Carrier 3. The data is compatible with selective degradation of variant-derived Δ10 transcripts in the absence of a NMD inhibitor (Cyclo-experiments). Equivalent experiments performed with Carriers 4 and 5 showed almost identical results, including the slight imbalance observed with forward primers located in junctions for exons 8/11 and exons 12/13 (data not shown).



Supplemental Figure 6. dPCR relative quantification of Δ9,10 and FL in-frame transcripts in LCLs from one *BRCA1* c.[594-2A>C; 641A>G] carrier (Carrier 1), one c.591C>T carrier, and 7 Controls. LCLs were treated with the NMD inhibitor Puromycin (Puro+ experiments) prior to RNA extraction and analysis. The boxplot in **Panel A** (displaying low, Q1, median, Q3, and high values) shows digital PCR measures of *BRCA1* Δ9,10 relative expression level (exons8/10 junction), using *BRCA2* as a reference). The data is expressed as the fold-increase relative to the average of 7 Controls. *BRCA1* Δ9,10 relative expression level in Carrier 1 (2 technical replicates) is in the upper limit of control samples, while a >2-fold increase is observed in the c.591C>T carrier. The boxplot in **Panel B** (displaying low, Q1, median, Q3, and high values) shows digital PCR measures of *BRCA1* FL relative expression level (exons 9/10 junction), using *BRCA2* as a reference. The data is expressed as the fold-increase relative to the average of 7 Controls. A 50% decrease of *BRCA1* FL relative expression is observed both in Carrier 1 (2 technical replicates) and c.591C>T carrier.



Supplemental Figure 7. *BRCA1* c.641A>G is predicted to induce exon 10 skipping by altering potential splicing regulatory elements. Panel A displays results from a comparative *in silico* analysis of *BRCA1* WT and *BRCA1* c.641A>G based on ESRseq scores, which was performed as described under Supplemental Methods. The bars indicate the ESRseq scores of hexamer stretches starting at the positions indicated in the X-axis. All hexamer sequences overlapping the c.641 position were taken into account. Of note, positive ESRseq scores are indicative of potential exonic splicing enhancer elements (ESE) whereas negative ESRseq scores indicate potential exonic silencer elements (ESSs). The negative nature of the Change in Total ESRseq score produced by the variant relative to WT (ΔtESRseq=-1.72) is predictive of exon 10 skipping. **Panel B** shows a representative example of knockdown experiments performed in MDA-MD231 breast cancer cells demonstrating a positive role of Tra2β in *BRCA1* exon10 inclusion. **Panel C** shows representative examples of $pB1_{long}$(WT) splicing reporter minigene experiments performed in MCF7 breast cancer cells. The $pB1_{long}$(WT) splicing reporter minigene is identical to pB1(WT) with the only exception that it includes the full intron 9 and intron 10 sequences. Even if extensive alternative splicing is observed for $pB1_{long}$(WT) in MCF7 cells, Δ10 appears to be the predominant alternative splicing event (and not Δ9,10, which is the predominant endogenous *BRCA1* splicing event both in blood-related samples and in MDA-MD231 cells). $pB1_{long}$(WT) results are similar to those obtained with pB1(WT)(Figure 3B). These observations may reflect *BRCA1* tissue-specific alternative splicing and/or a suboptimal capacity of splicing reporter minigenes to fully reproduce the alternative splicing pattern of the exon 8-exon 11 region.



Supplemental Figure 8. *BRCA1* alternative splicing landscape in blood, breast, and ovary related samples. **Panel A (left)** shows representative examples of capillary electrophoresis analysis o RT-PCR products generated with the E8.1-E11q.1 assay (33 cycle PCRCs) performed in fresh peripheral blood from healthy controls, non-malignant breast related samples, and non-malignant ovary related samples. The analyses detected full-length (FL), and up to four alternative splicing events, including two in-frame (Δ9,10 and ▼10p), and two out-of-frame (Δ9, and Δ10). ▼10p transcripts are not detected in the individual examples shown in panel A, but have been occasionally detected in blood and non-malignant breast samples (data not shown). The boxplots in **Panel A** (displaying low, Q1, median, Q3, and high values) represent $\Delta 9,10^{SF}$ and FL^{SF} as determined with the E8-E11p capillary electrophoresis assay (expressed as the % of the corresponding peak area to the Σ of all peak areas). The chart displays the SF observed in 23 blood samples (fresh peripheral blood from healthy controls), non-malignant breast-related samples ¥(N=2 stands for a pool of 10 non-malignant breast tissues measured in duplicate, plus one commercial source of non-malignant breast tissue measured in triplicate), and non-malignant ovary-related samples ¥(N=1 stands for one commercial pool of 3 non-malignant ovary tissues measured in duplicate). Normal outliers (>1.5 inter quartile range, IQR) display a small circle. Extreme outliers (>3 IQR) display an asterisk. Note that two c.591C>T carriers, positive controls known to increment $\Delta 9,10^{SF}$, are correctly detected as extreme outliers in this assay. **Panel B** displays $\Delta 9,10^{SF}$ as determined by digital PCR (using exon23-24 junction as a proxy for overall expression, see methods for further details). The precision of each measure (as determined by the QuantStudio 3D Analysis Cloud Software) is indicated. We included in the analysis fresh blood from 18 unrelated healthy controls, fresh blood from 5 positive controls (5 unrelated carriers of the *BRCA1* c.591C>T variant), one commercial source of non-malignant breast tissue, and one commercial pool of 3 non-malignant ovary tissues. For breast and ovarian samples, two technical replicas are shown.

Table 1: Characteristics of Carriers of *BRCA1* c.594-2A>C identified in BCAC

Description of Studies with carriers identified						Characteristics of carriers identified			
Study Acronym	**Study Design**	**Country of Origin**	***BRCA1/2* Sequencing of cases in this study**	**n Invasive Cases**	**n Controls**	**Case-Control Status**	**Age Interview/ Onset**	**Breast Tumour Pathology**	**Tumour Likelihood Ratio**
BBCS	Cancer registry and National Cancer Research network based cases (including bilateral cases), population based controls	United Kingdom	none tested	1446	1397	case control	43 58	Grade 3, ER-pos control	0,64 1,00
BSUCH	Hospital based cases, blood donor controls	Germany	none tested	815	954	case	40	NA	1,00
CGPS	Population-based cohort, nested case-control study	Denmark	none tested	2811	4086	case control	60 28	Grade 2, ER-neg control	2,34 1,00
KARBAC	Hospital-based familial and consecutive cases, geographically matched controls	Sweden	100% tested**	722	662	control*	61	control	1,00
MARIE	Population-based case-control study	Germany	none tested	1656	1778	control	54	control	1,00
MBCSG	Clinic-based familial/early onset breast cancer cases, population-based controls	Italy	100% tested**	189	400	control	35	control	1,00
MCBCS	Hospital-based case-control study	USA	4% tested**	1546	1931	control	31	control	1,00
MCCS	Population-based prospective cohort study	Australia	none tested	614	511	case	76	Grade NA, ER-pos	0,37
OFBCR	Population-based familial case-control study	Canada	68% tested**	1156	511	case case***	54 41	Grade 2, ER-pos Grade 3, ER-neg	0,34 3,16
pKARMA	Population-based cases, mammography screen study controls	Sweden	none tested	4553	5537	control	46	control	1,00
SEARCH	Population-based case-control study	United Kingdom	none tested	9097	8069	case control control	46 70 52	Grade NA, ER-pos control control	0,37 1,00 1,00
Totals				***24605***	***25836***				

For further details about participating BCAC studies, please see Michaelidou et al (1).
* Age data not available, mean diagnosis age for that study used for case-control likelihood analysis
** BCAC studies which had undergone genetic testing for *BRCA1/2* variation, and were excluded from final analyses to determine causality based on case-control presentation.
*** Case determined to overlap with a CIMBA proband. No segregation data was available, and pathology information was included only once in multifactorial likelihood

Supplementary Table 2: Characteristic of carriers of *BRCA1* c.594-2A>C identified in ENIGMA and CIMBA*

Consortium	Site	Country of Origin	Segregation Bayes Score for Family	Breast Tumour Pathology (age onset) for proband; other carrier relatives	Pathology LR
ENIGMA	Ambry Genetics	USA	Not informative	ER-pos (<50)	0.32
ENIGMA	Ambry Genetics	USA	Not informative	ER-neg (<50)	2.60
ENIGMA	Ambry Genetics	USA	Not informative	ER-pos (>50)	0.37
ENIGMA	Ambry Genetics	USA	Not informative	ER-neg (<50)	2.60
ENIGMA	Ambry Genetics	USA	Not informative	ER-pos (>50)	0.37
ENIGMA	Ambry Genetics	USA	Not informative	ER-pos (>50)	0.37
ENIGMA	Ambry Genetics	USA	Not informative	ER-neg (<50)	2.60
ENIGMA/CIMBA	Embrace	UK	1.64	Grade 3, ER-pos (>50)	0.90
ENIGMA/CIMBA	Embrace	UK	Not informative	Grade 1, ER NA (<50)	0.13
ENIGMA	French Consortium	France	0.73	Grade 3, ER-pos (<50)	0.64
ENIGMA	French Consortium	France	0.83	Grade NA, ER-neg (>50)	3.31
ENIGMA	French Consortium	France	1.00	NA	1.00
ENIGMA	GC-HBOC	Germany	0.42	NA	1.00
ENIGMA	GC-HBOC	Germany	0.55	NA	1.00
ENIGMA	GC-HBOC	Germany	2.38	Grade 3, ER-pos (<50); Grade 3, ER-pos (<50)	0.41
ENIGMA/CIMBA	GC-HBOC	Germany	Not informative	Grade 3, ER-pos (<50)	0.64
ENIGMA	GC-HBOC	Germany	Not informative	NA	1.00
ENIGMA	GC-HBOC	Germany	Not informative	Grade 3, ER-pos (>50)	0.90
ENIGMA	GC-HBOC	Germany	Not informative	Grade 3, ER-pos (>50)	0.90
ENIGMA	GC-HBOC	Germany	Not informative	Grade 3, ER-pos (>50)	0.90
ENIGMA	GC-HBOC	Germany	6.85	Grade 3, ER-pos (<50); Grade 3, ER-pos (>50); Grade 3, ER-pos (>50); Grade 3, ER-pos (>50)	0.47
ENIGMA/CIMBA	kConFab	Australia	4.37	Grade 3, ER-pos (>50)	0.90
ENIGMA/CIMBA	Leiden	Netherlands	0.10	NA	1.00
ENIGMA	New Zealand Familial Breast Cancer Study	New Zealand	0.65	NA	1.00
ENIGMA	Northshore	USA	1.82	Grade 1, ER-pos (>50); Grade NA, ER-pos (>50)	0.04
ENIGMA/CIMBA	Northshore	USA	0.02	NA	1.00
ENIGMA/CIMBA	Northshore	USA	1.18	NA	1.00
ENIGMA	Adult Genetics Unit, South Australia	Australia	Not informative	Grade 2, ER-pos (>50)	0.34
CIMBA	BCFR-AU	Australia	1.96	NA	1.00
CIMBA	NIH	USA	Not informative	NA (ovarian cancer patient)	1.00
CIMBA	MUV	Austria	Not informative	Grade 2, ER-pos (<50)	0.21

* *BRCA1* c.641 A>G (Asp214Gly) was reported to be present in all family probands. *BRCA1* c.641 A>G was also observed in an additional 13 carriers of *BRCA1* c.594-2A>C identified by Ambry Genetics, excluded from causality analysis because of unavailability of relevant information. NA=not available.

Supplemental Table 3: *BRCA1* and *BRCA2* exon boundary variants predicted/known to increase the level of naturally occurring in-frame RNA transcripts that may rescue gene functionality. *Variants at these positions should be considered class 3 (uncertain) unless proven otherwise.**

Gene	Alternative Splicing Event	Variants Implicated	Rationale
BRCA1	Δ8p	c.442-1 (IVS7-1) c.442-2 (IVS7-2)	Exon 8 acceptor site is an experimentally validated tandem acceptor site (NAGNAG) subject to alternative splicing (2). c.442-1,-2 variants are predicted to inactivate the proximal (5'), but not the distal (3') splice acceptor site, thus potentially producing Δ8p transcripts.
	Δ9,10	c.548-1 (IVS8-1) c.548-2 (IVS8-2) c.593 to non-G c.593+1 (IVS9+1) c.593+2 (IVS9+2) c.594-1 (IVS9-1) c.594-2 (IVS9-2) c.670 to non-G c.670+1 (IVS10+1) c.670+2 (IVS10+2)	Carriers of these variants are predicted to produce normal (or increased) levels of *BRCA1* Δ(9,10), a major in-frame alternative splicing event (2). *BRCA1* c.[594-2A>C; 641A>G] has been reported to demonstrate clinical characteristics inconsistent with a high risk of cancer expected for a pathogenic *BRCA1* variant (3).
	Δ13p	c.4186-1 (IVS12-1) c.4186-2 (IVS12-2)	Exon 13 acceptor site is an experimentally validated tandem acceptor site (NAGNAG) subject to alternative splicing (2). c.4186-1,-2 variants are predicted to inactivate the proximal (5'), but not the distal (3') splice acceptor site, potentially producing Δ13p transcripts.
	Δ14p	c.4358-1 (IVS13-1) c.4358-2 (IVS13-2)	Exon 14 acceptor site is an experimentally validated tandem acceptor site (NAGNAG) subject to alternative splicing (2). c.4358-1,-2 variants are predicted to inactivate the proximal (5´), but not the distal (3') splice acceptor site, potentially producing Δ14p transcripts.
BRCA2	Δ12	c.6842-1 (IVS11-1) c.6842-2 (IVS11-2) c.6937 to non-G c.6937+1 (IVS12+1) c.6937+2 (IVS12+2)	Carriers of these variants are predicted to produce exon12 skipping. BRCA2 Δ12 is a naturally occurring in-frame splicing event (ENIGMA Splicing Working group, unpublished data). BRCA2 exon12 is functionally redundant (4).

* Reference sequences: BRCA1 cDNA U14680.1/genomic NC_000017.11/(exon numbering according to U14680.1); BRCA2 cDNA U43746.1/genomic NC_000013/(exon numbering according to U43746).

References

1 Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K. *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*, **45**, 353-361, 361e351-352.

2 Colombo, M., Blok, M.J., Whiley, P., Santamarina, M., Gutierrez-Enriquez, S., Romero, A., Garre, P., Becker, A., Smith, L.D., De Vecchi, G. *et al.* (2014) Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum Mol Genet*, **23**, 3666-3680.

3 Rosenthal, E.T., Bowles, K.R., Pruss, D., van Kan, A., Vail, P.J., McElroy, H. and Wenstrup, R.J. (2015) Exceptions to the rule: Case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes. *Clin Genet*, in press.

4 Li, L., Biswas, K., Habib, L.A., Kuznetsov, S.G., Hamel, N., Kirchhoff, T., Wong, N., Armel, S., Chong, G., Narod, S.A. *et al.* (2009) Functional redundancy of exon 12 of BRCA2 revealed by a comprehensive analysis of the c.6853A>G (p.I2285V) variant. *Human mutation*, **30**, 1543-1550.

Supplemental Methods

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.

de la Hoya et al.

Splicing Analyses

1.1 Samples

RNA extraction protocols from lymphoblastoid cell lines (LCLs), leukocytes (LEU) (3-6 days cultured), and biopsy samples were performed with standard methods that have been described previously (1-4). In the case of peripheral blood collected in PAXgene tubes (Qiagen, Hilden, Germany), RNA extraction was performed with the PAXgene Blood RNA kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). When indicated, Puromycin (Puro) or Cycloheximide (Cyclo) was added to cell cultures (end concentration of 100µg/ml) 6-8 hours prior to cell harvest. In addition, we used commercially available total RNA (guanidium thiocyanate isolation method) from Human Mammary Gland and Human Ovary (Clontech, Mountain View, CA). More precisely, Human Mammary Gland RNA (Clontech, catalog No: 636576) was obtained from a 27 year-old Caucasian female. Human Ovary RNA (Clontech, catalog No: 636555) was pooled from 3 Caucasian females, aged 40, 55 and 51. In all cases, RNA samples were treated with RNase-Free DNase previous to Reverse Transcription (RT) reactions. For RT-PCR reactions, different contributors used different commercially available kits (SuperScriptII Reverse Transcriptase, Invitrogen, Carlsbas, CA; High Capacity RNA-to-cDNA Mater Mix, Applied Biosystems, Foster City, CA; Primescript RT reagent kit, Takara Biotechnology, Shiga, Japan; Transcriptor High Fidelity cDNA synthesis Kit, Roche Applied Science, Mannheim, Germany; One-Step RT-PCR kit, Qiagen, Hilden, Germany) in combination with oligodT and/or random hexamers. Previously, we have shown that different RNA extraction/cDNA synthesis protocols are indistinguishable for the purpose of BRCA1 RT-PCR analyses (2).

1.2 Capillary Electrophoresis

Capillary electrophoresis analysis of alternative splicing has been extensively described previously (2, 3). In the present study, we have analyzed the *BRCA1* alternative splicing landscape at the vicinity of exon 10 with up to five different RT-PCR assays: E7-E11q.1, E8.1-E11p, E8.1-E11q.1, E8.1-E11q.2, E8.2-E11q.2, and E8.1-E12 (see Table 1 below for further details). To perform semi-quantitative analyses, we kept a low number of PCR cycles (depending on the contributing laboratory, 28, 33, or 36 cycles). While 28 or 33 cycle PCRs are optimal for semi-quantification, 36 cycle PCRs were needed to detect the rather minor alternative splicing event ▼10p. Depending on the contributing laboratory, capillary electrophoresis analyses were performed with 3100, 3130, 3130XL, 3500XL, or 3730XL ABI PRISM Genetic Analyzers (Applied Biosystems, Foster City, CA) using POP-7 filled capillary arrays of 36 or 50 cm. Size-calling and peak areas were analyzed with GeneMapper v4.0 or GenScan v3.7 software (both from Applied Biosystems, Foster City, CA), or GeneMarker 2.4 (Softgenetics LLC, State College, PA). As internal size-standard, we used LIZ-500 or LIZ-1000 (Applied Biosystems, Foster City, CA). Splicing Fraction (SF) of each individual alternative splicing event were measured in all cases as the ratio between the peak area of the individual events and the Σ of all peak areas (all transcripts) detected by the corresponding assay. For instance, $\Delta 9,10^{SF}$ is determined in a E8-E11q assay as the ratio between Δ9,10 peak area and the Σ of Δ9+Δ10+▼10p+Δ9,10+full-length (FL) peak areas (depending on samples an PCR cycles, not all five peaks are necessarily observed in all determinations). Electropherograms with saturated peaks were not considered for SF analyses.

.

Table 1. Description of RT-PCR assays used in the present study.

RT-PCR assay	Forward primer		Reverse primer (FAM-labeled)		Capillary Electrophoresis Analysis of RT-PCR products (expected peak sizes expressed in bp)				
	mapping	seq	mapping	seq	Δ9,10	Δ10	Δ9	FL	▼10p
E7-E11q.1	c.372_c.393 (Exon 7)	CATCCAAAGTATGGGCTACAGA	c.799_c.819 (Exon 11q)	TGGCTCCACATGCAAGTTTG	324/327[a]	368/371[a]	401/404[a]	445/448[a]	466/469[a]
E8.1-E11p	c.459_c.478 (Exon 8)	TGTCCAACTCTCTAACCTTG	c.759_c.778 (Exon 11p)	TTTCTGGATGCCTCTCAGCT	199	243	276	320	341
E8.1-E11q.1	c.459_c.478 (Exon 8)	TGTCCAACTCTCTAACCTTG	c.799_c.819 (Exon 11q)	TGGCTCCACATGCAAGTTTG	240	284	317	361	382
E8.2-E11q.2	c.462_c.971 (Exon 11)	CCAACTCTCTAACCTTGGAACTGTG	c.949_c.971 (Exon 11q)	CTTCCAGCCCATCTGTTATGTTG	389	433	466	510	531
E8.1-E12	c.459_c.478 (Exon 8)	TGTCCAACTCTCTAACCTTG	c.4145_c.4164 (Exon 12)	CTGAGAGGATAGCCCTGAGC	276[b]	320[b]	353[b]	397[b]	418[b]

For each RT-PCR assay, we show the sequence of forward and reverse primers, mapping to the Ensembl sequence ENST00000357654 (NCBI NM_007294.3), and the expected size of the peaks corresponding to different *BRCA1* alternative splicing events. The actual size calling may vary ±2bp with respect to the expected size due to factors such as local density of size-standard peaks, capillary array length, and/or Taq polymerase addition of a 3'-adenine overhang. [a] Capillary electrophoresis analysis of RT-PCR products generated with the E7-E11q.1 assay produces ±3bp doublet peaks due to alternative splicing at BRCA1 exon 8 NAGNAG splicing acceptor site (3, 5). [b] With protocols used in the present study, the E8.1-E12 RT-PCR assay do not generate products spanning the long BRCA1 exon 11 (3426nt). For that reason, the expected size peaks displayed in the table correspond to *BRCA1* Δ11q transcripts (3, 5): Δ9,10+Δ11q, Δ10+Δ11q , Δ9+Δ11q and ▼10p+Δ11q transcripts.

1. 3 Digital PCR (dPCR)

All dPCR experiments were performed in a QuantStudio 3D Digital PCR 20K platform according to the manufacturer's instructions (Applied Biosystems, Foster City, CA). As indicated, we performed different assays combining FAM-labeled and VIC-labeled TaqMan assays. FAM-labeled assays included an Applied Biosystem pre-designed assay (Hs01556198) specific for the E8/E11 junction (Δ9,10 assay), and a custom designed TaqMan assay specific for the *BRCA1* E9/E10 junction (FL assay). The FL assay was designed with the Applied Biosystem proprietary on-line pipeline. VIC-labeled assays included an Applied Biosystem pre-designed assay (Hs01556193) specific for *BRCA1* E23/24 junction (a proxy for overall *BRCA1* expression), and a pre-designed assay (Hs00609073) specific for *BRCA2* E26/27 junction (a proxy for overall *BRCA2* expression). dPCR chips were analyzed in the cloud-based QuantStudio 3D Analysis Suit v2.0 (Applied Biosystem, Foster City, CA) to review quality (only green and yellow flag chips were considered for further analyses), and calculate copies/µl of FAM and VIC molecules (and the precision of these measures). Default settings were used in all cases. Subsequently, data was exported to an excel file to calculate the FAM/VIC ratio. To measure $\Delta 9,10^{SF}$, we performed experiments combining Δ9,10 and Hs01556193 assays. To measure FL^{SF} we performed experiments combining FL and Hs01556193 assays. To measure Δ9,10 and FL relative expression levels, we performed experiments combining Δ9,10 (or FL) assays with Hs00609073. In relative expression level experiments, the Δ9,10 (or FL) relative expression level of each sample was normalized to the average Δ9,10 (or FL) relative expression level as measured in control samples.

1.4 Absolute quantification of alternative splicing events by real-time PCR (qPCR)

In order to specifically amplify *BRCA1* full-length transcripts (transcripts containing exons 9,10, and 11q), transcripts containing six naturally occurring alternative splicing events (Δ9, Δ9,10, Δ9_11, Δ10, Δ11, and Δ11q), or transcripts combining two splicing events (Δ9,10+Δ11q), we designed various primers in specific exonic regions, or overlapping specific exon/exon junctions. Specificity of primer combinations (we tested 3 primer sets peralternative splicing event) were determined by temperature gradient PCR. Pooled cDNA from healthy controls was used as template to amplify all specific alternative splicing event fragments, with the single exception of Δ10 fragments, which were obtained using variant carrier cDNA as template. Cycling conditions were 95°C for 2mins followed by 35 cycles of 95°C for 30secs, gradient annealing temperature (annealing temperature of 55°C - 65°C at intervals of 2°C) for 30secs, 72°C for 60secs and a final extension time of 7mins. Gel electrophoresis was used to visualize the fragments. PCR products amplified using a primer combination and lowest annealing temperature for which only one PCR product could be observed were considered optimal (see table at the end of this section) and selected for cloning. PCR products were cloned using the pGEM-T Vector System (Promega, Auburn, Victoria, Australia). Recombinant clones were selected from a single colony and sequence confirmed. Plasmid preparations containing the PCR products for each of the eight fragments were quantified using a Qubit 3.0 Fluorometer (Aplied Biosystems, Foster City, CA) and a serial dilution was made. Dilutions were used as template for deriving a standard curve for quantitative PCR (qPCR).

A standard curve was determined using pGEM-T clones carrying the eight isoforms and the specific primers/annealing temperature for each transcript. Real-time PCR reactions were carried out in a Lightcycler 480 (Roche, Castle Hill, NSW, Australia) using Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen). Cycling conditions were: 50°C for 2mins, 45 cycles of 95°C for 2mins, 95°C for 20secs, optimized annealing temp (see table at the end of this section)15secs, 72°C for 20secs. Normalized expression values (using GAPDH as an internal reference) were obtained using the Lightcycler 480 Gene Scanning software for cDNA isolated from the variant carrier and a set of 11 non-variant carrying controls. Crossing point (CP) values were plotted against the standard curve. The number of molecules of the 3.2Kb plasmid was estimated at $2.9x10^8$/ng. Given that there are approximately 290 million molecules in 1ng plasmid, the starting number of molecules in each reaction could be estimated enabling a comparison within a sample and across the controls.

Table 2. Optimal primer sequences and annealing temperatures for absolute quantification of alternative splicing eventsby qPCR

	Forward		Reverse		product (bp)	Annealing Temp (C°)
Full-length	ex10F	CTCAAGGAACCAGGGATGAA	ex11pR	ACTGGGTTGATGATGTTCAGT	101	63
Δ11	ex10/12F	TGGATTCTGCAAAAAAGGGTGAA	ex12R	CTGAGTGGTTAAAATGTCACTCTGA	107	61
Δ11q	ex11p/12F	ATCCAGAAAAGTATCAGGGTGAA	ex12R	CTGAGTGGTTAAAATGTCACTCTGA	107	59
Δ9_11	ex8/12F	TGTCTACATTGAATTGGGTGAAGCA	ex12R	CTGAGTGGTTAAAATGTCACTCTGA	106	63
Δ9,10	ex8/11F	TGTCTACATTGAATTGGCTGCTTGT	ex11pR	GCACGCTTCTCAGTGGTGT	101	65
Δ10	ex9/11F	AATAAGGCAACTTATTGCAGCTGCTTGT	ex11pR	GCACGCTTCTCAGTGGTGT	104	61
Δ9	ex8/10F	GTCTGTCTACATTGAATTGGTGTGG	ex10/11R	CTCAGAAAATTCACAAGCAGCCTTT	113	61
Δ9,10+Δ11q	ex8/11pF	TCTGTCTACATTGAATTGGCTGCTT	ex11p/12R	AGATGCTGCTTCACCCTGAT	146	61

1.5 RNAseq experiments.

The TruSeq targeted RNA expression kit (Illumina, San Diego, CA) was used to target exon regions across *BRCA1*. Each probe pair targets a specific splice junction or coding SNP. Coding SNP probes are situated within the same exon, either side of the variant of interest, whereas splice junction probes are positioned near the 3' and 5' end of two adjacent exons, respectively. This design allows for the detection of certain alternative splicing events, and small splice junction aberrations, as the probes do not need to bind specifically with the other in their original pair. Any exon skipping event has the potential to be detected providing there is an upstream and a downstream probe flanking the deletion. Briefly, TruSeq Targeted RNA Expression chemistry involves pre-designed oligonucleotide probes that hybridise to the target *BRCA1* cDNA region followed by an extension-ligation reaction then takes place to connect the probes, and an amplification step to create the template strand. This is PCR amplified to add indices prior to sequencing. Sequencing was performed on Illuminas MiSeq platform. All *BRCA1* exons had predesigned probes situated on either end to allow detection of non-aberrant mRNA splice junctions, with the single exception of *BRCA1* 16-17. In addition, one probe in the *BRCA1* 3-5 pair spans the splice site inhibiting detection of splicing events involving this junction. TruSeq targeted RNA expression kit allowed us to also quantify each detected splice junction. After normalizing the read counts, the expression of each splicing event was compared across samples to determine expression differences. All samples were sequenced with and without treatment of Cycloheximide.

Splice junction *BRCA1* 2-3 was used as the full length reference transcript. We assumed that this junction is present in all alternative splicing events that don't overlap this junction. We also assumed that the alternative events do not co-occur. Under these assumptions we subtracted all alternative splicing reads, that didn't overlap the exon 2-3 junction, from the total 2-3 junction reads. Some junctions were exempt from this as they are common NAGNAG events (8p, 13p, 14p), which are likely to be present in the full length transcript. Δ9,10 was also excluded as it returned questionable read depths.

The resulting 2-3 read depth, together with the sum of all alternative events(excluding those mentioned above), gave the total expression, from which the proportions of each alternative

event were determined for each sample.The resulting proportions were back transformed (95% CI) for the log data. This provided the standard deviation and mean for the expression of the control samples for each junction, onto which the variant sample relative expression was overlayed. Delta 10 and FL data was extracted for the figures

1.6 RNA interference experiments.

All small interference RNAs (siRNAs) used represent validated sequences in several previous publications from our and other labs (6). 25 nM final concentration of A1 (hnRNP A1) siRNA (5'-CAGCUGAGGAAGCUCUUCAdTdT-3'), Tra2β siRNA (5'-GCAUGAAGACUUUCUGAAAdTdT-3'), SF2 (SRSF1) siRNA (5'-CCAAGGACAUUGAGGACGUdTdT-3'), or SC35 (SRSF2) siRNA (5'-AAUCCAGGUCGCGAUCGAAdTdT-3') were transfected into MDAMD231 cells with INTERFERin (Polyplus transfections, Illkirch, France) following the manufacturer's instructions. As a control, we transfected a scrambled sequence (the luciferase siRNA 5'-CGUACGCGGAAUACUUCGAdTdT-3). Forty-eight hours later, the RNA was extracted using RNAeasy plus kit (Qiagen, Hilden, Germany). The overall strategy has been described previously (6).

1.7 In silico predictions of variant-induced alterations in Exonic Splicing Regulatory sequences (ESRseq)

We analyzed the potential impact of *BRCA1* c.641A>G on RNA splicing by using an *in silico* approach based on the calculation of total ESRseq score changes (ΔtESRseq) as described previously (7).

1.8 Splicing Reporter Minigene Experiments

pCAS2-BRCA1-exon 10 minigene assay

The pCAS2 splicing vector has been previously described (8). The pCAS2-*BRCA1*-exon-10 minigene constructs (Figue 3A) were generated as follows. First, the wild-type genomic segment *BRCA1* c.594-147_c.670+173 was amplified by PCR from genomic DNA using forward primer BR1-10-BamHI-F (GACCGGATCCCTAAAGGAGAGAG) and reverse primer BR1-10-MluI-R (GACCACGCGTTTTAAATCTATCAG), carrying 5' tails with *BamHI* and *MluI* restriction sites, respectively (underlined). PCR-amplified genomic segments encompassed *BRCA1* exon 10 (77 bp) and part of the 5' and 3' flanking intronic sequences (147 bp and 173 bp, respectively). After digestion with *BamHI* and *MluI*, the PCR products were inserted into the *BamHI* and *MluI* cloning sites of pCAS2, a two-exon splicing reporter vector, in order to produce the three-exon pCAS2-BRCA1-exon10-WT minigene. Then, variants of interest were introduced into this construct by site-directed mutagenesis using the two-stage overlap extension PCR method (9). The inserts of the mutant minigenes were sequenced to ensure that no unwanted mutations were introduced during amplification or cloning.

Wild-type and variant minigene constructs were transiently transfected in parallel into HeLa cells using the FuGENE 6 transfection reagent, according to manufacturer's instructions (Roche Applied Science, Mannheim, Germany). Transfections were performed in 12-well plates by using $6x10^4$ cells/well (at ~60% confluence) and 400 ng of each minigene construct. Cells were then collected 24 h post-transfection. Total RNA was extracted using the NucleoSpin RNA II kit (Macherey Nagel, Düren, Germany), according to the manufacturer's instructions, including a DNase treatment. The RT-PCR reactions were performed in a 25 µl reaction volume by using the OneStep RT-PCR kit (Qiagen, Hilden, Germany), and 200 ng RNA as template. Reactions were performed using the vector-specific forward primer pCAS-KO1F (5'-TGACGTCGCCGCCCATCAC-3') and the reverse primer pCAS2R (5'-ATTGGTTGTTGAGTTGGTTGTC-3'), with 30 cycles of amplification. RT-PCR products were separated by electrophoresis on 2% agarose gels containing ethidium bromide and

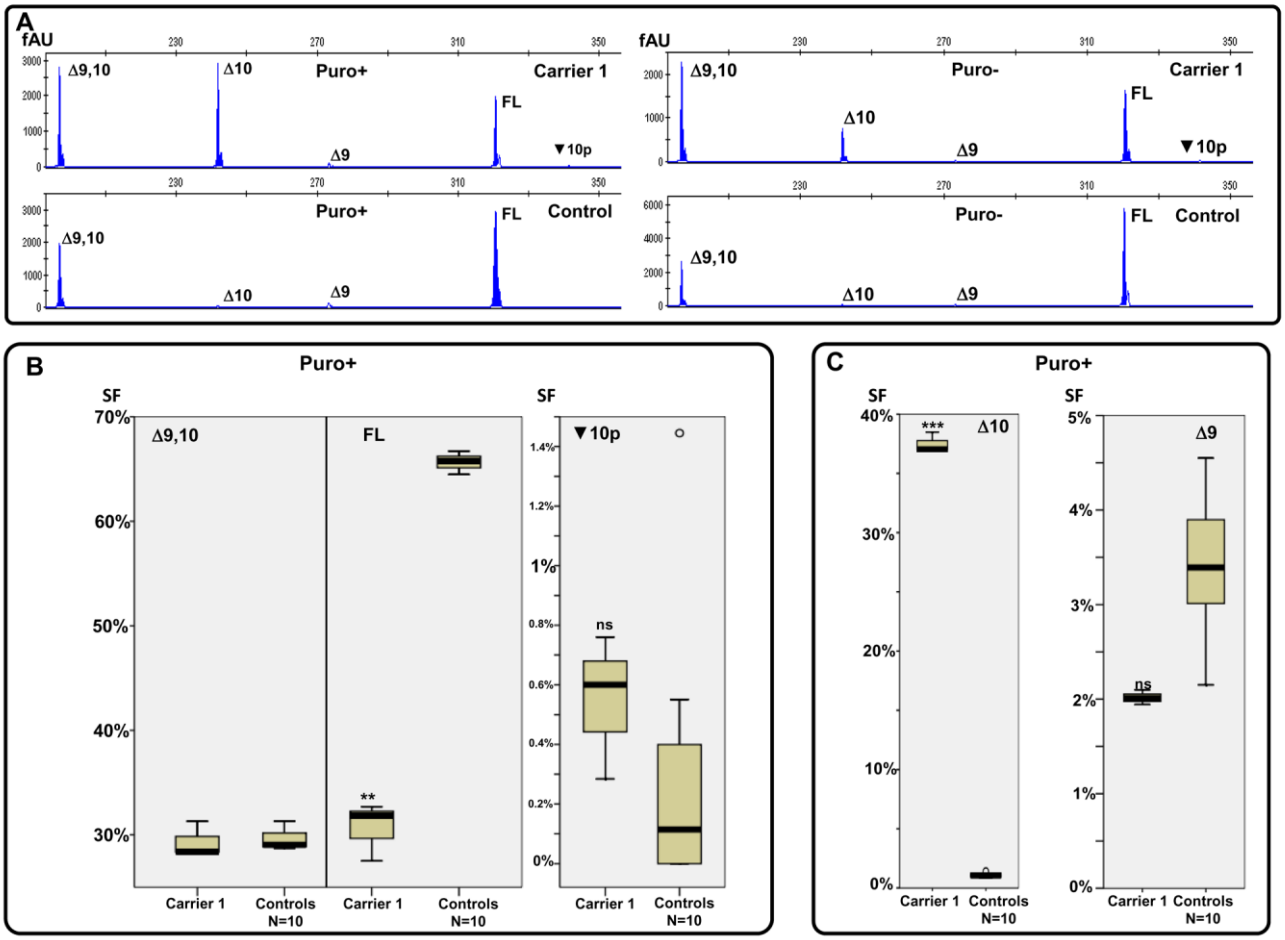
visualised by exposure to ultraviolet light under conditions of non-saturating exposure. RT-PCR products were gel-purified and fully sequenced to determine their identity.

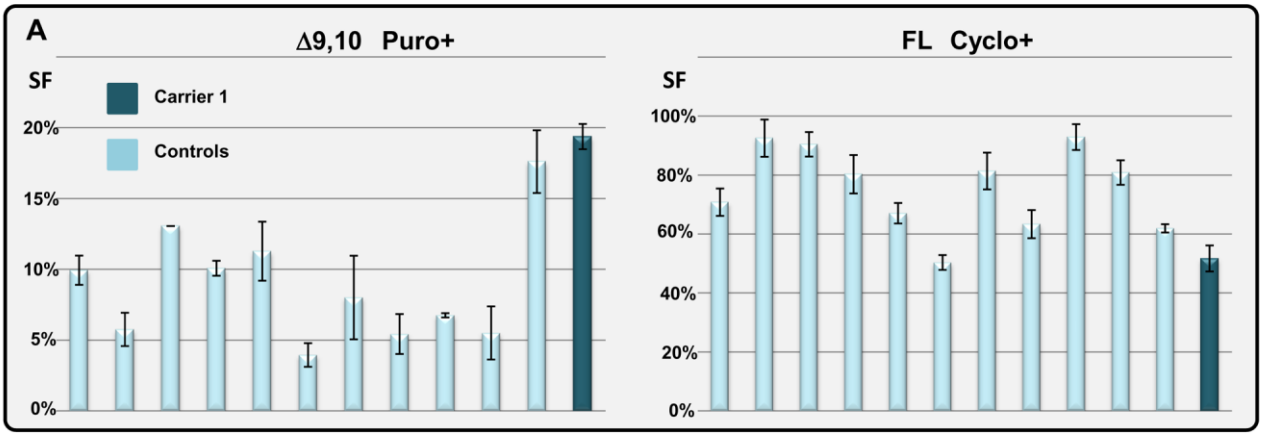
pB1 minigene assay

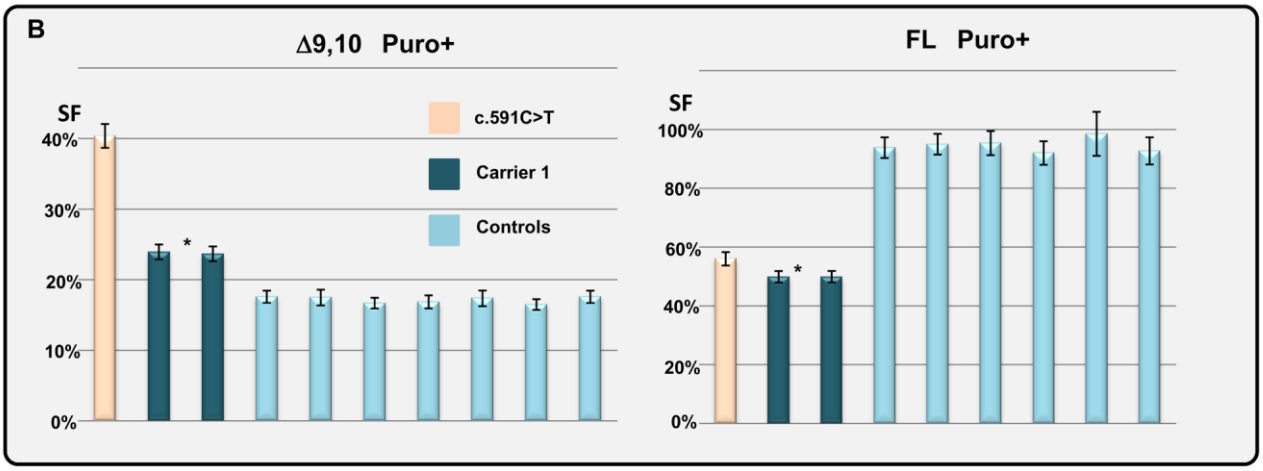
The pB1 minigene splicing reporter vector has been previously described (10). This construct has a pcDNA3(+) backbone and contains the exon 1 of α-globin followed by *BRCA1* exons 8 to 12 (and part of their flanking intronic regions) in under the control of the CMV promoter. Mutant pB1 minigenes were prepared by site-directed mutagenesis through a two-step PCR overlap extension method (9). After digestion with XhoI and HindIII, the *BRCA1* fragments (c.594-174_c.670+345) containing the variants of interest were inserted into pB1 *in lieu* of the corresponding WT sequence. Then, the inserts of the mutant minigenes were sequenced to ensure that no unwanted mutations were introduced during amplification or cloning. Wild-type and variant minigene constructs were transiently transfected into HeLa, MCF7, HBL100, and IgrOV1cells Transfection, RNA extraction and RT-PCR analysis were performed as described above for pCAS2-BRCA1-exon10, with the exception that the RT-PCR primers used here were: pB1-V.8(2)-F (‚5'-GAGGCCCTGGAGAGGACA-3',a vector-specific forward primer) and RT.1.8-11-R (, 5'-ACGCTTCTCAGTGGTGTTCA-3', a reverse primer on *BRCA1* exon 11).

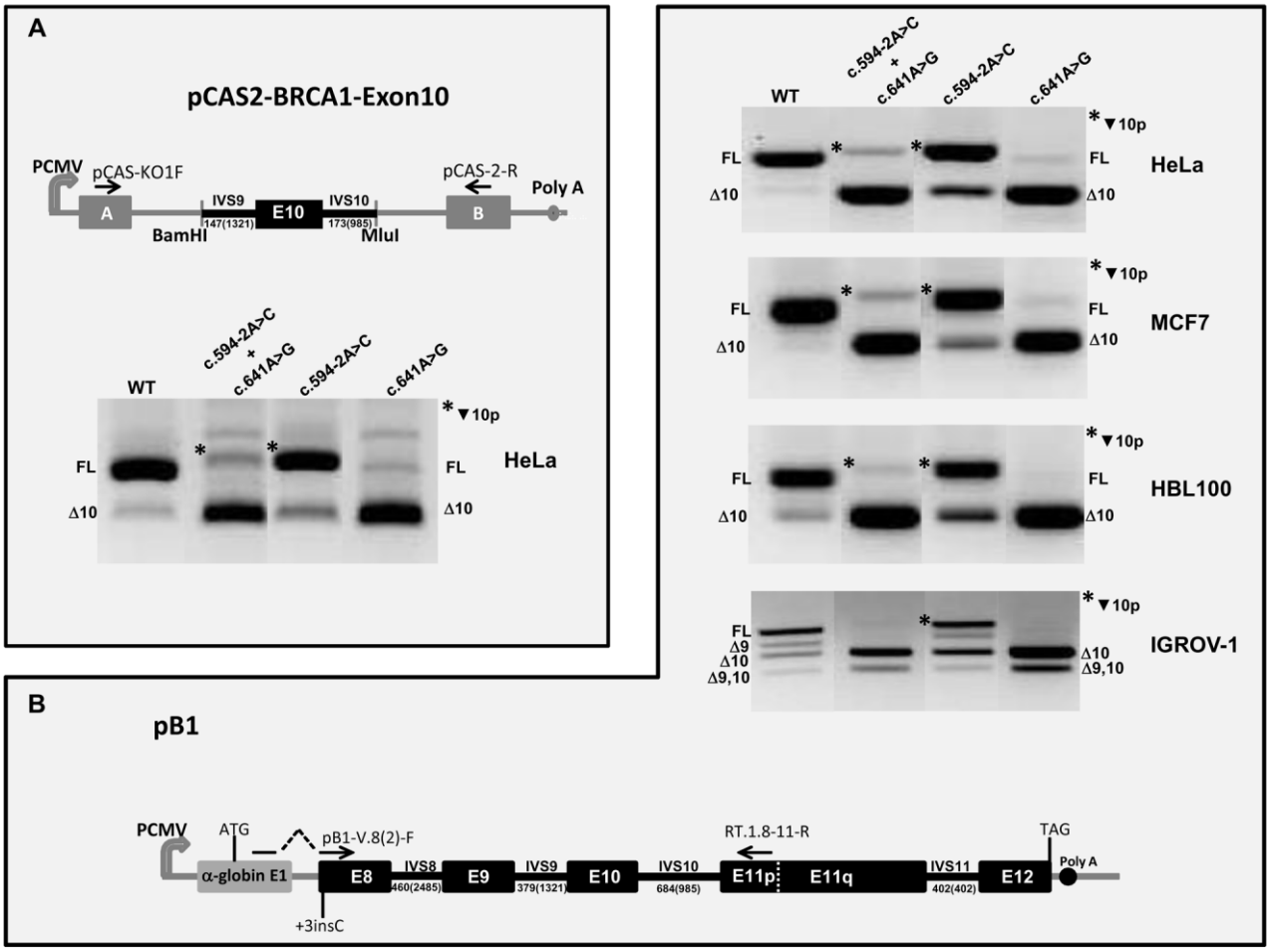
References

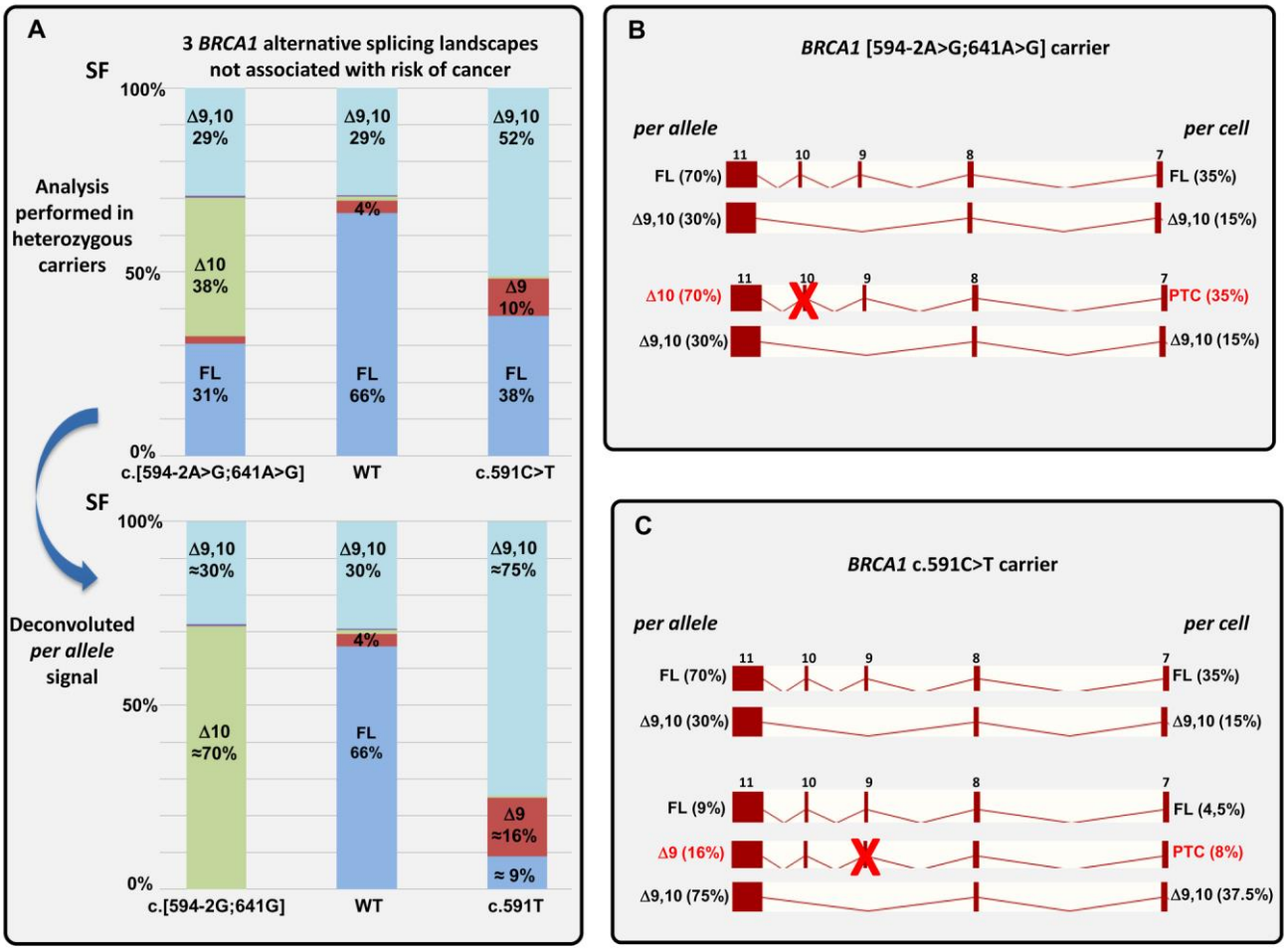
1 Vreeswijk, M.P. and van der Klift, H.M. (2012) Analysis and interpretation of RNA splicing alterations in genes involved in genetic disorders. *Methods Mol Biol*, **867**, 49-63.
2 Whiley, P.J., de la Hoya, M., Thomassen, M., Becker, A., Brandao, R., Pedersen, I.S., Montagna, M., Menendez, M., Quiles, F., Gutierrez-Enriquez, S. *et al.* (2014) Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin Chem*, **60**, 341-352.
3 Colombo, M., Blok, M.J., Whiley, P., Santamarina, M., Gutierrez-Enriquez, S., Romero, A., Garre, P., Becker, A., Smith, L.D., De Vecchi, G. *et al.* (2014) Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum Mol Genet*, **23**, 3666-3680.
4 Romero, A., Martin, M., Cheang, M.C., Lopez Garcia-Asenjo, J.A., Oliva, B., He, X., de la Hoya, M., Garcia Saenz, J.A., Arroyo Fernandez, M., Diaz Rubio, E. *et al.* (2011) Assessment of Topoisomerase II alpha status in breast cancer by quantitative PCR, gene expression microarrays, immunohistochemistry, and fluorescence in situ hybridization. *Am J Pathol*, **178**, 1453-1460.
5 Dosil, V., Tosar, A., Canadas, C., Perez-Segura, P., Diaz-Rubio, E., Caldes, T. and de la Hoya, M. (2010) Alternative splicing and molecular characterization of splice site variants: BRCA1 c.591C>T as a case study. *Clin Chem*, **56**, 53-61.
6 Raponi, M., Smith, L.D., Silipo, M., Stuani, C., Buratti, E. and Baralle, D. (2014) BRCA1 exon 11 a model of long exon splicing regulation. *RNA Biol*, **11**, 351-359.
7 Di Giacomo, D., Gaildrat, P., Abuli, A., Abdat, J., Frebourg, T., Tosi, M. and Martins, A. (2013) Functional analysis of a large set of BRCA2 exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. *Hum Mutat*, **34**, 1547-1557.
8 Gaildrat, P., Krieger, S., Di Giacomo, D., Abdat, J., Revillion, F., Caputo, S., Vaur, D., Jamard, E., Bohers, E., Ledemeney, D. *et al.* (2012) Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. *J Med Genet*, **49**, 609-617.
9 Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51-59.
10 Raponi, M., Douglas, A.G., Tammaro, C., Wilson, D.I. and Baralle, D. (2012) Evolutionary constraint helps unmask a splicing regulatory region in BRCA1 exon 11. *PLoS One*, **7**, e37255.


A
fAU
Δ9,10
Δ10
Puro+
Carrier 1
FL
Δ9
▼10p
Control
Puro-
B
Puro+
SF
70%
60%
50%
40%
30%
Δ9,10
FL
1.4%
1.2%
1%
0.8%
0.6%
0.4%
0.2%
0%
ns
**
Carrier 1
Controls
N=10
C
Puro+
40%
30%
20%
10%
0%
***
5%
4%
3%
2%
1%


A
Δ9,10 Puro+
SF
Carrier 1
Controls
20%
15%
10%
5%
0%
FL Cyclo+
SF
100%
80%
60%
40%
20%
0%


B
Δ9,10 Puro+
SF
c.591C>T
Carrier 1
Controls
40%
30%
20%
10%
0%
*
FL Puro+
SF
100%
80%
60%
40%
20%
0%
*


A
pCAS2-BRCA1-Exon10
PCMV
pCAS-KO1F
pCAS-2-R
Poly A
A
IVS9
147(1321)
E10
IVS10
173(985)
B
BamHI
MluI
WT
c.594-2A>C + c.641A>G
c.594-2A>C
c.641A>G
▼10p
FL
Δ10
HeLa
MCF7
HBL100
IGROV-1
Δ9
Δ10
Δ9,10
B
pB1
PCMV
ATG
pB1-V.8(2)-F
α-globin E1
E8
IVS8
460(2485)
E9
IVS9
379(1321)
E10
IVS10
684(985)
RT.1.8-11-R
E11p
E11q
IVS11
402(402)
E12
TAG
Poly A
+3insC


A
3 BRCA1 alternative splicing landscapes not associated with risk of cancer
SF
100%
50%
0%
Analysis performed in heterozygous carriers
Δ9,10 29%
Δ10 38%
FL 31%
Δ9,10 29%
4%
FL 66%
Δ9,10 52%
Δ9 10%
FL 38%
c.[594-2A>G;641A>G]
WT
c.591C>T
SF
100%
50%
0%
Deconvoluted per allele signal
Δ9,10 ≈30%
Δ10 ≈70%
Δ9,10 30%
4%
FL 66%
Δ9,10 ≈75%
Δ9 ≈16%
≈ 9%
c.[594-2G;641G]
WT
c.591T
B
BRCA1 [594-2A>G;641A>G] carrier
per allele
per cell
11 10 9 8 7
FL (70%)
FL (35%)
Δ9,10 (30%)
Δ9,10 (15%)
Δ10 (70%)
PTC (35%)
Δ9,10 (30%)
Δ9,10 (15%)
C
BRCA1 c.591C>T carrier
per allele
per cell
11 10 9 8 7
FL (70%)
FL (35%)
Δ9,10 (30%)
Δ9,10 (15%)
FL (9%)
FL (4,5%)
Δ9 (16%)
PTC (8%)
Δ9,10 (75%)
Δ9,10 (37.5%)

OXFORD UNIVERSITY PRESS | Human Molecular Genetics

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.

Journal:	*Human Molecular Genetics*
Manuscript ID	HMG-2015-TWB-01383.R1
Manuscript Type:	1 General Article - US Office
Date Submitted by the Author:	10-Feb-2016
Complete List of Authors:	de la Hoya, Miguel; Instituto de Investigación Sanitaria San Carlos, Laboratorio de Oncología Molecular Soukarieh, Omar ; Inserm U1079-IRIB, University of Rouen, Normandy Centre for Genomic and Personalized Medicine Lopez-Perolio, Irene; Instituto de Investigacion Sanitaria San Carlos (IdISSC), Hospital Clinico San Carlos, Molecular Oncology Laboratory Vega, Ana; SERGAS Grupo de Medicina Xenómica-USC, IDIS, CIBERER, Fundacion Publica Galega de Medicina Xenómica Walker, Logan; University of Otago, Department of Pathology van Ireland, Yvette; Leiden University Medical Center, Department of Human Genetics Baralle, Diana; Southampton General Hospital, Human Genetics Division Santamariña, Marta; Universidad de Santiago de Compostela, CIBERER, IDIS, Grupo de Medicina Xenómica-USC Lattimore, Vanessa; University of Otago, Department of Pathology Wijnen, Juul; Leiden University, Department of Human Genetics Whiley, Philip; QIMR Berghofer Medical Research Institute, Genetics and Computational Division Blanco, Ana; Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica-USC, CIBERER, IDIS Raponi, Michela; Academic Unit of Genetic Medicine, Division of Human Genetics Hauke, Jan; University of Cologne, Institute of Human Genetics Wappenschmidt, Barbara; Clinical Center University of Cologne, Division of Molecular Gynaeco-Oncology, Department of Gynaecology and Obstetrics Becker, Alexandra; University Hosptial Cologne, Center for Hereditary Breast and Ovarian Cancer, Center for Integrated Oncology (CIO), Medical Faculty van Overeem Hansen, Thomas; Copenhagen University hospital, Center for Genomic Medicine, Rigshospitalet Behar, Raquel; Instituto de Investigacion Sanitaria San Carlos (IdISSC), Hospital Clinico San Carlos, Molecular Oncology Laboratory Investigators, kConFab; Peter MacCallum Cancer Center, Peter MacCallum Cancer Center Niederacher, Dieter; Clinical Center University of Düsseldorf, Division of

	Molecular Genetics, Department of Gynaecology and Obstetrics; Arnold, Norbert; University Hospital Schleswig-Holstein, Division of Oncology, Department of Gynaecology and Obstetrics Dworniczak, Bernd; Westfalian Wilhelms-University of Muenster, Institute of Human Genetics Steinemann, Doris; Hannover Medical School, of Cell and Molecular Pathology Faust, Ulrike; University Hospital Tuebingen, Institute of Medical Genetics and Applied Genomics Rubinstein, Wendy; National Institute of Health, NIH Hulick, Peter; NorthShore University HealthSystem, Center for Medical Genetics Houdayer, Claude; Institut Curie and Université Paris Descartes, Service de Génétique and INSERM U830 Caputo, Sandrine; Institut Curie, Service de Génétique, Department de Biologie des Tumeurs Castera, Laurent; Centre Francois Baclesse, Laboratoire de Biologie et de Genetique du Cancer Pesaran, Tina; Ambry Genetics, Ambry Genetics Chao, Elizabeth ; Ambry Genetics, Ambry Genetics Brewer, Carole; Royal Devon and Exeter Hospital, Department of Clinical Genetics Southey, Melissa; University of Melbourne, Department of Pathology van Asperen, Christi; Leiden University Medical Center, Department of Human Genetics Singer, Christian; Medical University of Vienna, Dept of Obstetrics and Gynaecology Sullivan, Jan; Christchurch Hospital, Genetic Health Service NZ, South Island Hub Poplawski, Nicola; SA Pathology at the Women's and Children's Hospital, Adult Genetics Unit, South Australian Clinical Genetics Service Mai, Phuong; National Cancer Institute, Division of Cancer Epidemiology and Genetics Peto, Julian; LSHTM, NCDEU Johnson, Nichola; ICR, Breakthrough Breast Cancer Research Centre Burwinkel, Barbara; University Heidelberg, Department of Obstetrics and Gynecology; German Cancer Research Center, Molecular Epidemiology Surowy, Harald; University Heidelberg, Department of Obstetrics and Gynecology Bojesen, Stig; Herlev and Gentofte Hospital, Copenhagen General Population Study Flyger, Henrik; Herlev University Hospital, Department of Breast Surgery Lindblom, Annika; karolinska institutet, molecular medicine and suregery Margolin, Sara; Karolinska Institutet, Department of Oncology Pathology Chang-Claude, Jenny; German Cancer Research Center (DKFZ), Division of Cancer Epidemiology Rudolph, Anja; German Cancer Research Center (DKFZ), Division of Cancer Epidemiology Radice, Paolo; Fondazione IRCCS Istituto Nazionale dei Tumori, Department of Preventative and Predictive Medicine Galastri, Laura; Associazione Volontari Italiani Sangue (AVIS) comunale di Milano, Associazione Volontari Italiani Sangue (AVIS) comunale di Milano Olson, Janet; Mayo Clinic, Department of Health Science Research Hallberg, Emily; Mayo Clinic, Department of Health Sciences Research Giles, Graham; University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health Milne, Roger; Cancer Council Victoria, Cancer Epidemiology Centre Andrulis, Irene; Ontariao Cancer Genetics Network, Mount Sinai Hospital Glendon, Gord; Mount Sinai Hospital, Lunenfeld-Tanenbaum Research Institute Hall, Per; Karolinska Institutet, Department of Medical Epidemiology and

	Biostatistics Czene, Kamila; Karolinska Institutet, Department of Medical Epidemiology and Biostatistics Blows, Fiona; University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Oncology Shah, Mitul; University of Cambridge, Centre for Cancer Genetic Epidemilology and Dept of Public Health and Primary Care Wang, Qin; University of Cambridge, Centre for Cancer Genetic Epidemilology and Dept of Public Health and Primary Care Dennis, Joe; University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care Michailidou, Kyriaki; University of Cambridge, Department of Public Health and Primary Care McGuffog, Lesley; Strangeways Research Laboratory, Strangeways Research Laboratory Bolla, Manjeet; University of Cambridge, Centre for Cancer Genetic Epidemilology and Dept of Public Health and Primary Care Antoniou, Antonis; University of Cambridge, Department of Oncology Easton, Douglas; University of Cambridge, Centre for Cancer Genetic Epidemilology and Dept of Public Health and Primary Care Couch, Fergus; Mayo Clinic College of Medicine, Laboratory Medicine and Pathology Tavtigian, Sean V. ; University of Utah school of Medicine, Department of Oncological Sciences, Huntsman Cancer Institute Vreeswijk, Maaike; Leiden University Medical Center, Department of Human Genetics and Department of Pathology Parsons, Michael; Queensland Institute of Medical Research, Genetics and Computational Biology Meeks, Huong; University of Utah, Cancer Control and Population Sciences, Huntsman Cancer Institute Martins, Alexandra; Inserm U614, Faculty of Medicine Goldgar, David; University of Utah, Dermatology Spurdle, Amanda; QIMR Berghofer MRI, Genetics and Computational Biology Division
Key Words:	Splicing, BRCA1, consensus, classification, variant

Dr Anthony Wynshaw-Boris
Executive Editor
Human Molecular Genetics
2016-02-10

Dear Dr Wynshaw-Boris,

Re: HMG-2015-D-01383

Thank-you for the positive reviews of our manuscript entitled "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".

Please find below detailed responses to address the reviewer and editorial comments.

Reviewer 1

The study is well conducted, exhaustive and has important implications for diagnostics, making it necessary to reconsider the predicted pathogenicity of certain variants in the gene. I have only minor comments to be addressed.

The authors provide a similar frequency of c.594-2A>C in cases and controls from 11 studies in which there was at least one observation of the variant. The finding of 7 and 9 heterocigotes in 24,605 and 25,836 cases and controls respectively is considerably higher that the reported in ESP (1 heterocigote of 4300 individuals) or ExAc (2 of of 109096 alleles counted). The authors should comment on this, especially regarding the slightly higher frequency found in their control population.

While it does appear that the frequency of the c.594-2A>C variant is somewhat higher in the BCAC sample set than in public databases, it is impossible to say if we find too many or they find too few. We also note that quite different technologies to detect this variant are used in the BCAC data sets than in the two reference sets. This could also explain the lack of finding it in a higher frequency in the ExAC/ESP data sets. Further it is clear that the frequency of this variant may be somewhat population/geography dependent as it was observed in a subset of the BCAC European ancestry populations. We have observed this for other rare variants in BRCA1 and BRCA2. Although difficult to know for sure, we imagine that the sub-population structure of the non-Finnish European set of ExAC and the ESP dataset are similar to that of BCAC. For example it appears as if there is a higher proportion of American samples in the public reference datasets than in the BCAC set. We would imagine that the population structure of the US-based samples to have a higher degree of admixture than say, the BCAC set from Sweden and Denmark. Finally, we must admit that these differences could simply be due to chance. In any case, it is difficult to see how these minor frequency differences could explain our findings that the c.594-2A>C variant is not a pathogenic BRCA1 variant.

I would appreciate not to use the abbreviation "ASEV" to design an alternative splicing event. I don´t see the necessity of introducing a "new" term and is more confusing than helpful.

We have removed the abbreviation.

The discussion section is a bit long and reiterative. I would suggest making it more concise to highlight the important points; I don´t find it easy to follow in the present format.

A more concise discussion (1233 words vs. 1627 in the original manuscript) has been incorporated into the manuscript. We have eliminated references to a "previous rescue model vs. a new recue model", so that we hope now the discussion is less reiterative and easier to follow.

Reviewer 2

The study is thorough and well described and evidence present supports the conclusions drawn.

..parts of the manuscript use an overly conversational tone and some imprecise language e.g. "There was convincing evidence", "The vast majority" which the authors should remove to improve readability.

We have amended the text in response to this comment.

Background:

Use dbSNP IDs for both variants (rs80358033 and rs55680408)

dbSNP IDs have been incorporated into the results section (genetic studies), and also in the new supplemental Figure 1 panel A.

It would be useful to have a schematic figure to describe the position of the two variants discussed relative to exons 9 and 10.

A new schematic figure has been added as Supplemental Figure 1 panel A

Replace text "splicing aberrations and risk is not dichotomous" as is not entirely clear and (by my reading) suggests the opposite of what is meant.

We have amended the text to read: "These observations indicate that the relationship between splicing aberrations and increased risk is not straightforward…"

Results:

p6 - replace "the vast majority" with figures e.g. 21 of 25 tumours were ER positive.

We have replaced the text with the following: "The majority of tumors were ER positive (25/32)". Another tumor was grade 1 ER status unknown (thus also contributing to odds against pathogenicity), and only 6 tumors were ER-negative.

p7 - "2.77x10^6" should read "2.77x10^6:1"

Done.

p7 - again it would be helpful to have a schematic figure to describe the alternative splicing events being studied

Done. A new schematic figure has been added as Supplemental Figure 1 panel A.

p9 - "FL (transcripts containing" should be "FL transcripts (containing"

Amended.

p9 and throughout manuscript - replace "ex vivo" with "in vitro"

Since the terminology to describe such assays may be debated, we have rather elected to remove the term "ex vivo", and rather just refer to the technique as "minigene assays" without the need to qualify it as *ex vivo* or *in vitro*.

Discussion:

p11 - description of the delta9,10 rescue model is unclear - especially the use of "substituting" please clarify

p12 - replace "most predominant" with "predominant"

p12 - replace "highly predominant" with "predominant"

As suggested by reviewer 1, we have revised the discussion section. It is now shorter and, we believe, easier to follow. In this new version, the term "substituting" has been eliminated, and suggested replacements incorporated.

p12 - in the discussion of haplosufficiency the argument appears to be made that both the deletion resulting from the skipping of exons 9 and 10 and reduction of level of productive transcripts (assuming some degradation of delta10 transcripts by NMD) have no affect on the tumor suppressor activity of

BRCA1 - the evidence presented in the manuscript supports this for heterozygous cases - have any of the studies identified patients (or controls) homozygous for 594-2A>C;641A>G or compound heterozyotes with similar consequences to test this more rigorously?

We have not identified homozygotes for *BRCA1* c. 594-2A>C in our study, and would not expect to, given the rarity of this variant. *BRCA1* c. 594-2A>C was not identified in CIMBA as a compound heterozygote with a pathogenic *BRCA1* variant.

Methods:

p18 - replace “knock-downing” with “knocking down”

Replaced

Figure Legends:

p25 - replace “detected as well” with “detected”

Replaced

p25 - why is splicing fraction (SF) expression as a percentage? Splicing percentage (or an equivalent) is similar to the more widely used percent spliced in (PSI) metric and would be better description

PSI (percent Spliced In, or percent spliced in Index) is a metric developed for RNAseq data, which involves counting reads that align to known or predicted splice junctions, to estimate efficiency of splicing. It is calculated as the ratio between exon inclusion reads, and combined exon inclusion plus exclusion reads. This intron-centric method estimates the incidence of single-exon-skipping events. While PSI metrics can be probably used for dPCR data, we believe this metric is not appropriate for the data variables provided by capillary EP (peak areas) or qPCR (Ct values). Since we express capillary EP and qPCR as splicing fraction, we prefer for the sake of internal consistency to also express dPCR data as splicing fraction. Further, PSI is intended to be informative in simple situations such as cassette exons where only two splicing events (exclusion/inclusion) are possible. In our study, the situation is far more complex, with up to five different splicing events competing. Just one example, exon 9 inclusion can reflect up to 3 different transcripts (full-length, delta10, and ins10p), and exon 9 exclusion can reflect up to 2 different transcripts (delta10 and delta9,10). For that reason, exon9 inclusion/exclusion rate is not very informative.

p27 - replace “with the only exception” with “with the exception”

Replaced

p28 - remove “notoriously”

Removed

p28 - replace “events are highlighted in red” with “events are highlighted with a red cross”

Replaced

p29 - remove “last but not least”

Removed

Figures:

fig 1 - missing Y-axis labels in panel A

We have included a Y-axis label (fAU, fluorescence intensity in arbitrary units), and described this in the figure legend.

fig 3 - inconsistency in cell-line name - IGROV1 in main text and legend - IgOv1 in panel B label

Amended - now IGROV-1 both in text and legend.

Supplemental Methods:

p1 - “different commercially available kits” - please include details

Now we have included details as follows: “For RT-PCR reactions, different contributors used different commercially available kits (SuperScriptII Reverse Transcriptase, Invitrogen, Carlsbas, CA; High Capacity RNA-to-cDNA Mater Mix, Applied Biosystems, Foster City, CA; Primescript RT reagent kit,

Takara Biotechnology, Shiga, Japan; Transcriptor High Fidelity cDNA synthesis Kit, Roche Applied Science, Mannheim, Germany; One-Step RT-PCR kit, Qiagen, Hilden, Germany) in combination with oligodT and/or random hexamers."

p2 - "various ABIPRIS Genetic Analyzers" - please include details

Now we have included details as follows: "Depending on the contributing laboratory, capillary electrophoresis analyses were performed with 3100, 3130, 3130XL, 3500XL, or 3730XL ABI PRISM Genetic Analyzers (Applied Biosystems, Foster City, CA) using POP-7 filled capillary arrays of 36 or 50 cm. Size-calling and peak areas were analyzed with GeneMapper v4.0 or GenScan v3.7 software (both from Applied Biosystems, Foster City, CA), or GeneMarker 2.4 (Softgenetics LLC, State College, PA). As internal size-standard, we used LIZ-500 or LIZ-1000 (Applied Biosystems, Foster City, CA)."

Table 1 - reduce font size as text wrapping makes expected peak sizes confusing to read

Done

Supplemental Data:

p2 - colours and shadow on graph make it difficult to read text - please modify to use light shades and remove shadow

Modified. In addition, we have modified a similar graph in Figure 4 for consistency.

p3 - Panel A Carrier 8 - although this sample is not reported to include Puromycin or Cyclohexamide the delta10 SF is as high as all Puro+ and Cyclo+ samples - is this a missing label or genuinely high expression in Puro/Cyclo- sample - in which case it would be worth remarking on in text.

Carrier 8 corresponds to RNA extracted from fresh blood directly collected in tubes containing a RNA preserving solution (PAXgene tubes, Qiagen), hence not involving cell culture/puromycin treatment. Therefore, we conclude that our data represents genuinely high expression of delta10 in this particular sample. We have now specifically remarked on this fact in the text corresponding to supplemental figure 2 as follows: "one fresh blood sample (PAXgene system for blood sampling, no puromycin/cycloheximide treatment) named Carrier 8, and its corresponding control. Note the high level of Δ10 transcripts in Carrier 8, despite the fact that NMD is not inhibited by the PAXgene system"

Further, thanks to the reviewer comment, we realized that there was a minor error related with the Carrier 8 boxplot (related with data input into the SPSS data editor). This error has now has been fixed, and compared with the previous version of the box plot:

1) delta10 expression in the carrier is now slightly lower (36% vs. 39%), and

2) it is now apparent that the data corresponding to the 3 control samples is very similar, but not identical.

These corrections do not affect at all the interpretation of results.

Editorial and formatting changes:

TITLE PAGE:

___Complete author names (first name, middle initial, (if required) and last name) should be listed separately from their institutions/affiliations. Use numbers to reference one to the other. Do not use academic degrees (i.e. MD, PhD, MSc).

Compliant

___The corresponding author should be designated with an asterisk (*). Address, telephone, FAX, and email address should be listed. The publisher allows only one corresponding author.

We have replaced the superscript number with an asterisk. We ask respectfully for deviation of publishing restrictions to one corresponding author for this paper spanning splicing and genetics, since we think that the combined expertise of the two proposed corresponding authors will facilitate our ability to respond in most informative manner to questions arising from this publication.

___Financial/funding information should be listed in the acknowledgement section.

Amended

MANUSCRIPT:

___The section sequence order should be: Title Page, Abstract, Introduction, Results, Discussion, Materials and Methods, Acknowledgements, Conflict of Interest Statement, References, Legends to Figures, Tables and, finally, Abbreviations. Figures should be saved as separate high-resolution image files. Abbreviations are used for copy-editing purposes only.

Amended

___Running title, key words, word count, figure and/or table count should be deleted from the manuscript. The journal does not publish this information.

Compliant

___Use TIMES NEW ROMAN font for all text.

Amended

___Double-space the text.

Amended

REFERENCES:

___Type complete page numbers i.e. 345-349 NOT 345-9 or 345-49.
___Punctuate ALL abbreviated words in journal titles, i.e. Hum. Mol. Genet. NOT Hum Mol Genet.
___Abbreviate journal titles.
___Italicize journal titles.
___Volume number should be keyed in boldface.
___Do not use issue numbers.
___List 10 authors before adding et al.
___Use proper spacing between author names and initials.
___Use commas (,) not colons (:) in the references.
___Article titles of references are required.

Amended where relevant. Please note some journals do not provide page numbers, but rather e-references.

SUPPLEMENTAL DATA:

___Supplemental data should be uploaded in a separate file from that of the main manuscript.

Compliant

FIGURES

The four figures are provided in TIFF format with the proper size and dpi

We hope these responses and edits meet with your approval. For the main text, we have provided a document with changes tracked/noted, and another clean document. We have also uploaded scans of two conflict of interest statements, relevant to this manuscript.

Sincerely

Amanda Spurdle and Miguel de la Hoya

Corresponding authors

Amanda.Spurdle@qimrberghofer.edu.au; mdhoya@hotmail.com

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.

Miguel de la Hoya[1,*], Omar Soukarieh[2], Irene López-Perolio[1], Ana Vega[3], Logan C. Walker[4], Yvette van Ierland[5], Diana Baralle[6], Marta Santamariña[7], Vanessa Lattimore[4], Juul Wijnen[5, 8], Philip Whiley[9], Ana Blanco[3], Michela Raponi[6], Jan Hauke[10], Barbara Wappenschmidt[10], Alexandra Becker [10], Thomas v. O. Hansen[11], Raquel Behar[1], KConFaB investigators[12], Diether Niederacher[13], Norbert Arnold[14], Bernd Dworniczak[15], Doris Steinemann[16], Ulrike Faust[17], Wendy Rubinstein[18], Peter J. Hulick[19], Claude Houdayer[20], Sandrine M. Caputo[21], Laurent Castera[22], Tina Pesaran[23], Elizabeth Chao[23], Carole Brewer[24], Melissa C. Southey[25], Christi J. van Asperen[5], Christian F. Singer[26], Jan Sullivan[27], Nicola Poplawski[28, 29], Phuong Mai[30], Julian Peto[31], Nichola Johnson[32, 33], Barbara Burwinkel[34, 35], Harald Surowy[34, 35], Stig E. Bojesen[36-38], Henrik Flyger[39], Annika Lindblom[40], Sara Margolin[41], Jenny Chang-Claude[42, 43], Anja Rudolph[42], Paolo Radice[44], Laura Galastri[45], Janet E. Olson[46], Emily Hallberg[46], Graham G. Giles[47, 48], Roger L. Milne[47, 48], Irene L. Andrulis[49, 50], Gord Glendon[49], Per Hall[51], Kamila Czene[51], Fiona Blows[52], Mitul Shah[52], Qin Wang[53], Joe Dennis[53], Kyriaki Michailidou[53, 54], Lesley McGuffog[55], Manjeet K. Bolla[53], Antonis C. Antoniou[55], Douglas F. Easton[52, 53], Fergus J. Couch[46, 56], Sean Tavtigian[57], Maaike Vreeswijk[5], Michael Parsons[9], Huong Meeks[57], Alexandra Martins[2], David E. Goldgar [58, 59], Amanda B. Spurdle[9, 59, *]

[1] Molecular Oncology Laboratory, Instituto de Investigacion Sanitaria San Carlos (IdISSC), Hospital Clinico San Carlos, Madrid, Spain.

[2] Inserm U1079-IRIB, University of Rouen, Normandy Centre for Genomic and Personalized Medicine, Rouen, France.

[3] Fundacion Publica Galega de Medicina Xenómica-SERGAS Grupo de Medicina Xenómica-USC, IDIS, CIBERER, Santiago de Compostela, 15706, Spain.

[4] Department of Pathology, University of Otago, Christchurch, New Zealand.

[5] Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.

[6] Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

[7] CIBERER, Grupo de Medicina Xenómica-USC, Universidade de Santiago de Compostela, Fundacion Galega de Medicina Xenómica (SERGAS), Santiago de Compostela, Spain.

[8] Department of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands.

[9] Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, 4006, Australia.

[10] Center for Hereditary Breast and Ovarian Cancer, Center for Integrated Oncology (CIO), Medical Faculty, University Hosptial Cologne, Cologne, Germany.

[11] Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, DK-2100, Denmark.

[12] Peter MacCallum Cancer Center, The University of Melbourne, Melbourne, 3002, Australia.

[13] Department of Gynaecology and Obstetrics, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Germany.

[14] Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Campus Kiel, Christian-Albrechts University Kiel, Germany.

[15] Institute of Human Genetics, University of Münster, Münster, Germany.

[16] Institute of Human Genetics, Hannover Medical School, Hannover, Germany.

[17] Institute of Medical Genetics and Applied Genomics, Univeristy Hosptial Tuebingen, Tuebingen, Germany.

[18] National Institutes of Health, Bethesda, MD, USA.

[19] Center for Medical Genetics, NorthShore University Health System, University of Chicago Pritzker School of Medicine, 1000 Central Street, Suite 620, Evanston, IL, USA.

[20] Service de Génétique, Department de Biologie des Tumeurs, Institut Curie and INSERM U830, centre de recherche de l'Institut Curie, Paris, and Universite Paris Descartes, Sorbonne Paris Cite, Paris, France.

[21] Service de Génétique, Department de Biologie des Tumeurs, Institut Curie, Paris, France.

[22] Centre Francois Baclesse, Laboratoire de Biologie et de Genetique du Cancer, 14076, Caen, France.

[23] Ambry Genetics, 15 Argonaut, Aliso Viejo, CA 92656, USA.

[24] Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK.

[25] Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Vic, 3010, Australia.

[26] Dept of Obstetrics and Gynaecology, Medical University of Vienna, Vienna, Austria, Waehringer Guertel 18-20, A 1090 Vienna, Austria.

[27] Genetic Health Service NZ, South Island Hub, Christchurch Hospital, New Zealand.

[28] Adult Genetics Unit, South Australian Clinical Genetics Service, SA Pathology at the Women's and Children's Hospital, 72 King William Road, North Adelaide, SA 5067, Australia.

[29] University Department of Paediatrics, University of Adelaide, North Terrace, Adelaide, SA 5000, Australia.

[30] Clinical Genetics Branch,DCEG, NCI, NIH, 9609 Medical Center Drive, Room 6E-454, Bethesda, MD, USA.

[31] London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK.

[32] Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, SW3 6JB, UK.

[33] Division of Breast Cancer Research, The Institute of Cancer Research, London, SW3 6JB, UK.

[34] Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, University of Heidelberg, Heidelberg, 69120, Germany.

[35] Molecular Epidemiology Group, German Cancer Research Center, DKFZ, Heidelberg, 69120, Germany.

[36] Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen, Denmark.

[37] Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, 2730, Denmark.

[38] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 1165, Denmark.

[39] Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Denmark.

[40] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, SE-171 77, Sweden.

[41] Department of Oncology Pathology, Karolinska Institutet, Sweden.

[42] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany.

[43] University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

[44] Unit of "Molecular bases of genetic risk and genetic testing", Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy.

[45] Associazione Volontari Italiani Sangue (AVIS) comunale di Milano, Milano, Italy.

[46] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905, USA.

[47] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Vic, 3010, Australia.

[48] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Vic, 3004, Australia.

[49] Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada.

[50] Department of Molecular Genetics, University of Toronto, Canada.

[51] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE-171 77, Sweden.

[52] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK.

[53] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK.

[54] Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Cyprus.

[55] Strangeways Research Laboratory, Worts Causeway,Cambridge CB1 8RN, UK.

[56] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, 55905, USA.

[57] Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine.

[58] Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah, USA.

[59] These authors contributed equally.

[*] Corresponding authors.

Amanda Spurdle: QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, 4006, Australia; Tel +617 3362 0371; Email, amanda.spurdle@qimr.edu.au

Miguel de la Hoya: Laboratorio de Oncología Molecular, Instituto de Investigación Sanitaria San Carlos (IdISSC), Hospital Clínico San Carlos, c/Martín Lagos s/n, Madrid 28040, Spain; Tel +34 913303348; Fax +34 913303544; Email, mhoya@hotmail.com

Abstract

A recent analysis using family history weighting and co-observation classification modeling indicated that *BRCA1* c.594-2A>C (IVS9-2A>C), previously described to cause exon 10 skipping (a truncating alteration), displays characteristics inconsistent with those of a high risk pathogenic *BRCA1* variant. We used large-scale genetic and clinical resources from the ENIGMA, CIMBA and BCAC consortia to assess pathogenicity of c.594-2A>C. The combined odds for causality considering case-control, segregation, and breast tumor pathology information was 3.23×10^{-8}. Our data indicate that c.594-2A>C is always in *cis* with c.641A>G.

The spliceogenic effect of c.[594-2A>C;641A>G] was characterized using RNA analysis of human samples and splicing minigenes. As expected, c.[594-2A>C; 641A>G] caused exon 10 skipping, albeit not due to c.594-2A>C impairing the acceptor site but rather by c.641A>G modifying exon 10 splicing regulatory element(s). Multiple blood-based RNA assays indicated that the variant allele did not produce detectable levels of full-length transcripts, with a *per allele BRCA1* expression profile comprised of ≈70-80% truncating transcripts, and ≈20-30% of in-frame Δ9,10 transcripts predicted to encode a BRCA1 protein with tumor suppression function.

We confirm that *BRCA1*c.[594-2A>C;641A>G] should not be considered a high-risk pathogenic variant. Importantly, results from our detailed mRNA analysis suggest that BRCA-associated cancer risk is likely not markedly increased for individuals who carry a truncating variant in *BRCA1* exons 9 or 10, or any other *BRCA1* allele that permits 20-30% of tumor suppressor function. More generally, our findings highlight the importance of assessing naturally occurring alternative splicing for clinical evaluation of variants in disease-causing genes.

Introduction

Sequence variants that alter the highly conserved intronic dinucleotides at splice donor and acceptor sites of high-risk disease predisposition genes are often assumed to be pathogenic, due to their high likelihood to alter RNA splicing. Although such variants will almost certainly lead to disruption of normal splicing patterns, the exact nature of the resulting alternate splicing patterns cannot be reliably predicted. Indeed, a standardized classification scheme recently developed for mismatch repair gene variants through consensus across multiple international sites (1) proposes that mRNA assay and/or clinical data are necessary to upgrade dinucleotide donor and acceptor variant classification from "likely pathogenic" to "pathogenic".

The dinucleotide acceptor site variant *BRCA1* c.594-2A>C (also known as IVS9-2A>C) has recently been reported associated with clinical characteristics inconsistent with a high risk of cancer expected for a pathogenic *BRCA1* variant (2). Previous RNA analyses of carriers of *BRCA1* c.594-2A>C indicate that this variant is associated with an aberrant mRNA profile (3, 4), including production of exon 10 deleted out-of-frame transcripts. These observations indicate that the relationship between splicing aberrations and increased risk is not straightforward, and pose the question of which measures of mRNA transcript dysregulation best reflect variant pathogenicity, considering recommendations already published by the ENIGMA Splicing Working Group (5). We undertook a study to assess level of risk associated with *BRCA1* c.594-2A>C using segregation and large-scale case-control analysis, and detailed mRNA analyses correlating genotype with aberrant mRNA profiles.

Results

Genetic studies

Characteristics of *BRCA1* c.594-2A>C variant carriers identified in BCAC, CIMBA, and ENIGMA are detailed in Supplementary Tables 1 and 2.

BRCA1 c.594-2A>C (rs80358033) was identified in 7/24,605 invasive breast cancer cases and 9/25,836 controls, when including only the 11 studies with at least one observation (Supplementary Table 1). Standard case-control analysis yielded an odds ratio (OR) of 0.82 (95% CI 0.26-2.47), which was little different after adjustment for principle components (OR 0.83, 95% CI 0.41-2.24). However, some studies indicated that they had performed *BRCA1/2* mutation screening of cases and may have excluded cases with pathogenic variants. Since *BRCA1* c.594-2A>C has generally been assumed to be pathogenic on the basis of its location at a splice acceptor site, this could create a bias due to preferential exclusion of c.594-2A>C carriers cases but not controls. After exclusion of four studies that did such genetic testing, we were left with 5/20,992 cases and 6/22,332 controls that carried the c.594-2A>C variant (See Supplementary Table 1), yielding a revised OR of 0.87 (95% CI 0.26-2.86) after adjustment for principle components. The odds for causality based on carrier frequency and ages at diagnosis/interview in these cases and controls was 7.3×10^{-5} (equating to an odds against pathogenicity of 13770:1). The case-control findings demonstrate that the *BRCA1* c.594-2A>C variant is clearly not associated with a high risk of breast cancer, and is unlikely to be associated with even a moderate (~3-5-fold) risk of breast cancer. There were 15 *BRCA1* c.594-2A>C carrier individuals from 13 families identified in the CIMBA dataset through genotyping with the iCOGS array. It was confirmed with the submitting sites that none of these individuals carried another pathogenic variant in *BRCA1*, and that 8 of these families overlapped with those identified via ENIGMA while the proband for another family was also recruited into BCAC. Overall, information for segregation analysis was available for 14 probands from ENIGMA/CIMBA (Supplementary Table 1), and breast tumor pathology information for 32 cases from ENIGMA, CIMBA

or BCAC (Supplementary Tables 1 and 2). The combined odds for causality based on segregation analysis, assuming *BRCA1* age-specific risks of breast and ovarian cancer as estimated in the large study of Antoniou et al (6), was 0.10 (ranging from 0.02 to 6.85 for individual families). The breast tumor pathology features of variant carriers were not consistent with those found commonly for high-risk *BRCA1* pathogenic variant carriers. The majority of tumors were ER positive (25/32), and the odds for causality based on pathology information was 4.98×10^{-6} (200994:1 against causality).

After contacting the submitting centres and through re-investigation of original genetic test results, the *BRCA1* exonic variant rs55680408 (c.641A>G, p.Asp214Gly) was confirmed to be present in all ENIGMA/CIMBA c.594-2A>C families included in the final analysis, and another 13 c.594-2A>C carriers identified by Ambry Genetics that were excluded from analysis due to lack of relevant clinical information. Specifically, review of genetic testing data by Ambry Genetics identified a total of 20 carriers of *BRCA1* c.[594-2A>C; 641A>G] from >213,000 tests, including two siblings and a mother-daughter pair; there was clear evidence that the alleles were *in cis* from next generation sequencing reads, and neither allele was observed alone in 2636 unrelated parent exomes. Further, all carrier individuals from BCAC were shown to share the same *BRCA1* haplotype (data not shown). Based on the haplotype and genotype information, it was assumed that all *BRCA1* c.594-2A>C carriers in the BCAC dataset were also carriers of c.641A>G (p.Asp214Gly). Considering case-control, segregation and pathology information, the combined odds for causality was 3.61×10^{-11} (i.e. 2.77×10^{10}:1 *against* causality). These results indicate that individuals carrying *BRCA1*c.[594-2A>C; 641A>G] (Supplemental Figure 1A) should be counselled as *not* having a high risk of *BRCA1*-associated disease.

Splicing studies

Comprehensive characterization of BRCA1 alternative splicing landscape in c.[594-2A>C; 641 A>G] carriers by capillary electrophoresis and high throughput RNA sequencing (RNAseq).

To search for a plausible biological mechanism explaining the lack of evidence for an increased cancer risk in *BRCA1* c.[594-2A>C; 641A>G] carriers, we first performed a comprehensive characterization of the *BRCA1* alternative splicing landscape in the vicinity of exon 10.With this aim, we performed a series of capillary electrophoresis analyses on RNAs obtained from lymphoblastoid cell lines (LCLs) (see methods). We have shown previously that this approach is highly sensitive, allowing comprehensive identification, characterization, and semi-quantification of alternative splicing (4, 7). Experiments performed with two combinations of forward and reverse primers located in exons 8 and 11detected up to five different alternative splicing events both in LCLs from one c.[594-2A>C; 641 A>G] carrier (Carrier 1) and healthy controls (Figure 1), including 3 in-frame (full-length (FL), Δ9,10, and ▼10p), and two out-of-frame (Δ9 and Δ10). All but ▼10p (r.594-21_594-1*ins*) have been described previously as naturally occurring *BRCA1* alternative splicing events in control samples (7). No c.[594-2A>C; 641A>G] specific events were identified. Overall, experiments conducted in the presence of puromycin (Puro+ experiments in Figure 1 and Supplemental Figure 1) indicated that Δ9,10 splicing fraction ($\Delta 9,10^{SF}$) is similar in Carrier 1 and Controls (≈29%), $\Delta 10^{SF}$ is considerably higher (≈38% vs. ≈1%), and FL^{SF} much lower (≈31% vs. ≈66%). $\Delta 9^{SF}$ (<3%) and $▼10p^{SF}$ (<1%)were rather minor alternative splicing events in all tested samples. As expected, Puro- experiments measured higher $\Delta 9,10^{SF}$ in Carrier 1 than in Controls (Figure 1A and Supplemental Figure 1) due to a drop in $\Delta 10^{SF}$, probably reflecting nonsense mediated decay (NMD) degradation of out-of-frame Δ10 transcripts.

Overall, findings were confirmed by comparable experiments performed by other contributing centers (Supplemental Figure 2), assaying up to eight individual *variant allele* carriers identified in four unrelated families and 3 different types of samples (LCLs, Leukocytes (LEUs), and fresh whole blood); there was similar $\Delta 9,10^{SF}$ in Carriers and Controls (range 20-30% depending on specific protocols and/or sample used for experiments), and a significant increase of $\Delta 10^{SF}$ (with corresponding decrease of FL^{SF}) in Carriers. Complementary analyses performed in the subpopulation of *BRCA1*Δ11q transcripts were coincident, with similar $(\Delta 9,10+\Delta 11q)^{SF}$ in Carriers and Controls, and a significant increase of $(\Delta 10+\Delta 11q)^{SF}$ (with corresponding decrease of $\Delta 11q^{SF}$) in Carriers (Supplemental Figure 2C). Incidentally, our data supports ▼10p as a naturally occurring *BRCA1* alternative splicing event not previously reported, probably due to its very low SF. Capillary electrophoresis findings (in particular the lack of *variant allele* specific transcripts, and the detection of ▼10p in Controls) were confirmed by RNA-seq experiments (Supplemental Figure 3).

Quantitative analyses combined with alternative splicing event specific biallelic expression analyses confirms that c.[594-2A>C; 641A>G] modifies the BRCA1 alternative splicing landscape, but not the overall BRCA1 expression level.

The comprehensive analysis of the *BRCA1* alternative splicing landscape described above did not provide an obvious explanation for why c.[594-2A>C; 641A>G] carriers do not display features of a standard pathogenic *BRCA1* variant. Yet, the absence of carrier-specific transcripts prompted us to speculate that it is perhaps the actual level of naturally occurring in-frame transcripts in *variant allele* carriers that may explain the genetic findings, in particular levels of FL and Δ9,10 transcripts given that▼10p transcript levels were very low. Since capillary electrophoresis is a semi-quantitative approach, we decided to perform further analyses with quantitative PCR (qPCR) and digital PCR (dPCR) that, overall, confirmed

capillary electrophoresis findings (Figure 2). qPCR absolute quantification of individual alternative splicing events in Carrier 1 estimated for $\Delta 9,10^{SF}$ a value of 19%±0.9, in the upper-limit of Controls (ranging from 4% to 17%)(Figure 2A, left), together with an obvious reduction of FL transcripts (Figure 2A, right). In addition to Δ9,10, three other naturally occurring in-frame alternative splicing events involving exon 10 and/or nearby exon 11 have been described, namely Δ9_11, Δ11, and Δ11q (7). We used qPCR absolute quantification to estimate the SF of these alternative splicing events, detecting an increase of $(\Delta 9,10+\Delta 11q)^{SF}$ in Carrier 1 (9%±0.8) if compared with Controls (average of 7%). No differences were observed with regard to $\Delta 11^{SF}$ and $\Delta 9\_11^{SF}$ (Supplemental Figure 4). Similarly, dPCR analyses (Figure 2B) revealed a modest increase of $\Delta 9,10^{SF}$ in Carrier 1 (24%±0.9) if compared with Controls (average of 17%), together with a 50% reduction of FL^{SF} that is fully compatible with lack of FL transcripts arising from the *variant allele*.

Alternative splicing event specific reverse transcription and PCR amplification (RT-PCR) sequencing experiments (Supplemental Figure 5) performed in carriers 3 to 5 (from one Dutch family) confirmed that Δ9,10 expression is biallelic, whereas Δ10 expression is essentially monoallelic (arising from the *variant allele*). Neither qPCR absolute quantification (Supplemental Figure 4A), nor biallelic expression analysis (Supplemental Figure 5B) suggested higher overall *BRCA1*expression level in c.[594-2A>C; 641A>G] carriers. Yet, to further exclude this possibility we performed dPCR analyses of *BRCA1*Δ9,10 and FL expression relative to *BRCA2* (Supplemental Figure 6). The data indicated that Δ9,10 relative expression level is similar in LCLs from Carrier 1 and Controls, while FL expression level shows a 50% reduction, again supporting that the *variant allele* is not producing FL transcripts.

Taken together, capillary electrophoresis analyses of RT-PCR products, RNAseq, qPCR, dPCR and alternative splicing event specific sequencing experiments supported a model in which the *variant allele* does not produce novel *BRCA1* transcripts, nor increases overall *BRCA1*expression level, but rather

substitutes FL transcripts (containing exons 9 and 10) with out-of-frame Δ10 transcripts, such that the contribution of in-frame Δ9,10 transcripts to the overall expression level is *similar* or *slightly higher* (see Figure 2, Supplemental Figure 2) to that observed in wild-type (WT) alleles. Of note, according to our data the overall model is also probably true in the subset of *BRCA1*Δ11q transcripts (see Supplemental Figures 2C and 4B). According to this model, *BRCA1*Δ9 (out-of-frame) and *BRCA1* ▼10p (in-frame) contribution to the overall expression level are very low both in *variant* and *WT* alleles (see Figure 2B and 2C), and hence irrelevant to explain the lack of risk observed in *variant allele* carriers.

Splicing reporter minigene analyses reveal that c.641A>G is causing exon 10 skipping in c.[594-2A>C; 641 A>G] carriers.

We also performed minigene assay experiments to dissect the contribution of the individual variants c.594-2A>C and c.641A>G to the splicing pattern observed in *variant allele* carriers. Experiments were performed with two minigene assays (pCAS2-BRCA1-Exon10, and pB1). A schematic representation of these reporter minigenes is shown in Figure 3. pCAS2-BRCA1-Exon10 and pB1 experiments performed in HeLa cells, as well as pB1 experiments performed in breast (MCF7 and HBL100) and ovarian (IGROV-1) cell lines, revealed that both c.594-2A>C and c.641A>G impair normal exon 10 splicing, albeit with different outcomes (Figure 3). pCAS2-BRCA1-Exon10 c.594-2A>C and pB1 c.594-2A>C predominantly produced▼10p transcripts, but also a minor amount of Δ10 transcripts (Figure 3A), a finding confirming previous pSPL3-BRCA1-Exon10 experiments performed in COS-7 cells (8). By contrast, pCAS2-BRCA1-Exon10 c.641A>G and pB1 c.641A>G mostly produced Δ10 but no detectable▼10p. The finding that c.641A>G causes exon 10 skipping albeit being located outside the splice site, suggests that this variant disturbs the regulation of exon 10 splicing, probably by destroying splicing enhancer elements and/or by creating splicing silencer elements, a hypothesis supported by an *in silico* analysis based on ESRseq scores (Supplemental Figure 7A). The presence of regulatory

mechanisms underlying *BRCA1* exon 10 splicing was further supported by small interfering RNA (siRNA) experiments performed in MDA-MB231 cells showing that endogenous *BRCA1* depends on Tra2-β for exon 10 inclusion (Supplemental Figure 7B). Double mutant pCAS2-BRCA1-Exon10 c.[594-2A>C; 641 A>G] and pB1 c.[594-2A>C; 641 A>G] experiments mimicking the *variant* allele observed *in vivo* produced detectable levels of both Δ10 and ▼10p, with Δ10 being the predominant outcome in all cell lines tested (Figure 3).

Discussion

In the present study we have demonstrated that c.[594-2A>C; 641A>G] carriers (but not necessarily carriers of a potential *BRCA1* allele in which c.594-2A>C is not linked to c.641A>G) should not be considered at high-risk of developing *BRCA1*-associated cancers. The finding is remarkable, since the variant allele causes exon 10 skipping, a frame-shift alteration. In addition, we propose a plausible biological mechanism underlying the finding, the so-called *BRCA1* Δ9,10 *rescue model*, and we show the relevance of the findings for developing disease gene variant classification algorithms.

The first study addressing the spliceogenic impact of *BRCA1*c.594-2A>C demonstrated an association with exon 10 skipping (3), supporting the initial pathogenic classification by Myriad Genetics (2). Here we confirm exon 10 skipping in c.594-2A>C carriers, and we show that contrary to expectations this splicing alteration is not driven by c.594-2A>C, but rather by the linked variant c.641A>G. Further, we show that the *variant allele* does not produce full-length (FL) transcripts, nor other in-frame transcripts apart from *normal* levels of Δ9,10 and residual levels of ▼10p transcripts. These findings lead us to conclude that Δ9,10 transcripts arising from the *variant allele* confer sufficient tumor suppressor activity *in vivo* to compensate for the lack of FL transcripts. To be more precise, the combined genetic and splicing data lead us to formulate a Δ9,10 *rescue model* in which *BRCA1* alleles with an associated $\Delta 9,10^{SF}$ of ≈20%-30% (as measured in blood related samples) confer tumor suppressor haplosufficiency (Figure 4). The actual value is probably closer to 20% than to 30% (according both to qPCR and dPCR estimations in Carrier 1, and to capillary electrophoresis estimations in Carriers 2 to 8), but at any rate is very similar to that observed in control samples. The finding that Δ9,10 is a predominant alternative splicing event not only in blood derived samples but also in clinically relevant tissues such as breast and ovary (Supplemental Figure 8) is critical to support our *rescue model* for both breast and ovarian cancer. Indeed, family history of breast and/or ovarian cancer is a key criterion for genetic testing for most

participating ENIGMA and CIMBA sites, and segregation analysis modelled both breast and ovarian cancer risk, providing no indication that *BRCA1* c.594-2A>C (IVS9-2A>C) could be associated with increased ovarian cancer risk only. Further, similar to our findings reported for breast cancer, case-control data from a parallel study by the Ovarian Cancer Association Consortium does not support an association with ovarian cancer risk, with *BRCA1* c.594-2A>C identified in 2/16,121 cases and 4/26,167 controls (OCAC, unpublished data). Note that the *BRCA1* Δ9,10 *rescue model* predicts lack of breast and ovarian cancer risk not only for *BRCA1* variants causing exon 10 skipping (or exon 9 skipping)*, but* to any loss-of-function mutation in exons 9 or 10 (nonsense or frame shift mutations), provided that the mutant allele produces *normal* levels of Δ9,10 transcripts (Figure 4).

Evidently, the *BRCA1* Δ9,10 *rescue model* presumes that Δ9,10 transcripts encode a protein isoform ($BRCA1^{p.Gly183\_Lys223del}$) that has tumor suppressor activity. To our knowledge, this BRCA1 isoform (lacking only 41 out of 1863 amino acid residues) has not been detected *in vivo*, nor functionally characterized *in vitro*, but tumor suppressor activity is fully compatible with structural considerations: 1) the 41 missing residues are unlikely to affect protein folding, since they are embedded in an intrinsically disordered protein region spanning amino acids 170-1649 (9); 2) $BRCA1^{Gly183\_Lys223del}$ includes all known functional domains/residues critical for tumor suppression, including the RING domain (spanning amino acids 2-103) that mediates binding to BARD1, an obligated heterodimer partner *in vivo* (10). Interestingly, $BRCA1^{p.Gly183\_Lys223del}$ lacks some residues critical for E3 ligase activity (11), a *BRCA1* function that appears to be dispensable for tumor suppression (12, 13). Yet, the most compelling argument supporting $BRCA1^{p.Gly183\_Lys223del}$ tumor suppressor activity stems from combined genetic and splicing analyses of *BRCA1* c.591C>T (rs1799965). This variant, *also* not associated with the high risk of cancer expected for a pathogenic *BRCA1* variant (current odds for causality of 8.50×10^{-16} based on segregation and pathology information, ENIGMA unpublished data), expresses mostly Δ9,10 transcripts, a significant proportion of out-of-frame Δ9 transcripts, and very few FL transcripts (14), strongly pointing to $BRCA1^{p.Gly183\_Lys223del}$ as a protein with tumor suppressor function. As far as we know, the only cancer

predisposition gene for which a similar alternative splicing rescue model has been proposed is the tumor suppressor *adenomatous poliposis coli* (*APC*) gene, albeit in this case loss of function variants in the alternatively spliced region of *APC* exon 9 are not associated with lack of risk, but with a milder phenotype, termed attenuated familial adenomatous polyposis(15).

The *BRCA1*Δ9,10 *rescue model* highlights the often neglected relevance of naturally occurring alternative splicing in the clinical arena, and has obvious implications for variant classification algorithms. The ENIGMA consortium has developed and documented criteria for the 5-tier classification of *BRCA1/2* genetic variants based on qualitative and quantitative information (http://www.enigmaconsortium.org/). According to these rules, and consistent with those proposed by InSiGHT for Mismatch Repair gene variants (1), *BRCA1/2* variants considered extremely likely to alter splicing based on position (typically IVS±1 or IVS±2) were initially all considered Class-4 (likely pathogenic) if untested for splicing alterations. However, the findings presented in this study have been pivotal to support amendment to these classification criteria, specifying need for particular caution in interpreting variants in instances where Δ9,10 (or other known naturally occurring in-frame alternative splicing events) might rescue gene functionality (see Supplemental Table 3). Hence, we also recommend caution in interpreting coding sequence variants that lead to premature termination codons in *BRCA1* exons 9 and 10. This conservative stance is consistent with recent American College of Medical Genetics (ACMG) guidelines (16), which recommend considering the presence of alternative gene transcripts, understanding which are biologically relevant, and in which tissues the products are expressed. Thus, caution should be exercised when interpreting the impact of truncating variants confined to only a subset of transcripts, given the presence of other protein isoforms.

Of note, our results have additional implications unrelated to alternative splicing. More precisely, our study suggests that *BRCA1* tumor suppressor activity tolerates a substantial reduction in expression level *in vivo*. Indeed, results shown in Figure 4 indicate that a *BRCA1* allele producing as much as ≈70-80% of

transcript encoding tumor suppressor deficient protein (as measured in blood-related samples) may not necessarily confer high-risk of developing cancer. This observation supports the conservative viewpoint of the ENIGMA consortium that, in the absence of other information, a variant can be considered pathogenic due to an effect on mRNA integrity if it only produces transcripts carrying a premature stop codon or an in-frame deletion disrupting known functional domain(s), as determined by semi-quantitative or quantitative methods.

In brief, there are several broad messages arising from the present study. Our results confirm that mRNA and genetic studies are warranted to inform the clinical significance of sequence alterations at the highly conserved intronic dinucleotides of splice donor and acceptor sites, and highlight the need to consider both variant haplotype and alternative splicing events in the design and interpretation of assays assessing the functional consequences of variants of uncertain clinical significance. We have also shown that comprehensive understanding of alternative splicing, paired with clinical genetic studies, is critical to understand the clinical consequences of complex splicing profiles observed for certain spliceogenic variants. Lastly, we provide a baseline hypothesis for future investigation and interpretation of other likely spliceogenic *BRCA1/2* variants, a hypothesis that has implications for informing standards for generic variant classification guidelines.

Materials and Methods

1. *Genotyping and Sample Sets*

We undertook screening of *BRCA1* c.594-2A>C by direct genotyping, as part of the iCOGS experiment detailed elsewhere (17, 18). This study included genotype and pathology results from breast cancer cases and controls participating in the Breast Cancer Association Consortium (BCAC; http://apps.ccge.medschl.cam.ac.uk/consortia/bcac//), and from carriers of *BRCA1* assumed pathogenic

variants participating in the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA; http://apps.ccge.medschl.cam.ac.uk/consortia/cimba//). In addition, via the Evidence-based Network for Investigating Germline Mutant Alleles (ENIGMA, http://enigmaconsortium.org/, (19)), we identified probands recruited through familial cancer clinics who were found to be positive for *BRCA1* c.594-2A>C via clinical genetic testing. All study participants were enrolled into national or regional studies under ethically-approved protocols.

Information was recorded for all variant carriers regarding cancer status, age at diagnosis/interview, breast tumor pathology (grade, and Estrogen Receptor (ER), Progesterone Receptor (PR), and Herceptin-2 (HER2) status), and also pedigree and segregation information where available. For carriers identified though CIMBA and ENIGMA, the genotype for exonic variant c.641A>G (p.Asp214Gly) was sought from the original clinical testing report.

The BCAC dataset included 53,354 breast cancer cases and 49,720 controls and documented age at diagnosis/interview from 45 studies, detailed in (17). The denominator reduced to 24,605 cases and 25,836 controls when including only invasive breast cancer cases and controls from the 11 studies with at least one observation (Supplementary Table 1). These 11 studies included only individuals of European ancestry, and four (MCBS, MBCCSG, KARBAC, OFBCR) had undergone testing for germline *BRCA1/2* pathogenic variants (4% - 100% of samples, depending on the BCAC study), including two of four studies which sampled cases on the basis of reported family history or presence of bilateral disease.

The CIMBA dataset included 11,105 female *BRCA1* pathogenic variant carriers aged ≥ 18 y from 46 studies in CIMBA recruited through cancer genetics clinics. There were 4,845 females without report of cancer, 4,713 breast cancer cases, 933 ovarian cancer cases, and 614 individuals reporting both breast and ovarian cancer.

By contact with submitters and examination of clinical information, it was established that 11 of the 15 CIMBA probands overlapped with individuals included in the ENIGMA dataset, and one of proband was

also a participant in BCAC site (Supplementary Table 2). Only non-overlapping data was included in multifactorial likelihood analysis.

2. *Statistical methods:*

We evaluated the effect of the *BRCA1* c.594-2A>C variant on breast cancer risk in BCAC, using logistic regression models with adjustment for censoring age and population structure, based on six principal components which defined any residual population sub-structure. Censoring age was defined as age at breast cancer diagnosis, or age at last interview/follow-up. Only case-control studies in which the variant was observed at least once were included in the analysis.

In order to place case-control data into the same likelihood ratio (LR) framework as the other lines of evidence used for multifactorial likelihood analysis (20, 21), we compared the likelihood of the distribution of *BRCA1* c.594-2A>C variant carriers among cases and controls under the hypothesis that the variant has the same age specific relative risks as the "average" pathogenic *BRCA1* mutation compared to that under the hypothesis that it is not associated with any increased breast cancer risk. Specifically, we used the age at diagnosis of cases and age at interview for controls together with the relative risks of breast cancer estimated from case series unselected for family history (6) to calculate the probability that each individual carrying *BRCA1* c.594-2A>C in the sample is a *BRCA1* pathogenic variant carrier given their affected status and age. Under the hypothesis that *BRCA1* c.594-2A>C is a benign variant and does not confer increased breast cancer risk, we calculated the probability of the distribution of cases and controls among *BRCA1* c.594-2A>C carriers as a simple binomial probability with p=proportion of cases in the sample. These two likelihoods were then compared to derive the appropriate LR.

Bayes scores for segregation were derived as described previously (22), and pathology LRs were applied as indicated in Spurdle et al (23). The segregation scores, pathology LRs and case-control LRs are

mutually independent and were combined to derive a combined odds for causality as described previously (20).

We used the program Phase 2.0 (24) to estimate the most likely haplotypes of the BCAC cases and controls based on 29 variants in the region within and surrounding the *BRCA1* locus, in order to examine if all c.594-2A>C variant carriers were observed on the same haplotypic background. Variants used for phasing were those submitted by ENIGMA for inclusion on the iCOGS chip design, the most common of which were rs8176258, rs1799967, rs1799950, rs4986852, and rs1799966.

3. *mRNA Analysis methods:*

3.1. Nomenclature. We use as reference sequences to describe *BRCA1* genetic variants the GenBank reference sequences U14680.1 (cDNA) and NC_000017.11 (genomic). When referring to *BRCA1* exons, we use exon numbering according to U14680.1. To characterize the *BRCA1* alternative splicing landscape in c.[594-2A>C; 641A>G] carriers (sometimes referred throughout the text as *variant allele* carriers), we performed different RNA splicing analyses at the immediate vicinity of *BRCA1* exon 10 (defined as the gene region spanning exons 8 to 11). Since our methodology do not allow analysis of complete transcripts (from 5'-end to poly(A) tail), we refer throughout the text to alternative splicing event containing transcripts, or alternative splicing events, rather than to alternative splicing transcripts or RNA isoforms (7) . For the very same reason, full-length (FL) refers throughout the text to *BRCA1* exons 9- and 10-containing transcripts (transcripts containing the exons9/10 junction defined in the GenBank reference sequenceU14680.1), and not necessarily to the complete 5711nt mRNA described in U14680.1. We have designated alternative splicing events by combining U14680.1 exon numbering with the following symbols: Δ (exon skipping), ▼ (intron retention), p (proximal, or 5'), and q (distal, or 3'). 3.2 *RNA analysis of human samples.* Up to seven contributing laboratories (sites 1 to 7) performed RNA splicing analyses with various methodologies, including fluorescent RT-PCR followed by capillary

electrophoresis, real-time quantitative PCR (qPCR), digital PCR (dPCR), Sanger sequencing, and RNAseq (see Supplemental Methods for further details). Experiments were performed in RNAs extracted from lymphoblastoid cell lines (LCLs), short-term (3-6 days) cultured Leukocytes (LEU), or fresh peripheral blood. RNAs were derived from 8 individual c.[594-2A>C; 641A>G] carriers (hereafter referred as Carriers 1 to 8) identified in four unrelated families from Australia (Carrier 1, LCL), Germany (Carrier 2, LCL), The Netherlands (Carriers 3 to 7, LEUs), and France (Carrier 8, peripheral blood), and healthy controls. We conducted several experiments designed to characterize the *BRCA1* alternative splicing landscape observed in *variant carriers*. We used as quantitative description the splicing fraction (*SF*), defined here as the contribution of individual alternative splicing events to the overall *BRCA1* expression level (expressed as a percentage). As proxies for overall expression level, we used the Σ of all peak areas detected (capillary electrophoresis), or the signal obtained with a TaqMan assay recognizing the *BRCA1* exons 23-24 junction (dPCR). The latter was selected since both *BRCA1* exons 23 and 24 are likely constitutive exons (7). Note that *SF* is a relative measure between signals arising from the same locus (in this case *BRCA1*), so that it is neither directly related to the actual expression level on individual splicing events, nor with the overall expression level from that locus. It is formally possible that increments in the *SF* of one particular alternative splicing event correlate with actual reductions in the expression level of that splicing event. For that reason, we determine the absolute expression level of individual alternative splicing events by qPCR with standard curves (see supplemental methods for further details), and we performed relative expression analyses by dPCR, using as a reference a TaqMan assay recognizing the *BRCA2* exons 26-27 junction. When indicated, we used as a positive control RNA extracted from LCLs carrying the *BRCA1* variant c.591C>T [p.= (Cys197Cys)], known to increase $\Delta 9^{SF}$ and $\Delta 9,10^{SF}$(14). Many experiments were performed in parallel with cultured cells treated/untreated with a nonsense mediated mRNA decay pathway (NMD) inhibitor, either Puromycin (Puro+/- experiments), or Cycloheximide (Cyclo+/- experiments). RNA from Carrier 8 was directly extracted from fresh peripheral blood. Biallelic expression was assessed by alternative splicing eventspecific RT-PCR followed by

Sanger sequencing through rs1060915 (an informative exonic SNP located at *BRCA1* exon 13), using primers and protocols previously described (14). In addition, we searched for *BRCA1* tissue specific alternative splicing landscape in clinically relevant samples by comparing RNAs extracted from healthy control fresh peripheral blood, a pool of 10 healthy breast tissues (enriched normal epithelial areas selected by a pathologist) adjacent to breast tumor samples, and commercial RNAs from healthy breast and ovarian human tissues. Experiments were performed by capillary electrophoresis of RT-PCR products, and by dPCR. Depending on the contributing laboratories, different RNA isolation and cDNA synthesis approaches were used (see Supplemental Methods for further details).

3.3. *Minigene Splicing Assays*. To dissect the contribution of the individual BRCA1 variants c.594-2A>C and c.641A>G to the splicing alteration observed in c.[594-2A>C; 641A>G] carriers, we performed splicing assays with 2 different types of reported minigenes: pCAS2-BRCA1-Exon10 and pB1 (a minigene spanning BRCA1 exons 8 to 12). See Supplemental Methods and Figure 3 for further details.

3.4. *RNA interference experiments*. To identify splicing regulatory proteins involved in *BRCA1* exon 10 splicing, we performed a series of RNA interference experiments knocking down diverse splicing regulatory factors (hnRNPA1, Tra2β, SF2/ASF, and SC35). Experiments were performed in the breast cancer cell line MDAMD231 (see Supplemental Methods for further details).

Acknowledgements

Funding

The research described was supported by 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 [PI12/00539 and PI15/00059 to MdH, PI13/02030 to AV]; the French Ministry of Higher Education and Research [to OS]; the University of Otago, Mackenzie Charitable Foundation, Maria Lupton, and .Health Research Council of New Zealand [to LW]; UK Higher Education Funding Council Senior Fellowship Scheme, the University of Southampton [to DB]; Cancer research UK [to DB, MR]; FamilienHede Nielsen Foundation fund [to TvOH]; Cancer Research-UK Senior Cancer Research Fellowship [to ACA]; National Institute of Health [CA128978 and CA11616 to FJC]; an NIH specialized program of research excellence in breast cancer to the Mayo Clinic [P50 CA116201 to FJC]; and the US Breast Cancer Research Foundation [to FJC]; translational grant from the French National Cancer Institute and Direction Générale de l'Offre des Soins (INCa-DGOS AAP/CFB/CI) and a grant from the French North-West Canceropole (CNO) [to AM]; The Cancer Council Queensland [APP1086286 to ABS]; the NHMRC Senior Research Fellowship Scheme [ID 1061779 to ABS]; NHMRC Project grant scheme [ID #1010719 to ABS].

Additional infrastructure support to consortium members is as follows:

kConFab infrastructure has been supported by funding from the National Breast Cancer Foundation, National Health and Medical Research Council, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia [to kConFab, and the kConFab Clinical Follow-up study].

The German Consortium of Hereditary Breast and Ovarian Cancer (GC-HBOC) is supported by the German Cancer Aid (grant no 109076, Rita K. Schmutzler) and by the Center for Molecular Medicine

Cologne (CMMC)

The French consortium is supported by the French National Cancer Institute.

EMBRACE is supported by Cancer Research UK Grants C1287/A10118 and C1287/A11990.

BCFR was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR.

The BBCS is funded by Cancer Research UK and Breakthrough Breast Cancer (recently merged with Breast Cancer Campaign forming Breast Cancer Now) and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN).

The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ).

The CGPS was supported by the Chief Physician Johan Boserup and LiseBoserup Fund, the Danish Medical Research Council and Herlev Hospital

KARBAC was supported financially through the regional agreement on medical training and clinical research (ALF) between Stockholm City Council and KarolinskaInstitutet, and from the Stockholm Cancer Foundation and the Swedish Cancer Society.

KARBAC was supported financially through the regional agreement on medical training and clinical research (ALF) between Stockholm City Council and KarolinskaInstitutet, and from the Stockholm Cancer Foundation and the Swedish Cancer Society.

The MARIE study was supported by the Deutsche Krebshilfee.V. [70-2892-BR I], the Hamburg Cancer Society, the German Cancer Research Center and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402].

The MARIE study was supported by the Deutsche Krebshilfee.V. [70-2892-BR I], the Hamburg Cancer Society, the German Cancer Research Center and the Federal Ministry of Education and Research

(BMBF) Germany [01KH0402].

MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment according to Italian laws in support of the Fondazione IRCCS IstitutoNazionaleTumori.

The MCBCS was supported by the NIH grant CA128978 and a Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation.

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria.

OFBCR was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR.

The pKARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer, and the Swedish Medical Research Counsel.

SEARCH was supported by grants CRUK A490/A11021, C490/A16561.

Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research

Foundation, and the Ovarian Cancer Research Fund.

CIMBA data management was supported by Cancer Research-UK grant C12292/A11174 and C1287/A10118.

BCAC is funded by Cancer Research UK [C1287/A10118, C1287/A12014] and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS).

We thank all the families and individuals that participated in this research. We thank Paul Pharoah and the Ovarian Cancer Association Consortium for providing summary information on the frequency of the *BRCA1* c.594-2A>C variant in ovarian cancer cases and controls.

We acknowledge the contributions of Georgia Chenevix-Trench to CIMBA and the kConFaB resource, and additional study-specific acknowledgements as noted below.

NZBCS: Anne Smith, Bridget Robinson, Caroline Lintott, John Pearson, Yen Phung George Wiggins and the family members for their valuable contributions.

*The Netherlands Consortium:*FransHogervorst for assistance in data collation, Dr. M. Olderode-Berends from the University Medical Centre Groningen and Dr. E. van Riel from University Medical Centre Utrecht for providing blood samples for RNA studies, and Elsa Bik from the Leiden University Medical Centre, The Netherlands for excellent technical assistance.

kConFab: Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study for their contributions to this resource.

GC-HBOC: Dieter Schäfer, Center Frankfurt, for providing DNA samples and JulianeKöhler for excellent technical assistance.

Northshore: Tina Selkirk for assistance in data collation.

French Consortium: the French oncogeneticists, and the UNICANCER Genetic Group (UGG) lead by

Dr Catherine Noguès.

BCFR-AU: Maggie Angelakos, Judi Maskiell, Gillian Dite, Helen Tsimiklis.

MUV: Daniela Muhr for assistance in data collation.

BBCS: Eileen Williams, Elaine Ryder-Mills, Kara Sargus

BSUCH: Peter Bugert, Medical Faculty Mannheim

CGPS: Staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: DortheUldall Andersen, Maria BirnaArnadottir, Anne Bank, DortheKjeldgård Hansen. The Danish Breast Cancer Group (DBCG) is acknowledged for the tumor information. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases.

MARIE: Alina Vrieling, Katharina Buck, MuhabbetCelik, Ursula Eilber and Sabine Behrens.

MBCSG: SiranoushManoukian, Bernard Peissel, Jacopo Azzollini and Fernando Ravagnani of the Fondazione IRCCS IstitutoNazionaleTumori (INT), Milan, Italy; BernandoBonanni, Monica Barile and Irene Feroce of the IstitutoEuropeo di Oncologia (IEO), Milan, Italy; and the personnel of the Cogentech Cancer Genetic Test Laboratory, Milan, Italy.

OFBCR: Teresa Selander, NayanaWeerasooriya

SEARCH: Marie Mack

COGS: Paul Pharoah, Andrew Berchuck (OCAC), Georgia Chenevix-Trench, Ken Offit (CIMBA), Alison M. Dunning, Andrew Lee,Ed Dicks, Craig Luccarini, the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira, the staff of the CNIO genotyping unit, Jacques Simard, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière, Frederic Robidoux, the staff of the McGill University and Génome Québec Innovation Centre, Sune F. Nielsen, Borge G. Nordestgaard, the staff of the Copenhagen DNA laboratory, Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer, the staff of Mayo Clinic Genotyping Core Facility.

AM and INSERM: Dr Sophie Krieger for contributing with DNA samples, Prof Thierry Frébourg for providing patient's samples for RNA analysis, and AurélieDrouet and Gaia Castelain for technical

assistance.

Conflicts of Interest

Tina Pesaran and Elizabeth Chao are paid employees of Ambry Genetics.

References

1 Thompson, B.A., Spurdle, A.B., Plazzer, J.P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capella, G., den Dunnen, J.T. *et al.* (2014) Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.*, **46**, 107-115.

2 Rosenthal, E.T., Bowles, K.R., Pruss, D., van Kan, A., Vail, P.J., McElroy, H. and Wenstrup, R.J. (2015) Exceptions to the rule: Case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes. *Clin. Genet.*, **88**, 533-541.

3 Tesoriero, A.A., Wong, E.M., Jenkins, M.A., Hopper, J.L., Brown, M.A., Chenevix-Trench, G., Spurdle, A.B., Southey, M.C. and kConFab. (2005) Molecular characterization and cancer risk associated with BRCA1 and BRCA2 splice site variants identified in multiple-case breast cancer families. *Hum. Mutat.*, **26**, 495.

4 Whiley, 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.* (2014) Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin. Chem.*, **60**, 341-352.

5 Walker, L.C., Whiley, P.J., Houdayer, C., Hansen, T.V., Vega, A., Santamarina, M., Blanco, A., Fachal, L., Southey, M.C., Lafferty, A. *et al.* (2013) 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.*, **34**, 1424-1431.

6 Antoniou, A., Pharoah, P.D., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A. *et al.* (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.*, **72**, 1117-1130.

7 Colombo, M., Blok, M.J., Whiley, P., Santamarina, M., Gutierrez-Enriquez, S., Romero, A., Garre, P., Becker, A., Smith, L.D., De Vecchi, G. *et al.* (2014) Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum. Mol. Genet.*, **23**, 3666-3680.

8 Steffensen, A.Y., Dandanell, M., Jonson, L., Ejlertsen, B., Gerdes, A.M., Nielsen, F.C. and Hansen, T. (2014) Functional characterization of BRCA1 gene variants by mini-gene splicing assay. *Eur. J. Hum. Genet.*, **22**, 1362-1368.

9 Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabartty, A. and Arrowsmith, C.H. (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? *J. Mol. Biol.*, **345**, 275-287.

10 Clark, S.L., Rodriguez, A.M., Snyder, R.R., Hankins, G.D. and Boehning, D. (2012) Structure-Function Of The Tumor Suppressor BRCA1. *Comput. Struct. Biotechnol. J.*, **1**, pii: e201204005.

11 Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J. and Fields, S. (2015) Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics*, **200**, 413-422.

12 Shakya, R., Reid, L.J., Reczek, C.R., Cole, F., Egli, D., Lin, C.S., deRooij, D.G., Hirsch, S., Ravi, K., Hicks, J.B. *et al.* (2011) BRCA1 tumor suppression depends on BRCT phosphoprotein binding, but not its E3 ligase activity. *Science*, **334**, 525-528.

13 Savage, K.I. and Harkin, D.P. (2015) BRCA1, a 'complex' protein involved in the maintenance of genomic stability. *FEBS J.*, **282**, 630-646.

14 Dosil, V., Tosar, A., Canadas, C., Perez-Segura, P., Diaz-Rubio, E., Caldes, T. and de la Hoya, M. (2010) Alternative splicing and molecular characterization of splice site variants: BRCA1 c.591C>T as a case study. *Clin. Chem.*, **56**, 53-61.

15 Nieuwenhuis, M.H. and Vasen, H.F. (2007) Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. *Critical reviews in oncology/hematology*, **61**, 153-161.

16 Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.*, **17**, 405-424.

17 Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K. *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, **45**, 353-361, 361e351-352.

18 Couch, F.J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K.B., Soucy, P., Fredericksen, Z., Barrowdale, D., Dennis, J. *et al.* (2013) Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet.*, **9**, e1003212.

19 Spurdle, A.B., Healey, S., Devereau, A., Hogervorst, F.B., Monteiro, A.N., Nathanson, K.L., Radice, P., Stoppa-Lyonnet, D., Tavtigian, S., Wappenschmidt, B. *et al.* (2012) 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.*, **33**, 2-7.

20 Goldgar, D.E., Easton, D.F., Deffenbaugh, A.M., Monteiro, A.N., Tavtigian, S.V., Couch, F.J. and Breast Cancer Information Core Steering, C. (2004) Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am. J. Hum. Genet.*, **75**, 535-544.

21 Goldgar, D.E., Easton, D.F., Byrnes, G.B., Spurdle, A.B., Iversen, E.S., Greenblatt, M.S. and Group, I.U.G.V.W. (2008) Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum. Mutat.*, **29**, 1265-1272.

22 Thompson, D., Easton, D.F. and Goldgar, D.E. (2003) A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am. J. Hum. Genet.*, **73**, 652-655.

23 Spurdle, A.B., Couch, F.J., Parsons, M.T., McGuffog, L., Barrowdale, D., Bolla, M.K., Wang, Q., Healey, S., Schmutzler, R., Wappenschmidt, B. *et al.* (2014) 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.*, **16**, 3419.

24 Delaneau, O., Marchini, J. and Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179-181.

Legends to Figures

Figure 1. Capillary Electrophoresis analyses of *BRCA1*alternative splicing landscape in LCLs from one *BRCA1*c.[594-2A>C; 641A>G] carrier (Carrier 1) and 10 Controls. Panel A shows representative examples of capillary electrophoresis analysis of RT-PCR products generated with the E8.1-E11p assay in LCLs treated (Puro+) or untreated (Puro-) with the nonsense mediated decay inhibitor puromycin. The fluorescence intensity of each peak (Y-axis) is expressed in arbitrary units (AU). The analyses detected the full-length transcript (FL), and up to four alternative splicing events, two in-frame (Δ9,10 and ▼10p) and two out-of-frame (Δ9, and Δ10). In these particular examples, ▼10p transcripts are detected only in Carrier 1, but we have detected ▼10p transcripts in Controls, as summarized in panel B. The presence of ▼10p in Controls has been further confirmedby RNAseq (see Supp. Figure 3).The boxplots in **Panel B** (displaying low, Q1, median, Q3, and high values) show the splicing fraction (SF) of in-frame transcripts (Δ9,10, FL and ▼10p) observed in Carrier 1 (3 technical replicas) and 10 Controls. SF expressed as the % of the corresponding peak area to the Σ of all five peak areas detected by capillary electrophoresis. This particular experiment was performed with the E8.2-E11q.2 assay. Note that the $▼10p^{SF}$ is rather minor (<1%) regardless of the LCL tested. The FL^{SF} was much lower in Carrier 1 than in Control samples. The boxplots in **Panel C** (displaying low, Q1, median, Q3, and high values) show the SF of out-of-frame transcripts (Δ9 and Δ10) observed in Carrier 1 (3 technical replicates) and 10 Controls. The relative contribution of Δ10 to the overall signal was much higher in Carrier 1 than in Control samples. Normal outliers (>1.5 interquartile range, IQR) display small circles. (** represents P≤0.01) (*** represents P≤0.001) (ns=non-significant).

Figure 2. Quantification of major in-frame transcripts Δ9,10 and full-length (FL) in LCLs from one *BRCA1*c.[594-2A>C; 641A>G] carrier (Carrier 1) and Controls. Experiments were performed in LCLs treated with Puromycin (Puro+). **Panel A** displays $\Delta 9,10^{SF}$ and FL^{SF}, estimated as the ratio between the GADPH normalized absolute numbers of Δ9,10 (or FL) molecules and absolute number of *all BRCA1* transcripts, as determined by qPCR analysis performed with standard curves (see Supplemental Methods and Supplemental Figure 4). Standard deviation of 3 independent measures is shown. **Panel B** displays dPCR data measuring $\Delta 9,10^{SF}$ and FL (inclusion of exons 9 and 10)^{SF}, using exon23-24 junction as a proxy for overall *BRCA1* expression level. The precision of each measure (as determined by the QuantStudio 3D Analysis Cloud Software) is indicated. Two technical replicates of Carrier 1 are shown. We included as positive control a LCL carrying the *BRCA1* c.591C>T variant, known to increase $\Delta 9,10^{SF}$. The $\Delta 9,10^{SF}$ in Carrier 1 was higher than in Controls (24% in two technical replicates of Carrier 1 vs. an average of 17% in 7 control samples, Mann-Whitney U test; p=0.028 for difference between groups), but a 50% reduction of FL^{SF} (50% in two technical replicas of Carrier 1 vs. an average of 94% in 6 control samples, Mann-Whitney U test; p=0.036 for difference between groups).

Figure 3. Analysis of *BRCA1* c.594-2A>C and c.641A>G variants with splicing reporter minigene assays. The figure shows schematic non-scale representations of the splicing reporter minigenes pCAS2-BRCA1-exon10 (panel A) and pB1 (panel B) used for splicing assays. Minigenes were constructed as described under Supplemental Methods. PCMV indicates the cytomegalovirus promoter, boxes represent exons and lines in between indicate introns. *BRCA1* sequences are highlighted in black. Arrows represent primers used in RT-PCR reactions. With the exception of pB1 *BRCA1* intron 11 (402 nt-long full-length IVS11), minigenes harbor partial segments of *BRCA1* introns. For comparative purposes, the size in nucleotides of each segment is shown together with the size corresponding to the endogenous full-length *BRCA1* introns shown in brackets. As indicated, pB1 carries an additional cytosine (+3insC) in exon 8 to keep the ORF with α-globin exon 1 (Raponi et al., 2012). Splicing assays were performed by analyzing the splicing pattern of WT and mutant minigenes (c.594-2A>C, c.641A>G, and c.[594-2A>C; 641A>G]) transiently expressed in human cells (HeLa, COS-7, MCF7, HBL100 or IGROV-1) as described under Supplemental Methods. The images show RT-PCR products separated in ethidium bromide-stained agarose gels. FL, full-length; Δ9, exon 9 skipping; Δ10, exon 10 skipping; Δ9,10, skipping of both exons 9 and 10; *, retention of 21 intronic nucleotides immediately upstream exon 10 (▼10p). One can note that: (i) the relative level of alternatively spliced pB1(WT) transcripts is higher in IGROV-1 than in HeLa, MCF-7, or HBL100 cells, and (ii) the predominant alternative splicing event of pB1(WT) in these cell lines is Δ10, whereas that of endogenous wild-type *BRCA1* in blood related samples is Δ9,10 (Figure 4 and Supplemental Figures 1 and 2).

Figure 4. Combined genetic and splicing analyses of *BRCA1* c.[594-2A>C; 641A>G] and *BRCA1* c.591C>T supports a *BRCA1*Δ9,10 rescue model with far-reaching clinical implications. Panel A (top) shows the splicing fraction (SF) of five alternative splicing events detected by capillary electrophoresis analysis of RT-PCR products generated with the E8.2-E11q.2 assay (Puro+ experiments, 36 cycle PCRs, see Figure 1 and Supplemental Figure 1 for further details). As shown, this description of the *BRCA1* alternative splicing landscape in the vicinity of exon 10 is different in healthy control samples, c.[594-2A>C; 641A>G] carriers, and c.591C>T carriers. Yet, we show in the present study that none of these 3 *BRCA1* splicing landscapes is associated with high risk of developing *BRCA1* related cancers. The chart displays SFs that, in carriers, represent a combined signal from the variant allele and the accompanying WT allele. **Panel A (bottom).** Deduced *per allele* SFs are shown. Assuming that SFs arising from the accompanying WT allele equal to the average SFs observed in 10 Control samples (as shown in the central chart bar), we deconvoluted the SFs corresponding to c.[594-2G; c.641G] (left chart bar) and c.591T (right chart bar) alleles. **Panel B.** The cartoon represents the relative *per allele* (100% equals to the overall expression level arising from one individual allele) and *per cell* (100% equals to the overall expression arising from a diploid genome) expression (BRCA1 exons 7 to 11) in a c.[594-2G; c.641G] carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL and Δ9,10 transcripts are shown, albeit Δ9 and ▼10p transcripts account for ≈5% of the *per cell* expression. Truncating (out-of-frame) events are highlighted with a red cross. The analysis suggests that expressing up to ≈35% of *BRCA1* PTC-NMD transcripts (*per* diploid genome) is not associated with high-risk of developing cancer. The analysis suggests as well that a *BRCA1* allele expressing up to ≈70% (*per allele*) *BRCA1* PTC-NMD transcripts is not associated with high-risk of developing cancer (a relevant finding in the context of the two-hit model). **Panel C.** The cartoon represents the relative *per allele* (100% equals to the overall expression level arising from one individual allele) and *per cell* (100% equals to the overall expression arising from a diploid genome) expression (*BRCA1* exons 7 to 11) in a c.591C>T carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL, Δ9,10 and Δ9 (variant allele) are shown, albeit Δ9 (wt allele), Δ10 (wt and variant allele), and ▼10p (wt and variant allele) transcripts account for ≈5% of the *per cell* expression. The data strongly suggests that *BRCA1*Δ9,10 transcripts, representing up to 51% (per diploid genome) and up to 71% (*per* allele) of the overall *BRCA1* expression code for a BRCA1 protein with tumor suppressor activity. The model displayed in this figure is intended to illustrate the most relevant findings of our study. Yet, some limitations should be highlighted. First, the model assumes (based on 36-cycle PCR capillary EP data) that $\Delta 9,10^{SF}$ in Controls and c.[594-2A>C; 641A>G] carriers is ≈29%, while other experiments suggests that the actual value is probably lower in both instances (Figure 2, Supplemental Figure 2), albeit slightly increased in Carriers

vs. Controls. The model has been elaborated with data obtained in LCLs, not in clinically relevant tissues such as breast or ovarian.

Abbreviations

Cyclo- Cycloheximide absent

Cyclo+ Cycoeximide present

dPCR digital PCR

qPCR quantitative PCR

FL full-length

LCL lymphoblastoid cell line

LEU leukocyte

NMD nonsense mediated decay

PTC premature termination codon

Puro- Puromycin absent

Puro+ Puromycin present

qPCR quantitative PCR

RNAseq high-throughput RNA sequencing

RT reverse transcription

SF splicing fraction

siRNA small interference RNA

WT wildtype

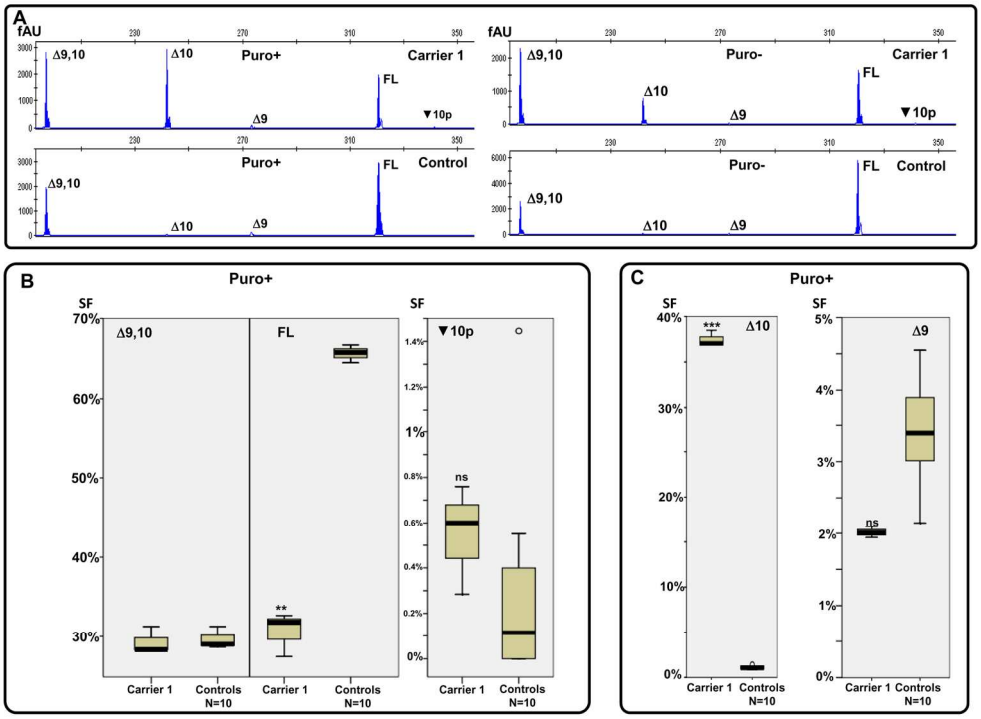


Figure 1. Capillary Electrophoresis analyses of BRCA1alternative splicing landscape in LCLs from one BRCA1c.[594-2A>C; 641A>G] carrier (Carrier 1) and 10 Controls. Panel A shows representative examples of capillary electrophoresis analysis of RT-PCR products generated with the E8.1-E11p assay in LCLs treated (Puro+) or untreated (Puro-) with the nonsense mediated decay inhibitor puromycin. The fluorescence intensity of each peak (Y-axis) is expressed in arbitrary units (AU). The analyses detected the full-length transcript (FL), and up to four alternative splicing events, two in-frame (D9,10 and ▼10p) and two out-of-frame (D9, and D10). In these particular examples, ▼10p transcripts are detected only in Carrier 1, but we have detected ▼10p transcripts in Controls, as summarized in panel B. The presence of ▼10p in Controls has been further confirmedby RNAseq (see Supp. Figure 3).The boxplots in Panel B (displaying low, Q1, median, Q3, and high values) show the splicing fraction (SF) of in-frame transcripts (D9,10, FL and ▼10p) observed in Carrier 1 (3 technical replicas) and 10 Controls. SF expressed as the % of the corresponding peak area to the Σ of all five peak areas detected by capillary electrophoresis. This particular experiment was performed with the E8.2-E11q.2 assay. Note that the ▼10pSF is rather minor (<1%) regardless of the LCL tested. The FLSF was much lower in Carrier 1 than in Control samples. The boxplots in Panel C (displaying low, Q1, median, Q3, and high values) show the SF of out-of-frame transcripts (D9 and D10) observed in Carrier 1 (3 technical replicates) and 10 Controls. The relative contribution of D10 to the overall signal was much higher in Carrier 1 than in Control samples. Normal outliers (>1.5 interquartile range, IQR) display small circles. (** represents P≤0.01) (*** represents P≤0.001) (ns=non-significant).
173x130mm (300 x 300 DPI)

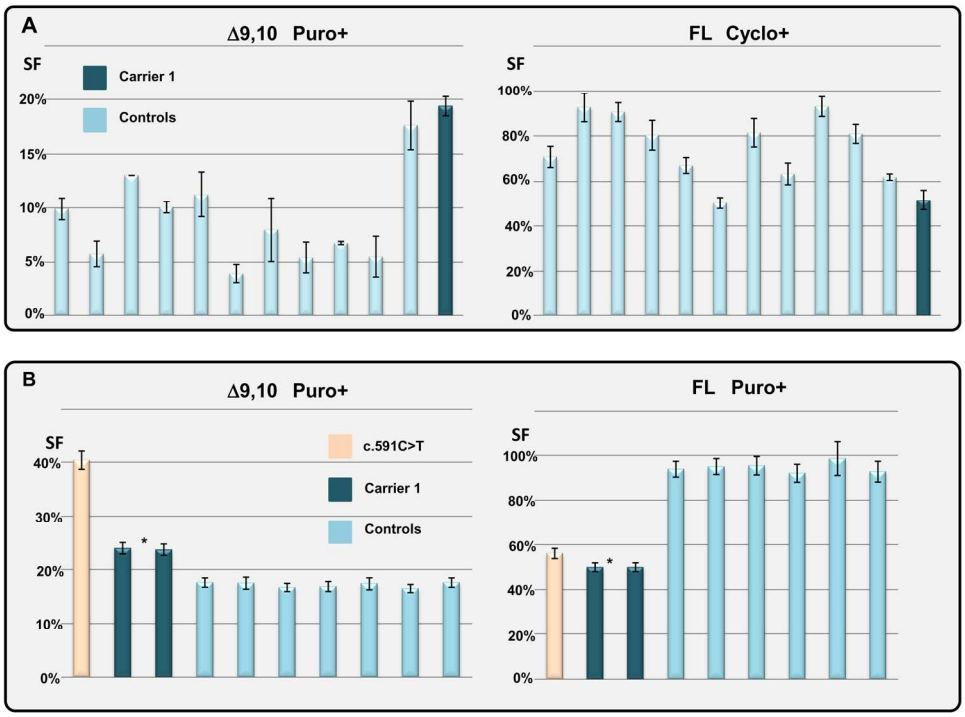


Figure 2. Quantification of major in-frame transcripts D9,10 and full-length (FL) in LCLs from one BRCA1c.[594-2A>C; 641A>G] carrier (Carrier 1) and Controls. Experiments were performed in LCLs treated with Puromycin (Puro+). Panel A displays D9,10SF and FLSF, estimated as the ratio between the GADPH normalized absolute numbers of D9,10 (or FL) molecules and absolute number of all BRCA1 transcripts, as determined by qPCR analysis performed with standard curves (see Supplemental Methods and Supplemental Figure 4). Standard deviation of 3 independent measures is shown. Panel B displays dPCR data measuring D9,10SF and FL (inclusion of exons 9 and 10)SF, using exon23-24 junction as a proxy for overall BRCA1 expression level. The precision of each measure (as determined by the QuantStudio 3D Analysis Cloud Software) is indicated. Two technical replicates of Carrier 1 are shown. We included as positive control a LCL carrying the BRCA1 c.591C>T variant, known to increase D9,10SF. The D9,10SF in Carrier 1 was higher than in Controls (24% in two technical replicates of Carrier 1 vs. an average of 17% in 7 control samples, Mann-Whitney U test; p=0.028 for difference between groups), but a 50% reduction of FLSF (50% in two technical replicas of Carrier 1 vs. an average of 94% in 6 control samples, Mann-Whitney U test; p=0.036 for difference between groups).
173x130mm (300 x 300 DPI)

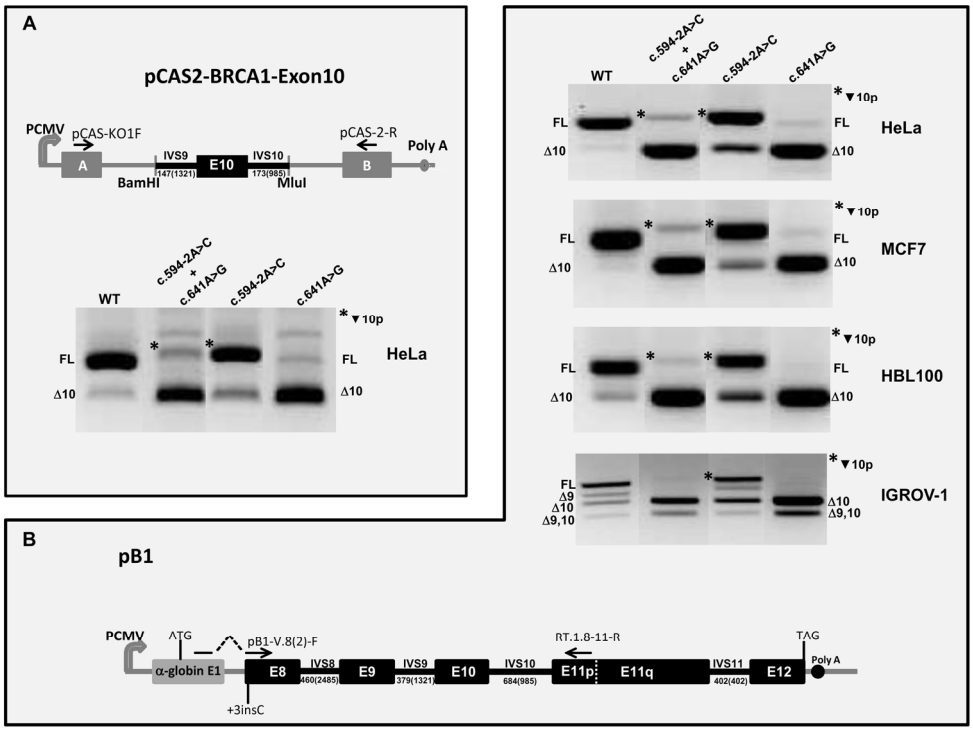


Figure 3. Analysis of BRCA1 c.594-2A>C and c.641A>G variants with splicing reporter minigene assays. The figure shows schematic non-scale representations of the splicing reporter minigenes pCAS2-BRCA1-exon10 (panel A) and pB1 (panel B) used for splicing assays. Minigenes were constructed as described under Supplemental Methods. PCMV indicates the cytomegalovirus promoter, boxes represent exons and lines in between indicate introns. BRCA1 sequences are highlighted in black. Arrows represent primers used in RT-PCR reactions. With the exception of pB1 BRCA1 intron 11 (402 nt-long full-length IVS11), minigenes harbor partial segments of BRCA1 introns. For comparative purposes, the size in nucleotides of each segment is shown together with the size corresponding to the endogenous full-length BRCA1 introns shown in brackets. As indicated, pB1 carries an additional cytosine (+3insC) in exon 8 to keep the ORF with a-globin exon 1 (Raponi et al., 2012). Splicing assays were performed by analyzing the splicing pattern of WT and mutant minigenes (c.594-2A>C, c.641A>G, and c.[594-2A>C; 641A>G]) transiently expressed in human cells (HeLa, COS-7, MCF7, HBL100 or IGROV-1) as described under Supplemental Methods. The images show RT-PCR products separated in ethidium bromide-stained agarose gels. FL, full-length; D9, exon 9 skipping; D10, exon 10 skipping; D9,10, skipping of both exons 9 and 10; *, retention of 21 intronic nucleotides immediately upstream exon 10 (▼10p). One can note that: (i) the relative level of alternatively spliced pB1(WT) transcripts is higher in IGROV-1 than in HeLa, MCF-7, or HBL100 cells, and (ii) the predominant alternative splicing event of pB1(WT) in these cell lines is D10, whereas that of endogenous wild-type BRCA1 in blood related samples is D9,10 (Figure 4 and Supplemental Figures 1 and 2).
173x130mm (300 x 300 DPI)

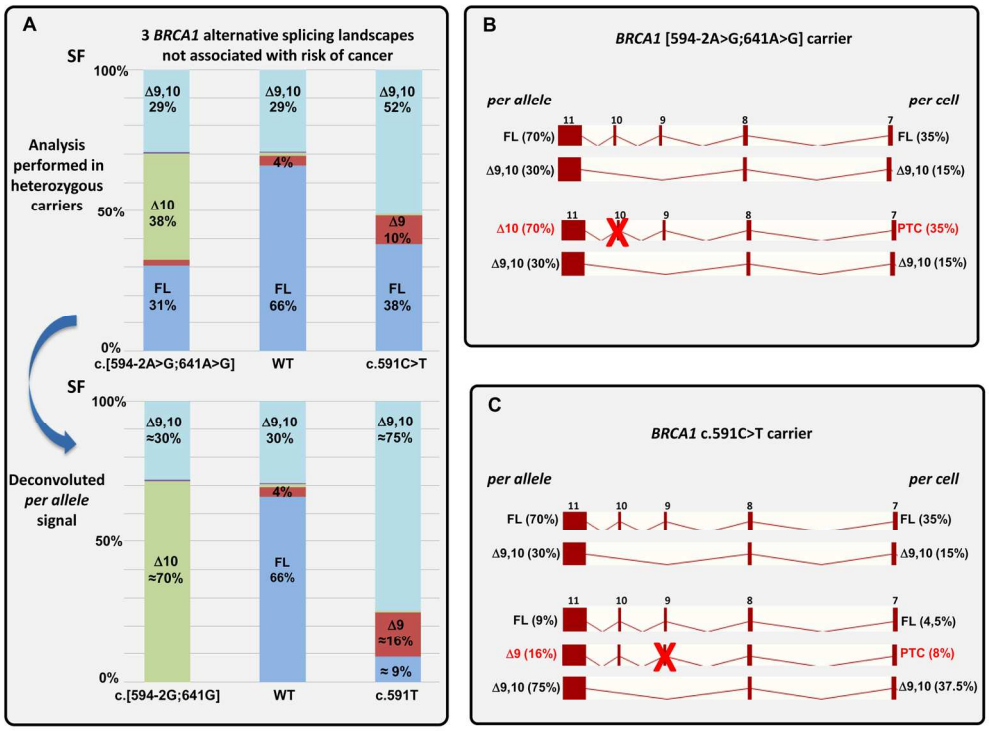


Figure 4. Combined genetic and splicing analyses of BRCA1 c.[594-2A>C; 641A>G] and BRCA1 c.591C>T supports a BRCA1D9,10 rescue model with far-reaching clinical implications. Panel A (top) shows the splicing fraction (SF) of five alternative splicing events detected by capillary electrophoresis analysis of RT-PCR products generated with the E8.2-E11q.2 assay (Puro+ experiments, 36 cycle PCRs, see Figure 1 and Supplemental Figure 1 for further details). As shown, this description of the BRCA1 alternative splicing landscape in the vicinity of exon 10 is different in healthy control samples, c.[594-2A>C; 641A>G] carriers, and c.591C>T carriers. Yet, we show in the present study that none of these 3 BRCA1 splicing landscapes is associated with high risk of developing BRCA1 related cancers. The chart displays SFs that, in carriers, represent a combined signal from the variant allele and the accompanying WT allele. Panel A (bottom). Deduced per allele SFs are shown. Assuming that SFs arising from the accompanying WT allele equal to the average SFs observed in 10 Control samples (as shown in the central chart bar), we deconvoluted the SFs corresponding to c.[594-2G; c.641G] (left chart bar) and c.591T (right chart bar) alleles. Panel B. The cartoon represents the relative per allele (100% equals to the overall expression level arising from one individual allele) and per cell (100% equals to the overall expression arising from a diploid genome) expression (BRCA1 exons 7 to 11) in a c.[594-2G; c.641G] carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL and D9,10 transcripts are shown, albeit D9 and ▼10p transcripts account for ≈5% of the per cell expression. Truncating (out-of-frame) events are highlighted with a red cross. The analysis suggests that expressing up to ≈35% of BRCA1 PTC-NMD transcripts (per diploid genome) is not associated with high-risk of developing cancer. The analysis suggests as well that a BRCA1 allele expressing up to ≈70% (per allele) BRCA1 PTC-NMD transcripts is not associated with high-risk of developing cancer (a relevant finding in the context of the two-hit model). Panel C. The cartoon represents the relative per allele (100% equals to the overall expression level arising from one individual allele) and per cell (100% equals to the overall expression arising from a diploid genome) expression (BRCA1 exons 7 to 11) in a c.591C>T carrier, inferred from capillary EP analyses shown in Panel A. For simplicity, only FL, D9,10 and D9 (variant allele) are shown, albeit D9 (wt allele), D10 (wt and variant allele), and ▼10p (wt and variant allele) transcripts account for ≈5% of the per cell expression. The data strongly suggests that BRCA1D9,10 transcripts, representing up to 51% (per diploid genome) and up to 71% (per allele) of the overall BRCA1 expression code for a BRCA1 protein with tumor suppressor activity. The model displayed in this figure is intended to illustrate the most relevant findings of our study. Yet, some limitations should be highlighted. F

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

173x130mm (300 x 300 DPI)

Human Molecular Genetics CONFLICT OF INTEREST FORM

Title:

Author(s): Elizabeth Chao *Ms number:* HMG-2015-D-01383

PLEASE NOTE: As an integral part of the online submission process, Corresponding authors are required to confirm whether they or their co-authors have any conflicts of interest to declare, and to provide details of these. If the Corresponding author is unable to confirm this information on behalf of all co-authors, the authors in question are required to complete and fax this Conflict of Interest form to the Editorial Office at +44 (0)1865 355901 . It is the Corresponding author's responsibility to ensure that all authors adhere to this policy.

MANUSCRIPT AUTHORS: If you have been asked to do so by the Corresponding author, please complete **Part I** or **II**. At the point of submission, *Hum. Mol. Gen.* policy requires that each author reveal any financial interests or connections, direct or indirect, or other situations *that might raise the question of bias* in the work reported or the conclusions, implications, or opinions stated - including pertinent commercial or other sources of funding for the individual author(s) or for the associated department(s) or organization(s), personal relationships, or direct academic competition.

When considering whether you should declare a conflicting interest or connection please consider the conflict of interest test: Is there any arrangement that would embarrass you or any of your co-authors if it was to emerge after publication and you had not declared it?

If the manuscript is published, this information will be communicated in a statement in the published paper.

I. There have been no involvements that might raise the question of bias in the work reported or in the conclusions, implications, or opinions stated.

_____
Printed name

_____
Signature Date

OR

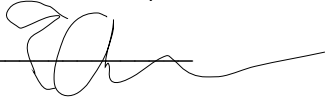
II. Conflict of interest statement.
Sample statement: I hold stock* in [business name], the makers of [product], and am currently conducting research sponsored by this company. I am also a member of the speakers' bureau for [business name].

* Please provide details of stock where this reveals a pertinent conflict of interest.

My statement is as follows:

I hold stock and previously held a leadership position at Ambry Genetics. Ambry is a commercial testing laboratory where patients/mutations described in the manuscript were identified through the course of routine clinical care.

Elizabeth Chao, MD
Printed name


Signature Date 2/1/2016

Human Molecular Genetics CONFLICT OF INTEREST FORM

Title:

Author(s): Tina Pesaran *Ms number:* HMG-2015-D-01383

PLEASE NOTE: As an integral part of the online submission process, Corresponding authors are required to confirm whether they or their co-authors have any conflicts of interest to declare, and to provide details of these. If the Corresponding author is unable to confirm this information on behalf of all co-authors, the authors in question are required to complete and fax this Conflict of Interest form to the Editorial Office at +44 (0)1865 355901 . It is the Corresponding author's responsibility to ensure that all authors adhere to this policy.

MANUSCRIPT AUTHORS: If you have been asked to do so by the Corresponding author, please complete **Part I** or **II**. At the point of submission, *Hum. Mol. Gen.* policy requires that each author reveal any financial interests or connections, direct or indirect, or other situations *that might raise the question of bias* in the work reported or the conclusions, implications, or opinions stated - including pertinent commercial or other sources of funding for the individual author(s) or for the associated department(s) or organization(s), personal relationships, or direct academic competition.

When considering whether you should declare a conflicting interest or connection please consider the conflict of interest test: Is there any arrangement that would embarrass you or any of your co-authors if it was to emerge after publication and you had not declared it?

If the manuscript is published, this information will be communicated in a statement in the published paper.

I. There have been no involvements that might raise the question of bias in the work reported or in the conclusions, implications, or opinions stated.

_____
Printed name

_____
Signature Date

OR

II. Conflict of interest statement.
Sample statement: I hold stock* in [business name], the makers of [product], and am currently conducting research sponsored by this company. I am also a member of the speakers' bureau for [business name].

* Please provide details of stock where this reveals a pertinent conflict of interest.

My statement is as follows:

I am a paid employee at Ambry Genetics. Ambry is a comerical testing laboratory where the some of the patients described in the manuscript were identified through the course of routine clinical care.

Tina Pesaran
Printed name

Tina Pesaran
Signature Date 2/1/2016