Title

A Likelihood Ratio Approach for Utilizing Case-Control Data in the Clinical Classification of Rare Sequence Variants: Application to BRCA1 and BRCA2

Journal

Human Mutation, 2023(1)

ISSN

1059-7794

Authors

Zanti, Maria

O′Mahony, Denise G

Parsons, Michael T

et al.

Publication Date

2023-09-14

DOI

10.1155/2023/9961341

Peer reviewed

HHS Public Access

Author manuscript

Hum Mutat. Author manuscript; available in PMC 2024 May 09.

Published in final edited form as:

Hum Mutat. 2023 ; 2023: . doi:10.1155/2023/9961341.

A likelihood ratio approach for utilizing case-control data in the clinical classification of rare sequence variants: application to *BRCA1* and *BRCA2*

A full list of authors and affiliations appears at the end of the article.

Abstract

A large number of variants identified through clinical genetic testing in disease susceptibility genes, are of uncertain significance (VUS). Following the recommendations of the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP), the frequency in case-control datasets (PS4 criterion), can inform their interpretation. We present a novel case-control likelihood ratio-based method that incorporates gene-specific age-related penetrance. We demonstrate the utility of this method in the analysis of simulated and real datasets. In the analyses of simulated data, the likelihood ratio method was more powerful compared to other methods. Likelihood ratios were calculated for a case-control dataset of *BRCA1* and *BRCA2* variants from the Breast Cancer Association Consortium (BCAC), and compared with logistic regression results. A larger number of variants reached evidence in favor of pathogenicity, and a substantial number of variants had evidence against pathogenicity - findings that would not have been reached using other case-control analysis methods. Our novel method provides greater power to classify rare variants compared to classical case-control methods. As an initiative from the ENIGMA Analytical Working Group, we provide user-friendly scripts and pre-formatted excel calculators for implementation of the method for rare variants in *BRCA1*, *BRCA2* and other high-risk genes with known penetrance.

Keywords

likelihood ratio; case-control; VUS; PS4; ACMG/AMP; variant classification; BRCA

Introduction

Clinical genetic testing of disease susceptibility genes often identifies variants of uncertain significance (VUS), complicating the clinical management of carriers and their families (Eccles et al., 2015). The assessment of the clinical significance of these rare sequence variants, including missense substitutions, in-frame deletions and insertions, and intronic variants, is essential to direct the clinical management of carriers and their relatives towards appropriate prevention, early detection and personalized treatments.

Corresponding Author: Kyriaki Michailidou, kyriakimi@cing.ac.cy.
†Maria Zanti and Denise G. O'Mahony contributed equally to this work

The most widely used method for the interpretation of germline variants is via application of the standards and guidelines recommended by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) (Richards et al., 2015). Strength levels (very strong, strong, moderate and supporting) are assigned to independent lines of evidence for or against variant pathogenicity. These strength levels are then combined and used in a scoring system to provide a clinical class, expressed as pathogenic, likely pathogenic, likely benign, benign, or VUS. These guidelines integrate various sources of information including the variant's nature and position (e.g., nonsense, frameshift, missense) and clinical data (e.g., prevalence in affected individuals and controls), and the combination of this information is interpreted to establish the significance of the variant under investigation with respect to risk. These criteria were recently reinterpreted in a quantitative Bayesian framework, which derived ranges of likelihood ratios (LRs) consistent with each of the evidence strength levels (Tavtigian et al., 2018). For case-control data, the specific criterion (PS4) states that a Relative Risk (RR) or Odds Ratio (OR) > 5.0 with nominal statistical significance (i.e., the confidence interval of the RR or OR does not include 1), provides strong evidence in favor of pathogenicity (Richards et al., 2015).

A significant advance in the classification of variants in cancer and other disease genes was the development of the multifactorial integrated likelihood ratio model (Goldgar et al., 2004); this model combines multiple features, under the assumption that each of them is an independent predictor of variant pathogenicity, in a Bayesian framework, thus providing a quantitative estimate of the pathogenicity of a variant (Goldgar et al., 2008). The ENIGMA consortium (Spurdle et al., 2012) has been applying and extending this multifactorial likelihood model. To date, application of this model has included clinically calibrated prior probabilities of pathogenicity derived from bioinformatic prediction of variant effect and location, along with a combined LR derived from clinical data (Goldgar et al., 2008), such as family history of cancer (Easton et al., 2007), breast cancer tumor pathology (Spurdle et al., 2014), variant co-segregation with disease (Belman, Parsons, Spurdle, Goldgar, & Feng, 2020; Thompson, Easton, & Goldgar, 2003), and variant co-occurrence in trans with a pathogenic variant (PV) in the same gene (Easton et al., 2007). This model can also incorporate LRs derived from variant frequency in cases and controls. Recently, case-control information derived from genotype data for 20 variants was incorporated into a comprehensive multifactorial likelihood analysis of *BRCA1* and *BRCA2* variants by ENIGMA (Parsons et al., 2019), using a method incorporating gene- and age-specific penetrance of PV carriers only. Such case-control LR calculations take into consideration gene- and age-specific penetrance values, and hence they might be expected to outperform the statistical measures currently recommended by ACMG/AMP for the analysis of case-control-data (i.e. OR or RR estimates).

In this paper, we present a novel case-control LR method, based on the same principle as used in the Parsons et al. (Parsons et al., 2019), that incorporates age information in both carriers and non-carriers in the dataset. The method can be used to obtain evidence in favor or against pathogenicity for rare variants in any gene for which there exist known age-specific penetrance estimates, based on data obtained from case-control studies. We illustrate the use of this method to calculate LRs for 24 *BRCA1* and 68 *BRCA2* variants from breast cancer case-control genotype data generated by the Breast Cancer Association

Consortium (BCAC) as part of the large-scale OncoArray project (Michailidou et al., 2017). We further demonstrate the utility of this case-control LR approach to aid interpretation of the clinical significance of variants using evidence aligned to ACMG/AMP code strengths or other classification methods.

Methods

Editorial Policies and Ethical Considerations

All participating studies were approved by the relevant ethics committees and informed consent was obtained from study participant (Michailidou et al., 2017). This research has been approved by the Cyprus National Bioethics Committee.

Case-control datasets

Simulated case-control dataset—Genotype data simulations were performed using the R (v3.6.1) (https://www.r-project.org/) statistical computing language. To create case-control datasets, genotypes for cases and controls were simulated using a Poisson distribution with *lambda* (λ) equal to the mean number of events (variant carriers) in the given interval, expressed as:

$$\lambda_{Cases} = N \times RR \times MAF$$

$$\lambda_{Controls} = N \times MAF$$

where N denotes the sample size, RR denotes the relative breast cancer risk of the causal variant and MAF denotes the minor allele frequency of the variant in the general population. Ages were simulated using normal distribution, with mean and standard deviation following the gene-specific age distribution in the CARRIERS population-based study (Hu et al., 2021).

Genotype data simulations were carried out for variants conferring a RR of 1 (indicating no increased risk), 2, 3, 4, 5, 6, 7, 8, 9 or 10, minor allele frequency in controls of 0.0001, 0.00005 or 0.00003 and sample size of N = 20,000 (20,000 breast cancer cases and 20,000 controls), 30,000 (30,000 breast cancer cases and 30,000 controls) or 50,000 (50,000 breast cancer cases and 50,000 controls). For each of these 90 scenarios, we simulated 10,000 replicates.

Additionally, in order to account for the possibility that age information is not available, we repeated the analysis using same age for all individuals.

BCAC OncoArray dataset—Genotype data were generated as part of the BCAC component of the OncoArray project (Michailidou et al., 2017) (studies included in the analysis are listed in Supplementary Table S1) and were available for 75,657 breast cancer cases and 52,987 controls of European ancestry. The majority of studies were population-based case–control studies, or case–control studies nested within population-based cohorts.

However, a subset of studies oversampled cases with a family history of breast cancer. Of these, 464 breast cancer cases and 1,347 controls had missing information regarding their age at diagnosis or interview, respectively and were excluded from the analyses. Another 1,445 cases and 858 controls were removed because their ages fell outside the interval of 21-80 years (the age range for which penetrance estimates were available). Cluster plots of 56 *BRCA1* and 127 *BRCA2* variants nominated by ENIGMA researchers for inclusion in the OncoArray project, were manually checked to review the automated calls. This was performed since automated genotype calling for rare variants from GWAS chips has been shown to be suboptimal (Coleman et al., 2016). Genotypes were adjusted for 41 *BRCA1* and 91 *BRCA2* variants, while 3 *BRCA1* and 2 *BRCA2* variants genotypes were determined to have been called correctly by automated clustering. Genotype recalling was not performed for 12 *BRCA1* and 34 *BRCA2* variants due to the low quality of the genotype data; these variants were not considered further.

After genotype cluster review and recalling, 16 *BRCA1* and 19 *BRCA2* variants were excluded from further analysis due to their high frequency (> 0.1%). Additionally, case-control LR calculations were not possible for four *BRCA1* and six *BRCA2* variants due to absence of variant carriers in the post filtering dataset. After these exclusions, case-control LR and logistic regression analysis was performed for 24 *BRCA1* and 68 *BRCA2* variants. It should be noted that some of the variants selected for the array have subsequently been classified or were those whose pathogenicity status were known and were included as positive or negative controls.

Statistical analyses

Case-control likelihood ratio method—This method (detailed in Supplementary File 1) compares the likelihood of the distribution of the variant of interest among cases and controls, under the hypothesis that the variant is associated with similar risks of the disease in question, as the “average” pathogenic variant (H_p), compared to the likelihood under the hypothesis that it is a benign variant not associated with increased risk (H_b). These risks may be age-, sex- and/or country-specific. Thus:

$$LR = \frac{Pr\,(Data \mid H_p)}{Pr(Data \mid H_b)}$$

where *Data* denotes observed data on carrier status of a variant of interest, case-control status and age at diagnosis or interview, combined over all individuals in the dataset.

In order to calculate the above LR, we follow a survival analysis framework. We first determine the probability that an individual with genotype k remains unaffected at age t, $S_k(t)$; and the corresponding probability that an individual with genotype k is affected at age t, $f_k(t)$ (where $k = 0$ or 1 for non-carriers and carriers, respectively). These probabilities can be computed from the age-specific baseline incidence, λ_0(t); and the age specific log-relative risk of an assumed pathogenic variant in the gene of interest, β(t). These probabilities are given by:

$$S_k(t) = \exp(-\int_0^t \lambda_0(t) e^{\beta(t)k} dt)$$

$$f_k(t) = S_k(t)\, e^{\beta(t)k}$$

As detailed in Supplementary File 1, the likelihood ratio is, to close approximation, given by:

$$ccLR = \frac{\prod_{v_j=1} S_1(t_j) e^{\beta(t) d_j} \;/\; S_0(t_j)}{\left(\sum_j^N S_1(t_j) e^{\beta(t_j) d_j} \;/\; S_0(t_j)\right)^K} \;/\; \frac{1}{N^K}$$

where N is the total number of individuals, K is the number of variant carriers, v_j is the variant status (0 for non-carriers and 1 for variant carriers) and d_j is the disease status (0 for controls and 1 for cases) for individual *j*.

The baseline incidence rates $\lambda_0(t)$ were taken from the age-specific background rates for England and Wales (1998-2002) (https://ci5.iarc.fr/CI5I-X/Default.aspx) and the age-specific breast cancer relative risks for pathogenic variant carriers $\beta(t)$ were taken from the recent large-scale BRIDGES (Breast Cancer Risk after Diagnostic Gene Sequencing) project (Dorling et al., 2021). To allow for possible carrier frequency differences by country, stratified LR calculations were performed within each country and then multiplied to provide a final LR.

Likelihood ratios are further translated into ACMG/AMP code strength categories according to published recommendations (Tavtigian et al., 2018). Likelihood ratio estimates in favor of variant pathogenicity are scored as: very strong, LR ≥ 350; strong, 350 > LR ≥ 18.7; moderate, 18.70 > LR ≥ 4.33 and supporting, 4.33 > LR ≥ 2.08. Likelihood ratio evidence for benign variant status are scored as: very strong, LR ≤ 0.0029; strong, 0.0029 < LR ≤ 0.053; moderate, 0.053 < LR ≤ 0.231 and supporting, 0.231 < LR ≤ 0.48. No evidence strength corresponded to estimates of 0.48 ≤ LR < 2.08.

In a series of sensitivity analyses, the method was applied using three other published RR estimates: from case series unselected for family history of breast cancer (Antoniou et al., 2003), cohort series of *BRCA1* and *BRCA2* carriers (Kuchenbaecker et al., 2017) and breast cancer hazard ratio estimates for missense *BRCA1* and *BRCA2* variants (Li et al., 2022). In order to account for country-specific effects, the stratified analysis was also performed using age- and country-specific incidence rates, derived from the Cancer Incidence in Five Continents, volume 9, 1998-2002, (https://ci5.iarc.fr/CI5I-X/Default.aspx). Age-specific breast cancer incidences for Greece and North Macedonia were retrieved from the 2020 cancer registry (European cancer information system, ECIS, https://ecis.jrc.ec.europa.eu/) since cancer incidence data were not available for the years 1998-2019. Unstratified analyses were also performed for comparison.

Detailed R scripts and pre-formatted excel calculators (user can either input individual-level data or tabulated by age groups) for the calculation of case-control LRs can be found using the following GitHub link: (https://github.com/BiostatUnitCING/ccLR). The files provided can be used to derive estimates based on the RR from Dorling et al. (Dorling et al., 2021), Kuchenbaecker et al. (Kuchenbaecker et al., 2017) or Antoniou et al. (Antoniou et al., 2003). In addition, this method can also be used to compute case-control LRs for variants in other disease susceptibility genes, by using age-specific penetrance estimates for the gene of interest (indicated by "Custom" gene in the pre-formatted excel calculators and R script). Furthermore, to allow for the possibility that age information is not available (or is only available for a subset of the dataset) user can incorporate individuals with unknown age at diagnosis or interview in any of the age groups specified in the tabulated calculator.

Odds ratio analysis—Odds ratio analysis was performed using logistic regression adjusted by age and country (if applicable) and Fisher's exact test (corrected using Haldane's method when simulations resulted in zero variant carriers in cases or controls (Agresti, 2003)). Logistic regression p-values were estimated using the likelihood ratio test. Based on the original ACMG/AMP recommendations (Richards et al., 2015), an OR estimate greater than 5.0, with the confidence interval not including 1.0, was used to define strong evidence of pathogenicity (PS4).

Evaluation and application of the case-control analyses methods—The simulated datasets were analyzed using the novel case-control LR method, logistic regression (adjusted by age), and Fisher's exact test. The case-control LR method was applied using age-specific breast cancer ORs for *BRCA1* and *BRCA2* PVs (Dorling et al., 2021). For causal variants with relative risk of 2 to 10, the power of the case-control LR method was estimated either as the probability of reaching at least supporting ($LR \geq 2.08$) or at least strong pathogenic ($LR \geq 18.7$) evidence. For benign variants with relative risk of 1, the power of the case-control LR method was estimated either as the probability of reaching at least supporting ($LR \leq 0.48$) or at least strong ($LR \leq 0.053$) benign ACMG/AMP evidence. Correspondingly, type I error for pathogenicity was calculated as the probability of obtaining at least supporting or at least strong pathogenic ACMG/AMP evidence when the relative risk was set to 1. Equivalently, type I error for evidence against pathogenicity was calculated as the probability of obtaining at least supporting or at least strong benign ACMG/AMP evidence when the relative risk was greater than one. The power of the OR methods was estimated as the probability of reaching the ACMG/AMP PS4 criterion (OR > 5.0, CI not including 1.0, p-value < 0.05). Following the analyses results of the simulated datasets, optimal LR cut-offs (to maximize power and minimize type I error) are used to define ACMG/AMP evidence strengths for the 92 variants included in the BCAC OncoArray dataset.

Results

Simulated datasets

Based on the simulation results for high-risk *BRCA1* (RR > 9) and *BRCA2* (RR > 5) variants, LR of strong and very strong evidence in favor of pathogenicity ($LR \geq 18.7$) and

of at least supporting evidence against pathogenicity (LR ≤ 0.48), should be used in order to maintain a high power (> 80%) and low type I error (< 0.05) (Supplementary Table S2).

Results for all measures in all simulated datasets show that the power to achieve strong evidence in favour of pathogenicity is consistently greater for the case-control LR method using age-specific breast cancer risks compared to standard OR analysis methods (Figure 1, Supplementary Table S2). The power to correctly categorize variants with a RR comparable to a typical *BRCA1* PV was >80% in all scenarios except for small datasets (N ≤ 30,000) with causal variants present at the lower frequency (MAF = 0.00003) (Figure 1A).

In addition, the case-control LR method can also be used to obtain evidence against pathogenicity, something that cannot be achieved using standard OR analysis methods. Results from simulated case-control datasets of benign variants (RR of 1, Figure 2), show that the case-control LR method using the age-specific RRs of the "average" *BRCA1* PV, exhibits adequate power (> 80%) to identify variants with evidence against pathogenicity (LR ≤ 0.48) for larger datasets (N ≥ 30,000) and MAF of 0.0001.

The implementation of the method to account for datasets with missing information, assuming the same age for all individuals, demonstrated reduced power and increased type I error in all simulations. However, the type I error was still less than 0.05 in all cases (Supplementary Figures S1 and S2, Supplementary Table S3).

BCAC OncoArray dataset

Logistic regression results—Using logistic regression, two *BRCA2* variants (2%) (Table 1) reached strong pathogenic evidence following the ACMG/AMP classification criterion (PS4 criterion, OR > 5, p-value < 0.05, CI not including 1.0) (Richards et al., 2015). Detailed logistic regression results for all variants are shown in Supplementary Table S4.

Case-control LRs and ACMG/AMP Code Strengths—In the country-stratified baseline analysis (using the breast cancer ORs estimated from BRIDGES (Dorling et al., 2021)), evidence in favor of pathogenicity (defined as LR ≥ 18.70 following the simulation cut-offs) was achieved for 6 variants (6.5%) (Table 2), of which 3 variants were assigned very strong and another 3 strong strengths. Evidence against pathogenicity (defined as LR ≤ 0.48) was observed for 59 variants (64.1%), of which 26 were assigned very strong, 14 strong, 7 moderate and 12 supporting strengths. The results for the remaining 27 variants (29.3%) were uninformative. Case-control LRs and corresponding ACMG/AMP code strengths for all 92 *BRCA1* and *BRCA2* variants are shown in Supplementary Table S4. The different sensitivity analyses did not show any major discrepancies in the estimated LRs (Supplementary Table S5).

Discussion

This study provides a detailed description of the methodology to calculate case-control LRs for rare variants using case-control data, based on age- and gene-specific relative risks and age information for non-carriers. The LRs are calculated by comparing the likelihood

of the distribution of the variant of interest in cases and controls, under the hypothesis that the variant has similar age-specific relative risks as the “average” pathogenic variant, compared to the hypothesis that it is not associated with increased (or decreased) disease risk. We evaluated the method using simulated datasets and further applied it to derive LRs for pathogenicity for individual variants from analysis of genotype data from a large case-control study. These can now be used in combination with other evidence to inform variant classification - either according to ACMG/AMP classification standards and guidelines (Richards et al., 2015; Tavtigian et al., 2018), or using multifactorial likelihood modelling approaches (Goldgar et al., 2004; Parsons et al., 2019). Further, we provide user-friendly scripts and a pre-formatted excel calculator to facilitate the future implementation of this method for the calculation of case-control LRs. These resources may be readily applied for the calculation of LRs to be used in classification of VUS in the *BRCA1* and *BRCA2*, and other disease susceptibility genes with known penetrance values.

Notably, our results demonstrate the improved performance of our LR-based method for assessing variant pathogenicity as it considers gene- and age-specific penetrance for carriers and age information for non-carriers. Using simulated case-control datasets we show that the case-control LR method using age-specific breast cancer ORs from high-penetrance genes (e.g. *BRCA1* and *BRCA2*) outperforms other OR analysis methods. These observations reflect the fact that the method presented here is more suitable for the analysis of rare variants in a case-control setting. We further provide cut-offs of LRs in favor or against pathogenicity, to be used in a real setting.

Analysis of the BCAC OncoArray data using our proposed method provided informative pathogenic ACMG/AMP classification evidence for six out of 92 variants analyzed. Furthermore, 59 variants reached evidence against pathogenicity, something that is not directly measured as a code strength through classical calculations of ORs. Given that *a priori*, the vast majority of rare sequence variants (e.g., *BRCA1* and *BRCA2*) will be neutral with respect to risk, this is a key advantage of our approach. In contrast, using logistic regression analysis, informative ACMG/AMP classification criterion PS4 (OR > 5.0, p-value < 0.05, CI not including 1.0) was reached only for two variants.

There are possible caveats that should be recognized. Selection of cases or controls for a family history of cancer would affect the carrier probabilities. The likelihood ratios would then be inaccurate, but in principle this could be considered by incorporating family history in the likelihoods, if known. Depletion of cases with known pathogenic variants by prior clinical sequencing could also bias the likelihood ratios, therefore the method is best applied to population-based case-control studies. For these reasons we highlight the ACMG/AMP recommendation to review all available evidence for/against pathogenicity for a given variant, and to denote obviously conflicting findings for different evidence types, before assigning a final classification. A conservative approach may be to assign case-control weight with a cap, for example at moderate strength for or against pathogenicity.

Our method gains power in part because it leverages data on individual-level age, but we have to acknowledge that age is not always available. The method can be implemented more approximately by assuming that individuals with unknown information are of the same age,

but this reduces power, because the expectation that carriers of risk variants develop the disease at a younger age is then not utilised. It may also increase type I error, because the likelihood ratio may be calculated for an age that is not appropriate for the dataset (for example, if the dataset consists predominantly of older individuals), although the type I error was still low in the simulations we considered. In the tabulated pre-formatted calculator, we allow user to incorporate individuals with unknown age at diagnosis or interview, in any of the age groups specified. A conservative approach would be to include individuals with unknown age in the oldest age-group. In this way, case-control genotypes from both existing data and new series, with and without age data, can be incorporated. However, we would like to emphasize that pooling series, particularly from different populations with different age/ethnicity structures or with different genotyping technologies can lead to biased results. Ideally, datasets should be analysed separately and the overall likelihood ratio generated by multiplying the study-specific likelihood ratios.

Conclusions

This manuscript describes in detail a novel method used for the calculation of the case-control LR to provide evidence of variant pathogenicity. This LR method is more informative compared to logistic regression analysis (or an OR calculation based on contingency tables and Fisher's exact test). It improves power as it considers age- and gene-specific penetrance values and age information for non-carriers, and can provide both evidence in favor and against pathogenicity. In addition, this method can be also implemented towards the classification of VUS in any disease susceptibility gene for which disease-penetrance has been reliably estimated. Open access scripts and pre-formatted excel calculators with code and instructions on how to use the method, are available at the following address: https://github.com/BiostatUnitCING/ccLR.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Maria Zanti[1,†], Denise G. O'Mahony[1,†], Michael T. Parsons[2], Hongyan Li[3], Joe Dennis[4], Kristiina Aittomäkkiki[5], Irene L. Andrulis[6,7], Hoda Anton-Culver[8], Kristan J. Aronson[9], Annelie Augustinsson[10], Heiko Becher[11], Stig E. Bojesen[12,13,14], Manjeet K. Bolla[4], Hermann Brenner[15,16,17], Melissa A. Brown[18], Saundra S. Buys[19], Federico Canzian[20], Sandrine M. Caputo[21,22], Jose E. Castelao[23], Jenny Chang-Claude[24,25], GC-HBOC study Collaborators[26], Kamila Czene[27], Mary B. Daly[28], Arcangela De Nicolo[29], Peter Devilee[30,31], Thilo Dörk[32], Alison M. Dunning[33], Miriam Dwek[34], Diana M. Eccles[35], Christoph Engel[36,37], D. Gareth Evans[38,39], Peter A. Fasching[40], Manuela Gago-Dominguez[41], Montserrat García-Closas[42], José A. García-Sáenz[43], Aleksandra Gentry-Maharaj[44], Willemina R.R. Geurts - Giele[45], Graham G. Giles[46,47,48], Gord Glendon[6], Mark S. Goldberg[49,50], Encarna B. Gómez Garcia[51], Melanie Güendert[52,53,54], Pascal Guénel[55], Eric Hahnen[26,56], Christopher A. Haiman[57], Per Hall[27,58], Ute Hamann[59], Elaine

F. Harkness[60,61,62], Frans B.L. Hogervorst[63], Antoinette Hollestelle[64], Reiner Hoppe[65,66], John L. Hopper[47], Claude Houdayer[67], Richard S. Houlston[68], Anthony Howell[69], ABCTB Investigators[70], Milena Jakimovska[71], Anna Jakubowska[72,73], Helena Jernström[10], Esther M. John[74,75], Rudolf Kaaks[24], Cari M. Kitahara[76], Stella Koutros[42], Peter Kraft[77,78], Vessela N. Kristensen[79,80], James V. Lacey[81,82], Diether Lambrechts[83,84], Melanie Léoné[85], Annika Lindblom[86,87], Jan Lubiński[72], Michael Lush[4], Arto Mannermaa[88,89,90], Mehdi Manoochehri[59], Siranoush Manoukian[91], Sara Margolin[58,92], Maria Elena Martinez[93,94], Usha Menon[44], Roger L. Milne[46,47,48], Alvaro N. Monteiro[95], Rachel A. Murphy[96,97], Susan L. Neuhausen[98], Heli Nevanlinna[99], William G. Newman[38,39], Kenneth Offit[100,101], Sue K. Park[102,103,104], Paul James[105,106], Paolo Peterlongo[107], Julian Peto[108], Dijana Plaseska-Karanfilska[71], Kevin Punie[109], Paolo Radice[110], Muhammad U. Rashid[59,111], Gad Rennert[112], Atocha Romero[113], Efraim H. Rosenberg[114], Emmanouil Saloustros[115], Dale P. Sandler[116], Marjanka K. Schmidt[117,118,119], Rita K. Schmutzler[26,56,120], Xiao-Ou Shu[121], Jacques Simard[122], Melissa C. Southey[46,48,123], Jennifer Stone[47,124], Dominique Stoppa-Lyonnet[21,125,126], Rulla M. Tamimi[78,127], William J. Tapper[35], Jack A. Taylor[116,128], Soo Hwang Teo[129,130], Lauren R. Teras[131], Mary Beth Terry[132], Mads Thomassen[133], Melissa A. Troester[134], Celine M. Vachon[135], Ana Vega[136,137,138], Maaike P.G. Vreeswijk[31], Qin Wang[4], Barbara Wappenschmidt[26,56], Clarice R. Weinberg[139], Alicja Wolk[140,141], Wei Zheng[121], Bingjian Feng[142], Fergus J. Couch[143], Amanda B. Spurdle[2], Douglas F. Easton[4,33], David E. Goldgar[142], Kyriaki Michailidou[1,4]

Affiliations

[1]Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus

[2]Population Health Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[3]Cancer Control and Population Science, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

[4]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[5]Department of Clinical Genetics, Helsinki University Hospital, University of Helsinki, Helsinki, Finland

[6]Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada

[7]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

[8]Department of Medicine, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA

[9]Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada

[10]Oncology, Department of Clinical Sciences in Lund, Lund University, Lund, Sweden

[11]Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[12]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark

[13]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark

[14]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[15]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

[16]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

[17]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

[18]School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia

[19]Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

[20]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany

[21]Service de Génétique, Institut Curie, Paris, France

[22]Paris Sciences Lettres Research University, Paris, France

[23]Oncology and Genetics Unit, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain

[24]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

[25]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[26]Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[27]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[28]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA

[29]Center for Omics Sciences, IRCCS San Raffaele Scientific Institute, Milan, Italy

[30]Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands

[31]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands

[32]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany

[33]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

[34]School of Life Sciences, University of Westminster, London, UK

[35]Faculty of Medicine, University of Southampton, Southampton, UK

[36]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany

[37]LIFE - Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany

[38]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, UK

[39]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, UK

[40]Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, University Hospital Erlangen, Erlangen, Germany

[41]Genomic Medicine Group, International Cancer Genetics and Epidemiology Group, Fundación Pública Galega de Medicina Xenómica, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain

[42]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA

[43]Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain

[44]MRC Clinical Trials Unit, Institute of Clinical Trials & Methodology, University College London, London, UK

[45]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands

[46]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia

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

[48]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia

[49]Department of Medicine, McGill University, Montréal, QC, Canada

[50]Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, Canada

[51]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, the Netherlands

[52]Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg, Germany

[53]Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Germany

[54]Institute of Diabetes Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[55]Team 'Exposome and Heredity', CESP, Gustave Roussy, INSERM, University Paris-Saclay, UVSQ, Villejuif, France

[56]Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[57]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

[58]Department of Oncology, Södersjukhuset, Stockholm, Sweden

[59]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany

[60]Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

[61]Nightingale & Genesis Prevention Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK

[62]NIHR Manchester Biomedical Research Unit, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

[63]Family Cancer Clinic, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands

[64]Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands

[65]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany

[66]University of Tübingen, Tübingen, Germany

[67]Department of Genetics, F76000 and Normandy University, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen University Hospital, Rouen, France

[68]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

[69]Division of Cancer Sciences, University of Manchester, Manchester, UK

[70]Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia

[71]Research Centre for Genetic Engineering and Biotechnology 'Georgi D. Efremov', MASA, Skopje, Republic of North Macedonia

[72]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland

[73]Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland

[74]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA

[75]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

[76]Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

[77]Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[78]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[79]Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway

[80]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

[81]Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA, USA

[82]City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA, USA

[83]Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium

[84]VIB Center for Cancer Biology, VIB, Leuven, Belgium

[85]Genetic and Cancer Medical Laboratory HCL-CLB, Hospices Civils de Lyon, Bron, France

[86]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

[87]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden

[88]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland

[89]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland

[90]Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland

[91]Unit of Medical Genetics, Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy

[92]Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden

[93]Moores Cancer Center, University of California San Diego, La Jolla, CA, USA

[94]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, USA

[95]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA

[96]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[97]Cancer Control Research, BC Cancer, Vancouver, BC, Canada

[98]Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA

[99]Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland

[100]Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[101]Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[102]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

[103]Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, South Korea

[104]Cancer Research Institute, Seoul National University, Seoul, Korea

[105]Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Victoria, Australia

[106]Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre and Royal Melbourne Hospital, Melbourne, Victoria, Australia

[107]Genome Diagnostics Program, IFOM - ETS the AIRC Institute of Molecular Oncology, Milan, Italy

[108]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

[109]Department of General Medical Oncology and Multidisciplinary Breast Centre, Leuven Cancer Institute and University Hospitals Leuven, Leuven, Belgium

[110]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy

[111]Department of Basic Sciences, Shaukat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Lahore, Pakistan

[112]Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel

[113]Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain

[114]Department of Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands

[115]Department of Oncology, University Hospital of Larissa, Larissa, Greece

[116]Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA

[117]Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, the Netherlands

[118]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, the Netherlands

[119]Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands

[120]Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[121]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA

[122]Genomics Center, Centre Hospitalier Universitaire de Québec – Université Laval Research Center, Québec City, QC, Canada

[123]Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia

[124]Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, Western Australia, Australia

[125]Department of Tumour Biology, INSERM U830, Paris, France

[126]Université Paris Descartes, Paris, France

[127]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

[128]Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA

[129]Breast Cancer Research Programme, Cancer Research Malaysia, Subang Jaya, Selangor, Malaysia

[130]Department of Surgery, Faculty of Medicine, University of Malaya, UM Cancer Research Institute, Kuala Lumpur, Malaysia

[131]Department of Population Science, American Cancer Society, Atlanta, GA, USA

[132]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA

[133]Department of Clinical Genetics, Odense University Hospital, Odence C, Denmark

[134]Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[135]Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA

[136]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain

[137]Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain

[138]Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain

[139]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA

[140]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[141]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

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

[143]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

Acknowledgments

This work was co-funded by the Republic of Cyprus through the Research and Innovation Foundation (Project: CULTURE/AWARD-YR/0418/0017). ABS and MTP are supported by Australian National Health and Medical Research Funding (APP177524). DGO is funded by Telethon Cyprus (Telethon Cyprus: 33173233) through the Cyprus Institute of Neurology and Genetics.

BCAC is funded by the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Additional funding for BCAC is provided via the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health and via the CanRisk project which is funded from the Cancer Research UK (grant PPRPGM-Nov20/100002).

Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer Research UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for iCOGS came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). 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 ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]. The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute (grant number Z01-CP010119), and the National Institute of Environmental Health Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BCEES was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). For the BCFR-NY, BCFR-PA, BCFR-UT this work was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. The BCINIS study is supported in part by the Breast Cancer Research Foundation (BCRF). 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). CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. 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 American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The California Teachers Study (CTS) and the research reported in this publication were supported by the National Cancer Institute of the National Institutes of Health under award number U01-CA199277; P30-CA033572; P30-CA023100; UM1-CA164917; and R01-CA077398. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. The collection of cancer incidence data used in the California Teachers Study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885; Centers for Disease Control and Prevention's National Program of Cancer Registries, under cooperative agreement 5NU58DP006344; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201800032I awarded to the University of California, San Francisco, contract HHSN261201800015I awarded to the University of Southern California, and contract HHSN261201800009I awarded to the Public Health Institute. The opinions, findings, and conclusions expressed herein are those of the author(s) and do not necessarily reflect the official views of the State of California, Department of Public Health, the National Cancer Institute, the National Institutes of Health, the Centers for Disease Control and Prevention or their Contractors and Subcontractors, or the Regents of the University of California, or any of its programs. The University of Westminster curates the DietCompLyf database funded by Against Breast Cancer Registered Charity No. 1121258 and the NCRN. The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. 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 and 70114178, 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. The GEPARSIXTO study was conducted by the German Breast Group GmbH. 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 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 mega grant from the Government of Russian Federation (2020-220-08-2197), 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. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. The MABCS study is funded by the Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov", MASA. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC). The MCBCS was supported by the NIH grants R35CA253187, R01CA192393, R01CA116167, R01CA176785 a NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [P50CA116201], and the Breast Cancer Research Foundation. 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. The MEC was supported by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society CAN 2018/675, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation FBKS 2021-19, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. The work of MTLGEBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program – grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade – grant # PSR-SIIRI-701. 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 NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were 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 Carolina Breast Cancer Study (NCBCS) was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. 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 POSH study is funded by Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, C1275/A15956 and Breast Cancer Campaign 2010PR62, 2013PR044. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). 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. 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. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The Two Sister Study (2SISTER) was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES102245), and, also by a grant from Susan G. Komen for the Cure, grant FAS0703856. SKKDKFZS is supported by the DKFZ. The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). 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. The TNBCC was supported by: a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), a grant from the Breast Cancer Research Foundation, a generous gift from the David F. and Margaret T. Grohne Family Foundation. The UCIBCS component of this research was supported by the NIH [CA58860, CA92044] and the Lon V Smith Foundation [LVS39420]. The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The UKOPS study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre and MRC Core Funding (MC_UU_00004/01). The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA.

Contract grant sponsor AV: supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds (INT20/00071; PI19/01424); the Autonomous Government of Galicia (Consolidation and structuring program: IN607B); the Fundación Mutua Madrileña (call 2018) and the AECC (PRYES211091VEGA).

We thank all the women who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. ABCFS thank Maggie Angelakos, Judi Maskiell,

Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Deborah Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirmala Pathmanathan, Peter Simpson, J. Dinny Graham, Mythily Sachchithananthan. Samples are made available to researchers on a non-exclusive basis. BBCS thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, Rachel Lloyd. The BCINIS study would not have been possible without the contributions of Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. F. Lejbkowicz, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. 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. CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. 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. Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. The authors would like to thank the California Teachers Study Steering Committee that is responsible for the formation and maintenance of the Study within which this research was conducted. A full list of California Teachers Study (CTS) team members is available at https://www.calteachersstudy.org/team. DIETCOMPLYF thanks the patients, nurses and clinical staff involved in the study. The DietCompLyf study was funded by the charity Against Breast Cancer (Registered Charity Number 1121258) and the NCRN. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach. FHRISK and PROCAS thank NIHR for funding. The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [RH, Hiltrud Brauch, Wing-Yee Lo], 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 [UH], 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]. HEBCS thanks Johanna Kiiski, Carl Blomqvist, Taru A. Muranen, Kirsimari Aaltonen, Karl von Smitten, Irja Erkkilä. HUBCS thanks Darya Prokofyeva and Shamil Gantsev. KARMA and SASBAC thank the Swedish Medical Research Counsel. KBCP thanks Eija Myöhänen. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. MABCS thanks Milena Jakimovska (RCGEB “Georgi D. Efremov”), Snezhana Smichkoska, Emilija Lazarova, Marina Iljoska (University Clinic of Radiotherapy and Oncology), Katerina Kubelka-Sabit, Dzengis Jasar, Mitko Karadjozov (Adzibadem-Sistina Hospital), Andrej Arsovski and Liljana Stojanovska (Re-Medika Hospital) for their contributions and commitment to this study. MARIE thanks Petra Seibold, Nadia Obi, Sabine Behrens, Ursula Eilber and Muhabbet Celik. MBCSG (Milan Breast Cancer Study Group): Paolo Radice, Paolo Peterlongo, Siranoush Manoukian, Bernard Peissel, Jacopo Azzollini, Claudia Monaco, 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. The MISS study group acknowledges the former Principal Investigator, professor Håkan Olsson. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. MSKCC thanks Marina Corines, Lauren Jacobs. MTLGEBCS would like to thank Martine Tranchant (CHU de Québec – Université Laval Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management and skilful technical assistance. J.S. is Chair holder of the Canada Research Chair in Oncogenetics. The following are NBCS Collaborators: Kristine K. Sahlberg (PhD), Anne-Lise Børresen-Dale (Prof. Em.), Lars Ottestad (MD), Rolf Kåresen (Prof. Em.) Dr. Ellen Schlichting (MD), Marit Muri Holmen (MD), Toril Sauer (MD), Vilde Haakensen (MD), Olav Engebråten (MD), Bjørn Naume (MD), Alexander Fosså (MD), Cecile E. Kiserud (MD), Kristin V. Reinertsen (MD), Åslaug Helland (MD), Margit Riis (MD), Jürgen Geisler (MD), OSBREAC and Grethe I. Grenaker Alnæs (MSc). NBHS and SBCGS thank study

participants and research staff for their contributions and commitment to the studies. For NHS and NHS2 the study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We would like to thank the participants and staff of the NHS and NHS2 for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. The OFBCR thanks Teresa Selander, Nayana Weerasooriya, Anna Marie Mulligan and Steve Gallinger. ORIGO thanks E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. PBCS thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. The ethical approval for the POSH study is MREC /00/6/69, UKCRN ID: 1137. We thank staff in the Experimental Cancer Medicine Centre (ECMC) supported Faculty of Medicine Tissue Bank and the Faculty of Medicine DNA Banking resource. The authors wish to acknowledge the roles of the Breast Cancer Now Tissue Bank in collecting and making available the samples and/or data, and the patients who have generously donated their tissues and shared their data to be used in the generation of this publication. PREFACE thanks Sonja Oeser and Silke Landrith. The RBCS thanks Jannet Blom, Saskia Pelders, Wendy J.C. Prager – van der Smissen, and the Erasmus MC Family Cancer Clinic. We thank the SEARCH and EPIC teams. SKKDKFZS thanks all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to this study. We thank the SUCCESS Study teams in Munich, Duessldorf, Erlangen and Ulm. SZBCS thanks Ewa Putresza. UCIBCS thanks Irene Masunaka. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

REFERENCES

Agresti A. (2003). Categorical Data Analysis (2nd ed.): Wiley.

Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, … Easton DF (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet, 72(5), 1117–1130. doi:10.1086/375033 [PubMed: 12677558]

Belman S, Parsons MT, Spurdle AB, Goldgar DE, & Feng BJ (2020). Considerations in assessing germline variant pathogenicity using cosegregation analysis. Genet Med, 22(12), 2052–2059. doi:10.1038/s41436-020-0920-4 [PubMed: 32773770]

Coleman JR, Euesden J, Patel H, Folarin AA, Newhouse S, & Breen G (2016). Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. Brief Funct Genomics, 15(4), 298–304. doi:10.1093/bfgp/elv037 [PubMed: 26443613]

Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C, Wahlstrom C, … Easton DF (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med, 384(5), 428–439. doi:10.1056/NEJMoa1913948 [PubMed: 33471991]

Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, … Goldgar DE (2007). A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet, 81(5), 873–883. doi:10.1086/521032 [PubMed: 17924331]

Eccles DM, Mitchell G, Monteiro AN, Schmutzler R, Couch FJ, Spurdle AB, … Group ECW (2015). BRCA1 and BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. Ann Oncol, 26(10), 2057–2065. doi:10.1093/annonc/mdv278 [PubMed: 26153499]

Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS, & Group, I. U. G. V. W. (2008). Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum Mutat, 29(11), 1265–1272. doi:10.1002/humu.20897 [PubMed: 18951437]

Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ, & Breast Cancer Information Core Steering, C. (2004). Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet, 75(4), 535–544. doi:10.1086/424388 [PubMed: 15290653]

Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, … Couch FJ (2021). A Population-Based Study of Genes Previously Implicated in Breast Cancer. N Engl J Med, 384(5), 440–451. doi:10.1056/NEJMoa2005936 [PubMed: 33471974]

Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, … Olsson H (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA, 317(23), 2402–2416. doi:10.1001/jama.2017.7112 [PubMed: 28632866]

Li H, Engel C, Hoya M, Peterlongo P, Yannoukakos D, Livraghi L, … Consortium, C. (2022). Risks of breast and ovarian cancer for women harboring pathogenic missense variants in BRCA1 and BRCA2 compared with those harboring protein truncating variants. Genet Med, 24(10), 2208. doi:10.1016/j.gim.2022.08.005 [PubMed: 36205748]

Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, … Easton DF (2017). Association analysis identifies 65 new breast cancer risk loci. Nature, 551(7678), 92–94. doi:10.1038/nature24284 [PubMed: 29059683]

Parsons MT, Tudini E, Li H, Hahnen E, Wappenschmidt B, Feliubadalo L, … Spurdle AB (2019). Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. Hum Mutat, 40(9), 1557–1578. doi:10.1002/humu.23818 [PubMed: 31131967]

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, … Committee ALQA (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med, 17(5), 405–424. doi:10.1038/gim.2015.30 [PubMed: 25741868]

Spurdle AB, Couch FJ, Parsons MT, McGuffog L, Barrowdale D, Bolla MK, … kConFab I (2014). Refined histopathological predictors of BRCA1 and BRCA2 mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. Breast Cancer Res, 16(6), 3419. doi:10.1186/s13058-014-0474-y [PubMed: 25857409]

Spurdle AB, Healey S, Devereau A, Hogervorst FB, Monteiro AN, Nathanson KL, … Enigma. (2012). ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat, 33(1), 2–7. doi:10.1002/humu.21628 [PubMed: 21990146]

Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, … ClinGen Sequence Variant Interpretation Working, G. (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med, 20(9), 1054–1060. doi:10.1038/gim.2017.210 [PubMed: 29300386]

Thompson D, Easton DF, & Goldgar DE (2003). A full-likelihood method for the evaluation of causality of sequence variants from family data. Am J Hum Genet, 73(3), 652–655. doi:10.1086/378100 [PubMed: 12900794]

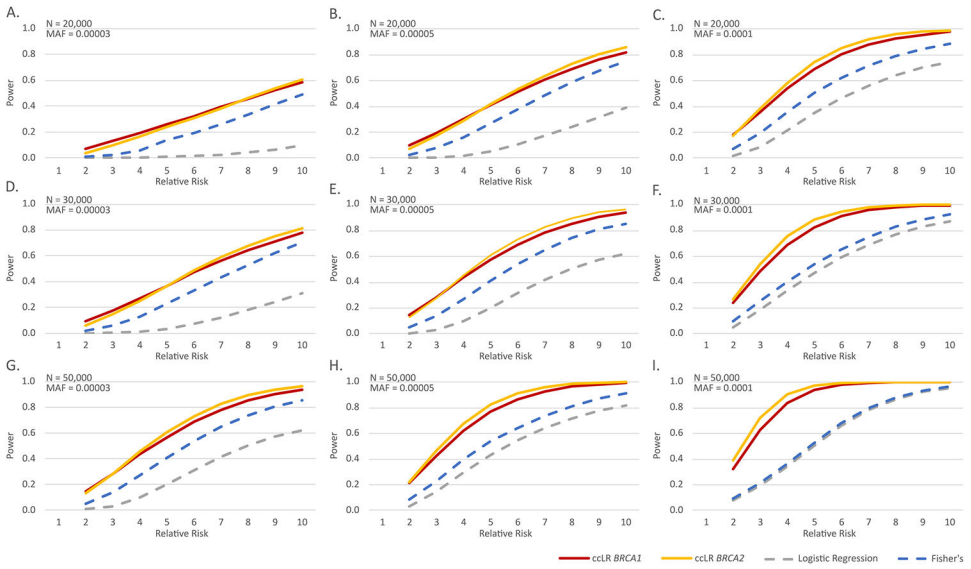

Figure 1.

Performance of the case-control likelihood ratio method and odds ratio analysis in providing at least strong ACMG/AMP evidence in favor of pathogenicity (LR ≥ 18.7), using simulated datasets. Power equals to the probability of reaching at least strong pathogenic ACMG/AMP evidence. Genotype data simulations were carried out for causal variants conferring disease relative risk between 2-10. We performed 10,000 simulations for each case scenario. Results represent simulated case-control data for 20,000 (panels A,B,C) or 30,000 (panels D,E,F) or 50,000 (G,H,I) breast cancer cases and controls and minor allele frequency of 0.00003 (panels A,D,G), 0.00005 (panels B,E,H) or 0.0001 (panels C,F,I). ccLR, case-control likelihood ratio; MAF, minor allele frequency; N, sample size.

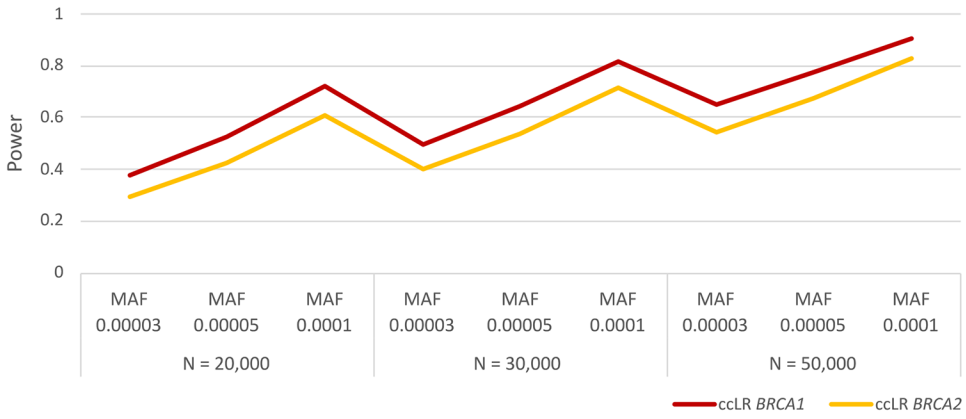

Figure 2.

Performance of the case-control likelihood ratio method in providing ACMG/AMP evidence against pathogenicity, using simulated datasets. Power equals to the probability of reaching at least supporting benign ACMG/AMP evidence (LR ≤ 0.48) when the relative risk was set to 1. We performed 10,000 simulations for each case scenario. Results represent simulated case-control data for 20,000, 30,000 or 50,000 breast cancer cases and controls and minor allele frequency of 0.00003, 0.00005 or 0.0001, ccLR, case-control likelihood ratio; MAF, minor allele frequency; N, sample size.

Table 1.

Statistically significant associated variants with breast cancer risk estimated by logistic regression (based on the ACMG/AMP PS4 criterion)

Gene	Variant_ID (GRCh37/hg19)	HGVS nucleotide	HGVS protein	Variant Carriers		OR (95%CI)	P-value
				Cases	Controls		
				N (Frequency)	N (Frequency)		
BRCA2	chr13_32937506_C_G	c.8167G>C	p. (Asp2723His)	18/72392 ($2.49x10^{-4}$)	1/50680 ($1.97x10^{-5}$)	12.30 (1.66-91.23)	0.014
	chr13_32954180_C_T	c.9154C>T	p. (Arg3052Trp)	10/72563 ($1.38x10^{-4}$)	1/50779 ($1.97x10^{-5}$)	8.32 (1.04-66.48)	0.045

Variant nomenclature according to: *BRCA2* (NM_000059.3, NP_000050.2)

Table 2.

Variants with informative LRs in favor of pathogenicity, estimated by the baseline analysis

Gene	Variant_ID (GRCh37/hg19)	HGVS nucleotide	HGVS protein	Variant Carriers		LR
				Cases	Controls	
				N (Frequency)	N (Frequency)	
BRCA1	chr17_41234451_A_G	c.4327C>T	p.(Arg1443*)	11/72558 ($1.52x10^{-4}$)	3/50781 ($5.91x10^{-5}$)	526.71
	chr17_41215947_T_G	c.5096G>T	p.(Arg1699Leu)	17/72560 ($2.34x10^{-4}$)	3/50780 ($5.91x10^{-5}$)	307.47
BRCA2	chr13_32937506_C_G	c.8167G>C	p.(Asp2723His)	18/72392 ($2.49x10^{-4}$)	1/50680 ($1.97x10^{-5}$)	8193.33
	chr13_32953453_A_G	c.8755-1G>A	p.?	3/72562 ($4.13x10^{-5}$)	-	41.18
	chr13_32954180_C_T	c.9154C>T	p.(Arg3052Trp)	10/72563 ($1.38x10^{-4}$)	1/50779 ($1.97x10^{-5}$)	86.82
	chr13_32968940_A_T	c.9371A>T	p.(Asn3124Ile)	16/72548 ($2.21x10^{-4}$)	-	3530.99

Variant nomenclature according to: *BRCA1* (NM_007294.4, NP_009225.1), *BRCA2* (NM_000059.3, NP_000050.2)

LR, Likelihood ratio.