**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 Hoya1,\*, Omar Soukarieh2, Irene López-Perolio1, Ana Vega3, Logan C. Walker4, Yvette van Ierland5, Diana Baralle6, Marta Santamariña7, Vanessa Lattimore4, Juul Wijnen5, 8, Philip Whiley9, Ana Blanco3, Michela Raponi6, Jan Hauke10, Barbara Wappenschmidt10, Alexandra Becker 10, Thomas v. O. Hansen11, Raquel Behar1, KConFaB investigators12, Diether Niederacher13, Norbert Arnold14, Bernd Dworniczak15, Doris Steinemann16, Ulrike Faust17, Wendy Rubinstein18, Peter J. Hulick19, Claude Houdayer20, Sandrine M. Caputo21, Laurent Castera22, Tina Pesaran23, Elizabeth Chao23, Carole Brewer24, Melissa C. Southey25, Christi J. van Asperen5, Christian F. Singer26, Jan Sullivan27, Nicola Poplawski28, 29, Phuong Mai30, Julian Peto31, Nichola Johnson32, 33, Barbara Burwinkel34, 35, Harald Surowy34, 35, Stig E. Bojesen36-38, Henrik Flyger39, Annika Lindblom40, Sara Margolin41, Jenny Chang-Claude42, 43, Anja Rudolph42, Paolo Radice44, Laura Galastri45, Janet E. Olson46, Emily Hallberg46, Graham G. Giles47, 48, Roger L. Milne47, 48, Irene L. Andrulis49, 50, Gord Glendon49, Per Hall51, Kamila Czene51, Fiona Blows52, Mitul Shah52, Qin Wang53, Joe Dennis53, Kyriaki Michailidou53, 54, Lesley McGuffog55, Manjeet K. Bolla53, Antonis C. Antoniou55, Douglas F. Easton52, 53, Fergus J. Couch46, 56, Sean Tavtigian57, Maaike Vreeswijk5, Michael Parsons9, Huong Meeks57, Alexandra Martins2, David E. Goldgar 58, 59, Amanda B. Spurdle9, 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](mailto: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](mailto: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 3.23x10-8. 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](#_ENREF_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](#_ENREF_2)). Previous RNA analyses of carriers of *BRCA1* c.594-2A>C indicate that this variant is associated with an aberrant mRNA profile ([3](#_ENREF_3), [4](#_ENREF_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](#_ENREF_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 x 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](#_ENREF_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 4.98x10-6 (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 3.61x10-11 (i.e. 2.77x1010:1 *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](#_ENREF_4), [7](#_ENREF_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](#_ENREF_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 (Δ9,10SF) is similar in Carrier 1 and Controls (≈29%), Δ10SF is considerably higher (≈38% vs. ≈1%), and FLSF much lower (≈31% vs. ≈66%). Δ9SF (<3%) and ▼10pSF (<1%)were rather minor alternative splicing events in all tested samples. As expected, Puro- experiments measured higher Δ9,10SF in Carrier 1 than in Controls (Figure 1A and Supplemental Figure 1) due to a drop in Δ10SF, 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 Δ9,10SF in Carriers and Controls (range 20-30% depending on specific protocols and/or sample used for experiments), and a significant increase of Δ10SF (with corresponding decrease of FLSF) in Carriers. Complementary analyses performed in the subpopulation of *BRCA1*Δ11q transcripts were coincident, with similar (Δ9,10+Δ11q)SF  in Carriers and Controls, and a significant increase of (Δ10+Δ11q)SF (with corresponding decrease of Δ11qSF) 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 Δ9,10SF 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](#_ENREF_7)). We used qPCR absolute quantification to estimate the SF of these alternative splicing events, detecting an increase of (Δ9,10+Δ11q)SF in Carrier 1 (9%±0.8) if compared with Controls (average of 7%). No differences were observed with regard to Δ11SF and Δ9\_11SF (Supplemental Figure 4). Similarly, dPCR analyses (Figure 2B) revealed a modest increase of Δ9,10SF in Carrier 1 (24%±0.9) if compared with Controls (average of 17%), together with a 50% reduction of FLSF 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](#_ENREF_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](#_ENREF_3)), supporting the initial pathogenic classification by Myriad Genetics ([2](#_ENREF_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 Δ9,10SF 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 (BRCA1p.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](#_ENREF_9)); 2) BRCA1Gly183\_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](#_ENREF_10)). Interestingly, BRCA1p.Gly183\_Lys223del lacks some residues critical for E3 ligase activity ([11](#_ENREF_11)), a *BRCA1* function that appears to be dispensable for tumor suppression ([12](#_ENREF_12), [13](#_ENREF_13)). Yet, the most compelling argument supporting BRCA1p.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 8.50x10-16 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](#_ENREF_14)), strongly pointing to BRCA1p.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](#_ENREF_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](#_ENREF_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](#_ENREF_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](#_ENREF_17), [18](#_ENREF_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](#_ENREF_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](#_ENREF_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](#_ENREF_20), [21](#_ENREF_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](#_ENREF_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](#_ENREF_22)), and pathology LRs were applied as indicated in Spurdle et al ([23](#_ENREF_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](#_ENREF_20)).

We used the program Phase 2.0 ([24](#_ENREF_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](#_ENREF_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](#_ENREF_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 Δ9*SF* and Δ9,10*SF*([14](#_ENREF_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](#_ENREF_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 ▼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 (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 Δ9,10SF and FLSF, 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 Δ9,10SF 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 Δ9,10SF. The Δ9,10SF 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).

**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 Δ9,10SF 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