Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Breast Cancer Association Consortium. Breast cancer risk genes — association analysis in more than 113,000 women. N Engl J Med 2021;384:428-39. DOI: 10.1056/NEJMoa1913948

Supplementary Material

Breast cancer risk genes: association analysis in more than 113,000 women.

Leila Dorling, Sara Carvalho, Jamie Allen, Anna González-Neira, Craig Luccarini, Cecilia Wahlström, Karen A. Pooley, Michael T. Parsons, Cristina Fortuno, Qin Wang, Manjeet K. Bolla MSc, Joe Dennis, Renske Keeman, M. Rosario Alonso, Nuria Álvarez, Belen Herraez, Victoria Fernandez, Rocio Núñez-Torres, Ana Osorio , Jeanette Valchich, Minerva Li, Therese Törngren, Patricia A. Harrington, Caroline Baynes, Don M. Conroy, Brennan Decker, Laura Fachal, Nasim Mavaddat, Thomas Ahearn, Kristiina Aittomäki, Natalia N. Antonenkova, Norbert Arnold, Patrick Arveux, Margreet G.E.M. Ausems, Päivi Auvinen, Heiko Becher, Matthias W. Beckmann, Sabine Behrens, Marina Bermisheva, Katarzyna Białkowska, Carl Blomqvist, Natalia V. Bogdanova, Nadja Bogdanova-Markov, Stig E. Bojesen DrMedSci, Bernardo Bonanni, Anne-Lise Børresen-Dale, Hiltrud Brauch, Michael Bremer, Ignacio Briceno MD, Thomas Brüning, Barbara Burwinkel, David A. Cameron, Nicola J. Camp, Archie Campbell, Angel Carracedo, Jose E. Castelao, Melissa H. Cessna, Stephen J. Chanock, Hans Christiansen, J. Margriet Collée, Emilie Cordina-Duverger, Sten Cornelissen, Kamila Czene, Thilo Dörk, Arif B. Ekici, Christoph Engel, Mikael Eriksson, Peter A. Fasching, Jonine Figueroa, Henrik Flyger, Asta Försti, Marike Gabrielson, Manuela Gago-Dominguez, Vassilios Georgoulias, Fabian Gil, Graham G. Giles, Gord Glendon, Encarna B. Gómez Garcia, Grethe I. Grenaker Alnæs, Pascal Guénel, Andreas Hadjisavvas, Lothar Haeberle, Eric Hahnen, Per Hall, Ute Hamann, Elaine F. Harkness, Jaana M. Hartikainen, Mikael Hartman, Wei He, Bernadette A.M. Heemskerk-Gerritsen, Peter Hillemanns, Frans B.L. Hogervorst, Antoinette Hollestelle, Weang Kee Ho, Maartje J. Hooning, Anthony Howell, Keith Humphreys, Faiza Idris, Anna Jakubowska, Audrey Jung, Pooja Middha Kapoor, Michael J. Kerin, Elza Khusnutdinova, Sung-Won Kim, Yon-Dschun Ko, Veli-Matti Kosma, Vessela N. Kristensen , Kyriacos Kyriacou, Inge M.M. Lakeman, Jong Won Lee, Min Hyuk Lee MD, Jingmei Li, Annika Lindblom, Wing-Yee Lo, Maria A. Loizidou, Artitaya Lophatananon, Jan Lubiński, Robert J. MacInnis, Michael J. Madsen, Arto Mannermaa, Mehdi Manoochehri, Siranoush Manoukian, Sara Margolin, Maria Elena Martinez, Tabea Maurer, Dimitrios Mavroudis, Catriona McLean, Alfons Meindl, Arjen R. Mensenkamp, Kyriaki Michailidou, Nicola Miller, Nur Aishah Mohd Taib, Kenneth Muir, Anna Marie Mulligan, Heli Nevanlinna, William G. Newman, Børge G. Nordestgaard, Pei Sze Ng, Jan C. Oosterwijk, Sue K. Park, Tjoung-Won Park-Simon, Jose I.A. Perez, Paolo Peterlongo, David J. Porteous, Karolina Prajzendanc, Darya Prokofyeva, Paolo Radice, Muhammad U. Rashid, Valerie Rhenius, Matti A. Rookus, Thomas Rüdiger, Emmanouil Saloustros, Elinor J. Sawyer, Rita K. Schmutzler, Andreas Schneeweiss, Peter Schürmann, Mitul Shah, Christof Sohn, Melissa C. Southey, Harald Surowy, Maija Suvanto, Somchai Thanasitthichai, Ian Tomlinson, Diana Torres, Thérèse Truong, Maria Tzardi, Yana Valova, Christi J. van Asperen, Rob M. Van Dam, Ans M.W. van den Ouweland, Lizet E. van der Kolk, Elke M. van Veen, Camilla Wendt, Justin A. Williams, Xiaohong R. Yang, Sook-Yee Yoon, M. Pilar Zamora, D. Gareth Evans, Miguel de la Hoya, Jacques Simard, Antonis C. Antoniou, Åke Borg, Irene L. Andrulis, Jenny Chang-Claude, Montserrat García-Closas, Georgia Chenevix-Trench, Roger L. Milne, Paul D.P. Pharoah, Marjanka K. Schmidt, Amanda B. Spurdle, Maaike P.G. Vreeswijk, Javier Benitez, Alison M. Dunning, Anders Kvist, Soo H. Teo, Peter Devilee, Douglas F. Easton PhD, on behalf of NBCS Collaborators, kConFab/AOCS Investigators, MyBrCa Investigators, and SGBCC Investigators.

Contents

Supplementary Methods

Studies

We included samples of female breast cancer (BC) patients (cases) and unaffected controls from 44 studies participating in the BCAC (http://bcac.ccge.medschl.cam.ac.uk/; Tables S1, S2). All studies were approved by the relevant ethical review boards and used appropriate consent procedures. Of these, 30 were population-based or hospital-based studies that included cases and controls independent of family history. A further 14 studies oversampled cases with a family history of BC (e.g. selecting cases attending cancer genetics clinics), while one study oversampled controls with a family history of cancer. Some studies, by design, included more than one woman from the same family, but for the analyses presented here, only data on the index cases were included. All women included were aged >18 years. In total, samples from 59,299 controls and 67,269 BC patients were included; after all quality control steps (see below), 53,461 controls and 60,466 cases with an invasive (54,624; 90.3%) or in situ (4,187; 6.9 %) tumor or tumor of unknown invasiveness (1,655; 2.7%) were included in the analyses.

Library preparation and sequencing

We defined a panel of 35 genes (Table S4; Supplementary File 5). We included 32 genes provided on commercial genetic testing panels at the time of design (in early 2016), for which breast cancer was an indication. We also included 3 other genes (*RINT1*, *BRE, RECQL*) suggested as susceptibility genes in the literature[1,2]. The analyses presented include the results of 34 genes, excluding *PPM1D*. Previous studies have shown an association between PTVs in *PPM1D* and breast and ovarian cancer risk, but for variants seen at low allelic fractions ("somatic mosaicism"). These variants are not inherited, are potentially due to treatment, and hence not relevant to the analysis of germline susceptibility variants presented here [3-5]. 365 carriers of the PTV c.1100delC in *CHEK2* (~1/3 of the total) overlapped with previous BCAC studies genotyping this variant.[6] Specific variants in *PALB2* (6), *ATM* (1) and *CHEK2* (6) were previously genotyped using the iCOGS array.[7]

Library preparation was conducted using the Fluidigm Juno 192.24 system in three laboratories (Human Cancer Genetics Programme, Human Genotyping Unit- Cegen. Spanish National Cancer Research Centre (CNIO), Madrid, Spain; Department of Clinical Sciences Lund, Lund University, Lund, Sweden; Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK). For all samples except those of SEARCH (Table S1), we used a sequencing panel of 1,349 fragments, designed to cover the coding sequence, intron/exon boundaries and UTRs of the 35 genes (Supplementary File 5). We attempted to cover alternative transcripts, but classified variants according to a canonical transcript (Table S4). We also included additional regulatory sequences for *BRCA1* and *BRCA2*, and 224 fragments that included known common breast cancer susceptibility variants (Supplementary File 5). For SEARCH, we used the same technology but designed an augmented panel that included 18 additional genes (not reported here).

Amplified products were combined into barcoded libraries of 768 samples, which were run on a single lane of an Illumina Hiseq4000. Samples were demultiplexed and then aligned to the reference genome (hg19) using BWA-MEM[8]. Each sample was sequenced to an average depth of 349 reads, in the target region. Depth, along with base quality, was used as part of the secondary quality control filtering.

Variant calling and quality control

Variant calling was performed using VarDict[9]; comparison with other callers indicated that this had much better specificity for this type of targeted sequencing[10]. We applied the following filters at the VCF level: phred scaled sequencing quality assessment of the bases contributing to the variant (QUAL) <30, allele fraction (AF) <0.2 and mean mapping quality (MQMEAN) <60, mean number of mismatches per read (NM) >2.0, AFxBase Depth < 7.5. Variants failing any of these filters were removed. We also removed any variants exhibiting amplicon bias (i.e. not present on all the amplicons covering the variant).

We next derived a callability matrix which indicated whether each position in the target region was callable in each sample, and eliminated positions and samples with low callable fraction. A callable position was defined as one with at least 15x coverage with base quality at least 20. We successively increased the callable fraction threshold from 0.01 to 0.95 in 0.01 increments, so in the final dataset all samples were callable in at least 95% of positions and all positions were callable in at least 95% of samples. The final callable sequenced region was 130.5kb, representing 91.1% of the target sequence. 107kb/114kb (93.8%) of the coding sequence was callable (Table S5).

As a final check, Integrative Genomics Viewer[11] was used to inspect read alignments for all 2,905 variants predicted to result in a truncated protein, including indels, nonsense substitutions, and canonical splice altering variants. Variant nomenclature errors were corrected (n=160) and likely miscalls were removed (n=623).

We excluded known or identified duplicates and close relatives identified through comparison of array genotypes from the iCOGs and OncoArray projects, and known close relatives based on pedigree data. We also excluded samples for which the genotypes were not consistent with the array genotyping, suggestive of sample swapping. We also excluded individuals who were from a minority ancestry for that study (that is, non-east Asian individuals from the 4 Asian studies and non-European individuals from the European studies). Ethnicity was defined genetically using principal components analysis from the array genotype data where this was available[12], otherwise by self-report. For Malaysia and Singapore (see below) we excluded admixed individuals, defined as not reaching a 50% threshold for a single ancestry (Chinese, Malay or Indian) based on genotyping.

PTVs were defined as frameshifting insertions/deletions, stop/gain or canonical splice variants as classified by the Emsembl Variant Effect Predictor (VEP)[13], with the exception of variants in the last exon of each gene, which were excluded from the primary analysis. We also exclude splice variants affecting the penultimate exon as these may lead to exon skipping and not result in nonsense mediated decay, with exception of 6 genes for which there is evidence that the truncating protein would still be pathogenic, irrespective of exon skipping (summarised in Table S6).We further excluded 7 canonical splice variants in *BRCA1* which are of uncertain significance according to ENIGMA guidelines: (c.594-2A>C[14], c.4096+1G>A, c.4096+2T>C, c.4096+1G>A and three variants within tandem acceptor sites: c.4186-2A>G, c.4358-1G>C, c.4358-2del). In-frame deletions/insertions, non-canonical splice variants, variants in UTRs and other intronic variants were not considered.

Missense variants were classified by protein domain location, principally as defined by UniProt (https://www.uniprot.org/), and, for *BRCA1*, *BRCA2* and *TP53*, by whether they were likely to be considered pathogenic according to commonly accepted guidelines. For *BRCA1* and *BRCA2*, subset analyses were conducted for variants considered pathogenic or likely pathogenic by either ClinVar

(https://www.ncbi.nlm.nih.gov/clinvar/) or ENIGMA *BRCA1/2* expert panel guidelines (https://enigmaconsortium.org/). For *TP53*, we also considered a definition of (likely) pathogenic, based on American College of Medical Genetics (ACMG) guidelines[15], augmented by variants classified as (likely) pathogenic based on a published quantitative model for *TP53* missense variant classification that utilizes a combination of bioinformatic prediction and reported germline:somatic ratio for a given variant[16].

Summary counts for PTVs and rare missense variants in population-based studies and all studies combined are provided as Supplementary Files.

Variant detection sensitivity and positive predictive value

Sensitivity was assessed by two approaches. First, we compared SNV calls for 75,059 samples previously genotyped using arrays (iCOGS and OncoArray)[12,17], based on samples and positions that passed quality control filters. Sensitivity was 89.7% (7,893/8,803 called variants) for variants with MAF<0.1%, and 94.8% (48,866/51,538) for variants with a frequency 0.1-1%. For common variants, genotype concordance was 97.3%. Second, we evaluated 130 samples that had previously been subject to sequencing in a clinical testing laboratory in Sweden, in which putative deleterious variants had been confirmed by Sanger validation, and 65 samples from carriers of deleterious *BRCA1/2* variants recruited into the EMBRACE study in the UK[18]. These samples were subject to the same library preparation and sequencing pipeline as the study samples. Of 207 confirmed variants within the filtered sequence, 198 (95.7%) were identified (77 SNVs, 92.8% and 121 indels, 98.4%).

Confirmatory Sanger sequencing was carried out on 160 PTVs and 145 missense variants that were called by VarDict and passed all QC filters above. The Positive Predictive Value (PPV) was 99.4% (159/160) and 93.1% (135/145), respectively.

Statistical analysis

The primary analyses were burden analyses in which the odds ratios (OR, with 95% confidence intervals) for carrying any variant in a given category were estimated using logistic regression. The primary analyses included covariates to adjust for country, except for Malaysia and Singapore, in which the three distinct ethnic groups (Chinese, Indian, Malay) were treated as different strata, and the UK, which was treated as three strata (SEARCH, from East Anglia, GENSCOT from Scotland and PROCAS and FHRISK from north-west England). We conducted separate analyses including only studies or substudies in which cases and controls were not selected for family history ("population-based studies"), and only studies in which the cases were oversampled for family history ("family-based studies"). One study (KOHBRA), in which controls were enriched for family history, was excluded from both these analyses. The odds ratios should provide consistent estimates of the incidence rate ratio (hazard ratio), but may overestimate the relative risk (ratio of the cumulative risk in carriers to non-carriers).

Heterozygous and homozygous carriers of variants in a gene were not distinguished as it was not always possible to do so with certainty, and the number of homozygotes was too small for separate analysis. "PTV carriers" and "missense variant carriers" therefore refer to either monoallelic (heterozygote) or bilallelic carriers throughout. Rare missense variants were defined as having a population frequency of less than 0.001, based on a weighted average of the frequencies in gnomAD non-Finnish Europeans (89%) and East Asian individuals (11%). If the variant could not be called in gnomAD, the weighted average allele frequency in the current dataset was used. Carriers of PTVs in

BRCA1 were excluded from the analysis of *BRCA2,* and vice-versa, and carriers of PTVs in *BRCA1* or *BRCA2* were excluded from the analysis of all other genes. Carriers of PTVs in other genes were excluded from the analysis of missense variants in that gene.

We conducted analyses for overall (invasive or in-situ) BC, BC by estrogen receptor (ER) - subtype and, among ER-negative cases, triple negative and non-triple-negative disease. Case-only analyses were used to evaluate the evidence for differences in OR by subtype and by age (assuming a linear trend in the log(OR) with age). Tests of the difference in effect size between population-based and familial enriched studies were performed by fitting multinomial logistic regression models with three outcomes (control, population-based case, familial case) and constructing likelihood ratio tests relative to the null model in which the effect sizes for population-based and familial studies were constrained to be equal.

To evaluate differences in the OR by ethnicity for those genes with a significant trend in OR by age, we computed age-specific ORs for each ethnicity, assuming the same linear trend in the log(OR) by age (as for the cumulative risk analyses below).

Bayesian False Discovery Probabilities

To determine Bayesian False Discovery Probabilities (BFDPs),[19] we assumed a prior probability of association of 0.99 for *BRCA1, BRCA2* and *CHEK2,* 0.8 for *PALB2* and *TP53*, and 0.5 for *ATM*. These probabilities were chosen to reflect the strong prior evidence for these genes (though the results for these genes were quite insensitive to the assumed prior and would have achieved a BFDP<5% for any plausible prior). We chose priors of 0.3 for *RAD51C* and *RAD51D*, reflecting their known associations with ovarian cancer, 0.2 for all remaining genes listed as probably disease associated in the overview by Easton *et al.*[20], and 0.1 for the remaining genes. We assumed a log-normally distributed prior effect size as described by Wakefield, except that we only considered positive associations as the prior evidence for all genes was in favour of PTVs being positively associated with risks. The variance of the prior log(OR) was determined by assuming a 95% probability that the OR was less than some bound K, where K=20 for *BRCA1* and *BRCA2*, K=6 for *PALB2* and K=3 for the other genes. (The results were insensitive to this latter assumption).

Absolute risk estimation

Cumulative risks, in the absence of other events, were calculated by combining age-specific relative risk estimates with the population incidence rates for the UK (2016) as a baseline, as previously described[6]. The age-specific relative risks were derived by assuming a linear trend in the log(relative risk) with age, estimated from the case-only analysis[6]. The age-specific ORs were all consistent with a log-linear decline in the OR with age. These relative risk estimates were derived from the population-based, European ancestry studies only. Revised calculations would be necessary for populations with different incidences (assuming the same relative risks). Cumulative risks were not computed for *TP53*, given the wide confidence interval on the relative risk estimate and the substantial childhood cancer risk.

Variant prevalences and familial relative risks

Adjusted population prevalences for the associated genes were computed from the observed prevalences in population controls, adjusted by the estimated sensitivity of the testing, using the formula $2p' = 2p/(cs(1-v))$, where $2p$ is the observed prevalence, *c* is the proportion of the coding sequence that was determined to be callable, *s* is the sensitivity of the testing for callable variants, as estimated by the comparison with known sequence variants and *v* is the proportion of deleterious variants that are assumed to be copy number variants (and hence not detectable). *c* was

estimated on a per gene basis (Table S5) while *s* was estimated across all genes as 0.957 (see above). v was assumed to be 0.15 for *BRCA1*[21]. For other genes, *v* was estimated from the proportion of unique variants annotated as pathogenic or likely pathogenic in ClinVar that were 50bp or larger: the assumed proportions were: *ATM*: 0.06, *BARD1*: 0.11, *BRCA2*: 0.02, *CHEK2*: 0.16, *RAD51C:* 0.22, *RAD51D: 0.14, PALB2: 0.08.* For *CHEK2* the adjustment was made only for the set of variants excluding c.1100delC.

The familial relative risk of breast cancer attributable to each gene was estimated using the formula :

$$f_j = {p_j\psi_j^2 + (1-p_j)(p_j\psi_j + 1 - p_j)^2} \Big/ {(2p_j\psi_j + 1 - 2p_j)^2}$$

Where p_j is the (combined) allele frequency of deleterious variants in gene *j* and ψ_j is the corresponding odds ratio. The combined effect of all genes was then derived as

$$ln\left(1 + \sum_j f_j - 1\right) \Big/ \ln(2)$$

That is, assuming an additive effect of the genes, and an overall familial relative risk to first degree relatives of 2.0.

Table S1. Description of studies included in the analyses.

Study	Abbreviation	Country	Study design	Case definition	Control definition	Selected familial cases	Design category	References
Amsterdam Breast Cancer Study	ABCS	Netherlands	Hospital-based consecutive cases; population-based controls (for iCOGS/OncoArray/BRIDGES from blood bank).	iCOGS/OncoArray/BRIDGES: Breast cancer patients diagnosed before age 50 in 1995-2011 at the Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital (NKI-AVL).	iCOGS/OncoArray/BRIDGES: Population-based cohort of women recruited through the Sanquin blood bank, all ages.	No	Mixed	[17,22]
Amsterdam Breast Cancer Study - Familial	ABCS-F	Netherlands	Clinical Genetic Center-based cases	iCOGS/OncoArray/BRIDGES: All non-BRCA1/2 breast cancer cases from the family cancer clinic of the NKI-AVL tested in the period 1995-2009; all ages and diagnosed with breast cancer in 1972-2010.	No controls. [Use controls of ABCS]	Yes	Case-only; clinical genetic center-based	[6]
Asia Cancer Program	ACP	Thailand	Hospital-based case-control study	Cases recruited 1999-2000 and 2008-present at The National Cancer Institute (Central region), The Prince Songkla University Research Centre (South region), The HRH Princess Maha Chakri Sirindhorn Medical Centre (MSMC)-Srinakarinviroj University (Eastern region), Khon-Kaen University Cancer Centre (North-eastern region). 1. Women who underwent biopsy and have been pathologically diagnosed as having breast cancer. 2. Aged less than 71 years of age.	Controls recruited 1999-2000 and 2008-present at The National Cancer Institute (Central region), The Prince Songkla University Research Centre (South region), The HRH Princess Maha Chakri Sirindhorn Medical Centre (MSMC)-Srinakarinviroj University (Eastern region), Khon-Kaen University Cancer Centre (North-eastern region). 1. Women aged less than 71 years of age without cancer history of any kinds 2. Women who attend the out-patient clinic under the minor injuries such as cuts, broken bones. 3. Women who are institutionalised at the hospital with diseases not related to cancer or metabolic syndromes such as diabetes, heart	No	Mixed	None

					diseases or conditions related to gynaecology and are well enough to give information to researchers.			
Bavarian Breast Cancer Cases and Controls	BBCC	Germany	Hospital-based cases; population based controls	Consecutive, unselected cases with invasive breast cancer recruited at the University Breast Centre, Franconia in Northern Bavaria during 1999-2013.	Healthy women with no diagnosis of cancer aged 55 or older. Invited by a newspaper advertisement in Northern Bavaria, and recruited during 1999-2013.	No	Mixed	**23,24**
Breast Cancer in Galway Genetic Study	BIGGS	Ireland	Hospital-based cases; population based controls	Unselected cases recruited from West of Ireland since 2001. Cases were recruited from University College Hospital Galway and surrounding hospitals	Women > 60 years with no personal history of any cancer and no family History of breast or ovarian cancer were identified from retirement groups in the West of Ireland (same catchment area as cases) during the period 2001-2008.	No	Mixed	**25-27**
Breast Oncology Galicia Network	BREOGAN	Spain	Population-based case-control	A population-based study conducted since 1997 in two cities in Galicia, Spain (Vigo and Santiago) covering approximately 700,000 inhabitants. The study currently includes over 1600 incident breast cancer cases diagnosed from 1997-2014 in two Galician hospitals with blood, tumor tissue and risk factor questionnaire.	Controls were frequency-matched to cases according to 5-year age group, inclusion in the universal Galician Public Health Service (SERGAS) registry database, and place of residence. They were healthy, unrelated female individuals from the same base population as cases randomly selected from SERGAS´ primary healthcare centers in the health areas of Santiago and Vigo. Recruitment began in 1997.	No	Population-based	**27-31**
Breast Cancer Study of the University of Heidelberg	BSUCH	Germany	Hospital-based cases;healthy blood donator controls	Cases diagnosed with breast cancer/breast cancer metastasis in 2008-2011 at the University Women`s Clinic Heidelberg.	Healthy, unrelated, ethnically matched female blood donors recruited in 2007, 2009 & 2012 by German Red Cross Blood Service of Baden-Württemberg-Hessen, Institute of Transfusion Medicine & Immunology, Mannheim.	No	Mixed	32
Crete Cancer Genetics Program	CCGP	Greece	Hospital-based case-control study	Incident breast cancer cases treated between 2004 and 2013 at the University Hospital of Heraklion on Crete; all enrolled within 6 months of diagnosis.	Healthy, unrelated, ethnically matched female blood donors recruited in 2014 by the laboratory of Hemostasis at the General Hospital of Heraklion "Venizelio".	No	Mixed	Unpublished

CECILE Breast Cancer Study	CECILE	France	Population-based case-control study	All incident cases of breast cancer diagnosed in 2005-2007 among women <75 years of age and residing in Ille-et-Vilaine or Côte d'Or. Cases were recruited from the main cancer treatment center (Centre Eugène-Marquis in Rennes and Centre Georges-François-Leclerc in Dijon) and from private or public hospitals in each area.	General population control women residing in the same geographic areas frequency-matched to the cases by 5-year age groups. Controls were recruited in 2005-2007 by phone using a random digit dialing procedure and predefined numbers by socioeconomic status to control for possible selection bias.	No	Population-based	[33]
Copenhagen General Population Study	CGPS	Denmark	Population-based case-control study	Consecutive, incident cases from 1 hospital with centralized care for a population of 400,000 women from 2001 to the present.	Community controls residing in the same region as cases and with no history of breast cancer were identified from the Copenhagen General Population Study recruited 2003-2007. All controls were known to still be breast cancer-free at the end of 2007.	No	Mixed	[34]
Spanish National Cancer Centre Breast Cancer Study	CNIO-BCS	Spain	Case-control study	Two groups of cases:1) 574 consecutive breast cancer patients, unselected for family history, from 3 public hospitals, 2 in Madrid and one in Oviedo, from 2000 to 2005. 2) 291 cases with at least one first degree relative also affected with breast cancer, recruited through the CNIO family cancer clinic in Madrid from 2000 to 2004.	Women attending the Menopause Research Centre between 2000 and 2004 and female members of the College of Lawyers attending a free, targeted medical check-up in 2005, all free of breast cancer and all in Madrid	Subset (N=291)	Mixed	[35]
Colombian Breast Cancer Case-Control Study	COLBCCC	Colombia	Case-control study	1,022 unselected women diagnosed with breast cancer after January 1, 2004; enrolled between 2007 and 2012.	1,023 healthy women attending the country-wide National Pap-Smear Screening Program in Colombia; enrolled between 2007 and 2012. Controls were matched to cases by +/- 2 years. Controls were women participating in the Colombian National Pap-Smear Screening Program (participation rate in 2005 was 77%)	No	Mixed	Unpublished

Family History Risk Study	FHRISK	UK	Clinic-based cohort study with a nested case-control study	Women diagnosed with breast cancer and attending the Family History Clinic in Manchester for increased risk of breast cancer. Recruitment period 2009-2012.	Women attending the same Family History Clinic as the cases but without a breast cancer diagnosis. Recruitment period is the same as for the cases.	Yes	Cohort and case-control	**36,37**
German Consortium for Hereditary Breast & Ovarian Cancer	GC-HBOC	Germany	Clinic-based case study and prospective cohort study	Women diagnosed with breast cancer in one of the GC-HBOC centres (Cologne, Munich, Kiel, Heidelberg, Düsseldorf, Ulm, Würzburg, Münster and Hannover). Recruitment period 1996-present.	Healthy, unrelated, ethnically and age-matched female control individuals (LIFE study, Leipzig, Germany).	Yes	Mixed	**38-41**
Gene Environment Interaction and Breast Cancer in Germany	GENICA	Germany	Population-based case-control study	Incident breast cancer cases enrolled between 2000 and 2004 from the Greater Bonn area (by of the hospitals within the study region); all enrolled within 6 months of diagnosis.	Selected from population registries from 31 communities in the greater Bonn area; matched to cases in 5-year age classes between 2001 and 2004.	No	Population-based	**42,43**
Generation Scotland	GENSCOT	Scotland	Prospective family-based cohort study; nested case-control	Incident and prevalent cases of histologically-confirmed breast cancer at the time of latest updated cancer registry linkage (currently 2013). Recruitment though the General Practitioners in the areas of Glasgow, Tayside, Ayrshire, Arran and Northeast Scotland.	Two groups of controls: (1) 2:1 unrelated individuals matched to cases on age in five-years at baseline and recruitment centre; (2) first-degree female relatives with no breast cancer diagnosis at the time of selection.	No	Prospective cohort	44
Genetic Epidemiology Study of Breast Cancer by Age 50	GESBC	Germany	Population-based study of women <50 years	All incident cases diagnosed <50 years of age in 1992-5 in two regions: Rhein-Neckar-Odenwald and Freiburg, by surveying the 38 clinics serving these regions	Selected from random lists of residents of the study regions supplied by population registries; two controls were selected for each case, matched by age and study region. Recruitment was carried out 1992-1998.	No	Population-based	45

Hannover Breast Cancer Study	HABCS	Germany	Hospital-based case-control study	Cases who received radiotherapy for breast cancer at Hannover Medical School between 1996-2003 (HaBCS I), or were diagnosed with breast cancer at a certified Breast Cancer Clinics in the Hannover region between 2012-2016 (HaBCS II), unselected for age or family history.	Anonymous female blood bank donors at Hannover Medical School, collected from 8/2005-12/2005, with known age and ethnic background.	No	Mixed	[46]
Helsinki Breast Cancer Study	HEBCS	Finland	Hospital-based case-control study, plus additional familial cases	(1) Consecutive cases (883) from the Department of Oncology, Helsinki University Central Hospital 1997-8 and 2000, (2) Consecutive cases (986) from the Department of Surgery, Helsinki University Central Hospital 2001 – 2004, (3) Familial breast cancer patients (536) from the Helsinki University Central Hospital, Departments of Oncology and Clinical Genetics (1995-)	Healthy females from the same geographical region in Southern Finland in 2003.	Subset (N=609)	Mixed	**[47-49]**
Hereditair Borst-en eierstokkanker Onderzoek Nederland	HEBON	Netherlands	Clinical genetic center-based recruitment of familial breast or ovarian cancer patients (cases)	Breast (or sometimes ovarian) cancer patients who were tested for mutations in BRCA1 and BRCA2 in one of the clinical genetic centers in the Netherlands between 1996 and 2016. All counselees received an invitation to participate in the HEBON study.	No controls. [Use of controls (bloodbank donors) from ORIGO, ABCS or RBCS].	Yes (All participants are familial cases)	Case-only; clinical genetic center-based	Unpublished

Hannover-Minsk Breast Cancer Study	HMBCS	Belarus	Hospital-based cases; population based controls	Ascertainment at the Byelorussian Institute for Oncology and Medical Radiology Aleksandrov N.N. in Minsk or at one of 5 regional oncology centers in Gomel, Mogilev, Grodno, Brest or Vitebsk through the years 2002-2008.	Controls from the same population aged 18-72 years. Healthy (without personally history of cancer) female probunds recruited from the same geographical regions as cases during the years 2002-2008. About 75% of controls were women invited for general medical examination at five regional gynecology clinics (in Gomel, Mogilev, Grodno, Brest or Vitebsk) and cancer-free volunteers ascertained at the Institute for Inherited Diseases in Minsk; 20% were cancer-free female blood bank donors recruited at Republic Blood Bank, Minsk, Belarus; finally 5% of controls were healthy cancer-free relatives of some breast cancer patients.	No	Mixed	[50]
Hannover-Ufa Breast Cancer Study	HUBCS	Russia	Hospital-based cases; population based controls	Consecutive Russian breast cancer patients aged 24-86 years ascertained at one of the two participating oncological centers in Bashkorstostan and Siberia through the years 2000-2008.	Population controls aged 18-84 years recruited from a population study of different populations of Russia. Healthy volunteers (without any malignancy) were selected from the same geographical regions during the years 2002-2008.	No	Mixed	[50]
Karolinska Breast Cancer Study	KARBAC	Sweden	Population and hospital-based cases; geographically matched controls	1. Familial cases from Department of Clinical Genetics, Karolinska University Hospital, Stockholm. 2. Consecutive cases from Department of Oncology, Huddinge & Söder Hospital, Stockholm 1998-2000	Blood donors of mixed gender from same geographical region. Excess material was received from all blood donors over a 3 month period in 2004 (approximately 3000) and DNA was extracted from a random sample of 1500	Subset (N=568)	Mixed	**[51,52]**
Karolinska Mammography Project for Risk Prediction of Breast Cancer - Cohort Study	KARMA	Sweden	Cohort study	Inclusion of 70,877 women Oct 2010 - March 2013. 3000 women had BC at cohort entry. In all, 800 women have been diagnosed with breast cancer since study entry (Oct 2015). Approximately 250 women are diagnosed with BC annually	Non - BC cases in the Karma Cohort	no	Prospective cohort	Submitted

Kuopio Breast Cancer Project	KBCP	Finland	Population-based prospective clinical cohort	1. Women seen at Kuopio University Hospital between 1990 and 1995 because of breast lump, mammographic abnormality, or other breast symptom who were found to have breast cancer. 2. Consecutive malignant breast cancer cases diagnosed at KUH from 2011 onwards.	Age and long-term area-of-residence matched controls selected from the National Population Register and interviewed in parallel with the cases	No	Population-based	**53,54**
Kathleen Cuningham Foundation Consortium for research into Familial Breast Cancer/Australian Ovarian Cancer Study	kConFab /AOCS	Australia and New Zealand	Clinic-based recruitment of familial breast cancer patients (cases); population-based case-control study of ovarian cancer (controls only)	Cases were from multiple-case breast and breast-ovarian families recruited though family cancer clinics from across Australia and New Zealand from 1998 to the present. Cases were selected for inclusion in BCAC studies if (i) family was negative for mutations in BRCA1 and BRCA2 (ii) case was the index for the family, defined as youngest breast cancer affected family member.	Female controls were ascertained by the Australian Ovarian Cancer Study identified from the electoral rolls from all over Australia from 2002-2006.	Yes	Mixed	**55,56**
Korean Hereditary Breast Cancer Study	KOHBRA	Korea	Population-based case-control study	Breast cancer patients at high risk were recruited from nationwide University Hospitals from May 2007 to May 2012. High-risk status mean 1) familial breast cancer, 2) early onset breast cancer (age <40), 3) breast and past/current ovarian cancer patients 4) cases with past/current double primary cancers, 5) bilateral breast cancer, 6) male breast cancer cases. All cases participated in the BCAC project were BRCA non-carriers and male breast cancers were not included.	Health examinee controls from communities were enrolled and individual matched to the cases on specific age. A part of the controls were recruited from unaffected family members of BRCA mutation carriers.	Subset (N=1192)	Mixed	57

Mammary Carcinoma Risk Factor Investigation	MARIE	Germany	Population-based case-control study	Incident cases diagnosed from 2001-2005 in the study region Hamburg in Northern Germany, and from 2002-2005 in the study region Rhein-Neckar-Karlsruhe in Southern Germany.	2 controls per case were randomly drawn from population registries and frequency matched by birth year and study region to the case. Controls were recruited from 2002 to 2006.	No	Population-based	[58]
Cyprus Breast Cancer Case Control Study	MASTOS	Cyprus	Population-based case-control study	Women between 40-70 years of age who had a histologically confirmed diagnosis of primary breast cancer between January 1999 and December of 2005. The majority of cases were ascertained from the Bank of Cyprus Oncology Centre, which operates as a referral centre and offers treatment and follow-up for up to 90% of all breast cancer cases diagnosed in Cyprus. The rest of the patients, were recruited at the Oncology Departments of the Nicosia, Limassol, Larnaca and Paphos district hospitals.	Cypriot women from the general population, who were invited to participate in the National programme for breast cancer screening with the use of mammography and received a negative result. Volunteers were enrolled in the study during the same calendar period as the cases, from the 5-district mammography screening centers that operate in Cyprus.	No	Population-based	[59]
Milan Breast Cancer Study Group	MBCSG	Italy	Clinic-based recruitment of familial/early onset breast cancer patients (cases); population-based controls	Familial and/or early onset breast cancer patients (aged 22-87) negative for mutations in BRCA genes, ascertained in two large cancer centres in Milan from 2000 to date.	Healthy blood donors aged 18-71 years, recruited at two blood centres in Milan from 2004 (centre 1) and 2007 (centre 2) to date	Yes (ca. 90%)	Mixed	[60,61]
Melbourne Collaborative Cohort Study	MCCS	Australia	Prospective cohort study: nested case-control study	Incident cases diagnosed between baseline (1990-1994) and last follow-up (2012) among the 24469 women participating in the cohort.	For each case a control was randomly selected from women from the cohort who did not develop breast cancer before the age at diagnosis of the case and matched the case on year of birth and country of birth.	No	Prospective cohort	[62]

Malaysian Breast Cancer Genetic Study	MYBRCA	Malaysia	Hospital-based case-control study	Breast cancer cases identified at the Breast Cancer Clinic in University Malaya Medical Centre Jan 2003-July 2014 and Subang Jaya Medical Centre Sep 2012-Sept 2014; cases are a mixture of prevalent and incident cases. Includes hospital-based and familial series.	Controls are cancer-free individuals (37-74 years) selected from women attending mammographic screening at the same hospitals.	Yes (subset)	Mixed	63,64
Norwegian Breast Cancer Study	NBCS	Norway	Hospital-based case-control study	Incidence cases from three different hospitals: 1) Cases (114) mean age 64 (28-92) at Ullevål Univ. Hospital 1990-94, 2) cases (182) mean age 59 (26-75) referred to Norwegian Radium Hospital 1975-1986, 3) cases (124), mean age 56 (29-82) with stage I or II disease, in the Oslo micro-metastases study at Norwegian Radium Hospital between 1995-1998, 4) Breast cancer cases referred to the Norwegian hospitals Akershus University Hospital in Lørenskog, Ullevaal university hospital in Oslo and Rikshospitalet-Radiumhospitalet in Oslo from 2007-2010. Mean age is 63 years. Consecutive series. 5) Breast cancer cases referred to the Norwegian Radium Hospital hospitalet 2010-2013. Neoadjuvantly treated with Avastin (Bevacizumab). 6) Consecutive series of Breast cancer incidents referred to Akershus university hospital 2004-2014.	Control subjects were healthy women, age 55-71, residing in Tromsø (440), and Bergen (109) attending the Norwegian Breast Cancer Screening Program. Healthy tissue from mammoplastic reduction surgery at a private clinic in Oslo.	No	Mixed	**65-68**

Ontario Familial Breast Cancer Registry	OFBCR	Canada	Population-based familial case-control study	Cases diagnosed between 1 Jan 1996-31 Dec 1998 were identified from the Ontario Cancer Registry which registers >97% of all cases residing in the province at the time of diagnosis. All women with invasive breast cancer aged 20–54 years who met the OFBCR definition for high genetic risk (family history of specific cancers particularly breast and ovarian, early onset disease, Ashkenazi ethnicity or a diagnosis of multiple breast cancer) were asked to participate by completing risk factor questionnaires and providing a blood sample. A 25% random sample of individuals in this age category who did not meet the OFBCR definition, 35% of those aged 55–69 at high risk and 8.75% aged 55–69 at low risk were also asked to participate. Individuals diagnosed in 2001 and 2002 were also included if they met high-risk criteria.	Unrelated, unaffected population controls were recruited by the Ontario Familial Breast and Colon Cancer Registries by calling randomly selected residential telephone numbers throughout the same geographical region. Eligible controls were women with no history of breast cancer and were frequency-matched by 5-year age group to the expected age distribution of cases.	Subset (N=628)	Mixed	69
Leiden University Medical Centre Breast Cancer Study	ORIGO	Netherlands	Hospital-based prospective cohort study	Consecutive cases diagnosed 1996-2006 in 2 hospitals of South-West Netherlands (Leiden & Rotterdam). No selection for family history; Rotterdam cases selected for diagnosis aged <70. Cases with in situ carcinomas eligible.	Three groups of controls: (1) Blood bank healthy donors from Southwest Netherlands recruited in 1996, 2000 or 2007; (2) People who married a person who was part of a family with high breast cancer risk (BRCA1/2/x). From the Southwest of the Netherlands, recruited 1990-1996; (3) Females tested at the local clinical genetics department for familial diseases, excluding familial cancer syndromes (no mutation found in gene(s) related to the disease being tested), recruited 1995-2007.	No	Mixed	**70,71**

NCI Polish Breast Cancer Study	PBCS	Poland	Population-based case-control study	Incident cases from 2000-2003 identified through a rapid identification system in participating hospitals covering ~ 90% of all eligible cases, and cancer registries in Warsaw and Łódź covering 100% of all eligible cases.	Randomly selected from population lists of all residents of Poland, stratified and frequency matched to cases by case city and age in 5 year categories. Recruited 2000-2003.	No	Population-based	[72]
The Prostate,Lung,Colorectal and Ovarian (PLCO) Cancer Screening Trial	PLCO	USA	Prospective cohort study: nested case-control	Incident cases arising in the sub-cohort of 78,232 women who gave a blood specimen in 1993-2001 are included if they were diagnosed with breast cancer. Recruitment via multiple screening centers across the US.	Controls were women in this sub-cohort who were not diagnosed with breast cancer. Controls were matched to cases on age at randomization (4 categories) and fiscal year of randomization (2 categories).	No	Prospective cohort	[73]
Predicting the Risk Of Cancer At Screening Study	PROCAS	UK	Population based study	Women diagnosed with breast cancer since joining the study of women attending the Breast Screening Programme (NHSBSP) in Greater Manchester. Recruitment period Oct 2009-May 2014.	Women attending routine NHS breast screening in Greater Manchester without a breast cancer diagnosis. Recruited during the same period as for the cases.	No	Population-based	[36]
Rotterdam Breast Cancer Study	RBCS	Netherlands	Hospital-based case-control study, Rotterdam area	Familial breast cancer patients selected from the Clinical Genetics Center at Erasmus MC Cancer Institute; recruited 1994 - 2005 (RBCS1) and 1995 - 2009 (RBCS2; for OncoArray).	Spouses or mutation-negative siblings of heterozygous Cystic Fibrosis mutation carriers selected from the Clinical Genetics Center at Erasmus MC Cancer Institute; recruited 1996 - 2006 (RBCS1) and 2005 - 2009 (RBCS2).	Yes	Mixed	[74]
Singapore and Sweden Breast Cancer Study	SASBAC	Sweden	Population-based case-control study	Incident cases from October 1993 to March 1995 identified via the 6 regional cancer registries in Sweden, to which reporting is mandatory.	Controls were randomly selected from the total population registry in 5-year age groups to match the expected age-frequency distribution among cases. Patients and controls were recruited from Oct 1993 through April 1995.	No	Population-based	[75]

Study of Epidemiology and Risk factors in Cancer Heredity	SEARCH	UK	Population-based case-control study	2 groups of cases identified through East Anglian Cancer Registry; 1) prevalent cases diagnosed 1991-1996 under 55 years of age at diagnosis, recruited 1996-2002; 2) incident cases diagnosed since 1996 under 70 years of age at diagnosis, recruited 1996-present.	Two groups of controls: (1) selected from the EPIC-Norfolk cohort study of 25,000 individuals age 45-74 recruited between 1992 and1994, based in the same geographic region as cases; (2) selected from GP practices from March 2003 to present, frequency matched to cases by age and geographic region	No	Mixed	[76]
Singapore Breast Cancer Cohort	SGBCC	Singapore	Hospital-based breast cancer cohort and population-based controls	Living breast cancer patients diagnosed with primary in situ or invasive breast cancer at 7 restructured hospitals in Singapore between 1980-2016. Cases are a mixture of prevalent and incident cases.	All community-dwelling individuals who are Singaporeans or Singaporean Permanent Residents, 21 years and older. Participants were recruited between 2006 and 2010 through word-of-mouth and personal recommendations. In some cases, recruiters also sought participants through "cold-calling" or through door-to-door invitations. Exclusion criteria were a medical history of cancer, acute myocardial infarction or stroke, or major psychiatric morbidity including schizophrenia, psychotic depression, and advanced Alzheimer's Disease.	No	Hospital-based	No refs.
Städtisches Klinikum Karlsruhe Deutsches Krebsforschungszentrum Study	SKKDKFZS	Germany	Hospital-based breast cancer cohort	Women diagnosed with primary *in situ* or invasive breast cancer at the Städtisches Klinikum Karlsruhe from March 1993 to July 2005.	No controls.	No	Patient cohort	[77]

IHCC-Szczecin Breast Cancer Study	SZBCS	Poland	Hospital-based case-control study	Prospectively ascertained cases of invasive breast cancer patients diagnosed at the Regional Oncology Hospital (Szczecin) in the years 2002, 2003, 2006 and 2007 or the University Hospital from 2002 to 2007 in Szczecin, West-Pomerania, Poland. Patients with pure intraductal or intralobular cancer were excluded (DCIS or LCIS) but patients with DCIS with micro-invasion were included.	Unaffected, matched to cases for year of birth, sex and region; from families with negative cancer family history; controls were part of a population-based study of the 1.3 million inhabitants of West Pomerania performed in 2003 and 2004 designed to identify familial aggregations of cancer by our centre	No	Mixed	**1,78-80**
Utah Breast Cancer Study	UBCS	USA	Mixed. (1) Pedigrees including multiple sampled breast cancer cases within 2 generations, also may include sampled, unaffected relatives; (2) hospital-based cases (from Huntsman Cancer Institute [HCI] or Intermountain Healthcare [IH]), and breast reduction controls; and (3) Population-based cases (from the Utah Cancer Registry [UCR]) and controls (from the Utah Drivers	Cases recruited from late 1970s to present (on-going). Ascertainment from: (1) UCR-confirmed breast cancer cases in high-risk pedigrees; (2) invasive breast cancer cases treated or surgery performed at HCI or IH clinics; (3) prevalent, population-based UCR-confirmed breast cancer cases.	Controls also recruited from late 1970s to present (on-going) from: (1) relatives in high-risk pedigrees; (2) hospital-based cancer-free women undergoing breast reductions; (3) Population-based controls selected from the UDLR to frequency match cases by sex and birth cohort.	Some	Mixed	81,82

			License Registry [UDLR])					

Table S2. Numbers of cases and controls, and age distributions, by study, after QC.

Study	Country	Cases sequenced in BRIDGES	Controls sequenced in BRIDGES	Cases in BRIDGES after QC	Controls in BRIDGES after QC	Case age at diagnosis		Control age at interview	
						Mean	Range	Mean	Range
ABCS	Netherlands	1075	1824	1007	1660	42.1	18-49	47.1	18-69
ABCS-F	Netherlands	313	0	208	0	45.2	22-86	-	-
ACP	Thailand	960	829	933	789	48.4	19-78	41.6	15-73
BBCC	Germany	357	234	244	159	61.2	27-90	57.5	22-84
BIGGS	Ireland	384	384	369	366	56.3	27-87	66.7	46-91
BREOGAN	Spain	973	570	598	398	56.0	30-88	55.1	30-86
BSUCH	Germany	263	697	241	549	56.8	32-88	57.8	30-69
CCGP	Greece	697	294	475	275	55.8	26-85	61.4	17-94
CECILE	France	988	979	941	943	54.3	25-74	54.6	25-74
CGPS	Denmark	3735	5202	3387	5076	61.5	26-98	56.3	20-94
CNIO-BCS	Spain	856	647	687	569	54.3	28-88	50.0	24-73
COLBCCC	Colombia	517	731	484	621	49.5	23-83	50.0	24-73
FHRISK	UK	311	1028	276	923	49.6	29-78	40.5	19-73
GC-HBOC	Germany	2742	1597	2566	1561	45.2	17-87	61.8	47-79
GENICA	Germany	1009	1005	848	894	58.2	23-80	58.4	24-80
GENSCOT	Scotland	478	1345	427	766	54.7	28-89	58.4	20-93
GESBC	Germany	635	1090	552	982	42.5	24-51	42.7	24-52
HABCS	Germany	1078	900	971	838	58.1	23-91	33.2	17-68
HEBCS	Finland	2154	1254	1905	1090	56.7	23-95	40.9	18-66
HEBON	Netherlands	2107	0	1953	0	47.2	22-91	-	-
HMBCS	Belarus	387	381	334	268	47.4	17-80	46.6	20-87
HUBCS	Russia	404	363	239	192	52.4	25-82	45.1	16-78
KARBAC	Sweden	421	539	376	471	59.1	27-88	-	-
KARMA	Sweden	3665	6221	3329	5633	55.5	23-94	60.2	29-82
KBCP	Finland	579	75	560	70	58.7	23-92	51.0	30-75
kConFab/ AOCS	Australia and New Zealand	1787	8	1463	7	52.9	20-94	51.9	41-77
KOHBRA	Korea	2019	2010	1956	1835	40.6	19-83	47.8	19-87

MARIE	Germany	2526	1981	2300	1768	62.1	49-75	61.8	49-75
MASTOS	Cyprus	1127	1177	990	1094	51.5	26-74	55.7	28-71
MBCSG	Italy	982	776	935	735	42.7	18-80	44.1	18-71
MCCS	Australia	1185	1139	1042	1029	63.6	31-88	63.3	39-88
MYBRCA	Malaysia	1168	1212	1076	1093	51.8	24-83	56.0	38-77
NBCS	Norway	623	614	565	600	60.3	24-96	61.4	55-71
OFBCR	Canada	562	494	505	416	58.8	24-83	55.1	25-81
ORIGO	Netherlands	0	960	0	919	-	-	-	-
PBCS	Poland	1899	1941	1757	1849	55.9	28-75-	55.6	24-75
PLCO	USA	2322	2574	2060	2221	68.4	55-87	62.3	54-74
PROCAS	UK	656	1653	518	1434	58.6	29-76	59.4	46-73
RBCS	Netherlands	1314	975	1043	899	44.4	22-99	-	-
SASBAC	Sweden	1152	1344	1131	1321	63.1	50-75	63.3	49-76
SEARCH	UK	13835	7251	12817	6486	54.5	23-87	53.3	16-87
SGBCC	Singapore	4588	4383	4271	4165	53.3	18-91	50.1	21-75
SKKDKFZS	Germany	1229	0	966	0	60.6	23-93		
SZBCS	Poland	372	204	357	191	59.2	26-91	56.7	25-85
UBCS	USA	1006	337	804	306	56.3	28-92	57.0	18-94

Table S3. Summary of other phenotypes established to be associated with deleterious germline variants in each gene on the BRIDGES panel

Gene	Other associated cancers	Other associated phenotypes	Syndrome
ABRAXAS1	-		
AKT1			
ATM	Leukemia, lymphoma (homozygotes)		Ataxia-telangiectasia (homozygotes)
BABAM2	-		
BARD1			
BRCA1	Ovary		
BRCA2	Ovary, prostate, pancreas, male breast, leukemia (homozygotes), brain tumors (homozygotes), Wilms' tumor (homozygotes)		Fanconi anaemia (homozygotes)
BRIP1	Ovary		Fanconi anaemia (homozygotes)
CDH1	Diffuse gastric, endometrial		
CHEK2			
EPCAM	Colorectal, endometrial, gastric, ovary	diarrhea-5 with congenital tufting enteropathy (DIAR5)	Lynch syndrome, Constitutional Mismatch Repair Syndrome (CMMRS)
FANCC	-		Fanconi anaemia (homozygotes)
FANCM	-		
GEN1	-		
MEN1	-	Neuroendocrine tumors, pituitary adenomas, insulinomas, parathyroid adenomas, prolactinomas	Multiple endocrine neoplasia type I
MLH1	Colorectal, endometrial, gastric, ovary		Lynch syndrome, Constitutional Mismatch Repair Syndrome (CMMRS)

MRE11	-		Ataxia-telangiectasia-like disorder (homozygotes)
MSH2	Colorectal, endometrial, gastric, ovary		Lynch syndrome, Constitutional Mismatch Repair Syndrome (CMMRS)
MSH6	Colorectal, endometrial, gastric, ovary		Lynch syndrome, Constitutional Mismatch Repair Syndrome (CMMRS)
MUTYH	Colorectal	Multiple colorectal adenomatous polyps	MUTYH associated polyposis
NBN	-	Aplastic anaemia (homozygotes)	Nijmegen breakage syndrome (homozygotes)
NF1	Neurofibrosarcomas, CNS tumors	Café-au-lait spots, neurofibromas, phaechromocytomas, paragangliomas	Neurofibromatosis Type I
PALB2	Pancreas		Fanconi anaemia (homozygotes)
PIK3CA	-		
PMS2	Colorectal, endometrial, gastric, ovary		Lynch syndrome, Constitutional Mismatch Repair Syndrome (CMMRS)
PTEN	Thyroid, colorectal, melanoma, endometrial, renal	Multiple hamartomas	Cowden's syndrome, PTEN tumor hamartoma syndrome, Bannayan-Riley-Ruvalcaba syndrome, macrocephaly-autism syndrome
RAD50	-		Nijmegen breakage syndrome-like disorder (homozygotes)
RAD51C	Ovary		
RAD51D	Ovary		
RECQL	-		

RINT1	-		Infantile liver failure syndrome 3 (homozygotes)
STK11	Colorectal, lung, pancreatic, thyroid, sertoli tumors	Hamartomatous polyps, hyperpigmented spots	Peutz-Jeghers syndrome
TP53	sarcoma, leukemia, brain, adrenocortical, choroid plexus carcinoma		Li-Fraumeni syndrome
XRCC2	-	poor growth, microcephaly, radial defects (homozygotes)	Fanconi anemia (homozygotes)

Table S4. Genes included on the BRIDGES panel, with canonical transcripts used in the analyses.

Gene	Ensembl transcript	NCBI transcript	Number of exons (Coding exons)	Comments
ABRAXAS1	ENST00000321945.7	NM_139076	9 (9)	Canonical, Havana gold flag, selected by HGMD.
AKT1	ENST00000555528.1	NM_005163	14 (13)	Non canonical (but joint largest protein), Havana gold flag, selected by HGMD.
ATM	ENST00000278616.4	NM_000051	63 (62)	Canonical, selected by HGMD.
BABAM2	ENST00000344773.2	NM_004899	13 (11)	Canonical, Havana gold flag.
BARD1	ENST00000260947.4	NM_000465	11 (11)	Canonical, Havana gold flag, selected by HGMD.
BRCA1	ENST00000357654.3	NM_007294	23 (22)	Non canonical (but largest protein), Havana gold flag, selected by HGMD.
BRCA2	ENST00000544455.1	NM_000059	28 (26)	Canonical, selected by HGMD.
BRIP1	ENST00000259008.2	NM_032043	20 (19)	Canonical, Havana gold flag, selected by HGMD.
CDH1	ENST00000261769.5	NM_004360	16 (16)	Canonical, Havana gold flag, selected by HGMD.
CHEK2	ENST00000328354.6	NM_007194	15 (14)	Non canonical, Havana gold flag, selected by HGMD.
EPCAM	ENST00000263735.4	NM_002354	9 (9)	Canonical, selected by HGMD.
FANCC	ENST00000289081.3	NM_000136	15 (14)	Canonical, Havana gold flag, selected by HGMD.
FANCM	ENST00000267430.5	NM_020937	23 (23)	Canonical, Havana gold flag, selected by HGMD.
GEN1	ENST00000317402.7	NM_182625	14 (13)	Non canonical (but same protein length), Havana gold flag, selected by HGMD.
MEN1	ENST00000312049.6	NM_130799	10 (9)	Non canonical, Havana gold flag, selected by HGMD.
MLH1	ENST00000231790.2	NM_000249	19 (19)	Canonical, Havana gold flag, selected by HGMD.
MRE11	ENST00000323929.3	NM_005591	20 (19)	Canonical, Havana gold flag.
MSH2	ENST00000233146.2	NM_000251	16 (16)	Non canonical (but longest protein), Havana gold flag, selected by HGMD.
MSH6	ENST00000234420.5	NM_000179	10 (10)	Canonical, Havana gold flag, selected by HGMD.
MUTYH	ENST00000450313.1	NM_001128425	16 (16)	Canonical, selected by HGMD.
NBN	ENST00000265433.3	NM_002485	16 (16)	Non canonical (but joint largest protein), Havana gold flag, selected by HGMD.
NF1	ENST00000356175.3	NM_000267	57 (57)	Non canonical, Havana gold flag, selected by HGMD.
PALB2	ENST00000261584.4	NM_024675	13 (13)	Canonical, Havana gold flag, selected by HGMD.
PIK3CA	ENST00000263967.3	NM_006218	21 (20)	Canonical, Havana gold flag, selected by HGMD.
PMS2	ENST00000265849.7	NM_000535	15 (15)	Canonical, Havana gold flag, selected by HGMD.
PPM1D	ENST00000305921.3	NM_003620	6 (6)	Non canonical (but largest protein), Havana gold flag, selected by HGMD.

PTEN	ENST00000371953.3	NM_000314	9 (9)	Canonical, Havana gold flag, selected by HGMD.
RAD50	ENST00000378823.3	NM_005732	25 (22)	Canonical, selected by HGMD.
RAD51C	ENST00000337432.4	NM_058216	9 (9)	Non canonical (but largest protein), Havana gold flag, selected by HGMD.
RAD51D	ENST00000345365.6	NM_002878	10 (10)	Canonical, Havana gold flag, selected by HGMD.
RECQL	ENST00000444129.2	NM_002907	15 (14)	Canonical, Havana gold flag.
RINT1	ENST00000257700.2	NM_021930	15 (15)	Canonical, Havana gold flag.
STK11	ENST00000326873.7	NM_000455	10 (9)	Non canonical (but largest protein), Havana gold flag, selected by HGMD.
TP53	ENST00000269305.4	NM_000546	11 (10)	Non canonical (but largest protein), Havana gold flag, selected by HGMD.
XRCC2	ENST00000359321.1	NM_005431	3 (3)	Non canonical (but only one producing protein), Havana gold flag, selected by HGMD.

Table S5. Coverage statistics by gene: bases targeted, callability and coverage for all targets, excluding samples failing QC (see Supplementary Methods).

Gene	Bases Targeted	Bases Callable	Callable Fraction	Mean coverage over targeted bases
ABRAXAS1	1,410	1,304	0.92	367
AKT1	1,703	1,640	0.96	348
ATM	10,411	10,102	0.97	411
BABAM2	1,468	1,468	1.00	335
BARD1	2,554	2,217	0.87	359
BRCA1	6,032	5,714	0.95	382
BRCA2	10,777	10,426	0.97	351
BRIP1	4,130	4,128	1.00	437
COH1	2,969	2,902	0.98	450
CHEK2	1,912	1,912	1.00	415
EPCAM	1,125	1,029	0.91	340
FANCC	1,957	1,957	1.00	435
FANCM	6,607	6,601	1.00	415
GEN1	2,987	2,951	0.99	455
MEN1	2,013	1,440	0.72	222
MLH1	2,651	2,285	0.86	300
MRE11	2,507	2,392	0.95	384
MSH2	3,125	3,063	0.98	411
MSH6	4,283	4,001	0.93	439
MUTYH	1,970	1,941	0.99	394
NBN	2,285	2,283	1.00	420
NF1	9,597	9,250	0.96	419
PALB2	3,821	3,692	0.97	444
PIK3CA	3,607	3,307	0.92	410
PMS2	2,889	2,460	0.85	390

PPM1D	1,938	1,447	0.75	434
PTEN	1,392	1,242	0.89	293
RAD50	4,439	4,334	0.98	451
RAD51C	1,311	1,103	0.84	382
RAD51D	1,187	891	0.75	269
RECQL	2,230	2,213	0.99	508
RINT1	2,679	2,670	1.00	499
STK11	1,482	767	0.52	170
TP53	1,382	1,382	1.00	504
XRCC2	903	845	0.94	384

Table S6 Summary of considerations for inclusion/exclusion of canonical splice variants affected the penultimate exon.

gene	HGVS c.DNA	penultimate exon skipping	predicted in-frame protein alteration	exon skipping predicted pathogenic, irrespective of NMD?	rationale	reference
ATM	c.8851-1,-2 c.8987+1,+2	r.8851_8987del (FS-alternative STOP)		yes	C-terminal residues 2957 to 2998 (PRD), and 3023 to 3056 (FATC) are critical to ATM structure/function. Exon skipping will introduce a FS alteration not preserving Val2951 to Val3056, and is therefore predicted LoF	[83]
BARD1	c.1904-1,-2 c.2001+1,+2	r.1904_2001del (FS-alternative STOP)		yes	The C-terminal BRCT domain p.(Ser616_Ser777) is critical for BARD1 function. Exon skipping will introduce a FS alteration not preserving residues Trp635 to Ser777, and is therefore predicted LoF	[84]
BRCA1	c.5407-1,-2 c.5467+1,+2	r.5407_5467del (FS-alternative STOP)		yes	The C-terminal BRCT domain p.(Leu1764_Pro1859) is critical for BRCA1 function. Exon skipping will introduce a FS alteration not preserving residues Gly1803 to Pro1859, and is therefore predicted LoF.	[85], ENIGMA classification rules
BRCA2	9502-1,-2 9648+1,+2	r.9502_9648del (no-FS)	p.(Asn3168_Leu3216del)	**unknown**	The DBD p.(2481-3186) is critical for BRCA2 function, but the clinical or functional relevance of p.(Asn3168_Leu3216del), eliminating only the C-terminal 18aa of the DBD as unknown (ENIGMA classification rules)	ENIGMA classification rules
BRIP1	c.2576-1,-2 c.2905+1,+2	r.2576_2905del (no-FS)	p.(Gly859_Lys967del)	**unknown**	p.Ser 990 is critical for BRCA1 binding but, as far as we know, there is no functional/clinical data demonstrating a critical role for the p.(Gly859_Lys967) region	

CHEK2	c.1462-1,-2 c.1542+1,+2	r.1462_1542del (no-FS)	p.(Pro488_Gln514del)	**unknown**	The kinase domains expands residues 220_486. The functional relevance of C-terminal residues 488_514 is unknown	[86]
RAD51C	c.966-1,-2 c.1026+1,+2	r.966_1026del (FS-alternative STOP)		yes	Exon skipping will not preserve the C-terminal end of the protein p.(Leu323_Leu373), including C-terminal B-strands 7, 8 and 9, considered structurally relevant , and is therefore predicted LoF	[87]
RAD51D	c.739-1,-2 c.903+1,+2	r.739_903del (no-FS)	p.(Val247_Gln301del)	yes	Exon skipping will eliminate residues Val247 to Gln301. These residues include RAD51D C-terminal domain B-stands 5,6, 7, considered structurally relevant, and is therefore predicted LoF	[87]
PALB2	c.3202-1,-3 c.3350+1,+2	r.3202_3350del (FS-alternative STOP)		yes	The C-terminal WD40 domain is critical for PALB2 function. Exon skipping is a frameshift alteration eliminating WD40 blades 5,6, and 7 (including residues critical for PALB2-BRCA2 interaction, and residues critical for toroidal structure 'sealed' in the seventh blade), and therefore predicted LoF	[88]

Table S7. Associations of protein truncating germline variants and overall breast cancer risk, separately for women of European and Asian descent, for genes showing overall evidence of association. Ethnicity defined by study and genotype (see Supplementary Methods).

	European studies					Asian studies					Asian vs. European				
	Carriers					Carriers					Unadjusted			Age adjusted*	
Gene	Cases	Controls	OR	(95% CI)	p.value	Cases	Controls	OR	(95% CI)	p-value	OR	(95%CI)	p-diff	OR	(95%CI)
ATM	266	138	2.07	(1.68-2.57)	2.0E-11	26	11	2.34	(1.15-4.76)	0.019	1.13	(0.54-2.37)	0.75		
BARD1	51	28	2.02	(1.26-3.24)	0.0038	11	4	2.54	(0.81-8.00)	0.11	1.26	(0.36-4.36)	0.72		
BRCA1	425	55	9.33	(7.00-12.43)	1.75E-52	65	3	22.07	(6.91-70.48)	1.8E-07	2.37	(0.72-7.82)	0.16	2.01	(0.60-6.67)
BRCA2	607	118	5.38	(4.38-6.59)	7.63E-59	136	17	8.16	(4.90-13.57)	6.2E-16	1.52	(0.88-2.63)	0.14	1.51	(0.85-2.69)
CHEK2	693	307	2.57	(2.23-2.95)	2.51E-39	11	8	1.51	(0.60-3.82)	0.39	0.59	(0.23-1.50)	0.27	0.53	(0.21-1.37)
MSH6	34	23	1.66	(0.96-2.86)	6.760E-02	3	0	0.00	(0-Inf)	0.91			0.92		
NF1	23	16	1.36	(0.71-2.62)	0.36	8	1	7.84	(0.98-62.74)	0.052	5.76	(0.65-50.9)	0.12		
PALB2	235	48	4.99	(3.62-6.86)	5.71E-23	35	7	4.52	(2.00-10.22)	2.9E-04	0.91	(0.38-2.18)	0.83	0.87	(0.36-2.08)
PTEN	11	5	2.14	(0.72-6.34)	0.17	3	1	2.81	(0.29-27.37)	0.38	1.31	(0.11-16.4)	0.83		
RAD51C	39	19	1.89	(1.08-3.31)	0.027	15	7	2.04	(0.83-5.02)	0.12	1.08	(0.37-3.13)	0.89		
RAD51D	37	19	1.65	(0.94-2.91)	0.082	14	6	2.27	(0.87-5.92)	0.09	1.38	(0.45-4.19)	0.58		
TP53	7	2	3.06	(0.63-14.92)	0.17	0	0								

*****Assuming the same linear trend in the log(OR) in both populations.

Table S8. Association analysis for PTVs in 34 genes by subtype of breast cancer, in population-based studies.

	ER-positive (30,466 cases)[1]				ER-negative (7,766 cases)[1]					Triple negative (2,841 cases)[1]				ER-, not triple negative (2,556 cases)[1]				
Gene	Case carriers	OR	95% CI	p-value	Case carriers	OR	95% CI	p-value	p-diff	Case carriers	OR	95% CI	p-value	Case carriers	OR	95% CI	p-value	p-diff
ABRAXAS1	8	0.78	(0.33-1.82)	0.56	2	0.81	(0.18-3.55)	0.78	0.96	*1*	0.98	(0.13-7.45)	0.98	0	0	(0 - Inf)	0.99	0.95
AKT1	3	0.81	(0.19-3.54)	0.78	0	0.00	(0-Inf)	0.99	0.90	*0*	0.00	(0-Inf)	1.00	0	0.00	(0 - Inf)	1.00	
ATM	196	2.33	(1.87-2.91)	9.4E-14	22	1.01	(0.64-1.59)	0.97	0.00055	7	0.91	(0.42-1.95)	0.81	9	1.25	(0.63 - 2.47)	0.53	0.50
BABAM2	5	0.71	(0.23-2.20)	0.55	0	0.00	(0-Inf)	0.99	0.91	0	0.00	(0-Inf)	0.99	0	0.00	(0 - Inf)	0.99	
BARD1	24	1.40	(0.81-2.42)	0.23	27	5.99	(3.51-10.21)	5.3E-11	4.8E-07	12	9.29	(4.58-18.85)	6.6E-10	3	2.44	(0.72 - 8.24)	0.15	0.044
BRCA1	120	3.92	(2.82-5.43)	3.2E-16	269	35.32	(26.21-47.60)	3.2E-121	2.5E-80	165	56.80	(41.18-78.34)	6.5E-134	30	11.18	(6.96– 17.95)	1.7E-23	9.6E-17
BRCA2	446	5.69	(4.65-6.96)	3.3E-57	149	7.53	(5.89-9.62)	1.2E-58	0.012	74	11.19	(8.27-15.16)	6.8E-55	29	4.85	(3.18– 7.41)	2.4E-13	7.8E-05
BRIP1	49	1.00	(0.69-1.45)	0.99	14	1.16	(0.65-2.07)	0.63	0.84	5	1.18	(0.47-2.95)	0.72	3	0.66	(0.21 - 2.13)	0.49	0.49
CDH1	8	1.05	(0.42-2.63)	0.93	2	1.11	(0.24-5.10)	0.89	0.99	1	1.44	(0.18-11.28)	0.73	0	0.00	(0 - Inf)	0.99	0.95
CHEK2	481	2.67	(2.30-3.11)	1.9E-37	67	1.64	(1.25-2.16)	0.00039	3.6E-05	16	1.06	(0.63-1.76)	0.83	33	2.53	(1.75 - 3.67)	9.3E-07	0.0047
EPCAM	10	0.83	(0.38-1.81)	0.64	2	0.67	(0.15-2.91)	0.59	0.68	0	0.00	(0-Inf)	0.99	0	0	(0 - Inf)	0.99	
FANCC	38	1.05	(0.69-1.60)	0.83	14	1.68	(0.93-3.04)	0.088	0.098	10	3.13	(1.58-6.18)	0.0011	2	0.73	(0.18 - 2.99)	0.66	0.046
FANCM	171	0.93	(0.76-1.13)	0.45	57	1.39	(1.04-1.86)	0.028	0.0094	23	1.64	(1.07-2.53)	0.025	11	0.90	(0.49 - 1.65)	0.73	0.057
GEN1	17	0.60	(0.34-1.07)	0.082	6	0.90	(0.38-2.15)	0.82	0.34	0	0.00	(0-Inf)	0.99	4	2.00	(0.70 - 5.70)	0.19	0.94
MEN1	1	0.28	(0.032-2.48)	0.25	0	0.00	(0-Inf)	0.99	0.91	0	0.00	(0-Inf)	1.00	0	0	(0 - Inf)	1.00	
MLH1	2	0.31	(0.067-1.48)	0.14	2	1.46	(0.31-6.87)	0.63	0.16	2	4.47	(0.93-21.53)	0.062	0	0	(0 - Inf)	0.99	0.93
MRE11	29	0.80	(0.50-1.26)	0.33	9	1.06	(0.52-2.18)	0.87	0.63	2	0.63	(0.15-2.60)	0.52	5	1.69	(0.67 - 4.30)	0.27	0.20
MSH2	7	1.08	(0.40-2.91)	0.88	4	2.54	(0.77-8.38)	0.13	0.18	2	3.37	(0.72-15.87)	0.12	1	1.60	(0.20 - 13.14)	0.66	0.63
MSH6	25	1.95	(1.09-3.51)	0.025	9	3.26	(1.47-7.21)	0.0036	0.32	3	3.36	(0.98-11.53)	0.054	2	2.41	(0.55 - 10.54)	0.24	0.78
MUTYH	145	1.02	(0.82-1.26)	0.85	52	1.09	(0.80-1.48)	0.60	0.68	18	1.17	(0.71-1.91)	0.55	22	1.03	(0.65 - 1.61)	0.91	0.72
NBN	65	1.02	(0.74-1.41)	0.89	14	0.74	(0.42-1.30)	0.29	0.40	5	0.73	(0.30-1.80)	0.50	6	0.93	(0.40 - 2.13)	0.86	0.71
NF1	15	1.25	(0.61-2.55)	0.54	7	2.46	(1.01-6.02)	0.048	0.22	2	2.02	(0.46-8.82)	0.35	2	2.10	(0.48 - 9.25)	0.33	0.80

PALB2	152	4.45	(3.23-6.14)	6.5E-20	56	6.72	(4.54-9.95)	1.6E-21	0.020	29	10.36	(6.42-16.71)	9.0E-22	20	7.35	(4.25 - 12.72)	9.8E-13	0.15
PIK3CA	2	0.22	(0.047-1.00)	0.049	0	0.00	(0-Inf)	0.98	0.91	0	0.00	(0-Inf)	0.99	0	0.00	(0 - Inf)	0.99	
PMS2	29	1.47	(0.883-2.46)	0.14	5	0.92	(0.36-2.38)	0.86	0.32	1	0.52	(0.07-3.81)	0.52	1	0.50	(0.067 - 3.69)	0.50	0.84
PTEN	9	2.42	(0.84-6.97)	0.10	0	0.00	(0-Inf)	0.99	0.88	0	0.00	(0-Inf)	1.00	0	0.00	(0 - Inf)	1.00	
RAD50	71	0.97	(0.71-1.31)	0.83	17	0.95	(0.57-1.60)	0.85	0.87	6	1.00	(0.44-2.30)	1.00	4	0.70	(0.25 - 1.91)	0.48	0.56
RAD51C	24	1.31	(0.74-2.30)	0.36	20	3.99	(2.20-7.26)	5.7E-06	0.00028	10	5.71	(2.69-12.13)	6.1E-06	4	2.17	(0.75 - 6.30)	0.16	0.098
RAD51D	26	1.52	(0.87-2.65)	0.15	13	2.92	(1.47-5.78)	0.0021	0.036	9	6.01	(2.73-13.24)	8.4E-06	2	1.38	(0.32 - 5.95)	0.67	0.050
RECQL	54	0.71	(0.51-0.99)	0.041	24	1.05	(0.67-1.64)	0.83	0.077	11	1.50	(0.80-2.81)	0.21	8	0.87	(0.42 - 1.80)	0.71	0.18
RINT1	20	0.72	(0.42-1.23)	0.23	6	0.76	(0.32-1.79)	0.53	0.87	2	0.74	(0.18-3.06)	0.67	3	0.93	(0.28 - 3.05)	0.91	0.80
STK11	3	1.56	(0.35-7.03)	0.56	0	0.00	(0-Inf)	0.99	0.89	0	0.00	(0-Inf)	1.00	0	0	(0 - Inf)	1.00	
TP53	3	1.95	(0.32-11.82)	0.47	2	5.42	(0.75-39.24)	0.094	0.22	0	0.00	(0-Inf)	1.00	1	9.67	(0.84 - 111.60)	0.069	0.95
XRCC2	9	1.03	(0.45-2.35)	0.95	5	1.72	(0.62-4.77)	0.30	0.34	1	0.96	(0.13-7.35)	0.97	3	3.05	(0.86 - 10.83)	0.084	0.30

[1]Total sample sizes after quality control. The analyses for genes other than *BRCA1* and *BRCA2* excluded PTVs in those genes and were hence slightly lower (ER-positive: 29,873 cases; ER-negative cases: 7,345; triple negative: 2,602 cases; ER-negative, non-triple-negative: 2,497 cases; 50,475 controls).

Table S9. Association analysis for PTVs in 34 genes by subtype of breast cancer, in all studies combined.

	ER-positive				ER-negative				Triple negative				ER-, not triple negative			
Gene	Case carriers	OR	95% CI	p-value	Case carriers	OR	95% CI	p-value	Case carriers	OR	95% CI	p-value	Case carriers	OR	95% CI	p-value
ABRAXAS1	10	0.85	(0.39-1.87)	0.69	2	0.65	(0.15-2.83)	0.56	1	0.80	(0.11-6.11)	0.83	0	0.00	(0-Inf)	0.99
AKT1	3	0.62	(0.15-2.52)	0.51	0	0.00	(0-Inf)	0.98	0	0.00	(0-Inf)	0.99	0	0.00	(0-Inf)	0.99
ATM	255	2.56	(2.09-3.14)	1.6E-19	30	1.08	(0.72-1.60)	0.71	7	0.76	(0.35-1.62)	0.48	16	1.57	(0.93-2.67)	0.095
BABAM2	5	0.67	(0.22-2.02)	0.47	0	0.00	(0-Inf)	0.98	0	0.00	(0-Inf)	0.99	0	0.00	(0-Inf)	0.99
BARD1	30	1.53	(0.92-2.54)	0.10	31	5.82	(3.50-9.69)	1.3E-11	12	8.15	(4.04-16.45)	4.7E-09	4	2.35	(0.79-6.96)	0.12
BRCA1	139	1.80	(1.41-2.28)	1.9E-06	350	16.36	(13.21-20.25)	5.8E-145	219	40.23	(31.31-51.70)	2.6E-183	51	4.17	(2.90-5.99)	1.5E-14
BRCA2	558	3.26	(2.81-3.79)	1.2E-54	184	3.72	(3.05-4.53)	1.3E-38	92	8.47	(6.51-11.02)	4.3E-57	43	1.40	(0.99-1.98)	0.06
BRIP1	61	1.00	(0.71-1.40)	0.99	17	1.06	(0.62-1.79)	0.84	8	1.61	(0.77-3.36)	0.20	3	0.45	(0.14-1.44)	0.18
CDH1	11	1.38	(0.59-3.21)	0.46	2	1.08	(0.24-4.95)	0.92	1	1.40	(0.18-10.96)	0.75	0	0.00	(0-Inf)	0.99
CHEK2	660	3.05	(2.66-3.50)	1.3E-56	98	1.90	(1.51-2.41)	7.3E-08	26	1.40	(0.93-2.11)	0.11	45	2.80	(2.03-3.87)	4.5E-10
EPCAM	11	0.70	(0.33-1.45)	0.33	2	0.47	(0.11-2.04)	0.32	0	0.00	(0-Inf)	0.99	0	0.00	(0-Inf)	0.99
FANCC	46	1.03	(0.70-1.51)	0.89	17	1.55	(0.90-2.68)	0.12	11	2.78	(1.45-5.32)	0.0021	2	0.53	(0.13-2.19)	0.38
FANCM	210	0.97	(0.81-1.16)	0.72	73	1.46	(1.12-1.90)	0.0050	27	1.60	(1.07-2.39)	0.023	16	1.00	(0.60-1.67)	1.00
GEN1	22	0.73	(0.43-1.23)	0.24	7	0.90	(0.40-2.03)	0.81	0	0.00	(0-Inf)	0.98	5	1.85	(0.71-4.82)	0.21
MEN1	3	0.92	(0.21-3.98)	0.91	0	0.00	(0-Inf)	0.98	0	0.00	(0-Inf)	0.99	0	0.00	(0-Inf)	0.99
MLH1	3	0.47	(0.13-1.78)	0.27	2	1.35	(0.29-6.37)	0.71	2	4.06	(0.84-19.66)	0.082	0	0.00	(0-Inf)	0.99
MRE11	35	0.81	(0.53-1.23)	0.32	11	0.98	(0.51-1.88)	0.94	2	0.52	(0.13-2.14)	0.37	7	1.55	(0.69-3.46)	0.29
MSH2	8	1.03	(0.40-2.63)	0.96	4	2.12	(0.65-6.91)	0.21	2	2.85	(0.61-13.30)	0.18	1	1.37	(0.17-11.05)	0.77
MSH6	26	1.79	(1.02-3.17)	0.044	9	2.75	(1.24-6.07)	0.013	3	2.81	(0.82-9.61)	0.10	2	1.94	(0.44-8.49)	0.38
MUTYH	178	0.98	(0.81-1.19)	0.86	76	1.16	(0.90-1.51)	0.26	22	1.28	(0.82-2.00)	0.28	42	1.14	(0.81-1.60)	0.45
NBN	76	1.06	(0.78-1.43)	0.72	18	0.86	(0.52-1.42)	0.55	7	0.92	(0.43-2.00)	0.84	8	1.09	(0.53-2.27)	0.81
NF1	19	1.53	(0.78-2.97)	0.22	8	2.55	(1.09-5.98)	0.032	2	1.82	(0.42-7.96)	0.43	2	1.87	(0.43-8.26)	0.41
PALB2	195	4.46	(3.33-5.98)	1.5E-23	80	7.16	(5.04-10.16)	3.1E-28	36	9.84	(6.35-15.25)	1.5E-24	30	7.37	(4.58-11.87)	2.1E-16

PIK3CA	4	0.44	(0.14-1.39)	0.16	1	0.49	(0.062-3.82)	0.49	0	0.00	(0-Inf)	0.99	1	1.58	(0.20-12.48)	0.66
PMS2	38	1.62	(1.01-2.58)	0.045	5	0.77	(0.30-1.98)	0.58	1	0.44	(0.06-3.24)	0.42	1	0.42	(0.06-3.10)	0.40
PTEN	12	2.89	(1.08-7.73)	0.035	2	1.94	(0.39-9.75)	0.42	0	0.00	(0-Inf)	0.99	0	0.00	(0-Inf)	0.99
RAD50	87	1.01	(0.76-1.34)	0.97	22	1.02	(0.64-1.62)	0.94	7	1.00	(0.46-2.15)	0.99	6	0.80	(0.35-1.85)	0.60
RAD51C	31	1.61	(0.95-2.74)	0.078	25	4.81	(2.75-8.44)	4.1E-08	13	7.35	(3.69-14.65)	1.5E-08	4	2.09	(0.72-6.07)	0.18
RAD51D	31	1.62	(0.97-2.73)	0.066	18	3.26	(1.78-5.99)	0.00014	9	5.25	(2.06-10.51)	0.00003	6	2.54	(1.00-6.46)	0.051
RECQL	76	0.86	(0.65-1.15)	0.30	35	1.23	(0.84-1.79)	0.29	13	1.58	(0.88-2.82)	0.12	17	1.24	(0.73-2.08)	0.43
RINT1	26	0.81	(0.50-1.31)	0.39	7	0.75	(0.34-1.67)	0.48	2	0.64	(0.16-2.66)	0.54	4	1.01	(0.36-2.86)	0.98
STK11	3	1.51	(0.33-6.81)	0.60	0	0.00	(0-Inf)	0.98	0	0.00	(0-Inf)	0.99	0	0.00	(0-Inf)	0.99
TP53	4	2.46	(0.44-13.64)	0.30	6	12.95	(2.58-65.05)	0.0019	1	6.68	(0.59-75.68)	0.13	4	18.59	(2.94-117.7)	0.0019
XRCC2	9	0.90	(0.40-2.03)	0.80	6	1.82	(0.71-4.68)	0.21	2	1.71	(0.39-7.55)	0.48	3	2.67	(0.76-9.37)	0.12

Table S10. Association analysis of protein truncating germline variants in 10 breast cancer associated genes, separately for invasive and in-situ breast cancer.

	DCIS				Invasive				
Gene	Case carriers	OR	95.CI	p.value	Case carriers	OR	95.CI	p.value	Case-only p.value
ATM	20	2.82	(1.72-4.609)	3.9E-05	268	2.06	(1.675-2.536)	8.2E-12	0.41
BARD1	1	0.56	(0.075-4.185)	0.57	58	2.19	(1.406-3.413)	0.00053	0.19
BRCA1	10	3.63	(1.774-7.427)	0.00042	486	10.82	(8.2-14.281)	1.5E-63	0.00053
BRCA2	28	3.47	(2.242-5.368)	2.3E-08	709	6.15	(5.088-7.438)	1.6E-78	0.0015
CHEK2	33	2.20	(1.51-3.211)	4.1E-05	664	2.52	(2.193-2.905)	4.3E-38	0.72
PALB2	9	2.53	(1.194-5.368)	0.015	255	5.02	(3.724-6.773)	3.9E-26	0.056
RAD51C	3	1.46	(0.426-5.001)	0.55	49	1.91	(1.176-3.113)	0.0089	0.57
RAD51D	2	1.21	(0.276-5.276)	0.80	49	1.90	(1.164-3.114)	0.010	0.34
TP53	2	13.65	(1.653-112.81)	0.015	5	2.40	(0.456-12.586)	0.3	0.026

Table S11. Associations of protein truncating germline variants in 34 genes and age at diagnosis in years, in population-based studies. OR is the interaction OR per year, derived from a case-only analysis. The baseline log(OR) is the estimated effect size at age 0 in the model used to generate the cumulative risks (see Supplementary Methods).

					All samples		European only	
Gene	Case carriers	OR	95.CI	p.value	Baseline log(OR)	s.e. (logOR)	Baseline log(OR)	s.e (logOR)
ABRAXAS1	16	0.954	(0.91-1.00)	0.06				
AKT1	3	1.003	(0.92-1.12)	0.96				
ATM	294	0.99	(0.98-1.00)	0.094				
BABAM2	7	1.007	(0.93-1.09)	0.87				
BARD1	61	0.978	(0.95-1.00)	0.073				
BRCA1	508	0.941	(0.94-0.95)	3.5E-65	5.260	0.144	5.057	0.15
BRCA2	744	0.968	(0.96-0.97)	4.2E-29	3.510	0.098	3.434	0.105
BRIP1	85	0.991	(0.97-1.01)	0.39				
CDH1	11	0.975	(0.92-1.03)	0.40				
CHEK2	701	0.986	(0.98-0.99)	1.9E-04	1.706	0.073	1.774	0.074
EPCAM	14	1.048	(1.00-1.10)	0.072				
FANCC	71	1.001	(0.98-1.03)	0.90				
FANCM	300	0.995	(0.98-1.01)	0.34				
GEN1	31	0.993	(0.96-1.03)	0.69				
MEN1	2	1.043	(0.91-1.19)	0.53				
MLH1	4	0.964	(0.88-1.06)	0.43				
MRE11	47	1.011	(0.98-1.04)	0.44				
MSH2	13	1.03	(0.97-1.09)	0.31				
MSH6	39	1.011	(0.98-1.04)	0.50				
MUTYH	225	0.993	(0.98-1.01)	0.31				
NBN	90	0.995	(0.98-1.02)	0.65				
NF1	31	0.993	(0.96-1.03)	0.68				
PALB2	271	0.984	(0.97-1.00)	0.0054	2.488	0.152	2.596	0.163
PIK3CA	3	1.005	(0.91-1.11)	0.92				
PMS2	39	1.007	(0.98-1.04)	0.67				
PTEN	14	0.915	(0.87-0.96)	5.4E-04				
RAD50	120	0.994	(0.98-1.01)	0.49				
RAD51C	54	1.021	(0.99-1.05)	0.14				
RAD51D	50	0.988	(0.96-1.02)	0.40				
RECQL	100	0.996	(0.98-1.02)	0.68				
RINT1	32	1.004	(0.97-1.04)	0.83				
STK11	4	0.887	(0.80-0.98)	0.021				
TP53	7	0.78	(0.70-0.87)	3.3E-06				
XRCC2	15	0.951	(0.91-1.00)	0.041				

Table S12. Association analysis of protein truncating germline variants in 9 breast cancer associated genes, by age at diagnosis.

Gene	age <40 years				age 40-49 years				age 50-59 years				age 60+ years			
	Cases	Controls	OR (95%CI)*	OR (95%CI)⁺	Cases	Controls	OR (95%CI)*	OR (95%CI)⁺	Cases	Controls	OR (95%CI)*	OR (95%CI)⁺	Cases	Controls	OR (95%CI)*	OR (95%CI)⁺
ATM	21	17	1.77 (0.87-3.59)	2.27 (1.40-3.68)	82	35	2.11 (1.39-3.21)	2.63 (2.00-3.51)	93	41	2.24 (1.53-3.28)	2.18 (1.65-2.87)	98	44	2.33 (1.61-3.38)	1.94 (1.48-2.53)
BARD1	6	3	4.30 (1.05-17.7)	3.44 (1.36-8.72)	17	8	1.91 (0.91-4.48)	2.68 (1.42-5.04)	22	8	2.73 (1.20-6.22)	2.65 (1.49-4.72)	16	8	1.84 (0.76-4.48)	1.58 (0.84-2.97)
BRCA1	175	10	32.8 (16.9-63.4)	46.3 (33.4-64.1)	176	14	12.4 (7.16-21.5)	14.2 (10.4-19.5)	109	20	5.63 (3.43-9.25)	6.52 (4.66-9.14)	48	12	3.98 (2.08-7.59)	2.33 (1.57-3.48)
BRCA2	156	20	11.9 (7.33-19.4)	18.7 (14.4-24.1)	229	28	7.94 (5.27-12.0)	7.85 (6.23-9.88)	214	39	5.39 (3.80-7.65)	5.14 (4.09-6.47)	145	42	3.05 (2.14-4.35)	2.81 (2.20-3.60)
CHEK2	77	28	4.54 (2.87-7.17)	4.23 (3.23-5.54)	171	64	2.25 (1.66-3.05)	2.63 (2.15-3.21)	228	94	2.41 (1.88-3.11)	2.64 (2.19-3.17)	225	96	2.22 (1.72-2.86)	2.06 (1.72-2.48)
PALB2	26	8	5.36 (2.26-12.7)	6.16 (3.69-10.3)	75	10	6.68 (3.38-13.2)	5.54 (3.81-8.05)	92	16	6.42 (3.55-11.60)	5.63 (3.91-8.10)	78	20	3.58 (2.11-6.06)	4.12 (2.84-5.97)
RAD51C	4	1	4.83 (0.52-45.2)	1.89 (0.65-5.51)	8	7	1.02 (0.36-2.85)	1.04 (0.47-2.34)	22	7	2.97 (1.25-7.05)	2.33 (1.30-4.17)	20	11	1.50 (0.70-3.23)	1.99 (1.10-3.60)
RAD51D	4	3	1.76 (0.38-8.17)	1.73 (0.72-6.29)	11	5	1.91 (0.64-5.71)	1.69 (0.81-3.51)	22	11	1.71 (0.82-3.60)	2.51 (1.39-4.54)	13	16	1.96 (0.73-5.31)	1.45 (0.74-2.88)

***** OR based on cases and controls in that age-group.

⁺ OR based on cases in that age-group vs. all controls.

Table S13. Association of missense variants with overall breast cancer risk, separately for variant within and outside domain, for eight genes with a statistically significant association between PTVs and breast cancer risk overall. Results are shown in in all studies and in population-based studies only.

(a) All samples.

				Domain vs. outside domain			Domain vs. non-carriers			Outside domain vs. non-carriers			LRT* among domains
Gene	Case carriers in domain	Case carriers outside domain	Case non-carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value	p-value
ATM	1040	2064	55064	1.17	(1.04-1.31)	0.0079	1.22	(1.11-1.34)	4.9E-05	1.04	(0.98-1.11)	0.21	0.022
BARD1	306	450	57873	1.06	(0.85-1.32)	6.00E-01	1.08	(0.91-1.29)	0.35	1.02	(0.89-1.17)	0.75	0.19
BRCA1	278	1395	56952	1.56	(1.26-1.92)	4.10E-05	1.59	(1.30-1.93)	4.0E-06	1.02	(0.94-1.11)	0.62	2.2E-04
BRCA2	965	2580	55055	1.04	(0.93-1.16)	0.45	1.02	(0.92-1.12)	0.73	0.97	(0.92-1.03)	0.39	0.31
CHEK2	852	354	56436	1.14	(0.94-1.40)	0.19	1.58	(1.41-1.77)	1.7E-15	1.39	(1.17-1.64)	1.3E-04	0.24
PALB2	805	247	57278	0.95	(0.78-1.17)	0.65	0.99	(0.89-1.10)	0.83	1.04	(0.87-1.24)	0.69	0.24
RAD51C	48	193	58414	1.15	(0.71-1.86)	0.58	1.06	(0.69-1.64)	0.79	0.93	(0.75-1.14)	0.48	NA
RAD51D	17	259	58382	0.71	(0.33-1.51)	0.37	0.76	(0.36-1.59)	0.46	1.07	(0.89-1.29)	0.45	NA

* Likelihood ratio test for difference in OR by domain.

(b) Population-based samples.

Gene	Case carriers in domain	Case carriers outside domain	Case non-carriers	Domain vs. outside domain			Domain vs. non-carriers			Outside domain vs. non-carriers			LRT* among domains
				OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value	p-value
ATM	803	1608	44705	1.15	(1.02-1.30)	0.028	1.17	(1.05-1.30)	0.0034	1.01	(0.94-1.09)	0.69	0.022
BARD1	236	355	46853	1.01	(0.79-1.28)	0.95	1	(0.83-1.21)	0.99	0.99	(0.85-1.15)	0.92	0.41
BRCA1	217	1176	46047	1.58	(1.25-1.99)	1.30E-04	1.65	(1.32-2.05)	7.7E-06	1.04	(0.96-1.14)	0.32	3.0E-06
BRCA2	792	2039	44596	1.06	(0.94-1.19)	0.35	1.02	(0.92-1.13)	0.68	0.97	(0.91-1.03)	0.29	0.27
CHEK2	618	277	45916	1.07	(0.86-1.34)	0.52	1.46	(1.28-1.65)	4.2E-09	1.35	(1.13-1.62)	0.001	0.40
PALB2	613	192	46424	0.86	(0.68-1.09)	0.21	0.93	(0.83-1.04)	0.18	1.07	(0.87-1.31)	0.50	0.48
RAD51C	40	156	47270	1.10	(0.66-1.84)	0.72	1.01	(0.63-1.60)	0.98	0.91	(0.73-1.14)	0.44	NA
RAD51D	17	207	47247	0.71	(0.33-1.53)	0.38	0.76	(0.36-1.60)	0.47	1.07	(0.88-1.31)	0.50	NA

* Likelihood ratio test for difference in OR by domain.

Table S14. Association of missense variants with overall and subtype-specific breast cancer risk, by domain, for *BRCA1.*

(a) All samples.

					All breast cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
RING	24-65	13	78	33	2.27	(1.50-3.42)	9.7E-05	1.94	(1.21-3.10)	0.0059	3.79	(2.22-6.46)	9.6E-07
Interaction with PALB2	1397-1424	17	11	25	0.38	(0.18-0.78)	0.0088	0.34	(0.14-0.83)	0.019	0.26	(0.04-1.95)	0.19
BRCT 1	1642-1736	48	139	74	1.80	(1.35-2.40)	6.3E-05	1.42	(1.01-2.00)	0.042	3.34	(2.22-5.02)	7.7E-09
BRCT 2	1756-1855	26	50	35	1.37	(0.88-2.14)	0.16	1.08	(0.62-1.87)	0.80	3.02	(1.69-5.40)	1.8E-04
No domain	NA	522	1395	1213	1.02	(0.94-1.10)	0.62	1.04	(0.95-1.14)	0.37	1.12	(0.97-1.30)	0.11

(b) Population-based samples

					All breast cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
RING	24-65	13	62	19	3.68	(2.18-6.21)	1.0E-06	2.44	(1.32-4.50)	0.0044	8.03	(4.33-14.90)	3.7E-11
Interaction with PALB2	1397-1424	16	10	24	0.38	(0.18-0.80)	0.011	0.36	(0.14-0.89)	0.027	0.28	(0.04-2.05)	0.21
BRCT 1	1642-1736	41	112	67	1.80	(1.32-2.46)	2.0E-04	1.52	(1.06-2.18)	0.022	3.28	(2.08-5.15)	2.9E-07
BRCT 2	1756-1855	23	33	30	1.11	(0.67-1.85)	0.69	1.06	(0.58-1.94)	0.84	2.20	(1.07-4.53)	0.032
No domain	NA	488	1176	1160	1.04	(0.96-1.14)	0.32	1.07	(0.97-1.17)	0.19	1.14	(0.97-1.33)	0.1

Table S15. Association of missense variants with overall and subtype-specific breast cancer risk, by domain, for *CHEK2.*

(a) All samples.

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
FHA	113-192	42	282	140	1.82	(1.48-2.24)	1.2E-08	1.95	(1.55-2.45)	1.00E-08	0.86	(0.53-1.38)	0.53
PKinase	220-486	142	570	362	1.49	(1.30-1.70)	7.8E-09	1.54	(1.33-1.79)	2.00E-08	0.90	(0.68-1.19)	0.45
No domain	NA	90	354	246	1.38	(1.17-1.64)	1.3E-04	1.44	(1.19-1.74)	1.60E-04	1.18	(0.85-1.62)	0.32

(b) Population-based samples.

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
FHA	113-192	38	190	135	1.60	(1.27-2.01)	7.0E-05	1.70	(1.32-2.20)	4.5E-05	1.00	(0.61-1.64)	0.99
PKinase	220-486	125	428	328	1.40	(1.21-1.62)	9.0E-06	1.50	(1.27-1.76)	1.5E-06	0.90	(0.65-1.24)	0.51
No domain	NA	83	277	234	1.35	(1.13-1.62)	0.001	1.40	1.14-1.71)	0.0012	1.16	(0.82-1.65)	0.41

Table S16. Association of missense variants with overall and subtype-specific breast cancer risk, by domain, for *ATM*.

(a) All samples.

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
TAN	8-165	38	95	87	0.94	(0.70-1.26)	0.68	0.97	(0.69-1.35)	0.84	0.93	(0.55-1.57)	0.78
Interaction with ABL1	1373-1382	3	2	3	0.47	(0.08-2.89)	0.42	0.46	(0.05-4.60)	0.51	0	(0-Inf)	0.99
FAT	1960-2566	204	690	512	1.25	(1.11-1.41)	1.9E-04	1.29	(1.13-1.47)	2.2E-04	1.17	(0.94-1.46)	0.16
PI3K/PI4K	2712-2962	79	216	136	1.45	(1.16-1.80)	9.2E-04	1.62	(1.27-2.06)	8.9E-05	1.24	(0.82-1.87)	0.30
FATC	3024-3056	11	37	45	0.77	(0.49-1.21)	0.26	0.81	(0.48-1.34)	0.41	0.61	(0.22-1.71)	0.34
None	NA	766	2064	1808	1.04	(0.98-1.11)	0.21	1.04	(0.96-1.12)	0.32	1.08	(0.96-1.22)	0.20

(b) Population-based samples.

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
TAN	8-165	35	73	81	0.89	(0.64-1.23)	0.46	0.92	(0.63-1.32)	0.64	0.81	(0.44-1.50)	0.50
Interaction with ABL1	1373-1382	3	2	3	0.51	(0.08-3.11)	0.46	0.49	(0.05-4.86)	0.54	0	(0-Inf)	0.99
FAT	1960-2566	184	545	497	1.2	(1.06-1.36)	0.0043	1.21	(1.05-1.40)	0.0077	1.19	(0.94-1.51)	0.14
PI3K/PI4K	2712-2962	71	155	125	1.41	(1.10-1.80)	0.0061	1.60	(1.23-2.10)	5.6E-04	1.30	(0.82-2.07)	0.27
FAT-C	3024-3056	10	28	45	0.71	(0.43-1.16)	0.17	0.80	(0.47-1.37)	0.41	0.68	(0.24-1.93)	0.47
None	NA	696	1608	1720	1.01	(0.94-1.09)	0.69	1.02	(0.94-1.11)	0.63	1.08	(0.94-1.23)	0.27

Table S17. Association of missense variants with overall and subtype-specific breast cancer risk, by domain, for *BRCA2.*

(a) All samples.

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
PALB2	10-40	13	29	23	1.35	(0.77-2.35)	0.29	1.34	(0.71-2.53)	0.36	1.02	(0.38-2.74)	0.96
DNA binding	2481-3186	258	936	872	1.01	(0.91-1.11)	0.87	0.97	(0.87-1.09)	0.62	1.10	(0.93-1.31)	0.28
No domain	NA	1072	2580	2359	0.97	(0.92-1.03)	0.39	0.98	(0.91-1.05)	0.51	0.98	(0.88-1.09)	0.69

(b) Population-based samples.

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
PALB2	10-40	11	19	15	1.49	(0.75-2.96)	0.26	1.37	(0.61-3.06)	0.44	0.85	(0.19-3.80)	0.84
DNA binding	2481-3186	242	773	831	1.01	(0.91-1.12)	0.80	0.96	(0.85-1.08)	0.51	1.18	(0.99-1.42)	0.071
No domain	NA	966	2039	2192	0.97	(0.91-1.03)	0.29	0.96	(0.89-1.03)	0.27	1.01	(0.90-1.14)	0.81

Table S18. Association of missense variants with overall and subtype-specific breast cancer risk, by domain, for *PALB2.*

(a) All samples.

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
DNA Binding	1-579	212	443	422	0.93	(0.81-1.07)	0.29	0.97	(0.82-1.13)	0.66	0.83	(0.64-1.09)	0.18
WD 1-7	853-1186	136	362	320	1.07	(0.92-1.25)	0.40	1.12	(0.94-1.33)	0.21	1.04	(0.79-1.37)	0.78
No domain	NA	92	247	244	1.04	(0.87-1.24)	0.69	1.04	(0.84-1.28)	0.72	0.99	(0.72-1.37)	0.95

(b) Population-based samples

					All cancer			ER positive			ER negative		
Domain	Amino acids	Unique protein positions	Case carriers	Control carriers	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
DNA Binding	1-579	199	341	401	0.89	(0.77-1.04)	0.14	0.92	(0.77-1.09)	0.34	0.84	(0.63-1.13)	0.25
WD 1-7	853-1186	130	272	291	0.97	(0.82-1.15)	0.71	1.02	(0.84-1.24)	0.84	0.85	(0.62-1.18)	0.34
No domain	NA	88	192	200	1.07	(0.87-1.31)	0.50	1.07	(0.85-1.36)	0.55	1.10	(0.76-1.60)	0.61

Table S19. Association of missense variants with overall and subtype-specific breast cancer, by pathogenicity, for *BRCA1*, *BRCA2* and *TP53.*

Subtype	Gene	All samples								Population samples							
		Pathogenic variant carriers		Pathogenic vs non-carriers			Non-pathogenic vs non-carriers			Pathogenic variant carriers		Pathogenic vs non-carriers			Non-pathogenic vs non-carriers		
		Cases	Controls	OR	95% CI	p-value	OR	95% CI	p-value	Cases	Controls	OR	95% CI	p-value	OR	95% CI	p-value
Overall	*BRCA1*	65	7	9.21	(4.19-20.25)	3.3E-08	1.05	(0.97-1.13)	0.22	61	4	16.11	(5.83-44.50)	8.4E-08	1.06	(0.98-1.15)	0.14
	BRCA2	69	10	6.26	(3.18-12.29)	1.0E-07	0.97	(0.92-1.02)	0.24	*43*	*8*	5.68	(2.62-12.29)	1.0E-05	0.97	(0.92-1.02)	0.26
	TP53	104	20	4.64	(2.86-7.52)	5.1E-10	1.05	(0.88-1.24)	0.59	52	19	2.91	(1.71-4.98)	9.0E-05	0.94	(0.77-1.14)	0.54
ER-positive	*BRCA1*	18	7	4.74	(1.92-11.73)	0.00074	1.05	(0.97-1.15)	0.23	18	4	8.03	(2.66-24.19)	2.2E-04	1.07	(0.98-1.18)	0.13
	BRCA2	40	10	6.32	(3.10-12.86)	3.8E-07	0.96	(0.91-1.02)	0.20	24	8	5.23	(2.28-12.02)	9.6E-05	0.95	(0.89-1.01)	0.12
	TP53	56	20	4.61	(2.72-7.83)	1.5E-08	1.14	(0.94-1.38)	0.18	33	19	3.21	(1.78-5.77)	1.0E-04	1.01	(0.81-1.26)	0.92
ER-negative	*BRCA1*	37	7	39.53	(16.79-93.07)	3.9E-17	1.19	(1.04-1.35)	0.012	34	4	53.72	(18.85-153.15)	9.1E-14	1.18	(1.02-1.37)	0.024
	BRCA2	9	10	5.65	(2.18-14.66)	3.7E-04	1.00	(0.91-1.09)	0.94	7	8	5.09	(1.81-14.35)	0.39	1.05	(0.94-1.16)	0.39
	TP53	22	20	6.78	(3.59-12.80)	3.5E-09	1.00	(0.73-1.36)	0.997	11	19	4.21	(1.95-9.10)	2.6E-04	0.81	(0.55-1.21)	0.30

Table S20. Comparison of results of association results for PTVs with the classification of Lee et al (2019). "Associated" is defined as a Bayesian False Discovery Probability of <5%. "Not moderate risk" is defined an upper 95% confidence limit on the OR for PTVs <2. "Uncertain" is defined as being in neither of these categories.

Gene	Lee et al (2019) classification	This paper
ABRAXAS1	N/D	Not moderate or high risk
AKT1	N/D	Uncertain
ATM	DEFINITIVE	Associated
BABAM2	N/D	Not moderate or high risk
BARD1	DEFINITIVE	Associated
BRCA1	DEFINITIVE	Associated
BRCA2	DEFINITIVE	Associated
BRIP1	REFUTED	Not moderate or high risk
CDH1	DEFINITIVE	Not moderate or high risk
CHEK2	DEFINITIVE	Associated
EPCAM	NO REPORTED EVIDENCE	Not moderate or high risk
FANCC	N/D	Not moderate or high risk
FANCM	N/D	Not moderate or high risk
GEN1	DISPUTED	Not moderate or high risk
MEN1	N/D	Not moderate or high risk
MLH1	DISPUTED	Not moderate or high risk
MRE11	DISPUTED	Not moderate or high risk
MSH2	DISPUTED	Uncertain
MSH6	DISPUTED	Uncertain
MUTYH	NO REPORTED EVIDENCE	Not moderate or high risk
NBN	LIMITED	Not moderate or high risk
NF1	NOT CURATED/NO REPORTED EVIDENCE	Uncertain
PALB2	DEFINITIVE	Associated
PIK3CA	NO REPORTED EVIDENCE	Not moderate or high risk
PMS2	DISPUTED	Not moderate or high risk
PTEN	DEFINITIVE	Uncertain
RAD50	LIMITED	Not moderate or high risk
RAD51C	DISPUTED	Associated
RAD51D	LIMITED	Associated
RECQL	MODERATE	Not moderate or high risk
RINT1	DISPUTED	Not moderate or high risk
STK11	DEFINITIVE	Uncertain
TP53	DEFINITIVE	Associated
XRCC2	LIMITED	Not moderate or high risk

Table S21. Association analysis for PTVs in 34 genes and overall breast cancer risk, for family-based studies (9,408 cases, 43,451 controls).

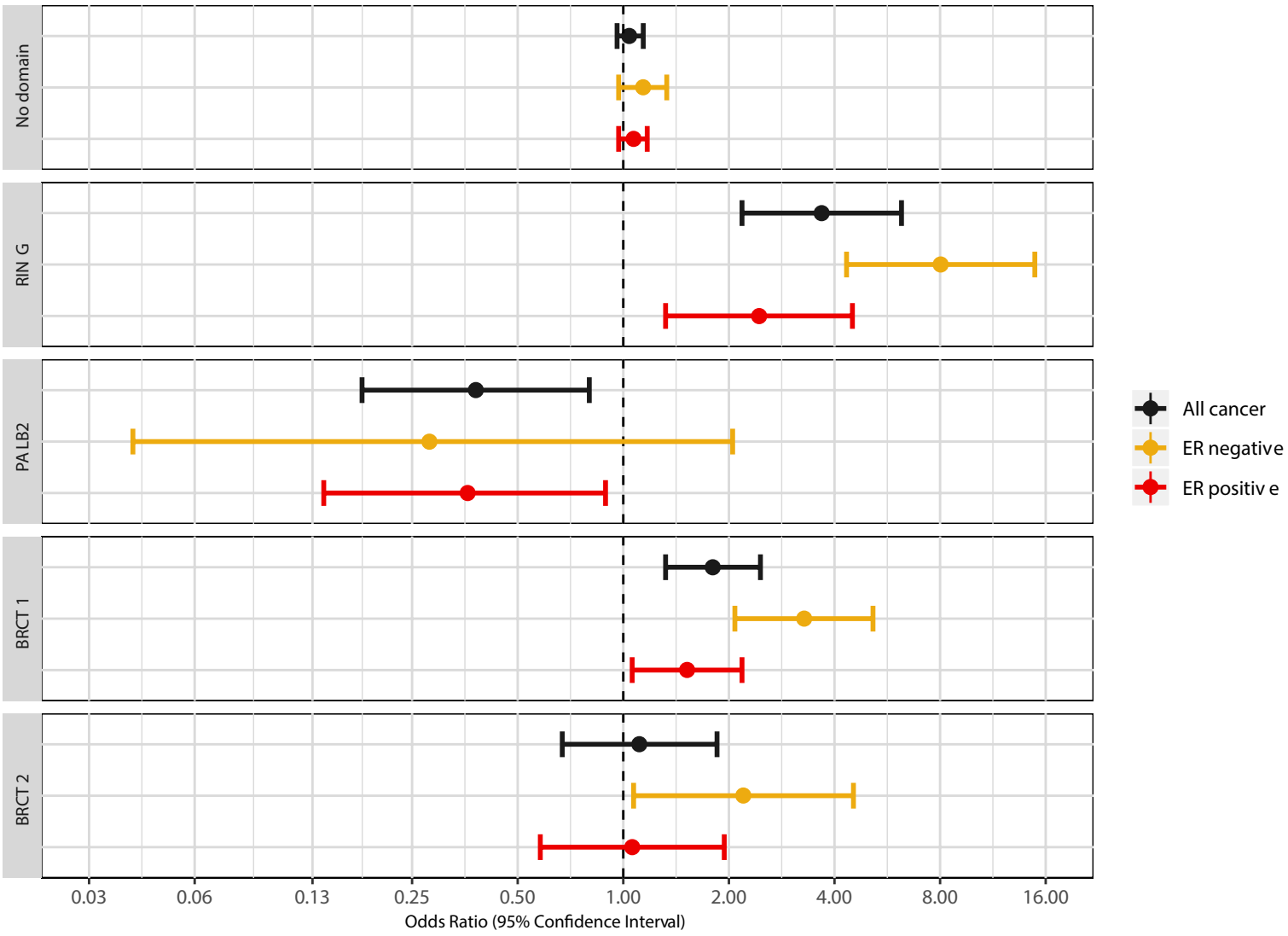
	Variant Carriers			
Gene	Cases	Controls	OR	95% CI
ABRAXAS1	4	1820	0.94	(0.27-3.31)
AKT1	0	3	0	(0-Inf)
ATM	117	130	3.38	(2.44-4.70)
BABAM2	1	6	0.78	(0.07-9.34)
BARD1	17	30	2.93	(1.33-6.44)
BRCA1	26	479	2.77	(1.49-5.14)
BRCA2	56	113	2.75	(1.80-4.20)
BRIP1	20	58	1.41	(0.75-2.64)
CDH1	6	10	6.99	(1.70-28.74)
CHEK2	361	277	5.19	(4.17-6.45)
c.1100delC	322	214	5.21	(4.13-6.59)
Other	39	63	4.77	(2.257-8.83)
EPCAM	2	168	0.6	(0.11-3.14)
FANCC	22	534	1.16	(0.64-2.10)
FANCM	79	261	1.38	(1.02-1.87)
GEN1	1	31	0.4	(0.05-3.52)
MEN1	2	45	8.59	(0.54-138.11)
MLH1	2	8	1.34	(0.21-8.46)
MRE11	8	46	0.59	(0.26-1.34)
MSH2	4	10	1.5	(0.40-5.65)
MSH6	4	224	0.79	(0.23-2.69)
MUTYH	13	228	1.33	(0.61-2.88)
NBN	22	8891	1.34	(0.77-2.35)
NF1	10	13	2.35	(0.87-6.38)
PALB2	87	48	8.11	(4.94-13.30)
PIK3CA	5	67	4.75	(1.21-18.63)
PMS2	123	345	1.66	(0.75-3.65)
PTEN	10	5	11.98	(2.56-55.97)
RAD50	301	102	1.52	(0.93-2.47)
RAD51C	15	19	5.76	(2.00-16.61)
RAD51D	14	23	3.78	(1.42-10.07)
RECQL	12	110	0.91	(0.44-1.88)
RINT1	11	48	1.22	(0.55-2.71)
STK11	0	4	0	(0-Inf)
TP53	3	2	4.93	(0.65-37.46)
XRCC2	3	17	1.16	(0.25-5.46)

Table S22. Association analysis for rare missense variants in 34 genes and overall breast cancer risk, for family-based studies (9,408 cases, 43,451 controls).

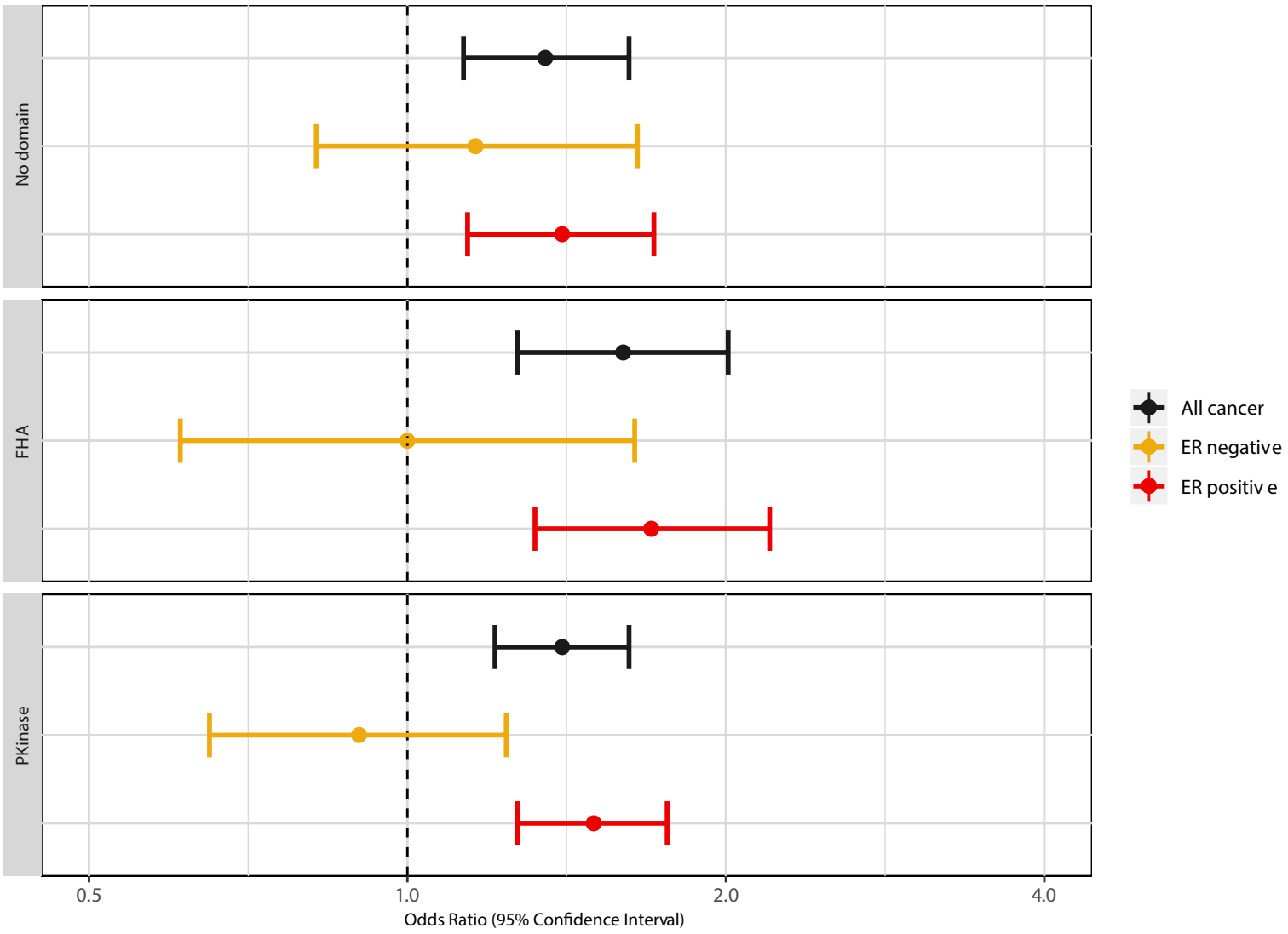
	Variant Carriers				
Gene	Cases	Controls	OR	95% CI	p-value
ABRAXAS1	62	213	1.44	(1.00-2.09)	0.051
AKT1	41	129	1.19	(0.74-1.91)	0.47
ATM	691	2139	1.32	(1.17-1.48)	2.7E-06
BABAM2	45	147	1.26	(0.82-1.94)	0.28
BARD1	162	525	1.33	(1.04-1.69)	0.022
BRCA1	276	1099	0.91	(0.76-1.08)	0.26
BRCA2	704	2616	1.03	(0.92-1.15)	0.64
BRIP1	248	825	1.02	(0.84-1.25)	0.81
CDH1	222	594	1.1	(0.86-1.39)	0.45
CHEK2	307	620	2.17	(1.79-2.63)	2.6E-15
EPCAM	66	302	0.75	(0.53-1.07)	0.11
FANCC	148	521	1.21	(0.96-1.52)	0.11
FANCM	400	1381	1.22	(1.05-1.42)	0.01
GEN1	172	607	1.2	(0.96-1.50)	0.12
MEN1	33	100	2.11	(1.23-3.62)	0.0064
MLH1	147	636	1.04	(0.83-1.31)	0.71
MRE11	150	527	1.4	(1.10-1.77)	0.0064
MSH2	201	880	0.92	(0.75-1.12)	0.38
MSH6	282	987	1.08	(0.90-1.28)	0.42
MUTYH	145	601	0.86	(0.67-1.09)	0.21
NBN	178	608	1.16	(0.92-1.46)	0.21
NF1	191	789	1.12	(0.91-1.37)	0.29
PALB2	237	806	1.12	(0.90-1.40)	0.3
PIK3CA	53	158	1.2	(0.83-1.75)	0.33
PMS2	178	789	1.05	(0.85-1.28)	0.66
PTEN	24	55	1.84	(1.00-3.36)	0.049
RAD50	307	941	1.21	(1.02-1.44)	0.031
RAD51C	45	182	1.01	(0.68-1.49)	0.98
RAD51D	52	173	1.17	(0.74-1.86)	0.5
RECQL	164	562	1.13	(0.89-1.43)	0.33
RINT1	206	660	1.39	(1.14-1.69)	0.0013
STK11	20	113	1.09	(0.57-2.08)	0.8
TP53	141	219	2.09	(1.49-2.93)	2.0E-05
XRCC2	63	182	1.22	(0.85-1.74)	0.28

Figure S1. Odds ratios with 95% confidence intervals for germline missense variants by domain for (a) *BRCA1* (b) *CHEK2* (c) *ATM* (d) *BRCA2* (e) *PALB2* and (f) *BARD1* in population-based studies.

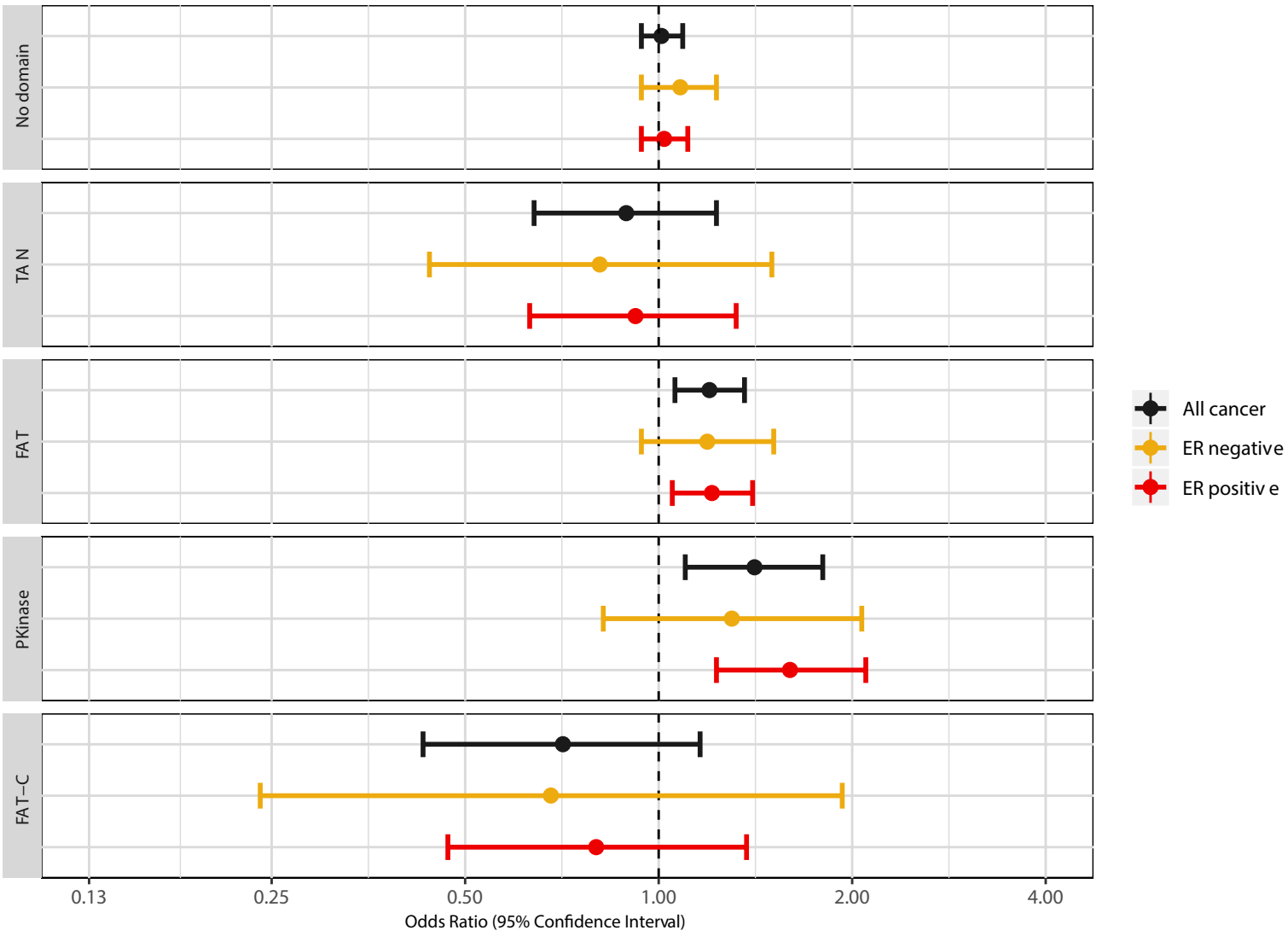
(a) *BRCA1*


No domain
RIN G
PA LB2
BRCT 1
BRCT 2
0.03
0.06
0.13
0.25
0.50
1.00
2.00
4.00
8.00
16.00
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positiv e

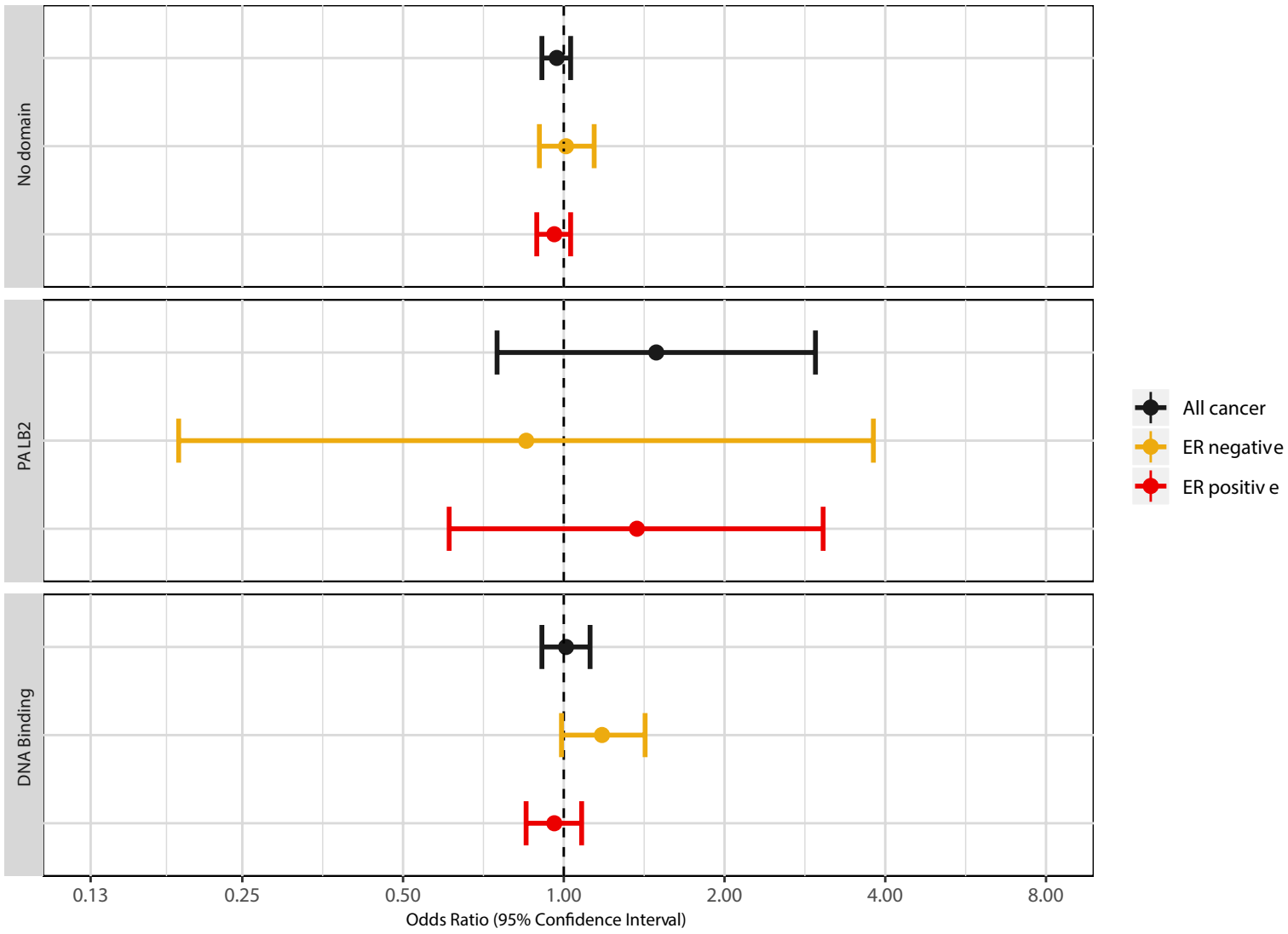
(b) *CHEK2*


No domain
FHA
PKinase
0.5
1.0
2.0
4.0
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positive

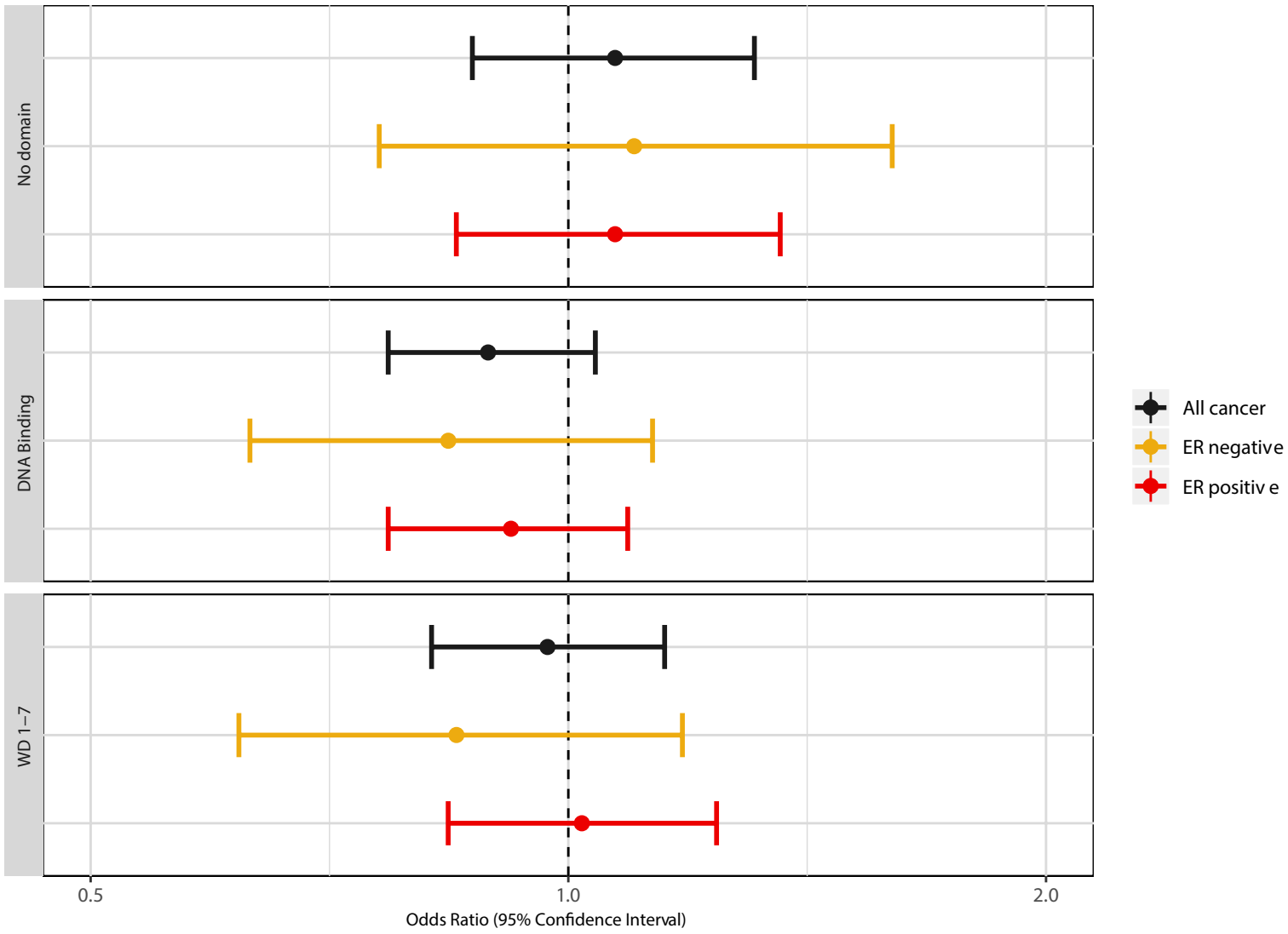
(c) *ATM*


No domain
TA N
FA T
PKinase
FA T–C
0.13
0.25
0.50
1.00
2.00
4.00
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positiv e

(d) *BRCA2*


No domain
PA LB2
DNA Binding
0.13
0.25
0.50
1.00
2.00
4.00
8.00
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positiv e

(e) *PALB2*


No domain
DNA Binding
WD 1–7
All cancer
ER negative
ER positiv e
0.5
1.0
2.0
Odds Ratio (95% Confidence Interval)

(f) *BARD1*

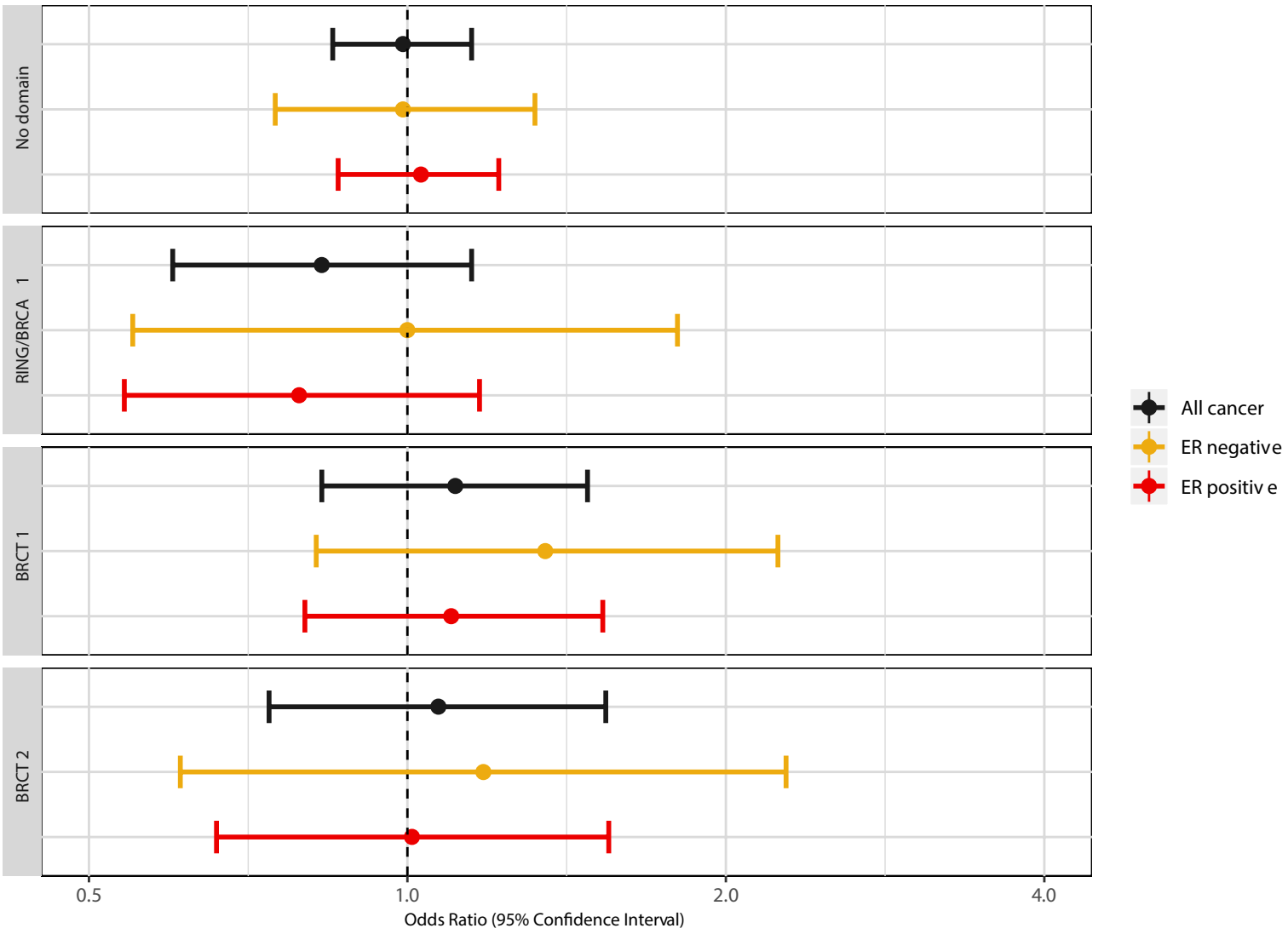
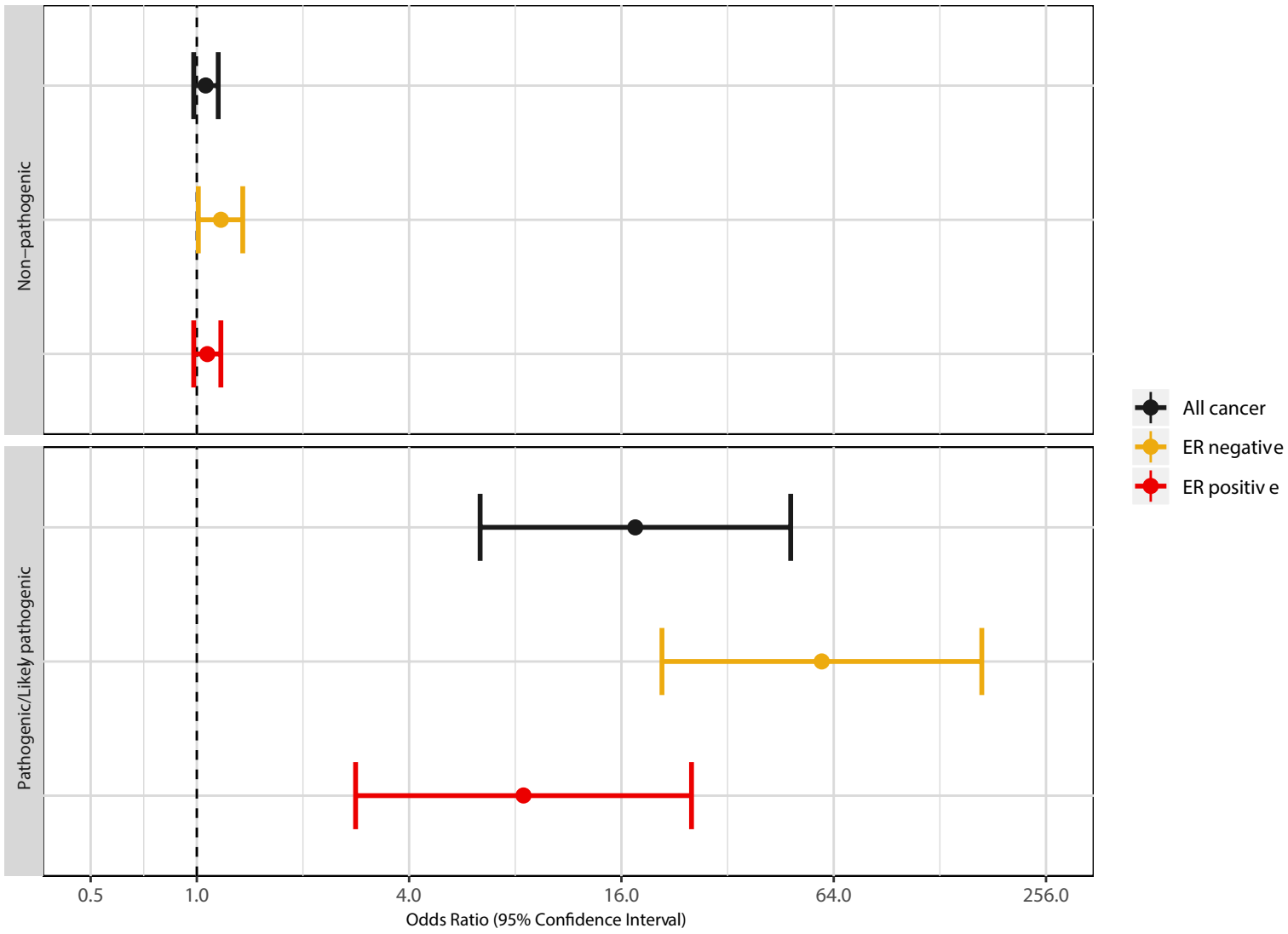
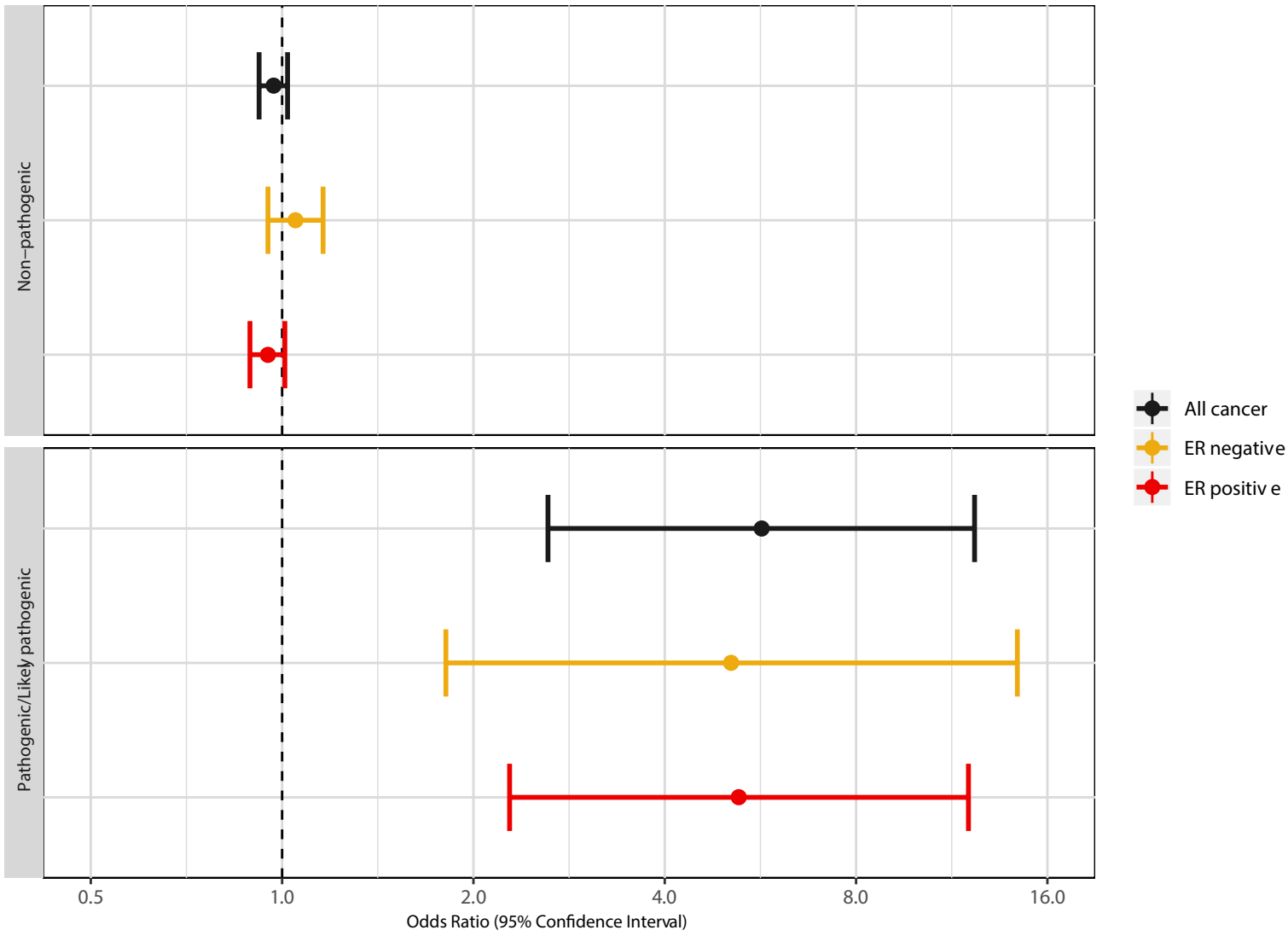

No domain
RING/BRCA 1
BRCT 1
BRCT 2
0.5
1.0
2.0
4.0
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positive

Figure S2. Odds ratios with 95% confidence intervals for missense germline variants by pathogenicity for (a) *BRCA1* (b) *BRCA2* and (c) *TP53* in population-based studies.

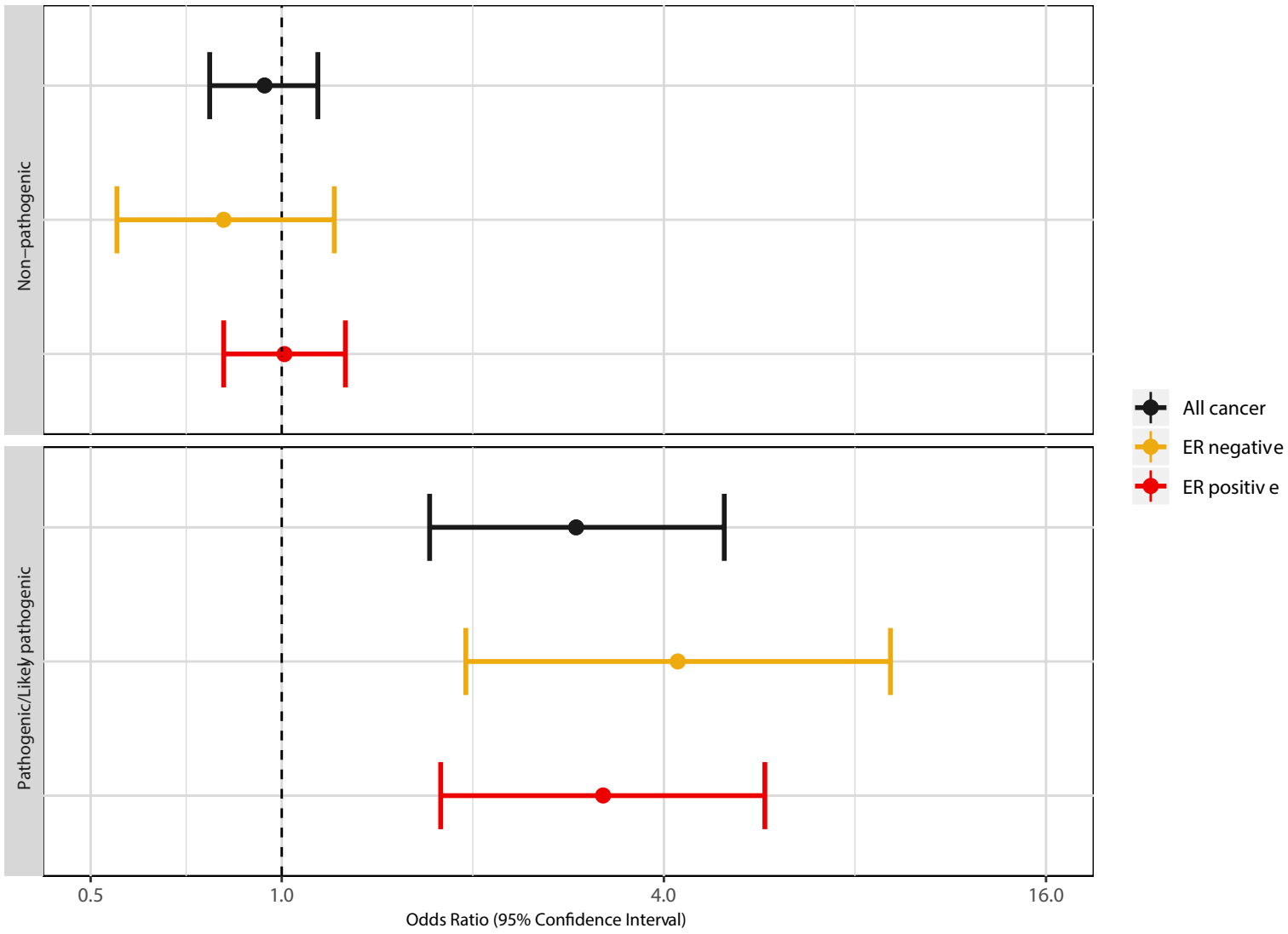
(a) *BRCA1*


Non−pathogenic
Pathogenic/Likely pathogenic
0.5
1.0
4.0
16.0
64.0
256.0
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positiv e

(b) *BRCA2*


Non−pathogenic
Pathogenic/Likely pathogenic
0.5
1.0
2.0
4.0
8.0
16.0
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positiv e

(c) *TP53*


Non−pathogenic
Pathogenic/Likely pathogenic
0.5
1.0
4.0
16.0
Odds Ratio (95% Confidence Interval)
All cancer
ER negative
ER positiv e

Supplementary Files Descriptions.

Four files give summary counts for the numbers of cases and controls carrying PTVs or rare missense variants, in each of the 34 genes, based on the dataset used in the analysis after quality control. Counts are presented separately for all cases and controls, and the subset of cases and controls in population-based studies only. Variants are described as chr<chr>_<pos>_<ref>_<alt>, where <chr> is chromosome, <pos> is the build37 (hg19) position, <ref> is the reference sequence and <alt> is the variant sequence.

Supplementary File 5 gives the primer designs for the Bridges panel.

Funding

BCAC was funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS), the PERSPECTIVE programme: The Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344), the *Ministère de l'Économie, de la Science et de l'Innovation du Québec* through Genome Québec, and the Quebec Breast Cancer Foundation and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the *Ministère de l'Économie et de l'Innovation du Québec* through Genome Québec, the Quebec Breast Cancer Foundation, and the Ontario Research Fund. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. The sequencing and analysis for this project was funded by the European Union's Horizon 2020 Research and Innovation Programme (BRIDGES: grant number 634935) and the Wellcome Trust [grant no: v203477/Z/16/Z].

The **ABCS** and **ABCS-F** studies were supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]. The **ACP** study is funded by the Breast Cancer Research Trust, UK. KM and AL are supported by the NIHR Manchester Biomedical Research Centre, by the Allan Turing Institute and by the ICEP (Cancer Research UK C18281/A19169). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. For **BIGGS**, ES is supported by NIHR Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, United Kingdom. IT is supported by the Oxford Biomedical Research Centre. The BREast Oncology GAlician Network (**BREOGAN**) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192, Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. The **BSUCH** study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). **CCGP** is supported by funding from the University of Crete. The **CECILE** study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The **CGPS** was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The **CNIO-BCS** was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). **COLBCCC** is supported by the German Cancer Research Center (DKFZ), Heidelberg, Germany. Diana Torres was in part supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation. The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. **FHRISK** is funded from NIHR grant PGfAR 0707-10031. DGE, AH and EvanV are supported by the NIHR Manchester Biomedical Research Centre. The **GC-HBOC** (German Consortium of Hereditary Breast and Ovarian Cancer) is supported by the German Cancer Aid (grant no 110837,

coordinator: Rita K. Schmutzler, Cologne). This work was also funded by the European Regional Development Fund and Free State of Saxony, Germany (LIFE - Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). The **GENICA** was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. Generation Scotland (**GENSCOT**) received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). Funding for identification of cases and contribution to BCAC funded in part by the Wellcome Trust Seed Award "Temporal trends in incidence and mortality of molecular subtypes of breast cancer to inform public health, policy and prevention" Reference 207800/Z/17/Z**.** The **GESBC** was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). The HABCS study was supported by the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. The HEBCS was financially supported by the Helsinki University Hospital Research Fund, the Finnish Cancer Society, and the Sigrid Juselius Foundation. **[HEBON]** The **HMBCS** was supported by a grant from the Friends of Hannover Medical School and by the Rudolf Bartling Foundation. The **HUBCS** was supported by a grant from the German Federal Ministry of Research and Education (RUS08/017), and by the Russian Foundation for Basic Research and the Federal Agency for Scientific Organizations for support the Bioresource collections and RFBR grants 14-04-97088, 17-29-06014 and 17-44-020498. Financial support for **KARBAC** was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and Bert von Kantzows foundation. The **KARMA** study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. The **KBCP** was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland. **kConFab** is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, 199600). G.C.T. is supported by the NHMRC. The **KOHBRA** study was partially supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (HI16C1127; 1020350; 1420190). The **MARIE** study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. **MBCSG** is supported by funds from the Italian Association for Cancer Research to P. Radice and P. Peterlongo, and Italian citizens who allocated the 5x1000 share of their tax payment in

support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects '5x1000') to S. Manoukian. The **MASTOS** study was supported by "Cyprus Research Promotion Foundation" grants 0104/13 and 0104/17, and the Cyprus Institute of Neurology and Genetics. The Melbourne Collaborative Cohort Study (**MCCS**) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. **MYBRCA** is funded by research grants from the Malaysian Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Malaysia. MYMAMMO is supported by research grants from Yayasan Sime Darby LPGA Tournament and Malaysian Ministry of Higher Education (RP046B-15HTM). The NBCS has received funding from the K.G. Jebsen Centre for Breast Cancer Research; the Research Council of Norway grant 193387/V50 (to A-L Børresen-Dale and V.N. Kristensen) and grant 193387/H10 (to A-L Børresen-Dale and V.N. Kristensen), South Eastern Norway Health Authority (grant 39346 to A-L Børresen-Dale) and the Norwegian Cancer Society (to A-L Børresen-Dale and V.N. Kristensen). The Ontario Familial Breast Cancer Registry (**OFBCR**) was supported by grants U01CA164920 and U01CA167551 from the USA National Cancer Institute of the National Institutes of Health. 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) or the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR or CCFR. The **ORIGO** study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The **PLCO** is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. **PROCAS** is funded from NIHR grant PGfAR 0707-10031. The **RBCS** was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The **SASBAC** study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. SEARCH is funded by Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. **SGBCC** is funded by the National Research Foundation Singapore, a NUS start-up Grant, National University Cancer Institute Singapore (NCIS) Centre Grant, Breast Cancer Prevention Programme, Asian Breast Cancer Research Fund and a NMRC Clinician Scientist Award. Additional controls were recruited by the Singapore Consortium of Cohort Studies-Multi-ethnic cohort (SCCS-MEC), which was funded by the Biomedical Research Council, grant number: 05/1/21/19/425. SKKDKFZS is supported by the DKFZ. The **SZBCS** was supported by Grant PBZ_KBN_122/P05/2004 and the program of the Minister of Science and Higher Education under the name "Regional Initiative of Excellence" in 2019-2022 project number 002/RID/2018/19 amount of financing 12 000 000 PLN. Ascertainment and data collection for the **UBCS** is supported by funding from National Cancer Institute grants R01 CA163353 (to N.J. Camp) and the Women's Cancer Center at the Huntsman Cancer Institute (HCI) which is funded in part by the Huntsman Cancer Foundation. Data collection is also made possible by the Utah Population Database (UPDB) and the Utah Cancer Registry (UCR). Support for the UPDB is

provided by the University of Utah, HCI, and the Comprehensive Cancer Center Support grant NCI P30 CA42014. The UCR is funded by the NCI's SEER Program, Contract No. HHSN261201800016I, with additional support from the US Center for Disease Control and Prevention's National Program of Cancer Registries, Cooperative Agreement No. NU58DP0063200, the University of Utah and Huntsman Cancer Foundation. A.B.S was supported by an NHMRC Senior Research Fellowship (ID1061779). C.F. was supported by a University of Queensland (UQ) International Scholarship from the UQ School of Medicine. M.T.P. was supported by NHMRC grant funding (ID1101400, ID1161589).

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. **ABCS** thanks the Blood bank Sanquin, The Netherlands. The **ACP** study wishes to thank the participants in the Thai Breast Cancer study. Special Thanks also go to the Thai Ministry of Public Health (MOPH), doctors and nurses who helped with the data collection process. Finally, the study would like to thank Dr Prat Boonyawongviroj, the former Permanent Secretary of MOPH and Dr Pornthep Siriwanarungsan, the Department Director-General of Disease Control who have supported the study throughout. **BIGGS** thanks Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones. The **BREOGAN** study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREOGAN), José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS; Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain. **BSUCH** thanks Peter Bugert, Medical Faculty Mannheim. **CCGP** thanks Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou, Georgia Sevastaki, Konstantinos Pompodakis. **CGPS** thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. **CNIO-BCS** thanks Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO). **COLBCCC** thanks all patients, the physicians Justo G. Olaya, Mauricio Tawil, Lilian Torregrosa, Elias Quintero, Sebastian Quintero, Claudia Ramírez, José J. Caicedo and Jose F. Robledo, and the technician Michael Gilbert for their contributions and commitment to this study. **FHRISK** thanks NIHR for funding. **GC-HBOC** thanks Stefanie Engert, Heide Hellebrand, Sandra Kröber and LIFE - Leipzig Research Centre for Civilization Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber). The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [HB, WYL], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) [HB], gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - EXC 2180 - 390900677 [HB], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [YDK, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident

Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [TB, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. **HMBCS** thanks Peter Hillemanns, Hans Christiansen and Johann H. Karstens. **HUBCS** thanks Shamil Gantsev. ICICLE thanks Kelly Kohut, Michele Caneppele, Maria Troy. **KARMA** and **SASBAC** thank the Swedish Medical Research Counsel. KBCP thanks Eija Myöhänen, Helena Kemiläinen. **kConFab/AOCS** wish to thank 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 (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. We thank all investigators of the **KOHBRA** (Korean Hereditary Breast Cancer) Study. **MARIE** thanks Petra Seibold, Dieter Flesch-Janys, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. **MASTOS** thanks all the study participants and express appreciation to the doctors: Yiola Marcou, Eleni Kakouri, Panayiotis Papadopoulos, Simon Malas and Maria Daniel, as well as to all the nurses and volunteers who provided valuable help towards the recruitment of the study participants. **MBCSG** thanks Bernard Peissel, Jacopo Azzollini, Dario Zimbalatti, Daniela Zaffaroni, Mariarosaria Calvello, Davide Bondavalli, Aliana Guerrieri Gonzaga, Monica Marabelli, Irene Feroce and the personnel of the Cogentech Cancer Genetic Test Laboratory. The **MCCS** was made possible by the contribution of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. **MYBRCA** thanks study participants and research staff (particularly Patsy Ng, Nurhidayu Hassan, Yoon Sook-Yee, Daphne Lee, Lee Sheau Yee, Phuah Sze Yee and Norhashimah Hassan) for their contributions and commitment to this study. The following are **NBCS** Collaborators: Kristine K. Sahlberg (PhD), Lars Ottestad (MD), Rolf Kåresen (Prof. Em.) Dr. Ellen Schlichting (MD), Marit Muri Holmen (MD), Toril Sauer (MD), Vilde Haakensen (MD), Olav Engebråten (MD), Bjørn Naume (MD), Alexander Fosså (MD), Cecile E. Kiserud (MD), Kristin V. Reinertsen (MD), Åslaug Helland (MD), Margit Riis (MD), Jürgen Geisler (MD) and OSBREAC. **ORIGO** thanks E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. The LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCDOC) with the help of Dr. J. Molenaar. **PBCS** thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. **PROCAS** thanks NIHR for funding. **RBCS** thanks Corine M. Beaufort, Jannet Blom, Renée Broeren-Foekens, Saskia Pelders, Wendy J.C. Prager van der Smissen, Kirsten Ruigrok – Ritstier, Anita M.A.C. Trapman –Jansen, Michelle van der Vlugt – Daane, Vanja de Weerd, and the Erasmus MC Family Cancer Clinic.We thank the **SEARCH** and EPIC teams. **SGBCC** thanks the participants and research coordinator Ms Tan Siew Li. **SKKDKFZS** thanks all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to this study. **SZBCS** thanks Ewa Putresza. **UBCS** thanks the Intermountain Healthcare Biorepository for its support and commitment to this project

Consortia Memberships

kConFab/AOBC Investigators

Adrienne Sexton, Alex Dobrovic, Alice Christian, Alison Trainer, Allan Spigelman , Andrew Fellows, Andrew Shelling , Anna De Fazio, Anneke Blackburn, Ashley Crook, Bettina Meiser, Briony Patterson, Christine Clarke, Christobel Saunders, Clare Hunt, Clare Scott, David Amor, David Gallego Ortega, Deb Marsh, Edward Edkins, Elizabeth Salisbury, Eric Haan, Finlay Macrea, Gelareh Farshid, Geoff Lindeman,

Georgia Trench, Graham Mann, Graham Giles, Grantley Gill, Heather Thorne, Ian Campbell, Ian Hickie, Liz Caldon, Ingrid Winship , James Cui, James Flanagan, James Kollias, Jane Visvader, Jennifer Stone, Jessica Taylor, Jo Burke, Jodi Saunus, John Forbes, John Hopper, Jonathan Beesley, Judy Kirk, Juliet French, Kathy Tucker, Kathy Wu, Kelly Phillips, Laura Forrest, Lara Lipton, Leslie Andrews, Lizz Lobb, Logan Walker, Maira Kentwell, Mandy Spurdle, Margaret Cummings, Margaret Gleeson, Marion Harris, Mark Jenkins, Mary Anne Young, Martin Delatycki, Mathew Wallis, Matthew Burgess, Melissa Brown, Melissa Southey, Michael Bogwitz , Michael Field, Michael Friedlander, Michael Gattas, Mona Saleh, Morteza Aghmesheh, Nick Hayward, Nick Pachter, Paul Cohen, Pascal Duijf, Paul James, Pete Simpson, Peter Fong, Phyllis Butow, Rachael Williams, Rick Kefford, Rodney Scott, Roger Milne, Rosemary Balleine, Sarah – Jane Dawson, Sheau Lok, Shona O'Connell, Sian Greening, Sophie Nightingale, Stacey Edwards, Stephen Fox, Sue-Anne McLachlan, Sunil Lakhani, Tracy Dudding, Yoland Antill**.**

MyBrCa Investigators

Cheng Har Yip, Sook-Yee Yoon, Weang Kee Ho, Pei Sze Ng, Shivaani Mariapun, Siti Norhidayu Hassan, Daphne Lee, Tiara Hasan, Meow Keong Thong, Min Min Tan, Joanna Lim, Shao Yan Lao, Chan Eng Chong, Eldarina Wijaya, Nadia Rajaram, Wei Xiong Wen, Mee Hong See, Suniza Jamaris, Mei Sze Teh, Li Ying Teoh, Kartini Rahmat, Farhana Fadzli, Anusya Vijayanathan, Faizah Harun, Hanani Che Halim, Ernie Azwa Yusop, Zurina Che Rohani.

NBCS Collaborators

Anne-Lise Børresen-Dale, Grethe I. Grenaker Alnæs, Kristine K. Sahlberg, Lars Ottestad, Rolf Kåresen, Ellen Schlichting, Marit Muri Holmen, Toril Sauer, Vilde Haakensen, Olav Engebråten, Bjørn Naume, Alexander Fosså, Cecile E. Kiserud, Kristin V. Reinertsen, Åslaug Helland, Margit Riis, Jürgen Geisler and OSBREAC.

SGBCC Investigators

Swee Ho Lim, Ern Yu Tan, Benita Kiat Tee Tan, Su-Ming Tan, Veronique Kiak Mien Tan, Ching Wan Chan, Siau-Wei Tang, Celene Wei Qi Ng, Geok Hoon Lim, Jinnie Siyan Pang, Jung Ah Lee, Patrick Mun Yew Chan, Juliana Chen, Sarah Qinghui Lu, Yirong Sim, Wei Sean Yong, Preetha Madhukumar, Fuh Yong Wong, Joanne Yuen Yie Ngeow, Tira Jing Ying Tan, Wai Peng Lee, Chi Wei Mok, Chin Mui Seah, Linda Tan, E Shyong Tai, Xueling Sim, Peh Joo Ho, Alexis Jiaying Khng.

References

1. Cybulski C, Carrot-Zhang J, Kluzniak W, et al. Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet.* 2015;47(6):643-646.
2. Park DJ, Tao K, Le Calvez-Kelm F, et al. Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. *Cancer Discov.* 2014;4(7):804-815.
3. Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. *Nature.* 2013;493(7432):406-410.
4. Swisher EM, Harrell MI, Norquist BM, et al. Somatic Mosaic Mutations in PPM1D and TP53 in the Blood of Women With Ovarian Carcinoma. *JAMA Oncol.* 2016;2(3):370-372.
5. Pharoah PDP, Song H, Dicks E, et al. PPM1D Mosaic Truncating Variants in Ovarian Cancer Cases May Be Treatment-Related Somatic Mutations. *J Natl Cancer Inst.* 2016;108(3).
6. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *J Clin Oncol.* 2016;34(23):2750-2760.
7. Southey MC, Goldgar DE, Winqvist R, et al. PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. *J Med Genet.* 2016;53(12):800-811.
8. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* 2013;1303:3997.
9. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 2016;44(11):e108.
10. Sandmann S, de Graaf AO, Karimi M, et al. Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Sci Rep.* 2017;7:43169.
11. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24-26.
12. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92-94.
13. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122.
14. de la Hoya M, Soukarieh O, Lopez-Perolio I, et al. Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum Mol Genet.* 2016;25(11):2256-2268.
15. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424.
16. Fortuno C, Cipponi A, Ballinger ML, et al. A quantitative model to predict pathogenicity of missense variants in the TP53 gene. *Hum Mutat.* 2019;40(6):788-800.
17. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet.* 2013;45(4):353-361, 361e351-352.
18. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med.* 2019.
19. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet.* 2007;81(2):208-227.
20. Easton DF, Pharoah PD, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med.* 2015;372(23):2243-2257.
21. Schmidt AY, Hansen TVO, Ahlborn LB, Jonson L, Yde CW, Nielsen FC. Next-Generation Sequencing-Based Detection of Germline Copy Number Variations in BRCA1/BRCA2: Validation of a One-Step Diagnostic Workflow. *J Mol Diagn.* 2017;19(6):809-816.

22. Schmidt MK, Tollenaar RA, de Kemp SR, et al. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. *J Clin Oncol.* 2007;25(1):64-69.
23. Fasching PA, Loehberg CR, Strissel PL, et al. Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients. *Breast Cancer Res Treat.* 2008;112(1):89-98.
24. Schrauder M, Frank S, Strissel PL, et al. Single nucleotide polymorphism D1853N of the ATM gene may alter the risk for breast cancer. *J Cancer Res Clin Oncol.* 2008;134(8):873-882.
25. Colleran G, McInerney N, Rowan A, et al. The TGFBR1*6A/9A polymorphism is not associated with differential risk of breast cancer. *Breast Cancer Res Treat.* 2010;119(2):437-442.
26. McInerney N, Colleran G, Rowan A, et al. Low penetrance breast cancer predisposition SNPs are site specific. *Breast Cancer Res Treat.* 2009;117(1):151-159.
27. Jiang X, Castelao JE, Chavez-Uribe E, et al. Family history and breast cancer hormone receptor status in a Spanish cohort. *PLoS One.* 2012;7(1):e29459.
28. Redondo CM, Gago-Dominguez M, Ponte SM, et al. Breast feeding, parity and breast cancer subtypes in a Spanish cohort. *PLoS One.* 2012;7(7):e40543.
29. Ali AM, Schmidt MK, Bolla MK, et al. Alcohol consumption and survival after a breast cancer diagnosis: a literature-based meta-analysis and collaborative analysis of data for 29,239 cases. *Cancer Epidemiol Biomarkers Prev.* 2014;23(6):934-945.
30. Cruz GI, Martinez ME, Natarajan L, et al. Hypothesized role of pregnancy hormones on HER2+ breast tumor development. *Breast Cancer Res Treat.* 2013;137(1):237-246.
31. Gago-Dominguez M, Castelao JE, Gude F, et al. Alcohol and breast cancer tumor subtypes in a Spanish Cohort. *Springerplus.* 2016;5:39.
32. Yang R, Dick M, Marme F, et al. Genetic variants within miR-126 and miR-335 are not associated with breast cancer risk. *Breast Cancer Res Treat.* 2011;127(2):549-554.
33. Menegaux F, Truong T, Anger A, et al. Night work and breast cancer: a population-based case-control study in France (the CECILE study). *Int J Cancer.* 2013;132(4):924-931.
34. Weischer M, Bojesen SE, Tybjaerg-Hansen A, Axelsson CK, Nordestgaard BG. Increased risk of breast cancer associated with CHEK2*1100delC. *J Clin Oncol.* 2007;25(1):57-63.
35. Milne RL, Ribas G, Gonzalez-Neira A, et al. ERCC4 associated with breast cancer risk: a two-stage case-control study using high-throughput genotyping. *Cancer Res.* 2006;66(19):9420-9427.
36. Evans DG, Astley S, Stavrinos P, et al. *Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and family history clinics: a dual cohort study*. Southampton (UK)2016.
37. Ingham SL, Warwick J, Buchan I, et al. Ovarian cancer among 8,005 women from a breast cancer family history clinic: no increased risk of invasive ovarian cancer in families testing negative for BRCA1 and BRCA2. *J Med Genet.* 2013;50(6):368-372.
38. Kast K, Rhiem K, Wappenschmidt B, et al. Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer. *J Med Genet.* 2016;53(7):465-471.
39. Rhiem K, Engel C, Graeser M, et al. The risk of contralateral breast cancer in patients from BRCA1/2 negative high risk families as compared to patients from BRCA1 or BRCA2 positive families: a retrospective cohort study. *Breast Cancer Res.* 2012;14(6):R156.
40. Graeser MK, Engel C, Rhiem K, et al. Contralateral breast cancer risk in BRCA1 and BRCA2 mutation carriers. *J Clin Oncol.* 2009;27(35):5887-5892.
41. Engel C, Rhiem K, Hahnen E, et al. Prevalence of pathogenic BRCA1/2 germline mutations among 802 women with unilateral triple-negative breast cancer without family cancer history. *BMC Cancer.* 2018;18(1):265.
42. Pesch B, Ko Y, Brauch H, et al. Factors modifying the association between hormone-replacement therapy and breast cancer risk. *Eur J Epidemiol.* 2005;20(8):699-711.

43. Justenhoven C, Pierl CB, Haas S, et al. The CYP1B1_1358_GG genotype is associated with estrogen receptor-negative breast cancer. *Breast Cancer Res Treat.* 2008;111(1):171-177.
44. Smith BH, Campbell A, Linksted P, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol.* 2013;42(3):689-700.
45. Chang-Claude J, Eby N, Kiechle M, Bastert G, Becher H. Breastfeeding and breast cancer risk by age 50 among women in Germany. *Cancer Causes Control.* 2000;11(8):687-695.
46. Dork T, Bendix R, Bremer M, et al. Spectrum of ATM gene mutations in a hospital-based series of unselected breast cancer patients. *Cancer Res.* 2001;61(20):7608-7615.
47. Syrjakoski K, Vahteristo P, Eerola H, et al. Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients. *J Natl Cancer Inst.* 2000;92(18):1529-1531.
48. Kilpivaara O, Bartkova J, Eerola H, et al. Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients. *Int J Cancer.* 2005;113(4):575-580.
49. Fagerholm R, Hofstetter B, Tommiska J, et al. NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nat Genet.* 2008;40(7):844-853.
50. Bogdanova N, Cybulski C, Bermisheva M, et al. A nonsense mutation (E1978X) in the ATM gene is associated with breast cancer. *Breast Cancer Res Treat.* 2009;118(1):207-211.
51. Wendt C, Lindblom A, Arver B, von Wachenfeldt A, Margolin S. Tumour spectrum in non-BRCA hereditary breast cancer families in Sweden. *Hered Cancer Clin Pract.* 2015;13(1):15.
52. Margolin S, Werelius B, Fornander T, Lindblom A. BRCA1 mutations in a population-based study of breast cancer in Stockholm County. *Genet Test.* 2004;8(2):127-132.
53. Hartikainen JM, Tuhkanen H, Kataja V, et al. An autosome-wide scan for linkage disequilibrium-based association in sporadic breast cancer cases in eastern Finland: three candidate regions found. *Cancer Epidemiol Biomarkers Prev.* 2005;14(1):75-80.
54. Hartikainen JM, Tuhkanen H, Kataja V, et al. Refinement of the 22q12-q13 breast cancer--associated region: evidence of TMPRSS6 as a candidate gene in an eastern Finnish population. *Clin Cancer Res.* 2006;12(5):1454-1462.
55. Mann GJ, Thorne H, Balleine RL, et al. Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the kConFab familial breast cancer resource. *Breast Cancer Res.* 2006;8(1):R12.
56. Beesley J, Jordan SJ, Spurdle AB, et al. Association between single-nucleotide polymorphisms in hormone metabolism and DNA repair genes and epithelial ovarian cancer: results from two Australian studies and an additional validation set. *Cancer Epidemiol Biomarkers Prev.* 2007;16(12):2557-2565.
57. Han SA, Park SK, Ahn SH, et al. The Korean Hereditary Breast Cancer (KOHBRA) study: protocols and interim report. *Clin Oncol (R Coll Radiol).* 2011;23(7):434-441.
58. Flesch-Janys D, Slanger T, Mutschelknauss E, et al. Risk of different histological types of postmenopausal breast cancer by type and regimen of menopausal hormone therapy. *Int J Cancer.* 2008;123(4):933-941.
59. Hadjisavvas A, Loizidou MA, Middleton N, et al. An investigation of breast cancer risk factors in Cyprus: a case control study. *BMC Cancer.* 2010;10:447.
60. De Vecchi G, Verderio P, Pizzamiglio S, et al. Evidences for association of the CASP8 -652 6N del promoter polymorphism with age at diagnosis in familial breast cancer cases. *Breast Cancer Res Treat.* 2009;113(3):607-608.
61. Catucci I, Verderio P, Pizzamiglio S, et al. SNPs in ultraconserved elements and familial breast cancer risk. *Carcinogenesis.* 2009;30(3):544-545; author reply 546.
62. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ.* 2002;156:69-70.

63. Phuah SY, Looi LM, Hassan N, et al. Triple-negative breast cancer and PTEN (phosphatase and tensin homologue) loss are predictors of BRCA1 germline mutations in women with early-onset and familial breast cancer, but not in women with isolated late-onset breast cancer. *Breast Cancer Res.* 2012;14(6):R142.
64. Mariapun S, Ho WK, Kang PC, et al. Variants in 6q25.1 Are Associated with Mammographic Density in Malaysian Chinese Women. *Cancer Epidemiol Biomarkers Prev.* 2016;25(2):327-333.
65. Aure MR, Jernstrom S, Krohn M, et al. Integrated analysis reveals microRNA networks coordinately expressed with key proteins in breast cancer. *Genome Med.* 2015;7(1):21.
66. Fleischer T, Edvardsen H, Solvang HK, et al. Integrated analysis of high-resolution DNA methylation profiles, gene expression, germline genotypes and clinical end points in breast cancer patients. *Int J Cancer.* 2014;134(11):2615-2625.
67. Fleischer T, Frigessi A, Johnson KC, et al. Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol.* 2014;15(8):435.
68. Quigley DA, Fiorito E, Nord S, et al. The 5p12 breast cancer susceptibility locus affects MRPS30 expression in estrogen-receptor positive tumors. *Mol Oncol.* 2014;8(2):273-284.
69. John EM, Hopper JL, Beck JC, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res.* 2004;6(4):R375-389.
70. de Bock GH, Schutte M, Krol-Warmerdam EM, et al. Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant. *J Med Genet.* 2004;41(10):731-735.
71. Huijts PE, Vreeswijk MP, Kroeze-Jansema KH, et al. Clinical correlates of low-risk variants in FGFR2, TNRC9, MAP3K1, LSP1 and 8q24 in a Dutch cohort of incident breast cancer cases. *Breast Cancer Res.* 2007;9(6):R78.
72. Garcia-Closas M, Egan KM, Newcomb PA, et al. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. *Hum Genet.* 2006;119(4):376-388.
73. Pfeiffer RM, Park Y, Kreimer AR, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. *PLoS Med.* 2013;10(7):e1001492.
74. Kriege M, Hollestelle A, Jager A, et al. Survival and contralateral breast cancer in CHEK2 1100delC breast cancer patients: impact of adjuvant chemotherapy. *Br J Cancer.* 2014;111(5):1004-1013.
75. Wedren S, Lovmar L, Humphreys K, et al. Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study. *Breast Cancer Res.* 2004;6(4):R437-449.
76. Lesueur F, Pharoah PD, Laing S, et al. Allelic association of the human homologue of the mouse modifier Ptprj with breast cancer. *Hum Mol Genet.* 2005;14(16):2349-2356.
77. Stevens KN, Fredericksen Z, Vachon CM, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. *Cancer Res.* 2012;72(7):1795-1803.
78. Jakubowska A, Cybulski C, Szymanska A, et al. BARD1 and breast cancer in Poland. *Breast Cancer Res Treat.* 2008;107(1):119-122.
79. Jakubowska A, Jaworska K, Cybulski C, et al. Do BRCA1 modifiers also affect the risk of breast cancer in non-carriers? *Eur J Cancer.* 2009;45(5):837-842.
80. Cybulski C, Kluzniak W, Huzarski T, et al. Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. *Lancet Oncol.* 2015;16(6):638-644.
81. Madsen MJ, Knight S, Sweeney C, et al. Reparameterization of PAM50 Expression Identifies Novel Breast Tumor Dimensions and Leads to Discovery of a Genome-Wide Significant Breast Cancer Locus at 12q15. *Cancer Epidemiol Biomarkers Prev.* 2018;27(6):644-652.

82. Camp NJ, Parry M, Knight S, et al. Fine-mapping CASP8 risk variants in breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2012;21(1):176-181.
83. Baretic D, Pollard HK, Fisher DI, et al. Structures of closed and open conformations of dimeric human ATM. *Sci Adv.* 2017;3(5):e1700933.
84. Irminger-Finger I, Ratajska M, Pilyugin M. New concepts on BARD1: Regulator of BRCA pathways and beyond. *Int J Biochem Cell Biol.* 2016;72:1-17.
85. Lee MS, Green R, Marsillac SM, et al. Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays. *Cancer Res.* 2010;70(12):4880-4890.
86. Berge EO, Staalesen V, Straume AH, Lillehaug JR, Lonning PE. Chk2 splice variants express a dominant-negative effect on the wild-type Chk2 kinase activity. *Biochim Biophys Acta.* 2010;1803(3):386-395.
87. Miller KA, Sawicka D, Barsky D, Albala JS. Domain mapping of the Rad51 paralog protein complexes. *Nucleic Acids Res.* 2004;32(1):169-178.
88. Oliver AW, Swift S, Lord CJ, Ashworth A, Pearl LH. Structural basis for recruitment of BRCA2 by PALB2. *EMBO Rep.* 2009;10(9):990-996.