Dorling *et al. Genome Medicine* (2022) 14:51
https://doi.org/10.1186/s13073-022-01052-8

Genome Medicine

RESEARCH **Open Access**

Check for updates

Breast cancer risks associated with missense variants in breast cancer susceptibility genes

Leila Dorling[1], Sara Carvalho[1], Jamie Allen[1], Michael T. Parsons[2], Cristina Fortuno[2], Anna González-Neira[3],
Stephan M. Heijl[4], Muriel A. Adank[5], Thomas U. Ahearn[6], Irene L. Andrulis[7,8], Päivi Auvinen[9,10,11],
Heiko Becher[12], Matthias W. Beckmann[13], Sabine Behrens[14], Marina Bermisheva[15], Natalia V. Bogdanova[16,17,18],
Stig E. Bojesen[19,20,21], Manjeet K. Bolla[1], Michael Bremer[16], Ignacio Briceno[22], Nicola J. Camp[23],
Archie Campbell[24,25], Jose E. Castelao[26], Jenny Chang-Claude[14,27], Stephen J. Chanock[6],
Georgia Chenevix-Trench[2], NBCS Collaborators[28,29,30,31,32,33,34,35,36], J. Margriet Collée[37], Kamila Czene[38],
Joe Dennis[1], Thilo Dörk[17], Mikael Eriksson[38], D. Gareth Evans[39,40,41,42], Peter A. Fasching[13,43],
Jonine Figueroa[6,25,44], Henrik Flyger[45], Marike Gabrielson[38], Manuela Gago-Dominguez[46,47],
Montserrat García-Closas[6], Graham G. Giles[48,49,50], Gord Glendon[7], Pascal Guénel[51], Melanie Gündert[52,53,54],
Andreas Hadjisavvas[55,56], Eric Hahnen[57,58], Per Hall[38,59], Ute Hamann[60], Elaine F. Harkness[41,42,61],
Mikael Hartman[62,63,64], Frans B. L. Hogervorst[5], Antoinette Hollestelle[65], Reiner Hoppe[66,67], Anthony Howell[42,68],
kConFab Investigators[69,70], SGBCC Investigators[62,63,71,72,73,74,75,76,77,78,79,80,81], Anna Jakubowska[82,83],
Audrey Jung[14], Elza Khusnutdinova[15,84], Sung-Won Kim[85], Yon-Dschun Ko[86], Vessela N. Kristensen[29,35],
Inge M. M. Lakeman[87,88], Jingmei Li[63,71], Annika Lindblom[89,90], Maria A. Loizidou[55,56], Artitaya Lophatananon[91],
Jan Lubiński[82], Craig Luccarini[92], Michael J. Madsen[23], Arto Mannermaa[9,93,94], Mehdi Manoochehri[60],
Sara Margolin[59,95], Dimitrios Mavroudis[96], Roger L. Milne[48,49,50], Nur Aishah Mohd Taib[97,98], Kenneth Muir[91],
Heli Nevanlinna[99], William G. Newman[39,40,42], Jan C. Oosterwijk[100], Sue K. Park[101,102,103], Paolo Peterlongo[104],
Paolo Radice[105], Emmanouil Saloustros[106], Elinor J. Sawyer[107], Rita K. Schmutzler[57,58,108], Mitul Shah[92],
Xueling Sim[62], Melissa C. Southey[48,50,109], Harald Surowy[52,53], Maija Suvanto[99], Ian Tomlinson[110,111],
Diana Torres[60,112], Thérèse Truong[51], Christi J. van Asperen[88], Regina Waltes[17], Qin Wang[1], Xiaohong R. Yang[6],
Paul D. P. Pharoah[1,92], Marjanka K. Schmidt[113,114], Javier Benitez[3,115], Bas Vroling[4,116], Alison M. Dunning[92],
Soo Hwang Teo[97,117], Anders Kvist[118], Miguel de la Hoya[119], Peter Devilee[87,120], Amanda B. Spurdle[2],
Maaike P. G. Vreeswijk[87] and Douglas F. Easton[1,92*]

---

*Correspondence: dfe20@medschl.cam.ac.uk

[1] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
Full list of author information is available at the end of the article

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Abstract

Background: Protein truncating variants in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, and *PALB2* are associated with increased breast cancer risk, but risks associated with missense variants in these genes are uncertain.

Methods: We analyzed data on 59,639 breast cancer cases and 53,165 controls from studies participating in the Breast Cancer Association Consortium BRIDGES project. We sampled training (80%) and validation (20%) sets to analyze rare missense variants in *ATM* (1146 training variants), *BRCA1* (644), *BRCA2* (1425), *CHEK2* (325), and *PALB2* (472). We evaluated breast cancer risks according to five in silico prediction-of-deleteriousness algorithms, functional protein domain, and frequency, using logistic regression models and also mixture models in which a subset of variants was assumed to be risk-associated.

Results: The most predictive in silico algorithms were Helix (*BRCA1*, *BRCA2* and *CHEK2*) and CADD (*ATM*). Increased risks appeared restricted to functional protein domains for *ATM* (FAT and PIK domains) and *BRCA1* (RING and BRCT domains). For *ATM*, *BRCA1*, and *BRCA2*, data were compatible with small subsets (approximately 7%, 2%, and 0.6%, respectively) of rare missense variants giving similar risk to those of protein truncating variants in the same gene. For *CHEK2*, data were more consistent with a large fraction (approximately 60%) of rare missense variants giving a lower risk (OR 1.75, 95% CI (1.47–2.08)) than *CHEK2* protein truncating variants. There was little evidence for an association with risk for missense variants in *PALB2*. The best fitting models were well calibrated in the validation set.

Conclusions: These results will inform risk prediction models and the selection of candidate variants for functional assays and could contribute to the clinical reporting of gene panel testing for breast cancer susceptibility.

Keywords: Breast cancer, Genetic epidemiology, Risk prediction, Missense variants

Background

Genetic testing for cancer susceptibility is now part of mainstream clinical practice. For breast cancer susceptibility, genetic testing generally focuses on high-risk genes, notably *BRCA1*, *BRCA2*, *PALB2*, and *TP53*, but testing of larger panels that include so-called "moderate-risk" genes is being increasingly offered [1]. While the evidence that many of these genes are risk associated is clear, for most this evidence is based on carrying a protein truncating variant (PTV). Besides PTVs, genetic testing also identifies missense variants for which the impact on protein function and associated cancer risk is generally unknown ("variants of uncertain significance" (VUS)), resulting in a major problem for genetic counselling. Some missense variants have been shown to confer risk [2, 3] with risk estimates comparable to PTVs, and it is possible that missense variants contribute substantially to risk [4, 5], at least in some genes. However, defining the set of missense variants in each gene that may confer risk, and their associated risk estimates, presents an ongoing problem.

Resolving this problem is complex as most variants are individually very rare, so the evidence must be based on combining data across multiple variants in a statistical model. To this end, efforts have been made to develop statistical algorithms that score missense variants according to in silico features that may predict pathogenicity. Here, we have compared the usefulness of five in silico algorithms in predicting breast cancer risk associated with missense variants using sequenced germline DNA from more than 59,000 cases and 53,000 controls from studies in the Breast Cancer Association Consortium (BCAC) [6] participating in the BRIDGES project [7]. We used the most predictive in silico algorithm to estimate the risks of breast cancer associated with subsets of rare missense variants, defined by categories of the in silico score, in *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, and *PALB2*. These predictions were then validated using an independent dataset.

Methods

Subjects

We included data from female breast cancer patients (cases) and unaffected controls from 44 studies participating in the BRIDGES project, as previously documented [7]. These studies are a subset of studies participating in the Breast Cancer Association Consortium (BCAC) for which targeted sequencing was performed using the BRIDGES panel (see below). Details of the participating studies, including the enrollment of cases and controls and sample sizes, are given in Additional File 1: Tables S1 and S2. Of these, 30 were population-based or hospital-based studies (hereafter: population studies) including cases and controls sampled independently of family history. A further 14 studies oversampled cases with a family history of breast cancer (hereafter: familial studies). All studies were approved by the relevant ethical review boards and used appropriate consent procedures. Five duplicated samples were identified and removed. After quality control procedures (see below), 53,165 controls and 59,639 cases with an invasive (53,838; 90.3%) or in situ (4,153; 7.0%) tumor, or tumor of unknown invasiveness (1648; 2.7%), were included in the analyses. Of these, 50,414 controls and 48,230 cases were from population studies.

Laboratory methods, variant calling, and classification

The BRIDGES project performed targeted sequencing on a panel of 34 genes [7]. Of these five (*ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*) were chosen for further analysis and presented here. These five genes, where the evidence for association with breast cancer risk is strongest, are most relevant to risk prediction and included in the current version of the BOADICEA/CanRisk risk prediction tool [8]. Details of library preparation, sequencing, variant calling, quality control procedures, and variant classification has been documented previously [7]. Missense variants in the entire gene were identified using the Ensembl Variant Effect Predictor (VEP; version 101.0) [9]. Rare variants for in silico analysis were defined as those with allele frequency < 0.1% (calculated as previously described [7]); in addition, variants with frequency < 5% were retained for a frequency-based analysis. Carriers of missense variants predicted to affect RNA splicing, according to the MaxEntScan tool [10] and SpliceAI scores [11], were removed (see Additional File 2: Table S3). Variants were annotated for functional protein domain location, defined according to published literature, the UniProt Knowledgebase [12], and for *BRCA1* and BRCA2, the ENIGMA *BRCA1/2* expert panel guidelines [13] (see Additional File 1: Table S4). Variants were also classified for disease pathogenicity assertion in ClinVar [14] with a filter for no conflicting interpretations; for *BRCA1* and *BRCA2*, variants were also reviewed against the ENIGMA *BRCA1/2* expert panel guidelines. The ENIGMA terminology report [15] reserves use of the word "pathogenic" to describe variants associated with at least a twofold cancer risk; however, for the purpose of this article, we describe any variant associated with risk as pathogenic.

Variants were scored using five in silico prediction algorithms: Align-GVGD [16], Combined Annotation Dependent Depletion (CADD; version 1.4) [17], Rare Exome Variant Ensemble Learner (REVEL) [18], BayesDel (without allele frequency; version 1) [19], and Helix (version 4.2.0) [20]. The first four are widely used for variant classification in cancer susceptibility genes. Align-GVGD classifies variants according to the level of cross-species conservation observed for a single missense substitution while considering the biophysical characteristics of the amino acids. CADD, BayesDel, and REVEL are ensemble methods that integrate several different annotations, including conservation metrics, regulatory information, transcript information, and protein-level scores, into a single score of deleteriousness. Helix combines structural, alignment, and gene data with a strict training regime where circularity is actively avoided to produce a variant score and certainty estimate. All variants were scored using default software settings. For Align-GVGD, the sequence alignment with the deepest phylogeny level was used. Variants in *BRCA1* and *BRCA2* were also annotated with the predictions of Hart et al. [21], who developed two BRCA-specific in silico algorithms (Random Forest (RF) and Naïve Voting Method (NVM)) to classify missense variants as functionally damaging or neutral. In addition, *BRCA1* variants were annotated using the prediction of loss-of-function made by the Saturation Genome Editing (SGE) experiments of Findlay et al. [22], which involved a comprehensive functional assessment of missense variants lying within the functional domain coding regions of *BRCA1*. *BRCA2* variants were annotated using homology-directed DNA repair (HDR) assay scores and predictions of pathogenicity from Richardson et al. [23]. For *PALB2*, variants were annotated with five different assay scores measuring HDR activity, PARPi sensitivity, and homologous recombination (HR) efficiency from the functional screening studies of Boonen et al. [24], Rodrigue et al. [25] and Wiltshire et al. [26].

Statistical analysis

The dataset was split into a training (80% of individuals) and a validation (20%) set. Samples for the validation set were selected randomly from population studies of cases unselected for family history of breast cancer and controls, in countries contributing a total of > 5000 samples (Denmark, Germany, Singapore (Chinese), Sweden, UK, USA). All remaining samples were included in the training set. The training set included 37,211 cases from population studies, 11,409 cases from familial studies, and 42,334 controls. Of these, 3818 individuals were carriers of PTVs in one or more of the five genes under consideration and were excluded from all analyses except the mixture models (see below). The validation set included 11,019 cases and 10,831 controls from population studies and did not include any carriers of PTVs. Oversampling of cases with a family history increases power but leads to biased effect sizes, so we chose this approach to maximize the power to discriminate between models in the training set, which could then be refit and tested on a dataset unselected for family history. All analyses were adjusted for country as a covariate; in addition, for Malaysia and Singapore, the three distinct ethnic groups (Chinese, Indian, Malay) were treated as different strata, and the UK was treated as three strata (SEARCH from East Anglia, GENSCOT from Scotland, and PROCAS and FHRISK from north-west England).

Training dataset analysis

An analysis flow diagram is presented in Fig. S1 (see Additional File 1). Analyses were performed in R version 4.0.3 (R: A Language and Environment for Statistical Computing; http://www.r-project.org). We first used logistic regression (LR) to explore which of the five in silico scores (Align-GVGD, BayesDel, CADD, Helix, and REVEL—all analyzed as continuous variables) were most strongly associated with risk of breast cancer. In addition, to assess the utility of gene-specific in silico tools, we analyzed the Hart et al. RF and NVM in silico predictions for *BRCA1* and *BRCA2*. To evaluate the usefulness of functional predictions, we also analyzed the BRCA1 SGE score; the Richardson et al. BRCA2 HDR score; and, for *PALB2*, five functional assay scores. These analyses were restricted to carriers of a rare (frequency < 0.1%) missense variant in the training set, with an endpoint of breast cancer occurrence (yes/no). The strongest predictors were used to test the association of different categories of the score(s) compared to a baseline category, in conjunction with functional protein domains, and hence create a set of risk categories. LR was then used in the training set (carriers and non-carriers) to estimate the odds ratios (OR) associated with different risk categories. As an alternative approach, we fitted mixture models in which only a proportion of variants (α) was assumed to be risk associated in the given gene; the OR was assumed to be the same for all risk associated variants, but the proportion of risk associated variants varied by risk category (as defined in the LR models). This model is motivated by the binary variant classification approach used in clinical genetics, where all variants are assumed to be either associated with moderate-high risk (likely pathogenic) or not (likely benign) [27]. We considered two types of mixture model: a constrained model in which the missense OR was equal to that of PTVs, and an unconstrained model in which the missense OR could differ from the PTV OR. Carriers of PTVs in the gene under consideration were re-included in the mixture models (to allow the risk associated missense OR to be constrained to the PTV OR). The mixture models were fitted using an expectation–maximization (EM) algorithm [28]. In the expectation step, the (posterior) probability that each variant was risk associated, given the case control data on that variant in the training set and the current parameter values was calculated. These probabilities were then used as weights in a logistic regression analysis in the maximization step. In a case–control dataset, the naïve proportions, α, will be biased because risk associated variants are more likely to be found in cases. For the final models, therefore, we also computed the proportions based only on variants reported in controls. To evaluate the overall fit of the models, we compared log-likelihoods.

The initial model selection was based on all samples, but final parameter estimates were obtained from population studies only. In the results, the ORs, *P*-values, and α presented are from population studies, unless indicated by the suffix "ALL".

Case-only analyses of age at diagnosis, with risk category as the outcome variable, were performed to evaluate trends in the ORs for variant risk category by age. We evaluated individual risk variants previously reported in literature and, in aggregate, those classified as "pathogenic" or "likely pathogenic" (hereafter, all termed: (likely) pathogenic) according to clinical guidelines. To examine whether rare variant frequency is associated with risk, we used a carrier-only LR analysis to test frequency up to 0.5% on a continuous scale and a log scale, and to compare rare variants in two groups: frequency < 0.1% versus frequency 0.1–0.5%. We also performed burden analyses within each gene comparing the risk for non-carriers to the risk for carriers of variants in one of four frequency groups: < 0.1%; 0.1–0.5%; 0.5–1%; and 1–5%. Variants with frequency between 0.1 and 5% were also evaluated individually.

Validation dataset analysis

To evaluate the calibration of the in silico training models, we performed case–control analyses using the validation dataset. In these analyses, OR estimates were fixed according to the population estimates from the training models (Table 1), but the other parameters (intercept and country covariates) were re-estimated, since the case–control proportions might differ between the training and validation datasets. From the validation model, we extracted the predicted probability that each individual was a case and hence derived expected numbers of cases and controls in each risk group. These were used to plot observed versus expected OR estimates and perform a goodness of fit chi-squared test.

The mixture models were assessed similarly, with the exception that both the OR parameter and the proportion of risk associated variants, α, were fixed. However, an adjustment to α was incorporated to allow for the different distribution of cases and controls within the

Table 1 Breast cancer risk association results from logistic regression and mixture models of population training samples

	N			**Logistic regression model**			**Mixture model**		
Risk group	**Variants**[a]	**Cases**	**Controls**	**OR**[b]	**95% CI**[c]	***P*-value**	**Missense OR (95% CI)**[d]	α[e]	**95% CI**[f]
ATM				Log-likelihood = −48,624.97			Log-likelihood = −48,624.64		
Non-carriers	–	33,351	37,001	1	–	–		0	–
Carriers							2.16 (1.78–2.63)[h]		
Variant outside FAT and PIK domains	714	1259	1443	0.98	(0.91–1.06)	0.67		0.0041	(0.001–0.02)
Variant inside FAT or PIK domain and CADD score quintiles 1–4[g]	171	317	333	1.10	(0.94–1.29)	0.24		0.055	(0.03–0.12)
Variant inside FAT or PIK domain and CADD score quintile 5[g]	103	239	162	1.64	(1.33–2.02)	3.1×10^{-6}		0.54	(0.41–0.68)
BRCA1				Log-likelihood = −48,652.14			Log-likelihood = −48,652.29		
Non-carriers	–	34,191	37,996	1	–	–		0	–
Carriers							10.61 (7.92–14.21)[h]		
Variant outside RING and BRCT domains	479	811	856	1.01	(0.92–1.12)	0.79		0.0015	(9.4×10^{-5}–0.025)
Variant inside RING or BRCT domain and low Helix score	79	120	103	1.18	(0.90–1.55)	0.23		1.0×10^{-11}	NA
Variant inside RING or BRCT domain and high Helix score	23	63	16	4.94	(2.83–8.61)	1.9×10^{-8}		0.48	(0.19–0.78)
BRCA2				Log-likelihood = −48,641.97			Log-likelihood = −48,638.78		
Non-carriers	–	33,006	36,517	1	–	–		0	–
Carriers							5.87 (4.75–7.24)[h]		
Variant with low Helix score	1160	2062	2323	0.98	(0.92–1.04)	0.47		5.1×10^{-5}	(2.4×10^{-9}–0.52)
Variant with high Helix score	62	114	94	1.28	(0.96–1.70)	0.087		0.11	(0.04–0.25)
CHEK2				Log-likelihood = −48,728.96			Log-likelihood = −48,728.70		
Non-carriers	–	34,582	38,480	1	–	–		0	–
Carriers							1.75 (1.47–2.08)[i]		
Variant with low Helix score	157	403	363	1.26	(1.08–1.46)	0.0025		0.33	(0.25–0.43)
Variant with high Helix score	121	265	177	1.73	(1.42–2.11)	4.7×10^{-8}		0.95	(0.86–0.98)
PALB2				Log-likelihood = −48,728.67			Log-likelihood = −48,729.17		
Non-carriers	–	34,622	38,291	1	–	–		0	–
Carriers	424	618	713	0.95	(0.85–1.06)	0.34	4.87 (3.50–6.77)[h]	1.1×10^{-4}	(1.6×10^{-9}–0.88)

[a] Number of unique missense substitutions in population dataset

[b] Logistic regression odds ratio estimate for missense variant carriers

[c] 95% confidence interval for logistic regression OR estimate for missense variant carriers

[d] Mixture model odds ratio and 95% confidence interval for missense variant carriers

[e] Alpha: estimated proportion of risk associated missense variants

[f] 95% confidence interval for alpha

[g] CADD quintiles 1–4 includes all CADD score values ≤ 3.736542; CADD quintile 5 includes all CADD score values > 3.736542

[h] Missense variant odds ratio constrained to equal odds ratio for protein truncating variants

[i] Missense variant odds ratio unconstrained

validation set compared to the training set. To do this, the proportions of cases and controls that were carrying a risk associated variant in the training set were estimated separately and α in the validation set was then computed as a weighted average of these two estimates. As an alternative approach, the predicted ORs in the validation set were computed using the posterior probabilities (PP) of each variant being risk associated (from the training set) as weights. This analysis was restricted to the subset of individuals carrying variants found in the training set or carrying no variant.

As a final analysis, a single unconstrained logistic regression model comprising all the defined risk groups across the five genes, with non-carriers of any missense variant as the baseline group, was fitted, and the risks in the validation set were evaluated.

The estimated familial relative risk λ_j due to deleterious missenses in each gene j was estimated using the formula $\lambda_j = \frac{(p_jr_j^2+q_j(p_jr_j+q_j)^2}{(2p_jr_j+1-2p_j)^2}$, where p_j is the estimated total frequency of deleterious missense variants, $q_j = 1 = p_j$ and r_j is the estimated relative risk conferred by deleterious variants. The total contribution of deleterious missense variants was estimated by assuming that the contribution of variants in the different genes is additive, i.e., $\lambda_{mis} = 1 + \sum(\lambda_j - 1)$. The proportion of the overall familial relative risk due to missense variants was then calculated as $\log(\lambda_{mis})/\log(2)$, that is assuming an overall familial relative risk of 2 and that variant combine multiplicatively with other genetic/familial factors, consistent with previous observations.

Results

ATM

The analysis of *ATM* missense variants included 4522 carriers of 1146 unique variants. In the carrier only analysis, BayesDel ($p_{\mathrm{ALL}} = 0.024$), CADD ($p_{\mathrm{ALL}} = 0.0022$), Helix ($p_{\mathrm{ALL}} = 0.0045$), and REVEL ($p_{\mathrm{ALL}} = 0.024$) scores were all predictive of risk (see Additional File 2: Table S5). For the most strongly associated score, CADD, the risk appeared to be restricted to the fifth quintile (Q5; CADD > 3.736542; $p = 0.033$ compared with third quintile). Functional protein domain was also predictive, with increased risks associated with the FRAP-ATM-TRRAP (FAT; $p_{\mathrm{ALL}} = 9.5 \times 10^{-4}$) and phosphatidylinositol 3-kinase and 4-kinase (PIK; $p_{\mathrm{ALL}} = 0.0016$) domains compared with variants outside a known domain. Including CADD and protein domain, only variants in the category that included CADD Q5 variants in the FAT or PIK domains (FAT/PIK + CADD5) were associated with risk relative to non-carriers (OR 1.64 (1.33–2.02), $p = 3.1 \times 10^{-6}$; Table 1, Figs. 1a and 2a). In the most parsimonious mixture model, risk associated variants conferred an equivalent risk to PTVs (OR 2.16 (1.78–2.63)); an estimated 54% (95% CI (41–68%)) of variants in the FAT/PIK + CADD5 risk group were risk associated, compared to less than 6% of variants in other risk categories (Table 1, Figs. 1a and 2a). There was no evidence that missense variants were associated with a different risk compared with PTVs ($p = 0.48$). The mixture model was a slightly better fit to the data than the LR model ($2 \times$ log-likelihood difference = 0.67). There was no association between age-at-diagnosis and risk category (see Additional File 1: Table S6).

Thirteen *ATM* missense variants were classified as (likely) pathogenic on the ClinVar database (see Additional File 1: Table S7). These variants, in aggregate, were associated with an increased risk (OR 1.85 (0.98–3.50, $p = 0.060$; $p_{\mathrm{ALL}} = 0.00053$)). However, the association of (likely) pathogenic variants was not present when the analysis was restricted to the five variants not in the FAT or PIK domains (OR = 0.97 (0.19–5.08)), though the carrier numbers were small and the confidence interval wide. Conversely, variants in the FAT/PIK + CADD5 risk group, in aggregate, remained risk associated, even when

(See figure on next page.)

Fig. 1 Odds ratios and alpha estimates for each of five genes in population training samples. **A** *ATM*. Odds ratios for breast cancer risk from logistic regression models. Alpha is the estimated proportion of risk associated variants from mixture models, based on variants in control samples. *ATM* risk categories: variants lying within the FAT or PI3K/PI4K protein domains with CADD score in the fifth quintile (FAT/PIK + CADD5); variants lying within the FAT or PI3K/PI4K protein domains with CADD score in any of the first four quintiles (FAT/PIK + CADD1-4); variants lying outside the FAT and PI3K/PI4K protein domains (Outside FAT/PIK). **B** *BRCA1*. Odds ratios for breast cancer risk from logistic regression models. Alpha is the estimated proportion of risk associated variants from mixture models, based on variants in control samples. *BRCA1* risk categories: variants lying within the RING or BRCT domains with a high Helix score (RING/BRCT + Helix-high); variants lying with the RING or BRCT domains with a low Helix score (RING/BRCT + Helix-low); variants lying outside the RING and BRCT domains (Outside RING/BRCT). **C** *BRCA2*. Odds ratios for breast cancer risk from logistic regression models. Alpha is the estimated proportion of risk associated variants from mixture models, based on variants in control samples. *BRCA2* risk categories: variants with a high Helix score (Helix-high); variants with a low Helix score (Helix-low). **D** *CHEK2*. Odds ratios for breast cancer risk from logistic regression models. Alpha is the estimated proportion of risk associated variants from mixture models, based on variants in control samples. *CHEK2* risk categories: variants with a high Helix score (Helix-high); variants with a low Helix score (Helix-low). **E** *PALB2*. Odds ratios for breast cancer risk from logistic regression models. Alpha is the estimated proportion of risk associated variants from mixture models, based on variants in control samples. *PALB2* risk categories: carriers of any missense variant (Carriers)

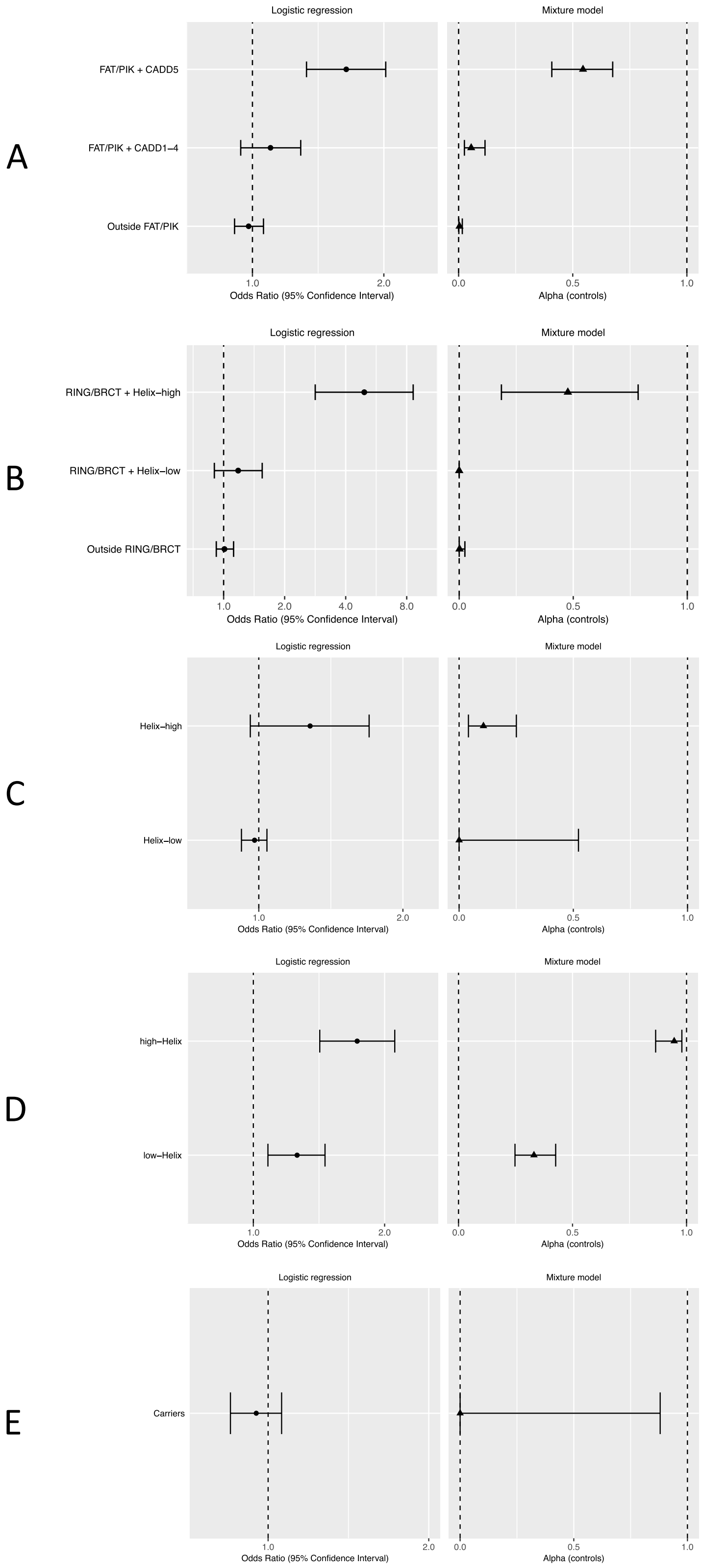

Fig. 1 (See legend on previous page.)

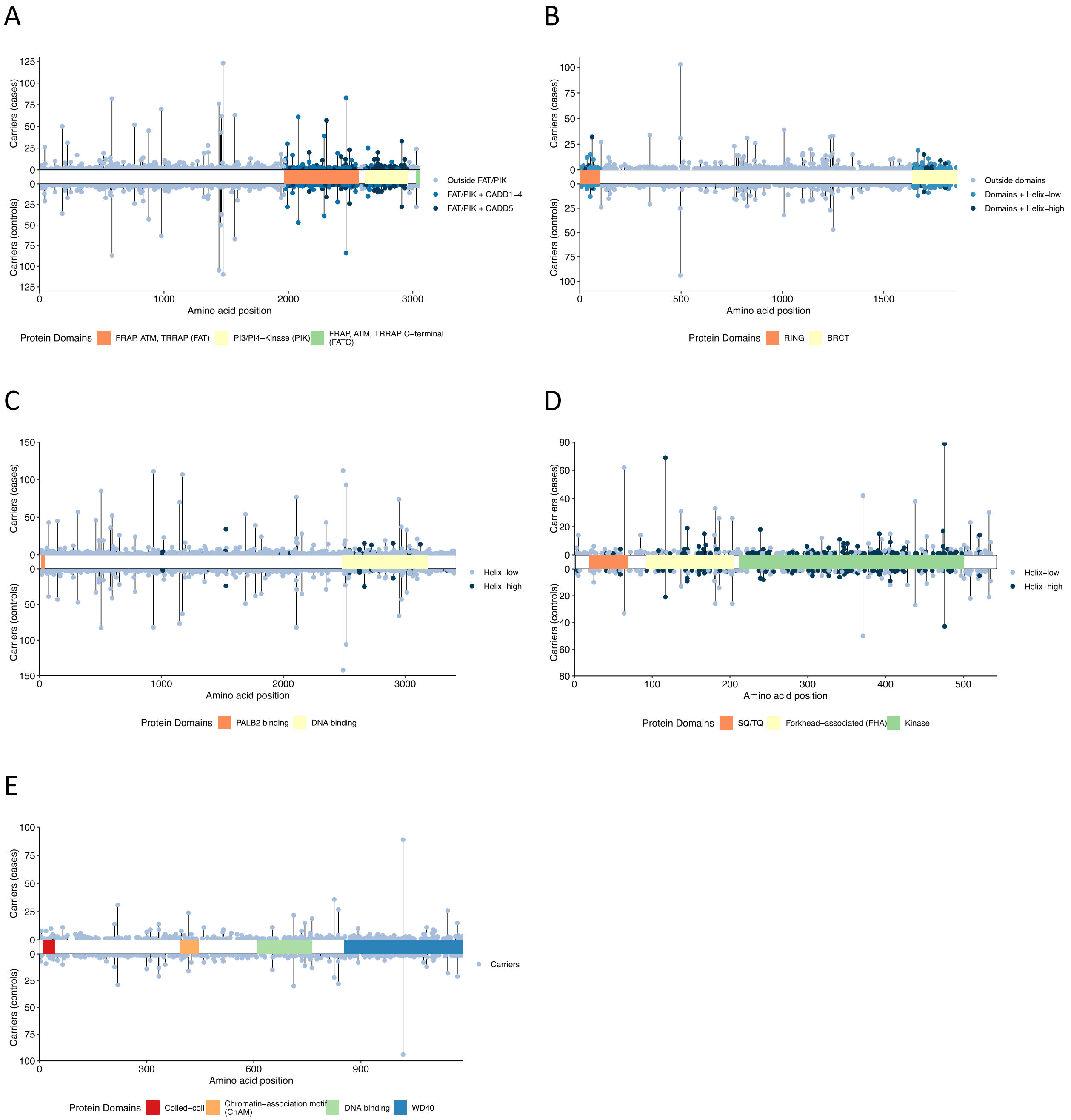

Fig. 2 Case and control carriers across all samples for each observed missense variant by gene. **A** *ATM*. *ATM* risk categories: variants lying within the FAT or PI3K/PI4K protein domains with CADD score in fifth quintile (FAT/PIK + CADD5); variants lying within the FAT or PI3K/PI4K protein domains with CADD score in any of first four quintiles (FAT/PIK + CADD1-4); variants lying outside the FAT and PI3K/PI4K protein domains (Outside FAT/PIK). **B** *BRCA1*. *BRCA1* risk categories: variants lying within the RING or BRCT domains with a high Helix score (RING/BRCT + Helix-high); variants lying with the RING or BRCT domains with a low Helix score (RING/BRCT + Helix-low); variants lying outside the RING and BRCT domains (Outside RING/BRCT). **C** *BRCA2*. *BRCA2* risk categories: variants with a high Helix score (Helix-high); variants with a low Helix score (Helix-low). **D** *CHEK2*. *CHEK2* risk categories: variants with a high Helix score (Helix-high); variants with a low Helix score (Helix-low). **E** *PALB2*. *PALB2* risk categories: carriers of any missense variant (Carriers)

variants defined as (likely) pathogenic were excluded (OR 1.60 (1.29–1.99)). Two of the variants classified as (likely) pathogenic were observed in controls only (Additional File 1: Table S7). One of these (c.8546G > C) is located in the PIK domain, the other (c.3848 T > C) is not within any domain; however, both have a Q5 CADD score.

The pathogenic variants listed on ClinVar include c.7271 T > G (p.Val2424Gly), previously reported as associated with high risk of breast cancer [29, 30]. In the training dataset, c.7271 T > G was identified in 12 cases (6 population-based) and 6 controls and was not associated with risk ($p = 0.37$, $p_{\mathrm{ALL}} = 0.081$); its population-based OR estimate of 1.63 (0.56–4.73) was lower than previous estimates (for example [31]). Another variant previously reported as risk associated, c.6919C > T (p.Leu2307Phe) [32], was associated with an increased population risk (OR = 3.71 (1.87–7.38), $p = 0.00018$). Both variants are located in the FAT domain and have a CADD score in Q5, but after excluding them from the model, there remained a significantly increased risk for carriers in the FAT/PIK + CADD5 risk group (OR 1.48 (1.18–1.85), $p = 0.00064$).

BRCA1

The analysis of *BRCA1* missense variants included 2288 carriers of 644 unique variants. For missense variant carriers, all five continuous in silico scores were associated with risk (Align-GVGD $p_{\mathrm{ALL}} = 1.3 \times 10^{-8}$, BayesDel $p_{\mathrm{ALL}} = 0.0013$, CADD $p_{\mathrm{ALL}} = 0.011$, Helix $p_{\mathrm{ALL}} = 2.1 \times 10^{-9}$, REVEL $p_{\mathrm{ALL}} = 1.5 \times 10^{-5}$). Variants in two protein domains were also significantly associated with risk compared with variants outside these domains (RING finger domain $p_{\mathrm{ALL}} = 3.5 \times 10^{-4}$; BRCA1 C-terminal domains (BRCT I-II) $p_{\mathrm{ALL}} = 0.0030$; see Additional File 2: Table S5). The Helix tool categorizes variants with a high score (> 0.5) as "deleterious" and variants with a low score (< 0.5) as "benign"; hereafter, we refer to these categories as Helix-high and Helix-low, respectively. Including Helix category and protein domain, we found that only variants that were inside the RING or BRCT I-II domains and also in the Helix-high category (RING/BRCT + Helix-high) were associated with risk (OR compared with non-carriers 4.94 (2.83–8.61), $p = 1.9 \times 10^{-8}$; $p_{\mathrm{ALL}} = 2.5 \times 10^{-9}$; Table 1, Figs. 1b and 2b). In a mixture model in which the OR for risk associated missense variants was constrained to that for PTVs (OR 10.61 (7.92–14.21)), the estimated proportions of risk associated variants in the RING/BRCT + Helix-high risk category was 48% (19–78%) and close to 0% for all other variants (Table 1, Figs. 1b and 2b). There was no evidence that the risk associated missense OR differed from the PTV OR ($p = 0.98$). The LR and mixture models were similarly good fits to the data (2 × log-likelihood difference = 0.30).

In a case-only analysis, the OR associated with variants in the RING/BRCT + Helix-high risk category reduced as age increased (per year OR 0.98 (0.96–1.00), $p = 0.036$; see Additional File 1: Table S6).

According to the ENIGMA guidelines and/or ClinVar classifications [13, 14], 13 of the *BRCA1* missense variants in the dataset (four in the RING domain and nine in the BRCT domains) would be classified as (likely) pathogenic (see Additional File 1: Table S7). In total, the 13 variants were carried by 60 cases and 6 controls and were strongly associated with risk in the subset of population samples (OR 16.68 (5.16–53.94), $p = 2.6 \times 10^{-6}$). In our dataset, the most frequent of these variants was c.181 T > G (p.Cys61Gly), carried by 29 cases and 2 controls (OR 15.06 (3.58–63.36)). After excluding all (likely) pathogenic variants, there also remained an increased risk associated with variants in the RING/BRCT + Helix-high category (OR 2.39 (1.19–4.78), $p = 0.014$)).

RF and NVM predictions from the analysis of Hart et al. were available for 577 unique *BRCA1* missense variants. Variants predicted to be damaging by the RF model (OR 1.82 (1.33–2.49), $p = 1.9 \times 10^{-4}$) or the NVM model (OR 2.14 (1.52–3.01), $p = 1.2 \times 10^{-5}$) were associated with increased risk of breast cancer but not as strongly as for variants in the Helix-high category (OR 2.76 (1.93–3.95), $p = 2.6 \times 10^{-8}$; see Additional File 2: Table S5).

BRCA1 Saturation Genome Editing (SGE) score [22] was available for 100 unique variants and was strongly associated with risk ($p_{\mathrm{ALL}} = 1.5 \times 10^{-4}$; see Additional File 2: Table S5). Carriers of variants with an SGE loss of function ($\mathrm{LOF}_{\mathrm{SGE}}$) consequence had a higher risk than carriers of variants with a functional ($\mathrm{FUNC}_{\mathrm{SGE}}$) consequence ($\mathrm{OR}_{\mathrm{ALL}}$ 10.79 (3.31–35.16)). Carriers of variants with an intermediate function ($\mathrm{INT}_{\mathrm{SGE}}$) consequence also had, on average, a higher risk than carriers of $\mathrm{FUNC}_{\mathrm{SGE}}$ variants ($\mathrm{OR}_{\mathrm{ALL}}$ 3.17 (0.32–31.15)) though the number of $\mathrm{INT}_{\mathrm{SGE}}$ carriers was small (total $n = 6$). Since the *BRCA1* SGE experiment specifically targeted the domain-coding regions of the gene, only four variants outside of the domains were scored. Thus, all *BRCA1* missense variants were assigned to one of four potential risk levels, with SGE score prioritized where available: $\mathrm{INT}_{\mathrm{SGE}}$/$\mathrm{LOF}_{\mathrm{SGE}}$; RING/BRCT + Helix-high (SGE score missing); RING/BRCT + Helix-low (SGE score missing); or $\mathrm{FUNC}_{\mathrm{SGE}}$ or carriers of variants outside of the domains. Compared with non-carriers, there was increased risk for carriers of variants in the $\mathrm{INT}_{\mathrm{SGE}}$/$\mathrm{LOF}_{\mathrm{SGE}}$ category (OR 7.22 (2.48–21.01), $p = 2.9 \times 10^{-4}$) and in the RING/BRCT + Helix-high category (OR 5.35 (2.48–11.57), $p = 2.0 \times 10^{-5}$; see Additional File 1: Table S8). In a mixture model in which the OR for risk associated missense variants was constrained to that for PTVs (OR 10.69 (7.97–14.33)), the estimated proportions of risk associated variants in the

INT_{SGE} /LOF_{SGE} and the RING/BRCT + Helix-high risk categories were 75% (24–97%) and 51% (6–94%), respectively (Additional File 1: Table S8). The SGE LR model and SGE mixture model were similarly good fits to the data (2 × log-likelihood difference = 0.12) and both were better fits to the data compared to the Helix-only models (LR models 2 × log-likelihood difference = 3.40, mixture models 2 × log-likelihood difference = 3.58).

BRCA2

The analysis of *BRCA2* missense variants included 5467 carriers of 1425 unique variants. Align-GVGD ($p_{ALL} = 0.0072$), BayesDel ($p_{ALL} = 0.059$), CADD ($p_{ALL} = 0.036$), and Helix ($p_{ALL} = 0.0016$) scores were associated with risk for carriers of *BRCA2* missense variants (see Additional File 2: Table S5). Risks did not differ by protein domain ($p_{ALL} = 0.91$). Compared with non-carriers, carriers of Helix-high variants had a modestly increased risk of breast cancer (OR 1.28 (0.96–1.70), $p = 0.087$; $p_{ALL} = 0.020$) whereas carriers of a Helix-low variant had no increased risk (OR 0.98 (0.92–1.04), $p = 0.47$; $p_{ALL} = 0.40$; Table 1, Figs. 1c and 2c). Under a mixture model in which risk associated missense variants conferred the same risk as PTVs (OR 5.87 (4.75–7.24)), an estimated 11% (4–25%) of the Helix-high variants were associated with risk, compared with < 0.1% of Helix-low variants (Table 1, Figs. 1c and 2c). A model that allowed the OR for missense variants to differ from that of PTVs did not converge. The constrained mixture model was a better fit to the data than the logistic regression model (2 × log-likelihood difference = 6.38). There was no association between age-at-diagnosis and risk category (see Additional File 1: Table S6).

Twelve *BRCA2* variants would be classified as (likely) pathogenic according to ENIGMA guidelines or ClinVar (see Additional File 1: Table S7). In aggregate, the relative risk estimate for these variants was similar to that for PTVs (OR 8.91 (2.61–30.42), $p = 4.8 \times 10^{-4}$). Ten of these variants were categorized as Helix-high and two as Helix-low. Two of the variants categorized as (likely) pathogenic and Helix-high were observed in controls only (see Additional File 1: Table S7). After excluding the (likely) pathogenic variants from the LR model, there remained no increased risk associated with variants classified as Helix-high (OR 0.60 (0.27–1.34)).

RF and NVM predictions were available for 1338 and 1339 unique *BRCA2* missense variants, respectively. There was no association with risk for variants predicted to be damaging by either the RF model ($p = 0.16$) or the NVM model ($p = 0.32$; see Additional File 2: Table S5).

BRCA2 HDR assay score was available for 82 unique variants and was strongly associated with risk ($p_{ALL} = 6.7 \times 10^{-4}$; see Additional File 2: Table S5). Carriers of variants with a prediction of likely pathogenic or pathogenic (LP/P) had a higher risk than carriers of variants with a prediction of likely benign or benign (LB/B) (OR_{ALL} 5.57 (2.36–13.17)). Since the BRCA2 HDR experiment specifically targeted the DNA binding domain-coding region of the gene, no variants outside of the domains were scored. Thus, all *BRCA2* missense variants were assigned to one of four potential risk levels, with functional classification prioritized where available: LP/P; Helix-high (no functional classification); Helix-low (no functional classification); or LB/B. Compared with non-carriers, there was increased risk for carriers of variants in the LP/P category only (OR 4.72 (1.88–11.84), $p = 9.3 \times 10^{-4}$); see Additional File 1: Table S9). In a mixture model in which the OR for risk associated missense variants was constrained to that for PTVs (OR 5.86 (4.75–7.24)), the estimated proportions of risk associated variants in LP/P risk category was 43% (11–82%) (see Additional File 1: Table S9). The functional LR model was a slightly better fit to the data than the mixture model (2 × log-likelihood difference = 1.85) but both were better fits to the data compared to the Helix-only models (LR models 2 × log-likelihood difference = 13.88, mixture models 2 × log-likelihood difference = 5.65).

CHEK2

The analysis of *CHEK2* missense variants included 1552 carriers of 325 unique variants. In the carrier-only analysis, BayesDel ($p_{ALL} = 0.0091$), CADD ($p_{ALL} = 0.0073$), Helix ($p_{ALL} = 0.0021$), and REVEL ($p_{ALL} = 0.016$) scores were associated with risk (see Additional File 2: Table S5). Compared with non-carriers, carriers of a Helix-high variant had a larger increased risk (OR 1.73 (1.42–2.11), $p = 4.7 \times 10^{-8}$) than carriers of Helix-low variants, but the latter were also associated with an increased risk (OR 1.26 (1.08–1.46), $p = 0.0025$; see Table 1, Figs. 1d and 2d). There was no significant association with protein domain ($p_{ALL} = 0.98$).

In the mixture model analysis, the constrained model in which risk associated missense variants conferred the same risk as PTVs could be rejected ($p = 0.027$). Under the best fitting model, the OR for missense variants was 1.75 (1.47–2.08), with 95% (86–98%) of Helix-high variants and 33% (25–43%) of Helix-low variants being risk associated (see Table 1, Figs. 1d and 2d). The mixture model was a similar fit to the LR model (2 × log-likelihood difference = 0.52). We also explored mixture models with two levels of risk variant: one with an OR equal to that of PTVs and another conferring a lower risk compared to that of PTVs. The two-level model fitted slightly better in the full training dataset (2 × log-likelihood difference = 1.10) but not in the population-based studies (two-level model converged to the one-level model). The

OR associated with Helix-high variants decreased as age increased (per year OR 0.99 (0.98–1.00), $p = 0.017$; see Additional File 1: Table S6).

Two variants, c.470 T > G (p.Ile157Ser) and c.433C > T (p.Arg145Trp), were listed as (likely) pathogenic on ClinVar; both variants have high Helix scores but the number of carriers in our population-based sample was too small to evaluate their association with risk (see Additional File 1: Table S7). One rare variant, c.349A > G (p.Arg117Gly), was previously identified as risk associated in BCAC samples, as part of the OncoArray genome-wide association study (GWAS) project [31]. In the current dataset, this variant, which is in the Helix-high category, had an OR 2.69 (1.46–4.94). After excluding the BCAC GWAS samples from the current dataset, the OR was 3.40 (1.52–7.61). Excluding c.349A > G from the LR model did not change the overall relative risk associated with the Helix-high category (OR 1.64 (1.33–2.02)).

PALB2

The analysis of *PALB2* missense variants included 1659 carriers of 472 unique variants. We found no overall evidence of risk associated with missense variants in *PALB2* (OR 0.95 (0.85–1.06), $p = 0.34$; $p_{\mathrm{ALL}} = 0.98$). In the carrier-only analysis, CADD was the only score associated with risk ($p_{\mathrm{ALL}} = 0.020$; see Additional File 2: Table S5); however, there was no significant difference in risk between CADD quintiles ($p_{\mathrm{ALL}} = 0.16$). There was no evidence for a difference in risk for carriers of variants inside any protein domain versus those outside ($p_{\mathrm{ALL}} = 0.25$). In a mixture model in which the missense variant risk was constrained to that for PTVs (OR 4.87 (3.50–6.77)), the estimated proportion of risk associated variants was 0.011% (95% CI 0–88%; Table 1, Figs. 1e and 2e). The log-likelihoods for the mixture model and logistic regression model were similar (2 × log-likelihood difference = 1.01).

Three (likely) pathogenic variants were listed on ClinVar but none of these were present in our samples. Another variant, c.104 T > C (p.Leu35Pro), has been suggested to be pathogenic based on evidence from one family and tumor genomic analysis [33], but this variant was also not found in our samples.

A subset of the variants from the functional screening studies were available in the training data set: 26 of the 48 assayed by Boonen et al. [24], 34 of the 84 assayed by Wiltshire et al. [26] and 18 of the 44 assayed by Rodrigue et al. [25]. None of the functional assay scores or the authors' corresponding classifications of pathogenicity were associated with risk in the BRIDGES samples (see Additional File 2: Table S5).

Frequency analysis

In burden analyses of variants with frequencies up to 5%, variants in *ATM* with frequency < 0.1% were associated, in aggregate, with risk ($p = 0.0024$) but no group of variants of greater frequency was associated (see Additional File 1: Table S10). For *CHEK2*, variants with frequency < 0.1% ($p = 1.1 \times 10^{-14}$) and those with frequency 0.1–0.5% ($p = 3.6 \times 10^{-5}$) were associated with risk; there were no variants with frequency 0.5–5%. None of the other genes showed an association between any variant frequency group and risk (see Additional File 1: Table S10).

When analyses were restricted to frequencies up to 0.5%, there was no association between risk and frequency, either on a continuous scale or as the difference in risk between the two frequency groups < 0.1% and 0.1–0.5%, for *BRCA2*, *CHEK2*, or *PALB2* (see Additional File 1: Table S11). For *ATM*, we found frequency inversely associated with risk (continuous $p_{\mathrm{ALL}} = 0.0098$) and a higher risk for variants with frequency < 0.1% compared with variants of frequency 0.1–0.5% ($p_{\mathrm{ALL}} = 0.031$). After adjusting for the CADD and domain risk groups, the associations remained statistically significant ($p_{\mathrm{ALL}} = 0.0097$ and $p_{\mathrm{ALL}} = 0.012$, respectively). For *BRCA1*, we found frequency inversely associated with risk (continuous $p_{\mathrm{ALL}} = 0.022$) and a significantly higher risk for variants with frequency < 0.1% compared with variants of frequency 0.1–0.5% ($p_{\mathrm{ALL}} = 0.0066$). However, after adjusting for the Helix and domain risk groups, neither of these associations remained statistically significant ($p_{\mathrm{ALL}} = 0.36$ and $p_{\mathrm{ALL}} = 0.39$, respectively).

We evaluated the risks for individual missense variants with frequency between 0.1 and 5% (see Additional File 1: Table S12). In *BRCA1*, one variant, c.2521C > T (p.Arg841Trp), was associated with a decreased risk of breast cancer (OR 0.67 (0.52–0.87), $p = 0.0027$). Two previously-reported variants in *CHEK2* were identified: c.470 T > C (p.Ile157Thr) and c.538C > T (p.Arg180Cys) [34]. c.470 T > C was associated with an OR of 1.24 (1.09–1.42), consistent with the estimate for the Helix-low risk category, while c.538C > T was associated with a higher OR 1.44 (1.12–1.84). No *ATM*, *BRCA2*, or *PALB2* missense variants were individually associated with increased risk.

Model validation

We evaluated the calibration of the best fitting models from the training set, for each gene, in the validation set: these included the LR models, the mixture model using the estimated proportions (α) from the training set, and the mixture model using the posterior probabilities derived from the training set. For each gene and each model, carriers of variants in the predicted risk groups

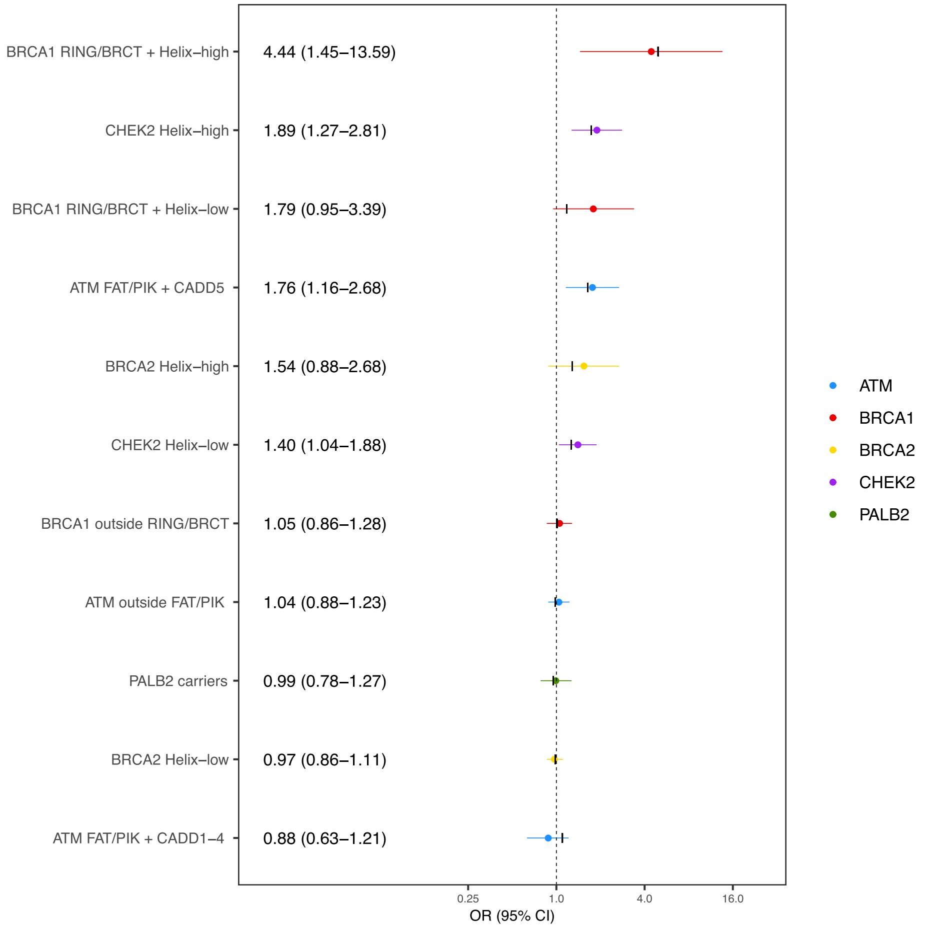

Fig. 3 Breast cancer risk estimates from composite gene model in validation samples. Black marks indicate corresponding ORs from training models. Risk categories: ATM FAT/PIK + CADD5: *ATM* variants lying within the FAT or PI3K/PI4K protein domains with CADD score in fifth quintile; ATM FAT/PIK + CADD1-4: *ATM* variants lying within the FAT or PI3K/PI4K protein domains with CADD score in any of first four quintiles; ATM outside FAT/PIK: variants lying outside the FAT and PI3K/PI4K protein domains; BRCA1 RING/BRCT + Helix-high: *BRCA1* variants lying within the RING or BRCT domains with a high Helix score; BRCA1 RING/BRCT + Helix-low: *BRCA1* variants lying with the RING or BRCT domains with a low Helix score; BRCA1 outside RING/BRCT: *BRCA1* variants lying outside the RING and BRCT domains; BRCA2 Helix-high: *BRCA2* variants with a high Helix score; BRCA2 Helix-low: *BRCA2* variants with a low Helix score; CHEK2 Helix-high: *CHEK2* variants with a high Helix score; CHEK2 Helix-low: *CHEK2* variants with a low Helix score; PALB2 carriers: carriers of any missense variant in *PALB2*

were associated with an increased risk, and there were no differences between the observed and predicted ORs (see Additional File 1: Table S13 and Figs. S2-S6). In silico scores, likelihood ratios and posterior probabilities for every variant included in the population training dataset are given in Additional File 2: Tables S14-18.

Using a composite five gene model, we estimated ORs for eleven risk categories (Fig. 3). In total, 184 samples carried a missense variant in more than one of the five genes and were excluded from this analysis. Four categories were significantly associated with an increased risk relative to non-carriers, consistent with the estimates derived from the training set: *ATM* FAT/PIK + CADD5 (OR 1.76 (1.16–2.68), $p = 0.0078$), *CHEK2* Helix-low (OR 1.40 (1.04–1.88), $p = 0.025$), *CHEK2* Helix-high (OR = 1.89 (1.27–2.81), $p = 0.0017$), and *BRCA1* within domain and Helix-high (OR 4.44 (1.45–13.59), $p = 0.0089$) risk groups. The OR estimate for *BRCA2* Helix-high variant carriers was higher than that in the training dataset, but the confidence interval was considerably wider (OR 1.54 (0.88–2.68)). As predicted, variants in the remaining categories were not associated with risk.

Discussion

To date, the risks associated with missense variants in breast cancer predisposition genes have been largely unclear. In this study of over 112,000 women, we were able to use a range of in silico scores produced by statistical algorithms and knowledge of functional protein domains to determine the risks associated with subsets of rare missense variants. We identified groups of missense variants conferring increased risks of breast cancer in *ATM*, *BRCA1*, *BRCA2*, and *CHEK2*, but not in *PALB2*. The ORs for *BRCA1* and *CHEK2* decreased with age at diagnosis, consistent with previous observations for PTVs [7]. Previous analysis of the full BRIDGES dataset showed that protein domains in *ATM* and *BRCA1* were predictive of risk [7]; the analysis presented here showed that in silico scores improved these predictions, in a formal model evaluation that allowed the models to be tested in an independent validation set. Under the best fitting mixture models, for *ATM*, *BRCA1*, and *BRCA2*, a small proportion of rare missense variants were associated with risks comparable to those for PTVs. In contrast, for *CHEK2*, a high proportion of *CHEK2* missense variants were risk associated and the estimated risk was markedly lower than that associated with PTVs. In *PALB2*, the evidence for association was weak; the mixture model analysis indicated that the proportion of missense variants associated with a high risk is likely to be very small. However, we cannot rule out the possibility that some variants are risk-associated since the power for detecting an association with risk for *PALB2* is lower than, for example, *BRCA1* and *BRCA2*. One variant in *BRCA1* (p.Arg841Trp) was individually associated with a reduced risk of breast cancer (0.67 (0.52–0.87)). Given that this finding is inconsistent with all the other associations and that the variant is not in any of the key functional domains, it seems quite likely that this is a chance association; further replication in other datasets will be required to confirm or refute the association.

We used five in silico scores to predict the pathogenicity of individual variants. Helix, BayesDel, and CADD were all predictive for the four genes for which we were able to identify subsets of risk-associated variants; Helix was most predictive for *BRCA1*, *BRCA2*, and *CHEK2* while CADD outperformed all the other scores for *ATM*. In addition to the in silico scores, we also tested the BRCA1 SGE functional assay score. We found that the SGE score slightly improved the performance of the model for predicting risk for *BRCA1* missense variant carriers, compared with the Helix-only model. Consistent with this, we observed two variants that were classified as loss-of-function variants by SGE but appeared in our low-risk group; these were present in three cases and no controls. Conversely, another four variants that were classified as normal function by SGE but appeared in our high-risk group were present in eight cases and five controls. Overall, variants categorized by SGE as disruptive to function, or lying within a protein domain and scored high by Helix, were strongly associated with increased risk. Under the mixture model, the proportions of risk-associated variants were also high, although the confidence intervals for the proportion of associated variants were wide. It is notable that 11 of the 31 variants in these categories have previously been identified as (likely) pathogenic by ClinVar and/or ENIGMA.

Similarly, for *BRCA2*, we also tested the HDR functional assay score and found it improved the performance of the model for predicting risk for *BRCA2* variant carriers, compared to the Helix-only model. Consistent with this, four variants in the Helix-high category were classified as benign by the functional study and observed in 22 cases and 35 controls. Conversely, one variant in the Helix-low category was classified as pathogenic by the functional study and observed in two cases and no controls. After accounting for variants predicted to be pathogenic by the functional assay, there remained no significant increase in risk for carriers of variants in the Helix-high variant category, compared to non-carriers, although the OR of 1.37 for the Helix-high category was higher than the ORs of 0.97 and 0.96 for the Helix-low and predicted benign categories, respectively. We note that the variants tested using the HDR assay were subsequently classified using a combination of the assay result and American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines; ACMG guidelines also used by the ENIGMA *BRCA1/2* expert panel and in evidence for pathogenicity in ClinVar. Consequently, there is considerable overlap between classifications; nine of the 14 variants classified as (likely) pathogenic by the functional study have been previously identified as (likely) pathogenic by ClinVar and/or ENIGMA.

The BRCA2 HDR functional assay included only variants lying in the DNA binding domain of *BRCA2*. The majority of high-Helix variants were also in the DNA binding domain (37/62) and fewer [21] in the "coldspot" regions of exons 10 and 11 as described by Dines et al. [35] (by definition, none of the *BRCA1* variants in the high-risk category fall in the corresponding exon 11 coldspot).

In *ATM*, the risk conferred by missense variants was confined to specific protein-coding domains, namely the FAT and PIK domains, consistent with previous studies [5] and as shown previously in BRIDGES [7]. Variants within these domains could be further distinguished using the CADD score; variants in the top quintile were associated with risk whereas variants in the first four quintiles were not. In a mixture model, 54% of variants in the top CADD quintile were estimated to be associated with risk. One variant in this group, c.7271 T > G (p.Val2424Gly), has been previously reported as a breast cancer risk variant but the OR estimate for this variant, 1.63 (0.56–4.73), was markedly lower than previously estimated (relative risks ranging from 8.0 to 12.7) [29–31]. The reasons for this difference are unclear but might be due, in part, to previous studies oversampling for cases with a family history of breast cancer.

The results for *CHEK2* were in marked contrast to those for *BRCA1*, *BRCA2*, and *ATM*. In the best fitting mixture model, the proportion of associated variants was high, and the estimated risk was clearly lower than for PTVs. A model in which there were two levels of risk, with the higher level equal to the PTV risk, fitted slightly better in the full training dataset but not in the population-based training studies. In addition, however, three individual *CHEK2* variants were associated with differing levels of risk: c.470 T > C (p.Ile157Thr) OR 1.24 (1.09–1.42); c.538C > T (p.Arg180Cys) OR 1.44 (1.12–1.84); and c.349A > G (p.Arg117Gly) OR 2.69 (1.46–4.94). The c.470 T > C variant was too common to be included in the main analyses, possibly explaining why the heterogeneity in risk was not readily detectable by the mixture models; however, the confidence interval for c.470 T > C from the individual-level analysis did not include the LR and mixture model OR estimates of 1.73 and 1.75, respectively, for the risk-associated variants. Taken together, these observations suggest that there is substantial variation in risk associated with *CHEK2* missense variants.

The relative performances of the in silico prediction algorithms are perhaps less marked than might appear; for example, Helix, which was the most predictive algorithm for three of the genes, was also predictive for *ATM*. Some of the differences in the associations may be due to chance. Align-GVGD was initially developed for *BRCA1/2* so it is perhaps not surprising that the algorithm does relatively well for *BRCA1* but less well for *CHEK2*, for example. Helix was not developed for a specific gene so may be a more useful tool in general.

We controlled for the potential effects of population stratification by stratifying analyses by country and by excluding individuals with the minority ancestry for that country. Thus, European studies excluded individuals of non-European ancestry and Asian studies excluded individuals of non-Asian ancestry. In addition, for the studies in Malaysia and Singapore, we further stratified into the three ethnic groups (Chinese, Malay, Indian). In previous analysis of PTVs, we found no differences in effect sizes when additionally correcting for ancestry informative principal components, suggesting that this correction was adequate, particularly since most of the associations were based on many variants [7]. Nevertheless, it remains possible that some estimates may be biased due to residual population stratification [36, 37].

Under the best fitting mixture model, approximately 7% of all rare missense variants in *ATM* were associated with similar risk to that of PTVs. The estimated carrier frequency of pathogenic missense variants in *ATM* was 0.0030, or approximately 89% of the PTV frequency. The corresponding proportion of associated rare missense variants for *BRCA1* and *BRCA2* was 2% and 0.6%, with an estimated carrier frequency of 0.00026 (~ 18%) and 0.00028 (~ 9%), respectively. Thus, missense variants add modestly to the contribution of *BRCA1* and *BRCA2* variants to breast cancer incidence, but make a relatively more substantial contribution for *ATM*. The differences between genes in the relative contributions of missense variants to risk presumably reflect the relative proportion of residues within functional domains in which disrupted function is associated with cancer risk, and the size of those domains. For *CHEK2*, approximately 60% of rare missense variants were risk associated and the estimated carrier frequency of pathogenic missense variants in *CHEK2* was comparable to the frequency of PTVs. The predicted proportion of breast cancer cases possessing pathogenic germline missense variants in these genes is approximately 0.6%, 0.3%, 0.2%, and 1.3% for *ATM*, *BRCA1*, *BRCA2*, and *CHEK2*, respectively. The estimated additional contribution to the familial relative risk of breast cancer made by pathogenic missense variants in these five genes is approximately 2.7%.

The task of identifying which specific individual missense variants are risk associated is a complex one and is difficult to resolve fully even with a large dataset, since most variants are rare and there are many possible models to consider. Despite the size of our study, it was difficult to distinguish, for any gene, between the LR models (in which all variants in a given category confer a given risk) and the mixture models (in which all risk-associated variants confer the same risk, but the proportion that are associated varies by category). This difficulty arises because the number of carriers for individual variants is small, and as a result, the estimated risk of pathogenic missense variants and probability of pathogenicity (α) are strongly confounded. Further, selecting the best models and estimating the risks based on these models is likely to result in overfitting and biased risk estimates. In order to strengthen the validity of our findings, we used a training-validation study design. We were able to replicate the predicted OR estimates in the validation dataset, suggesting that any bias due to overfitting was small. Nevertheless, the validation dataset was relatively small, so further validation of the best models reported here in large independent datasets is critical.

Ultimately, high-throughput functional assays that can evaluate all possible missense substitutions may provide more precise definitions of risk categories. The analyses of the BRCA1 SGE scores and the BRCA2 HDR assay scores suggest that this approach should be useful, although the scores for *BRCA1* were highly concordant with the best in silico score in this case. The available *PALB2* functional assays did not predict risk, but this may just reflect the low power of these analyses when the proportion of risk-associated variants is very low. As further prediction algorithms based on in silico and/or in vitro data are developed, large population-based epidemiological datasets such as BRIDGES can be used to validate their predictions. However, further large studies are likely to be required to provide more precise variant-specific risk estimates.

Conclusions

This study confirms that subsets of missense variants in established breast cancer susceptibility genes are associated with increased risks of the disease and provides estimates of relative risks for those subsets, as well as probabilities for association with risk at the variant level. The pattern of risk varies substantially by gene. Accurately and precisely defining these risks is critical to the counselling and management of women in whom these variants are identified.

Abbreviations

PTVs: Protein truncating variants; VUS: Variants of uncertain significance; BCAC : Breast Cancer Association Consortium; VEP: Ensembl Variant Effect Predictor; CADD: Combined Annotation-Dependent Depletion; REVEL: Rare Exome Variant Ensembl Learner; SGE: Saturation Genome Editing; LR: Logistic regression; OR: Odds ratio; EM: Expectation-maximization; PP: Posterior probability; Q5: Fifth quintile; LOF_{SGE}: Loss of function; $FUNC_{SGE}$: Functional; INT_{SGE}: Intermediate function; GWAS: Genome-wide association study.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-022-01052-8.

Additional file 1.

Additional file 2.

Acknowledgements

NBCS Collaborators, kConFab Investigators, and SGBCC Investigators are listed in Additional File 1: Additional Note.

Authors' contributions

DFE, PD, and SHT conceived the study and obtained funding. LD, SC, and JMA performed the statistical and bioinformatics analysis. MTP and CF performed variant annotation analysis. SMH and BV developed the Helix algorithm. A G-N, JB, AMD, CL, and AK led the laboratory analysis. LD and DFE drafted the manuscript. ABS, MPGV, PD, SHT, MKS, MdlH, and AMD revised the manuscript. All other authors generated study-specific data. All authors read and approved the final manuscript.

Funding

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]. BCAC co-ordination was additionally funded by the European Union's Horizon 2020 Research and Innovation Programme (BRIDGES: grant number 634935, BCAST: grant number 633784) and by Cancer Research UK [C1287/A16563]. Study specific funding is given in the Additional Note.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are not publicly available due to constraints by the ethics committees of individual studies. The datasets are available via the BCAC Data Access Co-ordinating Committee (bcac@medschl.cam.ac.uk), upon reasonable request. Summary-level genotype data are available via http://bcac.ccge.medschl.cam.ac.uk and in Additional File 2: Tables S14-18. Individual-level data are available via the BCAC Data Access Co-ordinating Committee (bcac@medschl.cam.ac.uk).

Declarations

Ethics approval and consent to participate

All contributing studies were approved by the relevant ethical review boards and used appropriate consent procedures. Details of the ethical review boards for each study are given in Additional File 2: Table S19. The research conformed to the principles of the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

BV and SMH are employees and shareholders of Bio-Prodict, Nijmegen, The Netherlands. The remaining authors declare that they have no competing interests.

Author details

[1]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [2]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia. [3]Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain. [4]Bio-Prodict, Nijmegen, The Netherlands. [5]Family Cancer Clinic, The Netherlands Cancer Institute - Antoni Van Leeuwenhoek Hospital, Amsterdam 1066 CX, The Netherlands. [6]Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD 20850, USA. [7]Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada. [8]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada. [9]Translational Cancer Research Area, University of Eastern Finland, 70210 Kuopio, Finland. [10]Institute of Clinical Medicine, Oncology, University of Eastern Finland, 70210 Kuopio, Finland. [11]Department of Oncology, Cancer Center, Kuopio University Hospital, 70210 Kuopio, Finland. [12]Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. [13]Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg (FAU), 91054 Erlangen, Germany. [14]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [15]Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa 450054, Russia. [16]Department of Radiation Oncology, Hannover Medical School, 30625 Hannover, Germany. [17]Gynaecology Research Unit, Hannover Medical School, 30625 Hannover, Germany. [18]N.N. Alexandrov Research Institute of Oncology and Medical Radiology, 223040 Minsk, Belarus. [19]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, 2730 Herlev, Denmark. [20]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, 2730 Herlev, Denmark. [21]Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark. [22]Medical Faculty, Universidad de La Sabana, 140013 Bogota, Colombia. [23]Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA. [24]Centre for Genomic and Experimental Medicine, Institute of Genetics & Cancer, The University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. [25]Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh EH16 4UX, UK. [26]Oncology and Genetics Unit,

Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, 36312 Vigo, Spain. [27]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. [28]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, 0379 Oslo, Norway. [29]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, 0450 Oslo, Norway. [30]Department of Research, Vestre Viken Hospital, 3019 Drammen, Norway. [31]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, 0379 Oslo, Norway. [32]Department of Oncology, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Radiumhospitalet, 0379 Oslo, Norway. [33]Department of Oncology, Akershus University Hospital, 1478 Lørenskog, Norway. [34]Breast Cancer Research Consortium, Oslo University Hospital, 0379 Oslo, Norway. [35]Department of Medical Genetics, Oslo University Hospital and University of Oslo, 0379 Oslo, Norway. [36]Department of Community Medicine, The Arctic University of Norway, 9037 Tromsø, Norway. [37]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam 3015 CN, The Netherlands. [38]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 65 Stockholm, Sweden. [39]Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester M13 9WL, UK. [40]North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK. [41]Nightingale & Genesis Prevention Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester M23 9LT, UK. [42]NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 9WL, UK. [43]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA 90095, USA. [44]Cancer Research UK Edinburgh Centre, The University of Edinburgh, Edinburgh EH4 2XR, UK. [45]Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, 2730 Herlev, Denmark. [46]Fundación Pública Galega de Medicina Xenómica, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, , 15706 Santiago de Compostela, Spain. [47]Moores Cancer Center, University of California San Diego, La Jolla, CA 92037, USA. [48]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC 3004, Australia. [49]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC 3010, Australia. [50]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC 3168, Australia. [51]Team "Exposome and Heredity", CESP, Inserm, Gustave Roussy, University Paris-Saclay, UVSQ, Villejuif, France. [52]Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [53]Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, 69120 Heidelberg, Germany. [54]Institute of Diabetes Research, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany. [55]Cancer Genetics, Therapeutics and Ultrastructural Pathology, The Cyprus Institute of Neurology & Genetics, 2371 Nicosia, Cyprus. [56]Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, 2371 Nicosia, Cyprus. [57]Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, 50937 Cologne, Germany. [58]Center for Integrated Oncology (CIO), Faculty of Medicine, University Hospital Cologne, University of Cologne, 50937 Cologne, Germany. [59]Department of Oncology, 118 83 Södersjukhuset, Stockholm, Sweden. [60]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [61]Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester M13 9PT, UK. [62]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore 117549, Singapore. [63]Department of Surgery, National University Health System, Singapore 119228, Singapore. [64]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore 119228, Singapore. [65]Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam 3015 GD, The Netherlands. [66]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, 70376 Stuttgart, Germany. [67]University of Tübingen, 72074 Tübingen, Germany. [68]Division of Cancer Sciences, University of Manchester, Manchester M13 9PL, UK. [69]Research Department, Peter MacCallum Cancer Center, Melbourne, VIC 3000, Australia. [70]Department of Oncology, Sir Peter MacCallum, The University of Melbourne, Melbourne, VIC 3000, Australia. [71]Human Genetics Division, Genome Institute of Singapore, Singapore 138672, Singapore. [72]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore 119077, Singapore. [73]Cancer Genetics Service, National Cancer Centre, Singapore 169610, Singapore. [74]Breast Department, KK Women's and Children's Hospital, Singapore 229899, Singapore. [75]SingHealth Duke-NUS Breast Centre, Singapore 168753, Singapore. [76]Department of General Surgery, Tan Tock Seng Hospital, Singapore 308433, Singapore. [77]Division of Surgical Oncology, National Cancer Centre, Singapore 169610, Singapore. [78]Department of General Surgery, Singapore General Hospital, Singapore 169608, Singapore. [79]Division of Breast Surgery, Department of General Surgery, Changi General Hospital, Singapore 529889, Singapore. [80]Division of Radiation Oncology, National Cancer Centre, Singapore 169610, Singapore. [81]Division of Medical Oncology, National Cancer Centre, Singapore 169610, Singapore. [82]Department of Genetics and Pathology, Pomeranian Medical University, 71-252 Szczecin, Poland. [83]Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, 71-252 Szczecin, Poland. [84]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa 450000, Russia. [85]Department of Surgery, Daerim Saint Mary's Hospital, Seoul 07442, Korea. [86]Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, 53113 Bonn, Germany. [87]Department of Human Genetics, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. [88]Department of Clinical Genetics, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. [89]Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 76 Stockholm, Sweden. [90]Department of Clinical Genetics, Karolinska University Hospital, 171 76 Stockholm, Sweden. [91]Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M13 9PL, UK. [92]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. [93]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, 70210 Kuopio, Finland. [94]Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland. [95]Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, 118 83 Stockholm, Sweden. [96]Department of Medical Oncology, University Hospital of Heraklion, 711 10 Heraklion, Greece. [97]Breast Cancer Research Unit, Faculty of Medicine, University Malaya Cancer Research Institute, University of Malaya, 50603 Kuala Lumpur, Malaysia. [98]Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia. [99]Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, 00290 Helsinki, Finland. [100]Department of Genetics, University Medical Center Groningen, University Groningen, Groningen 9713 GZ, The Netherlands. [101]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul 03080, Korea. [102]Convergence Graduate Program in Innovative Medical Science, Seoul National University College of Medicine, Seoul 03080, South Korea. [103]Cancer Research Institute, Seoul National University, Seoul 03080, Korea. [104]Genome Diagnostics Program, IFOM - the FIRC Institute of Molecular Oncology, 20139 Milan, Italy. [105]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale Dei Tumori (INT), 20133 Milan, Italy. [106]Department of Oncology, University Hospital of Larissa, 411 10 Larissa, Greece. [107]School of Cancer & Pharmaceutical Sciences, Comprehensive Cancer Centre, Guy's Campus, King's College London, London, UK. [108]Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, 50931 Cologne, Germany. [109]Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia. [110]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B15 2TT, UK. [111]Wellcome Trust Centre for Human Genetics and Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford OX3 7BN, UK. [112]Institute of Human Genetics, Pontificia Universidad Javeriana, 110231 Bogota, Colombia. [113]Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni Van Leeuwenhoek Hospital, Amsterdam 1066 CX, The Netherlands. [114]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni Van Leeuwenhoek Hospital, Amsterdam 1066 CX, The Netherlands. [115]Biomedical Network On Rare Diseases (CIBERER), 28029 Madrid, Spain. [116]Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Medical Center, Nijmegen, The Netherlands. [117]Breast Cancer Research Programme, Cancer Research Malaysia, Subang Jaya, 47500 Selangor, Malaysia. [118]Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, 22381 Lund, Sweden. [119]Molecular Oncology Laboratory, Hospital Clinico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), 28040 Madrid, Spain. [120]Department of Pathology, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands.

Received: 20 August 2021 Accepted: 4 May 2022
Published online: 18 May 2022

References

1. Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372(23):2243–57.
2. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet. 2004;75(4):535–44.
3. Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer–predisposition genes. Am J Human Gene. 2007;81(5):873–83.
4. Tavtigian SV, Chenevix-Trench G. Growing recognition of the role for rare missense substitutions in breast cancer susceptibility. Biomark Med. 2014;8(4):589–603.
5. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, et al. Rare, Evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. Am J Human Gene. 2009;85(4):427–46.
6. Breast Cancer Association Consortium [Available from: http://bcac.ccge.medschl.cam.ac.uk.
7. Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, et al. Breast cancer risk genes-association analysis in more than 113,000 women. N Engl J Med. 2021;384:428–39.
8. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction modelincorporating genetic and nongenetic risk factors. Genet Med. 2019;21(8):1708–18.
9. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.
10. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2–3):377–94.
11. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176(3):535-48.e24.
12. UniProt [Available from: https://www.uniprot.org/.
13. ENIGMA: Evidence-based Network for the Interpretation of Germline Mutant Alleles [Available from: https://enigmaconsortium.org/.
14. ClinVar [Available from: https://www.ncbi.nlm.nih.gov/clinvar/.
15. Spurdle AB, Greville-Heygate S, Antoniou AC, Brown M, Burke L, De La Hoya M, et al. Towards controlled terminology for reporting germline cancer susceptibility variants: an ENIGMA report. J Med Genet. 2019;56(6):347–57.
16. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Hum Mutat. 2008;29(11):1342–54.
17. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310–5.
18. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Ame J Human Gene. 2016;99(4):877–85.
19. Feng BJ. PERCH: a unified framework for disease gene prioritization. Hum Mutat. 2017;38(3):243–51.
20. Vroling B, Heijl S. White paper: the Helix Pathogenicity Prediction Platform. arXiv:210401033 [preprint]. 2021. Available from: https://doi.org/10.48550/arXiv.2104.01033.
21. Hart SN, Hoskin T, Shimelis H, Moore RM, Feng B, Thomas A, et al. Comprehensive annotation of BRCA1 and BRCA2 missense variants by functionally validated sequence-based computational prediction models. Genet Med. 2019;21(1):71–80.
22. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018;562(7726):217–22.
23. Richardson ME, Hu C, Lee KY, LaDuca H, Fulk K, Durda KM, et al. Strong functional data for pathogenicity or neutrality classify BRCA2 DNA-binding-domain variants of uncertain significance. Am J Human Gene. 2021;108(3):458–68.
24. Boonen RA, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M, et al. Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat Commun. 2019;10(1):1–15.
25. Rodrigue A, Margaillan G, Torres Gomes T, Coulombe Y, Montalban G, da Costa e Silva Carvalho S, et al. A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. Nucleic Acids Res. 2019;47(20):10662–77.
26. Wiltshire T, Ducy M, Foo TK, Hu C, Lee KY, Nagaraj AB, et al. Functional characterization of 84 PALB2 variants of uncertain significance. Genet Med. 2020;22(3):622–32.
27. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat. 2008;29(11):1282–91.
28. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc: Ser B (Methodol). 1977;39(1):1–22.
29. Goldgar DE, Healey S, Dowty JG, Da Silva L, Chen X, Spurdle AB, et al. Rare variants in the ATMgene and risk of breast cancer. Breast Cancer Res. 2011;13(4):R73.
30. Stankovic T, Kidd AMJ, Sutcliffe A, McGuire GM, Robinson P, Weber P, et al. ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. Am J Human Gene. 1998;62(2):334–45.
31. Southey MC, Goldgar DE, Winqvist R, Pylkäs K, Couch F, Tischkowitz M, et al. PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet. 2016;53(12):800.
32. Mangone FR, Miracca EC, Feilotter HE, Mulligan LM, Nagai MA. ATM gene mutations in sporadic breast cancer patients from Brazil. Springerplus. 2015;4(1):23.
33. Foo TK, Tischkowitz M, Simhadri S, Boshari T, Zayed N, Burke KA, et al. Compromised BRCA1-PALB2 interaction is associated with breast cancer risk. Oncogene. 2017;36(29):4161–70.
34. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallée M, Voegele C, Babikyan D, et al. Rare, evolutionarily unlikely missense substitutions in CHEK2contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. Breast Cancer Res. 2011;13(1):R6.
35. Dines JN, Shirts BH, Slavin TP, Walsh T, King M-C, Fowler DM, et al. Systematic misclassification of missense variants in BRCA1 and BRCA2 "coldspots." Genet Med. 2020;22(5):825–30.
36. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature. 2019;570(7759):71–6.
37 Feng Y-CA, Howrigan DP, Abbott LE, Tashman K, Cerrato F, Singh T, et al. Ultra-rare genetic variation in the epilepsies: a whole-exome sequencing study of 17,606 individuals. Am J Human Gene. 2019;105(2):267–82.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

Additional Note: Breast Cancer Association Consortium funding and acknowledgements, and additional references.

Funding

A.B Spurdle, M.T. Parsons and C. Fortuno were supported by funding from the Australian National Medical Research Council (IDs 1177524, 1101400, 1161589). The **ABCS** study was 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, the Allan Turing Institute under the EPSRC grant EP/N510129/1. 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 and FEDER PI17/00918/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. **FHRISK** and **PROCAS** are funded from NIHR grant PGfAR 0707-10031. DGE, AH and WGN are supported by the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007). 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) and the Federal Ministry of Education and Research, Germany (grant no 01GY1901). 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, Johanniter GmbH Bonn, 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 Sigrid Juselius Foundation and The Cancer Foundation Finland. The **HEBON** study is supported by the Dutch Cancer Society grants NKI1998-1854, NKI2004-3088, NKI2007-3756, the Netherlands Organisation of Scientific Research grant NWO 91109024, the Pink Ribbon grants 110005 and 2014-187.WO76, the BBMRI grant NWO 184.021.007/CP46 and the Transcan grant JTC 2012 Cancer 12-054. 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), B.M. was supported by grant 17-44-020498, 17-29-06014 of the Russian Foundation for Basic Research, D.P. was supported by grant 18-29-09129 of the Russian Foundation for Basic Research, E.K was supported by the program for support the bioresource collections №007-030164/2 and by the megagrant from the Government of Russian Federation No. 075-15-2021-595, and the study was performed as part of the assignment of the Ministry of Science and Higher Education of the Russian Federation (№AAAA-A16-116020350032-1). 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 (VTR) 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. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. 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]. The **MASTOS** study was supported by “Cyprus Research Promotion Foundation” grants 0104/13 and 0104/17, and the Cyprus Institute of Neurology and Genetics. **MBCSG** is supported by grants from the Italian Association for Cancer Research (AIRC). 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 Wellcome Trust (v203477/Z/16/Z), the Malaysian Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Malaysia. 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 grant U01CA164920 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), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. 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. 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 (NRF-NRFF2017-02), NUS start-up Grant, National University Cancer Institute Singapore (NCIS) Centre Grant, Breast Cancer Prevention Programme, Asian Breast Cancer Research Fund and the NMRC Clinician Scientist Award (SI Category). Population-based controls were from the Multi-Ethnic Cohort (MEC) funded by grants from the Ministry of Health, Singapore, National University of

Singapore and National University Health System, Singapore. **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. **UBCS** was supported by funding from National Cancer Institute (NCI) grant R01 CA163353 (to N.J. Camp) and the Women's Cancer Center at the Huntsman Cancer Institute (HCI). Data collection for UBCS was supported by the Utah Population Database (UPDB) and Utah Cancer Registry (UCR). The UPDB is supported by HCI (including the Huntsman Cancer Foundation), University of Utah program in Personalized Health and Center for Clinical and Translational Science, and NCI grant P30 CA2014. The UCR is funded by the NCI's SEER Program, Contract No. HHSN261201800016I, 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.

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 former Department Director-General of Disease Control who have supported the study throughout. **BBCS** thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. **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. The **BSUCH** study acknowledges the Principal Investigator, Barbara Burwinkel, and, 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** and **PROCAS** thank NIHR for funding. The **GENICA** Network: Dr. Margarete Fischer-Bosch-

Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [Hiltrud Brauch, Wing-Yee Lo, Reiner Hoppe], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Tübingen, 72074 Tübingen, Germany [Hiltrud Brauch], gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - EXC 2180 - 390900677 [Hiltrud Brauch], Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany [Yon-Dschun Ko, 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 [Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. **HABCS** thanks Michael Bremer. **HEBCS** thanks Carl Blomqvist, Taru A. Muranen, Kristiina Aittomäki, Outi Malkavaara. **HEBON** Investigators are J. Margriet Collée, Frans B. L. Hogervorst, Maartje J. Hooning, Carolien M. Kets, Peter Devilee, Christi J. van Asperen, Matti A. Rookus, Marjanka K. Schmidt, Cora M. Aalfs, Muriel A. Adank, Margreet G. E. M. Ausems, Marinus J. Blok, Encarna B. Gómez Garcia, Bernadette A. M. Heemskerk-Gerritsen, Antoinette Hollestelle, Agnes Jager, Linetta B. Koppert, Marco Koudijs, Mieke Kriege, Hanne E. J. Meijers-Heijboer, Arjen R. Mensenkamp, Thea M. Mooij, Jan C. Oosterwijk, Ans M. W. van den Ouweland, Frederieke H. van der Baan, Annemieke H. van der Hout, Lizet E. van der Kolk, Rob B. van der Luijt, Carolien H. M. van Deurzen, Helena C. van Doorn, Klaartje van Engelen, Liselotte P. van Hest, Theo A. M. van Os, Senno Verhoef, Maartje J. Vogel & Juul T. Wijnen. **HMBCS** thanks Peter Hillemanns, Hans Christiansen and Johann H. Karstens. **HUBCS** thanks Darya Prokofyeva and Shamil Gantsev. **KARMA** and **SASBAC** thank the Swedish Medical Research Counsel. **KBCP** thanks Eija Myöhänen. **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, Nadia Obi, Sabine Behrens, Ursula Eilber and Muhabbet Celik. **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** (Milan Breast Cancer Study Group): Manoukian Siranoush, Bernard Peissel, Jacopo Azzollini, Daniela Zaffaroni, Bernardo Bonanni, Irene Feroce, Mariarosaria Calvello, Aliana Guerrieri Gonzaga, Monica Marabelli, Davide Bondavalli 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. We thank the coordinators, the research staff and especially the **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 **OFBCR** thanks Teresa Selander, Nayana Weerasooriya and Steve Gallinger. **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. We thank the **SEARCH** and **EPIC** teams. **SGBCC** thanks the participants and all research coordinators for their excellent help with recruitment, data and sample collection. **SKKDKFZ**S thanks all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to this study. **SZBCS** thanks Ewa Putresza. **UBCS** thanks all study participants, the ascertainment, laboratory and research informatics teams at Huntsman Cancer Institute and Intermountain Healthcare, and Justin Williams, Brandt Jones, Melissa Cessna, Stacey Knight and Kerry Rowe for their important contributions to this study.

kConFab 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.

NBCS Collaborators

Kristine K. Sahlberg, Anne-Lise Børresen-Dale, Inger Torhild Gram, Olav Engebråten, Bjørn Naume, Jürgen Geisler, OSBREAC, Grethe I. Grenaker Alnæs.

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.

Additional References

1. Schmidt MK, Tollenaar RA, de Kemp SR, Broeks A, Cornelisse CJ, Smit VT, et al. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. J Clin Oncol. 2007;25(1):64-9.
2. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45(4):353-61, 61e1-2.
3. Schmidt MK, Hogervorst F, van Hien R, Cornelissen S, Broeks A, Adank MA, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. J Clin Oncol. 2016;34(23):2750-60.
4. Fasching PA, Loehberg CR, Strissel PL, Lux MP, Bani MR, Schrauder M, 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.
5. Schrauder M, Frank S, Strissel PL, Lux MP, Bani MR, Rauh C, 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):873-82.
6. Colleran G, McInerney N, Rowan A, Barclay E, Jones AM, Curran C, et al. The TGFBR1*6A/9A polymorphism is not associated with differential risk of breast cancer. Breast Cancer Res Treat. 2010;119(2):437-42.
7. McInerney N, Colleran G, Rowan A, Walther A, Barclay E, Spain S, et al. Low penetrance breast cancer predisposition SNPs are site specific. Breast Cancer Res Treat. 2009;117(1):151-9.
8. Jiang X, Castelao JE, Chavez-Uribe E, Fernandez Rodriguez B, Celeiro Munoz C, Redondo CM, et al. Family history and breast cancer hormone receptor status in a Spanish cohort. PLoS One. 2012;7(1):e29459.
9. Redondo CM, Gago-Dominguez M, Ponte SM, Castelo ME, Jiang X, Garcia AA, et al. Breast feeding, parity and breast cancer subtypes in a Spanish cohort. PLoS One. 2012;7(7):e40543.
10. Ali AM, Schmidt MK, Bolla MK, Wang Q, Gago-Dominguez M, Castelao JE, 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):934-45.
11. Cruz GI, Martinez ME, Natarajan L, Wertheim BC, Gago-Dominguez M, Bondy M, et al. Hypothesized role of pregnancy hormones on HER2+ breast tumor development. Breast Cancer Res Treat. 2013;137(1):237-46.
12. Gago-Dominguez M, Castelao JE, Gude F, Fernandez MP, Aguado-Barrera ME, Ponte SM, et al. Alcohol and breast cancer tumor subtypes in a Spanish Cohort. Springerplus. 2016;5:39.
13. Yang R, Dick M, Marme F, Schneeweiss A, Langheinz A, Hemminki K, et al. Genetic variants within miR-126 and miR-335 are not associated with breast cancer risk. Breast Cancer Res Treat. 2011;127(2):549-54.
14. Menegaux F, Truong T, Anger A, Cordina-Duverger E, Lamkarkach F, Arveux P, et al. Night work and breast cancer: a population-based case-control study in France (the CECILE study). Int J Cancer. 2013;132(4):924-31.

15. 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.
16. Milne RL, Ribas G, Gonzalez-Neira A, Fagerholm R, Salas A, Gonzalez E, et al. ERCC4 associated with breast cancer risk: a two-stage case-control study using high-throughput genotyping. Cancer Res. 2006;66(19):9420-7.
17. Evans DG, Astley S, Stavrinos P, Harkness E, Donnelly LS, Dawe S, 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. Programme Grants for Applied Research. Southampton (UK)2016.
18. Ingham SL, Warwick J, Buchan I, Sahin S, O'Hara C, Moran A, 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):368-72.
19. Kast K, Rhiem K, Wappenschmidt B, Hahnen E, Hauke J, Bluemcke B, et al. Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer. J Med Genet. 2016;53(7):465-71.
20. Rhiem K, Engel C, Graeser M, Zachariae S, Kast K, Kiechle 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.
21. Graeser MK, Engel C, Rhiem K, Gadzicki D, Bick U, Kast K, et al. Contralateral breast cancer risk in BRCA1 and BRCA2 mutation carriers. J Clin Oncol. 2009;27(35):5887-92.
22. Engel C, Rhiem K, Hahnen E, Loibl S, Weber KE, Seiler S, 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.
23. Pesch B, Ko Y, Brauch H, Hamann U, Harth V, Rabstein S, et al. Factors modifying the association between hormone-replacement therapy and breast cancer risk. Eur J Epidemiol. 2005;20(8):699-711.
24. Justenhoven C, Pierl CB, Haas S, Fischer HP, Baisch C, Hamann U, et al. The CYP1B1_1358_GG genotype is associated with estrogen receptor-negative breast cancer. Breast Cancer Res Treat. 2008;111(1):171-7.
25. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, 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.
26. 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):687-95.
27. Dork T, Bendix R, Bremer M, Rades D, Klopper K, Nicke M, et al. Spectrum of ATM gene mutations in a hospital-based series of unselected breast cancer patients. Cancer Res. 2001;61(20):7608-15.
28. Syrjakoski K, Vahteristo P, Eerola H, Tamminen A, Kivinummi K, Sarantaus L, et al. Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients. J Natl Cancer Inst. 2000;92(18):1529-31.
29. Kilpivaara O, Bartkova J, Eerola H, Syrjakoski K, Vahteristo P, Lukas J, 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):575-80.

30. Fagerholm R, Hofstetter B, Tommiska J, Aaltonen K, Vrtel R, Syrjakoski K, 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):844-53.

31. Bogdanova N, Cybulski C, Bermisheva M, Datsyuk I, Yamini P, Hillemanns P, et al. A nonsense mutation (E1978X) in the ATM gene is associated with breast cancer. Breast Cancer Res Treat. 2009;118(1):207-11.

32. 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.

33. 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):127-32.

34. Hartikainen JM, Tuhkanen H, Kataja V, Dunning AM, Antoniou A, Smith P, 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.

35. Hartikainen JM, Tuhkanen H, Kataja V, Eskelinen M, Uusitupa M, Kosma VM, 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):1454-62.

36. Mann GJ, Thorne H, Balleine RL, Butow PN, Clarke CL, Edkins E, 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.

37. Beesley J, Jordan SJ, Spurdle AB, Song H, Ramus SJ, Kjaer SK, 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):2557-65.

38. Han SA, Park SK, Ahn SH, Lee MH, Noh DY, Kim LS, et al. The Korean Hereditary Breast Cancer (KOHBRA) study: protocols and interim report. Clin Oncol (R Coll Radiol). 2011;23(7):434-41.

39. Flesch-Janys D, Slanger T, Mutschelknauss E, Kropp S, Obi N, Vettorazzi 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):933-41.

40. Hadjisavvas A, Loizidou MA, Middleton N, Michael T, Papachristoforou R, Kakouri E, et al. An investigation of breast cancer risk factors in Cyprus: a case control study. BMC Cancer. 2010;10:447.

41. De Vecchi G, Verderio P, Pizzamiglio S, Manoukian S, Barile M, Fortuzzi 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):607-8.

42. Catucci I, Verderio P, Pizzamiglio S, Manoukian S, Peissel B, Barile M, et al. SNPs in ultraconserved elements and familial breast cancer risk. Carcinogenesis. 2009;30(3):544-5; author reply 6.

43. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC Sci Publ. 2002;156:69-70.

44. Phuah SY, Looi LM, Hassan N, Rhodes A, Dean S, Taib NA, 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.

45. Mariapun S, Ho WK, Kang PC, Li J, Lindstrom S, Yip CH, et al. Variants in 6q25.1 Are Associated with Mammographic Density in Malaysian Chinese Women. Cancer Epidemiol Biomarkers Prev. 2016;25(2):327-33.
46. Aure MR, Jernstrom S, Krohn M, Vollan HK, Due EU, Rodland E, et al. Integrated analysis reveals microRNA networks coordinately expressed with key proteins in breast cancer. Genome Med. 2015;7(1):21.
47. Fleischer T, Edvardsen H, Solvang HK, Daviaud C, Naume B, Borresen-Dale AL, 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):2615-25.
48. Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, 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.
49. Quigley DA, Fiorito E, Nord S, Van Loo P, Alnaes GG, Fleischer T, et al. The 5p12 breast cancer susceptibility locus affects MRPS30 expression in estrogen-receptor positive tumors. Mol Oncol. 2014;8(2):273-84.
50. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, 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):R375-89.
51. de Bock GH, Schutte M, Krol-Warmerdam EM, Seynaeve C, Blom J, Brekelmans CT, et al. Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant. J Med Genet. 2004;41(10):731-5.
52. Huijts PE, Vreeswijk MP, Kroeze-Jansema KH, Jacobi CE, Seynaeve C, Krol-Warmerdam EM, 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.
53. Garcia-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, 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):376-88.
54. Pfeiffer RM, Park Y, Kreimer AR, Lacey JV, Jr., Pee D, Greenlee RT, 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.
55. Kriege M, Hollestelle A, Jager A, Huijts PE, Berns EM, Sieuwerts AM, et al. Survival and contralateral breast cancer in CHEK2 1100delC breast cancer patients: impact of adjuvant chemotherapy. Br J Cancer. 2014;111(5):1004-13.
56. Wedren S, Lovmar L, Humphreys K, Magnusson C, Melhus H, Syvanen AC, et al. Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study. Breast Cancer Res. 2004;6(4):R437-49.
57. Lesueur F, Pharoah PD, Laing S, Ahmed S, Jordan C, Smith PL, et al. Allelic association of the human homologue of the mouse modifier Ptprj with breast cancer. Hum Mol Genet. 2005;14(16):2349-56.
58. Stevens KN, Fredericksen Z, Vachon CM, Wang X, Margolin S, Lindblom A, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. Cancer Res. 2012;72(7):1795-803.

59. Jakubowska A, Cybulski C, Szymanska A, Huzarski T, Byrski T, Gronwald J, et al. BARD1 and breast cancer in Poland. Breast Cancer Res Treat. 2008;107(1):119-22.
60. Jakubowska A, Jaworska K, Cybulski C, Janicka A, Szymanska-Pasternak J, Lener M, et al. Do BRCA1 modifiers also affect the risk of breast cancer in non-carriers? Eur J Cancer. 2009;45(5):837-42.
61. Cybulski C, Kluzniak W, Huzarski T, Wokolorczyk D, Kashyap A, Jakubowska A, et al. Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. Lancet Oncol. 2015;16(6):638-44.
62. Cybulski C, Carrot-Zhang J, Kluzniak W, Rivera B, Kashyap A, Wokolorczyk D, et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet. 2015;47(6):643-6.
63. Madsen MJ, Knight S, Sweeney C, Factor R, Salama M, Stijleman IJ, 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):644-52.
64. Camp NJ, Parry M, Knight S, Abo R, Elliott G, Rigas SH, et al. Fine-mapping CASP8 risk variants in breast cancer. Cancer Epidemiol Biomarkers Prev. 2012;21(1):176-81.