

ORIGINAL ARTICLE

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 Soukarieh², Irene López-Perolio¹, Ana Vega³, Logan C. Walker⁴, Yvette van Ierland⁵, Diana Baralle⁶, Marta Santamariña⁷, Vanessa Lattimore⁴, Juul Wijnen^{5,8}, Philip Whaley⁹, Ana Blanco³, Michela Raponi⁶, Jan Hauke¹⁰, Barbara Wappenschmidt¹⁰, Alexandra Becker¹⁰, 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 Poplawski^{28,29}, Phuong Mai³⁰, Julian Peto³¹, Nichola Johnson^{32,33}, Barbara Burwinkel^{34,35}, Harald Surowy^{34,35}, Stig E. Bojesen^{36,37,38}, Henrik Flyger³⁹, Annika Lindblom⁴⁰, Sara Margolin⁴¹, Jenny Chang-Claude^{42,43}, Anja Rudolph⁴², Paolo Radice⁴⁴, Laura Galastri⁴⁵, Janet E. Olson⁴⁶, Emily Hallberg⁴⁶, Graham G. Giles^{47,48}, Roger L. Milne^{47,48}, Irene L. Andrulis^{49,50}, Gord Glendon⁴⁹, Per Hall⁵¹, Kamila Czene⁵¹, Fiona Blows⁵², Mitul Shah⁵², Qin Wang⁵³, Joe Dennis⁵³, Kyriaki Michailidou^{53,54}, Lesley McGuffog⁵³, Manjeet K. Bolla⁵³, Antonis C. Antoniou⁵³, Douglas F. Easton⁵³, Fergus J. Couch^{46,55}, Sean Tavitgian⁵⁶, Maaïke P. Vreeswijk⁵, Michael Parsons⁹, Huong D. Meeks⁵⁶, Alexandra Martins², David E. Goldgar^{57,†} and Amanda B. Spurdle^{9,†,*}

[†]The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint last Authors.

Received: November 23, 2015. Revised: February 10, 2016. Accepted: March 17, 2016

© The Author 2016. Published by Oxford University Press.

All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

¹Molecular Oncology Laboratory, Instituto de Investigacion Sanitaria San Carlos (IdISSC), Hospital Clinico San Carlos, Madrid, Spain, ²Inserm U1079-IRIB, University of Rouen, Normandy Centre for Genomic and Personalized Medicine, Rouen, France, ³Fundacion Publica Galega de Medicina Xenómica-SERGAS Grupo de Medicina Xenómica-USC, IDIS, CIBERER, Santiago de Compostela 15706, Spain ⁴Department of Pathology, University of Otago, Christchurch 8140, New Zealand, ⁵Department of Clinical Genetics, Leiden University Medical Centre, Leiden 2300, The Netherlands, ⁶Human Development and Health, Faculty of Medicine, University of Southampton, Southampton S016 5YA, UK, ⁷CIBERER, Grupo de Medicina Xenómica-USC, Universidade de Santiago de Compostela, Fundacion Galega de Medicina Xenómica (SERGAS), Santiago de Compostela 15706, Spain, ⁸Department of Clinical Genetics, Leiden University Medical Centre, Leiden 2300, The Netherlands, ⁹Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia, ¹⁰Center for Hereditary Breast and Ovarian Cancer, Center for Integrated Oncology (CIO), Medical Faculty, University Hospital Cologne, Cologne 50931, Germany, ¹¹Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen DK-2100, Denmark, ¹²Peter MacCallum Cancer Center, University of Melbourne, Melbourne, VIC 3002, Australia, ¹³Department of Gynaecology and Obstetrics, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Düsseldorf 40225, Germany, ¹⁴Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Campus Kiel, Christian-Albrechts University Kiel, Kiel 24105, Germany, ¹⁵Institute of Human Genetics, University of Münster, Münster 48149, Germany, ¹⁶Institute of Human Genetics, Hannover Medical School, Hannover 30625, Germany, ¹⁷Institute of Medical Genetics and Applied Genomics, University Hospital Tuebingen, Tuebingen 72076, Germany, ¹⁸National Institutes of Health, Bethesda, MD 20892-2152, USA, ¹⁹Center for Medical Genetics, NorthShore University Health System, University of Chicago Pritzker School of Medicine, Evanston, IL 60201, USA, ²⁰Service de Génétique, Department de Biologie des Tumeurs, Institut Curie and INSERM U830, Centre de Recherche de l'Institut Curie, Paris, and Université Paris Descartes, Sorbonne Paris Cité, Paris 75248, France, ²¹Service de Génétique, Department de Biologie des Tumeurs, Institut Curie, Paris 75248, France, ²²Centre Francois Baclesse, Laboratoire de Biologie et de Genetique du Cancer, 14076 Caen, Paris 75248, France, ²³Ambry Genetics, Aliso Viejo, CA 92656, USA, ²⁴Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK, ²⁵Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia, ²⁶Department of Obstetrics and Gynaecology, Medical University of Vienna, Vienna, Austria, Waehringer Guertel 18-20, A 1090 Vienna, Austria, ²⁷Genetic Health Service NZ, South Island Hub, Christchurch Hospital, Christchurch 8140, New Zealand, ²⁸Adult Genetics Unit, South Australian Clinical Genetics Service, SA Pathology at the Women's and Children's Hospital, North Adelaide, SA 5067, Australia, ²⁹University Department of Paediatrics, University of Adelaide, North Terrace, Adelaide, SA 5000, Australia, ³⁰Clinical Genetics Branch, DCEG, NCI, NIH, Bethesda, MD, USA, ³¹London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, ³²Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London SW3 6JB, UK, ³³Division of Breast Cancer Research, Institute of Cancer Research, London SW3 6JB, UK, ³⁴Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, University of Heidelberg, Heidelberg 69120, Germany, ³⁵Molecular Epidemiology Group, German Cancer Research Center, DKFZ, Heidelberg 69120, Germany, ³⁶Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen 2730, Denmark, ³⁷Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev 2730, Denmark, ³⁸Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 1165, Denmark, ³⁹Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, 2730 Denmark, ⁴⁰Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm SE-171 77, Sweden, ⁴¹Department of Oncology Pathology, Karolinska Institutet, Stockholm SE-171 77, Sweden, ⁴²Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, ⁴³University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany, ⁴⁴Unit of "Molecular bases of genetic risk and genetic testing", Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano 20139, Italy, ⁴⁵Associazione Volontari Italiani Sangue (AVIS) comunale di Milano, Milano 20139, Italy, ⁴⁶Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA, ⁴⁷Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, VIC 3010, Australia, ⁴⁸Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC 3004, Australia,

⁴⁹Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada, ⁵⁰Department of Molecular Genetics, University of Toronto, M5B 1W8, Canada, ⁵¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden, ⁵²Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK, ⁵³Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK, ⁵⁴Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, 1683, Nicosia, Cyprus, ⁵⁵Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA, ⁵⁶Department of Oncological Sciences and ⁵⁷Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84132, USA

*To whom correspondence should be addressed at: QIMR Berghofer Medical Research Institute, 300 Herston Rd, Brisbane, QLD 4006, Australia. Tel: +617 3362 0371; Email: amanda.spurdle@qimr.edu.au (A.B.S.); Laboratorio de Oncología Molecular, Instituto de Investigación Sanitaria San Carlos (IdISSC), Hospital Clínico San Carlos, c/Martin Lagos s/n, Madrid 28040, Spain. Tel: +34 913303348; Fax +34 913303544; Email: mhoya@hotmail.com (M.d.l.H)

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 composed of ≈ 70 – 80% truncating transcripts, and ≈ 20 – 30% of in-frame $\Delta 9,10$ 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 Material, Tables S1 and S2.

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 Material, Table S1[TQ1]). 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 Material, Table S1), 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 13 770:1). 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- to 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 eight 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 Material, Table S1), and breast tumor pathology information for 32 cases from ENIGMA, CIMBA or BCAC (Supplementary Material, Tables S1 and S2). 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 estrogen receptor (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 Amby Genetics that were excluded from analysis due to lack of relevant clinical information. Specifically, review of genetic testing data by Amby 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 BRCA1c.[594-2A > C; 641A > G] (Supplementary Material, Fig. S1A) 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; 641A > G] carriers by capillary electrophoresis and high throughput RNA sequencing

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 Materials and 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 11 detected up to five different alternative splicing events both in LCLs from one c.[594-2A > C; 641A > G] carrier (Carrier 1) and healthy controls (Fig. 1), including three in-frame (full-length (FL), $\Delta 9,10$, and $\nabla 10p$), and two out-of-frame ($\Delta 9$ and $\Delta 10$). All but $\nabla 10p$ (r.594-21_594-1ins) 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 Fig. 1 and Supplementary Material, Fig. S1) indicated that $\Delta 9,10$ splicing fraction (SF) ($\Delta 9,10^{SF}$) is similar in Carrier 1 and Controls ($\approx 29\%$), $\Delta 10^{SF}$ is considerably higher ($\approx 38\%$ versus $\approx 1\%$), and FL^{SF} much lower ($\approx 31\%$ versus $\approx 66\%$). $\Delta 9^{SF}$ ($< 3\%$) and $\nabla 10p^{SF}$ ($< 1\%$) 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 (Fig. 1A and Supplementary Material, Fig. S1) due to a drop in $\Delta 10^{SF}$, probably reflecting nonsense-mediated decay (NMD) degradation of out-of-frame $\Delta 10$ transcripts.

Overall, findings were confirmed by comparable experiments performed by other contributing centers (Supplementary Material, Fig. S2), assaying up to eight individual variant allele carriers identified in four unrelated families and three 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 $\Delta 11q$ 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 (Supplementary Material, Fig. S2C). Incidentally, our data supports $\nabla 10p$ 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 $\nabla 10p$ in Controls) were confirmed by RNA-seq experiments (Supplementary Material, Fig. S3).

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 $\Delta 9,10$ transcripts given that $\nabla 10p$ 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 (Fig. 2). qPCR

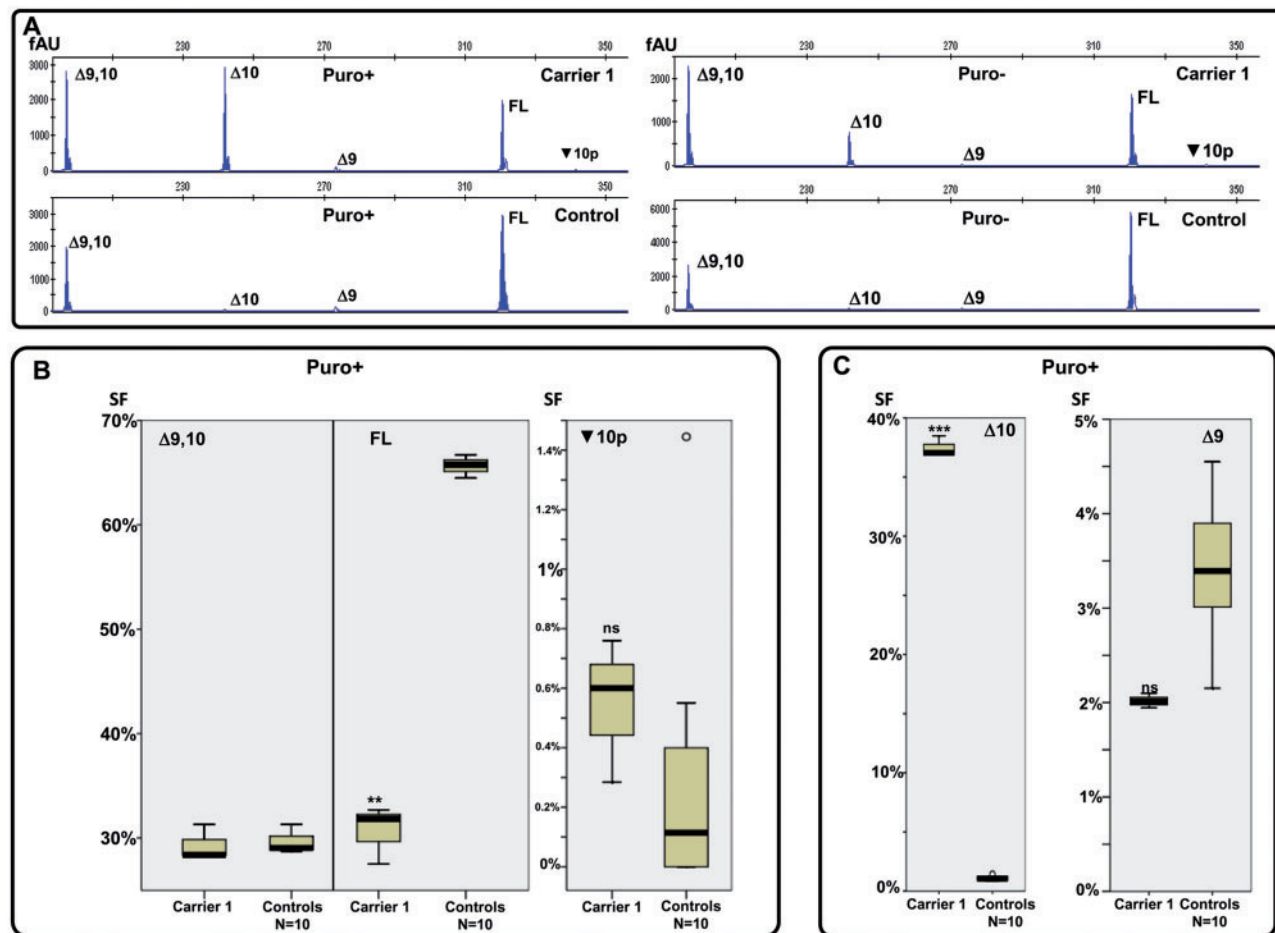


Figure 1. Capillary Electrophoresis analyses of BRCA1 alternative splicing landscape in LCLs from one BRCA1c.[594-2A>C; 641A>G] carrier (Carrier 1) and 10 Controls. Panel A shows representative examples of capillary electrophoresis analysis of RT-PCR products generated with the E8.1-E11p assay in LCLs treated (Puro+) or untreated (Puro-) with the NMD inhibitor puromycin. The fluorescence intensity of each peak (Y-axis) is expressed in arbitrary units (AU). The analyses detected the FL transcript, and up to four alternative splicing events, two in-frame ($\Delta 9,10$ and $\nabla 10p$) and two out-of-frame ($\Delta 9$ and $\Delta 10$). In these particular examples, $\nabla 10p$ transcripts are detected only in Carrier 1, but we have detected $\nabla 10p$ transcripts in Controls, as summarized in panel B. The presence of $\nabla 10p$ in Controls has been further confirmed by RNAseq (see Supplementary Material, Fig. S3). The boxplots in Panel B (displaying low, Q1, median, Q3 and high values) show the SF of in-frame transcripts ($\Delta 9,10$, FL and $\nabla 10p$) observed in Carrier 1 (three technical replicates) 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 $\nabla 10p^{SF}$ 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 ($\Delta 9$ and $\Delta 10$) observed in Carrier 1 (three technical replicates) and 10 Controls. The relative contribution of $\Delta 10$ 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 $P \leq 0.01$) (--- represents $P \leq 0.001$) (ns = non-significant).

absolute quantification of individual alternative splicing events in Carrier 1 estimated for $\Delta 9,10^{SF}$ a value of $19\% \pm 0.9$, in the upper-limit of Controls (ranging from 4% to 17%; Fig. 2A, left), together with an obvious reduction of FL transcripts (Fig. 2A, right). In addition to $\Delta 9,10$, three other naturally occurring in-frame alternative splicing events involving exon 10 and/or nearby exon 11 have been described, namely $\Delta 9_11$, $\Delta 11$ and $\Delta 11q$ (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 ($9\% \pm 0.8$) if compared with Controls (average of 7%). No differences were observed with regard to $\Delta 11^{SF}$ and $\Delta 9_11^{SF}$ (Supplementary Material, Fig. S4). Similarly, dPCR analyses (Fig. 2B) revealed a modest increase of $\Delta 9,10^{SF}$ in Carrier 1 ($24\% \pm 0.9$) 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

(Supplementary Material, Fig. S5) performed in carriers 3–5 (from one Dutch family) confirmed that $\Delta 9,10$ expression is biallelic, whereas $\Delta 10$ expression is essentially monoallelic (arising from the variant allele). Neither qPCR absolute quantification (Supplementary Material, Fig. S4A), nor biallelic expression analysis (Supplementary Material, Fig. S5B) 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 $\Delta 9,10$ and FL expression relative to BRCA2 (Supplementary Material, Fig. S6). The data indicated that $\Delta 9,10$ 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, RNA sequencing (RNAseq), qPCR, dPCR and alternative splicing event specific sequencing experiments supported a model in which the variant allele does not produce novel BRCA1

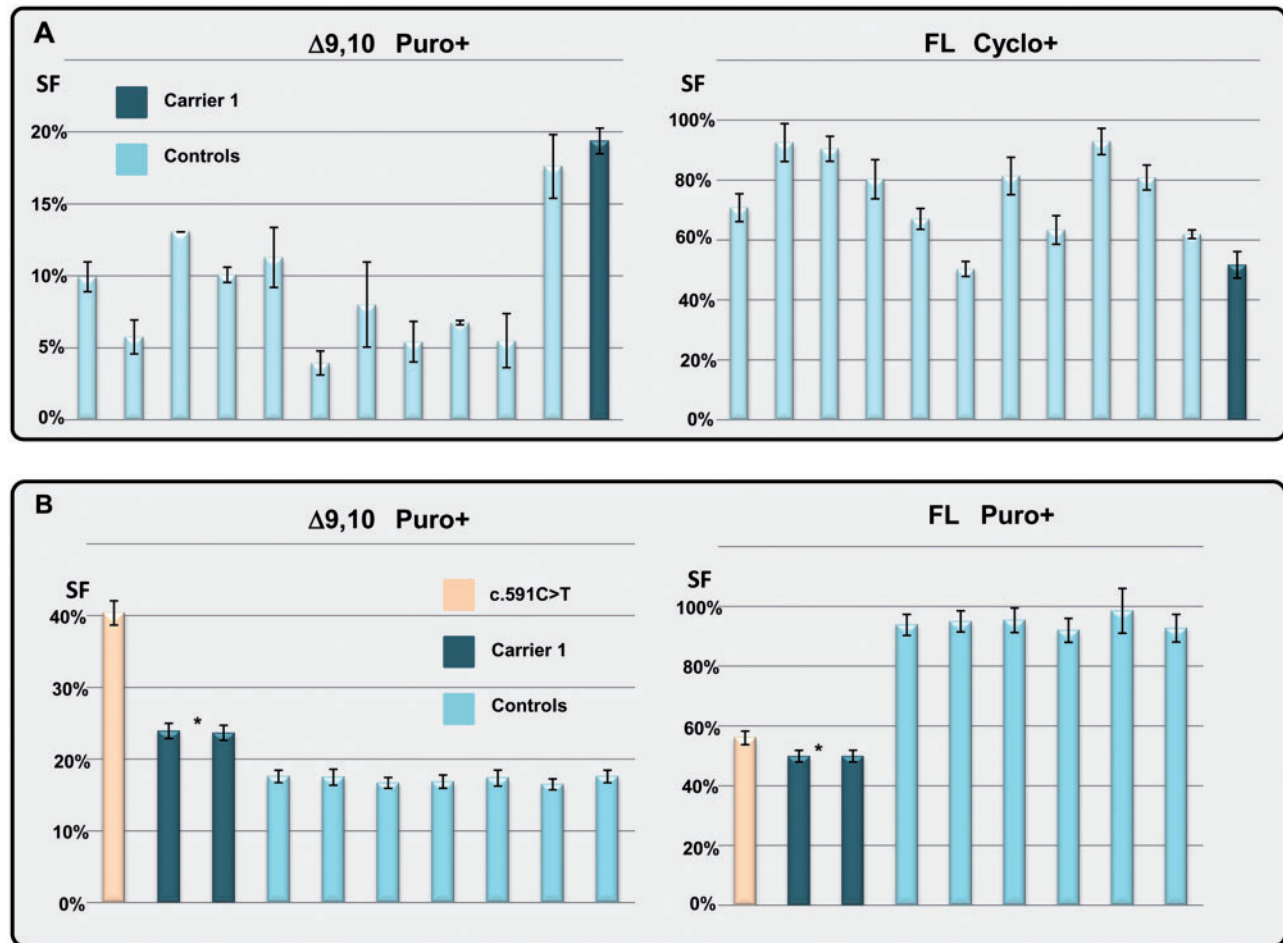


Figure 2. Quantification of major in-frame transcripts $\Delta 9,10$ and FL in LCLs from one BRCA1c.[594-2A>C; 641A>G] carrier (Carrier 1) and Controls. Experiments were performed in LCLs treated with Puromycin (Puro+). Panel A displays $\Delta 9,10^{\text{SF}}$ and FL^{SF}, estimated as the ratio between the GAPDH normalized absolute numbers of $\Delta 9,10$ (or FL) molecules and absolute number of all BRCA1 transcripts, as determined by qPCR analysis performed with standard curves (see Supplementary Material, Methods and Fig. S4). Standard deviation of 3 independent measures is shown. Panel B displays dPCR data measuring $\Delta 9,10^{\text{SF}}$ 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 $\Delta 9,10^{\text{SF}}$. The $\Delta 9,10^{\text{SF}}$ in Carrier 1 was higher than in Controls (24% in two technical replicates of Carrier 1 versus 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 replicates of Carrier 1 versus an average of 94% in six control samples, Mann-Whitney U test; $P=0.036$ for difference between groups).

transcripts, nor increases overall BRCA1 expression level, but rather substitutes FL transcripts (containing exons 9 and 10) with out-of-frame $\Delta 10$ transcripts, such that the contribution of in-frame $\Delta 9,10$ transcripts to the overall expression level is similar or slightly higher (see Fig. 2, Supplementary Material, Fig. S2) 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 $\Delta 11q$ transcripts (see Supplementary Material, Figs S2C and S4B). According to this model, BRCA1 $\Delta 9$ (out-of-frame) and BRCA1 $\nabla 10p$ (in-frame) contribution to the overall expression level are very low both in variant and WT alleles (see Fig. 2B and C), 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; 641A>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 (8). 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 (Fig. 3). pCAS2-BRCA1-Exon10 c.594-2A>C and pB1 c.594-2A>C predominantly produced $\nabla 10p$ transcripts, but also a minor amount of $\Delta 10$ transcripts (Fig. 3A), a finding confirming previous pSPL3-BRCA1-Exon10 experiments performed in COS-7 cells (9). In contrast, pCAS2-BRCA1-Exon10 c.641A>G and pB1 c.641A>G mostly produced $\Delta 10$ but no detectable $\nabla 10p$. 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 (Supplementary Material, Fig. S7A). The presence of regulatory mechanisms

underlying BRCA1 exon 10 splicing was further supported by small interfering RNA experiments performed in MDA-MB231 cells showing that endogenous BRCA1 depends on Tra2- β for exon 10 inclusion (Supplementary Material, Fig. S7B). 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 $\Delta 10$ and $\nabla 10p$, with $\Delta 10$ being the predominant outcome in all cell lines tested (Fig. 3).

Discussion

In this 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 $\Delta 9,10$ 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 BRCA1c.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 FL transcripts, nor other in-frame transcripts apart from normal levels of $\Delta 9,10$ and residual levels of $\nabla 10p$ transcripts. These findings lead us to conclude that $\Delta 9,10$ transcripts arising from the variant allele confer sufficient tumor suppressor activity *in vivo* to compensate for the lack of FL

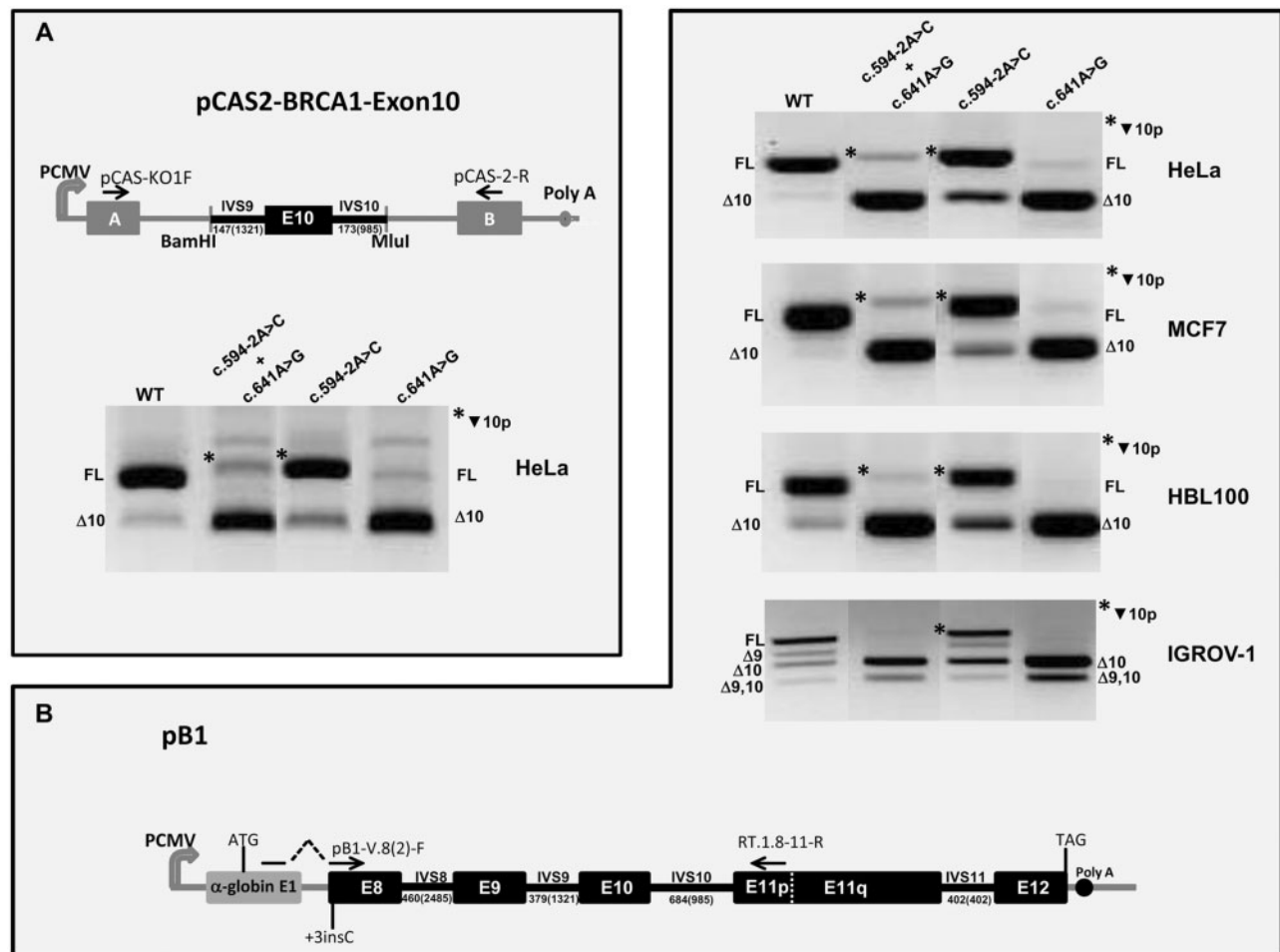


Figure 3. Analysis of BRCA1 c.594-2A > C and c.641A > G variants with splicing reporter minigene assays. The figure shows schematic non-scale representations of the splicing reporter minigenes pCAS2-BRCA1-exon10 (panel A) and pB1 (panel B) used for splicing assays. Minigenes were constructed as described under Supplementary Material, 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 FL 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 FL 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 (8). 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 Supplementary Methods. The images show RT-PCR products separated in ethidium bromide-stained agarose gels. FL, full-length; $\Delta 9$, exon 9 skipping; $\Delta 10$, exon 10 skipping; $\Delta 9,10$, skipping of both exons 9 and 10; asterisk, retention of 21 intronic nucleotides immediately upstream exon 10 ($\nabla 10p$). 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 $\Delta 10$, whereas that of endogenous wild-type BRCA1 in blood related samples is $\Delta 9,10$ (Fig. 4 and Supplementary Material, Figs 1 and 2).

transcripts. To be more precise, the combined genetic and splicing data lead us to formulate a $\Delta 9,10$ rescue model in which BRCA1 alleles with an associated $\Delta 9,10^{SF}$ of $\approx 20\text{--}30\%$ (as measured in blood related samples) confer tumor suppressor haplo-sufficiency (Fig. 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 2–8), but at any rate is very similar to that observed in control samples. The finding that $\Delta 9,10$ is a predominant alternative splicing event not only in blood derived samples but also in clinically relevant tissues such as breast and ovary (Supplementary Material, Fig. S8) 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 $\Delta 9,10$ 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 $\Delta 9,10$ transcripts (Fig. 4).

Evidently, the BRCA1 $\Delta 9,10$ rescue model presumes that $\Delta 9,10$ 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 (10); 2) BRCA1^{P.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* (11). Interestingly, BRCA1^{P.Gly183_Lys223del} lacks some residues critical for E3 ligase activity (12), a BRCA1 function that appears to be dispensable for tumor suppression (13,14). 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 $\Delta 9,10$ transcripts, a significant proportion of out-of-frame $\Delta 9$ transcripts, and very few FL transcripts (15), 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 polyposis 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 (16).

The BRCA1 $\Delta 9,10$ 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 five-tier classification of BRCA1/2 genetic variants based on qualitative and quantitative information (<http://www.enigmaconsortium.org/>, last accessed March 30, 2016). 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 $\Delta 9,10$ (or other known naturally occurring in-frame alternative splicing events) might rescue gene functionality (see Supplementary Material, Table S3). 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 (17), 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 $\approx 70\text{--}80\%$ 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

Genotyping and Sample Sets

We undertook screening of BRCA1 c.594-2A > C by direct genotyping, as part of the iCOGS experiment detailed elsewhere (18,19). This study included genotype and pathology results from breast cancer cases and controls participating in the

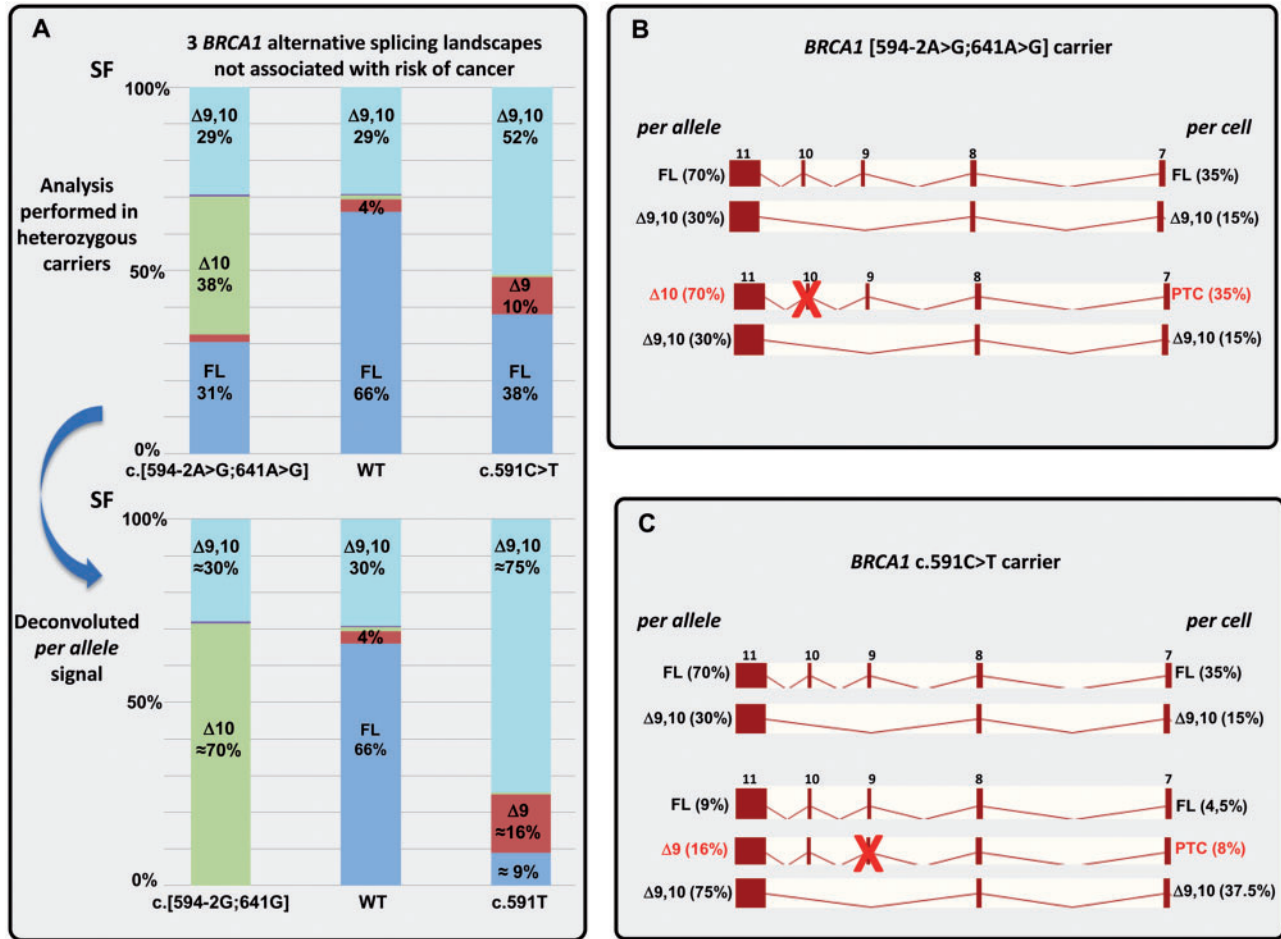


Figure 4. Combined genetic and splicing analyses of BRCA1 c.[594-2A>G; 641A>G] and BRCA1 c.591C>T supports a BRCA1Δ9,10 rescue model with far-reaching clinical implications. Panel A (top) shows the SF of five alternative splicing events detected by capillary electrophoresis analysis of RT-PCR products generated with the E8.2-E11q.2 assay (Puro+ experiments, 36 cycle PCRs, see Fig.1 and Supplementary Material, Fig. 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>G; 641A>G] carriers, and c.591C>T carriers. Yet, we show in the present study that none of these three 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 7–11) 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 7–11) 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 BRCA1Δ9,10 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>G; 641A>G] carriers is ≈29%, while other experiments suggests that the actual value is probably lower in both instances (Fig. 2, Supplementary Material, Fig. 2), albeit slightly increased in Carriers versus Controls. The model has been elaborated with data obtained in LCLs, not in clinically relevant tissues such as breast or ovarian. PTC, premature termination codon.

Breast Cancer Association Consortium (BCAC; <http://apps.ccgce.medschl.cam.ac.uk/consortia/bcac/>, last accessed March 30, 2016), and from carriers of BRCA1 assumed pathogenic variants participating in the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA; <http://apps.ccgce.medschl.cam.ac.uk/consortia/cimba/>, last accessed March 30, 2016). In addition, via the Evidence-based Network for Investigating Germline Mutant Alleles (ENIGMA, <http://enigmaconsortium.org/>, last accessed

March 30, 2016 (20)), we identified probands recruited through familial cancer clinics who were found to be positive for BRCA1 c.594-2A>G 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 ER, Progesterone Receptor (PR), and Herceptin-2 (HER2)

status), and also pedigree and segregation information where available. For carriers identified through 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 (18). 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 Material, Table S1). 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 years from 46 studies in CIMBA recruited through cancer genetics clinics. There were 4845 females without report of cancer, 4713 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 Material, Table S2). Only non-overlapping data was included in multifactorial likelihood analysis.

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 (21,22), 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 (23), and pathology LRs were applied as indicated in Spurdle et al. (24) 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 (21).

We used the program Phase 2.0 (25) 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.

mRNA analysis methods

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 8–11). 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, FL refers throughout the text to *BRCA1* exons 9- and 10-containing transcripts (transcripts containing the exons9/10 junction defined in the GenBank reference sequence U14680.1), 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), \blacktriangledown (intron retention), p (proximal, or 5') and q (distal, or 3').

RNA analysis of human samples

Up to seven contributing laboratories (sites 1–7) performed RNA splicing analyses with various methodologies, including fluorescent RT-PCR followed by capillary electrophoresis, real-time qPCR, dPCR, Sanger sequencing and RNAseq (see Supplementary Material, Methods for further details). Experiments were performed in RNAs extracted from LCLs, short-term (3–6 days) cultured LEU or fresh peripheral blood. RNAs were derived from eight individual c.[594-2A > C; 641A > G] carriers (hereafter referred as Carriers 1–8) identified in four unrelated families from Australia (Carrier 1, LCL), Germany (Carrier 2, LCL), The Netherlands (Carriers 3–7, 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 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 and 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 Supplementary Material, 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 $\Delta 9,10^{SF}$ (15). 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 event specific RT-PCR followed by Sanger sequencing through rs1060915 (an informative exonic SNP located at BRCA1 exon 13), using primers and protocols previously described (15). 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 Supplementary Material, Methods for further details).

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 two different types of reported minigenes: pCAS2-BRCA1-Exon10 and pB1 (a minigene spanning BRCA1 exons 8–12). See Supplementary Material, Methods and Figure 3 for further details.

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, Tra2 β , SF2/ASF and SC35). Experiments were performed in the breast cancer cell line MDAMD231 (see Supplementary Material, Methods for further details).

Supplementary material

Supplementary Material is available at HMG online.

Acknowledgements

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 Juliane Köhler 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: Dorthe Uldall Andersen, Maria Birna Arnadóttir, Anne Bank, Dorthe Kjeldgård 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; BernardoBonanni, 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, Nayana Weerasooriya.

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 La Boissière, 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, Professor 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. All other authors have declared no conflicts of interest.

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 M.d.H., PI13/02030 to A.V.]; the French Ministry of Higher Education and Research [to O.S.]; the University of Otago, Mackenzie Charitable Foundation, Maria Lupton, and Health Research Council of New Zealand [to L.W.]; UK Higher Education Funding Council Senior Fellowship Scheme, the University of Southampton [to D.B.]; Cancer research UK [to D.B., M.R.]; FamilienHede Nielsen Foundation fund [to T.V.O.H.]; Cancer Research-UK Senior Cancer Research Fellowship [to A.C.A.]; National Institute of Health [CA128978 and CA11616 to F.J.C.]; an NIH specialized program of research excellence in breast cancer to the Mayo Clinic [P50 CA116201 to F.J.C.]; and the US Breast Cancer Research Foundation [to F.J.C.]; 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 A.M.]; The Cancer Council Queensland [APP1086286 to A.B.S.]; the NHMRC Senior Research Fellowship Scheme [ID 1061779 to A.B.S.]; NHMRC Project grant scheme [ID 1010719 to A.B.S.].

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 Krebsforschungszentrum (DKFZ), 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 Krebsforschungszentrum (DKFZ), 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 Istituto Nazionale Tumori.

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 Märta and Hans Rausing Initiative Against Breast Cancer, and the Swedish Medical Research Council.

SEARCH is 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 no. 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).

References

- 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.
- 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. Raponi, M., Douglas, A.G., Tammara, C., Wilson, D.I. and Baralle, D. (2012) Evolutionary constraint helps unmask a splicing regulatory region in BRCA1 exon 11. *PLoS One.*, **7**, e37255.
9. Steffensen, A.Y., Dandanell, M., Jonson, L., Ejlersen, 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.
10. Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabarty, 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.
11. 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.
12. 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.
13. 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.
14. Savage, K.I. and Harkin, D.P. (2015) BRCA1, a 'complex' protein involved in the maintenance of genomic stability. *FEBS J.*, **282**, 630–646.
15. Dosit, 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.
16. 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. *Crit. Rev. Oncol. Hematol.*, **61**, 153–161.
17. 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.
18. 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.
19. 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.
20. 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.
21. Goldgar, D.E., Easton, D.F., Deffenbaugh, A.M., Monteiro, A.N., Tavtigian, S.V., Couch, F.J. and Breast Cancer Information Core (BIC) Steering Committee. (2004) Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am. J. Hum. Genet.*, **75**, 535–544.
22. Goldgar, D.E., Easton, D.F., Byrnes, G.B., Spurdle, A.B., Iversen, E.S., Greenblatt, M.S. and the IARC Unclassified Genetic Variants Working Group. (2008) Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum. Mutat.*, **29**, 1265–1272.
23. 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.
24. 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.
25. Delaneau, O., Marchini, J. and Zagury, J.F. (2012) A linear complexity phasing method for thousands of genomes. *Nat. Methods*, **9**, 179–181.