OPEN ACCESS

Original research

Evaluation of European-based polygenic risk score for breast cancer in Ashkenazi Jewish women in Israel

Hagai Levi,[1,2] Shai Carmi,[3] Saharon Rosset,[4] Rinat Yerushalmi,[5,6] Aviad Zick,[7,8] Tamar Yablonski-Peretz,[7,8] The BCAC Consortium, Qin Wang,[9] Manjeet K Bolla,[9] Joe Dennis,[9] Kyriaki Michailidou ,[9,10] Michael Lush,[9] Thomas Ahearn,[11] Irene L Andrulis,[12,13] Hoda Anton-Culver,[14] Antonis C Antoniou ,[9] Volker Arndt,[15] Annelie Augustinsson,[16] Päivi Auvinen,[17,18,19] Laura Beane Freeman,[11] Matthias Beckmann,[20] Sabine Behrens,[21] Marina Bermisheva,[22] Clara Bodelon,[23] Natalia V Bogdanova,[24,25,26] Stig E Bojesen,[27,28,29] Hermann Brenner,[15,30,31] Helen Byers,[32] Nicola Camp,[33] Jose Castelao,[34] Jenny Chang-Claude,[21,35] María-Dolores Chirlaque,[36] Wendy Chung,[37] Christine Clarke,[38] NBCS Collaborators,[39,40,41,42,43,44,45,46,47,48,49,50] Margriet J Collee,[51] Sarah Colonna,[33] CTS Consortium,[52,53] Fergus Couch,[54] Angela Cox,[55] Simon S Cross,[56] Kamila Czene,[57] Mary Daly,[58] Peter Devilee ,[59,60] Thilo Dork,[25] Laure Dossus,[61] Diana M Eccles ,[62] A. Heather Eliassen,[63,64,65] Mikael Eriksson,[57] Gareth Evans,[32,66] Peter Fasching,[20] Olivia Fletcher,[67] Henrik Flyger,[68] Lin Fritschi,[69] Marike Gabrielson,[57] Manuela Gago-Dominguez,[70] Montserrat García-Closas,[11] Jose Angel Garcia-Saenz,[71] Jeanine Genkinger,[72,73] Graham G Giles,[74,75,76] Mark Goldberg,[77,78] Pascal Guénel,[79] Per Hall,[57,80] Ute Hamann,[81] Wei He,[57] Peter Hillemanns,[25] Antoinette Hollestelle,[82] Reiner Hoppe,[83,84] John Hopper,[75] ABCTB Investigators,[85] Simona Jakovchevska,[86] Anna Jakubowska,[87,88] Helena Jernström,[16] Esther John,[89,90] Nichola Johnson,[67] Michael Jones,[91] Joseph Vijai,[92,93] Rudolf Kaaks,[21] Elza Khusnutdinova,[22,94] Cari Kitahara,[95] Stella Koutros,[11] Vessela Kristensen,[40,50] Allison W Kurian,[89,90] James Lacey,[52,53] Diether Lambrechts,[96,97] Loic Le Marchand,[98] Flavio Lejbkowicz,[99] Annika Lindblom,[100,101] Sibylle Loibl,[102] Adriana Lori,[23] Jan Lubinski,[87] Arto Mannermaa,[17,103,104] Mehdi Manoochehri,[81] Dimitrios Mavroudis,[105] Usha Menon,[106] AnnaMarie Mulligan,[107,108] Rachel Murphy,[109,110] Ines Nevelsteen,[111] William G Newman ,[32,66] Nadia Obi,[112] Katie O'Brien,[113] Ken Offit,[92,93] Andrew Olshan,[114] Dijana Plaseska-Karanfilska,[86] Janet Olson,[115] Salvatore Panico,[116] Tjoung-Won Park-Simon,[25] Alpa Patel,[23] Paolo Peterlongo,[117] Brigitte Rack,[118] Paolo Radice ,[119] Gad Rennert,[99] Valerie Rhenius,[120] Atocha Romero,[121] Emmanouil Saloustros,[122] Dale Sandler,[113] Marjanka K Schmidt,[123,124,125] Lukas Schwentner,[118] Mitul Shah,[120] Priyanka Sharma,[126] Jacques Simard,[127] Melissa Southey,[74,76,128] Jennifer Stone,[75,129] William J Tapper,[62] Jack Taylor,[113,130] Lauren Teras,[23] Amanda E Toland,[131] Melissa Troester,[114] Thérèse Truong,[79] Lizet E van der Kolk,[132] Clarice Weinberg,[133] Camilla Wendt,[134] Xiaohong Rose Yang,[11] Wei Zheng,[135] Argyrios Ziogas,[14] Alison M Dunning,[120] Paul Pharoah,[9,120,136] Douglas F Easton,[9,120] Shay Ben-Sachar,[6,137] Naama Elefant,[137,138] Ron Shamir,[1] Ran Elkon [2]

ABSTRACT

Background Polygenic risk score (PRS), calculated based on genome-wide association studies (GWASs), can improve breast cancer (BC) risk assessment. To date, most BC GWASs have been performed in individuals of European (EUR) ancestry, and the generalisation of EUR-based PRS to other populations is a major challenge. In this study,

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmg-2023-109185).

For numbered affiliations see end of article.

Correspondence to
Prof. Ran Elkon, Department of Human Molecular Genetics and Biochemistry, Tel Aviv University, Tel Aviv, Israel;
ranel@tauex.tau.ac.il
Prof. Ron Shamir;
rshamir@tau.ac.il
Dr. Naama Elefant;
naamael1@clalit.org.il

Received 26 January 2023
Accepted 28 May 2023

Check for updates

To cite: Levi H, Carmi S, Rosset S, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmg-2023-109185

we examined the performance of EUR-based BC PRS models in Ashkenazi Jewish (AJ) women.

Methods We generated PRSs based on data on EUR women from the Breast Cancer Association Consortium (BCAC). We tested the performance of the PRSs in a cohort of 2161 AJ women from Israel (1437 cases and 724 controls) from BCAC (BCAC cohort from Israel (BCAC-IL)). In addition, we tested the performance of these EUR-based BC PRSs, as well as the established 313-SNP EUR BC PRS, in an independent cohort of 181 AJ women from Hadassah Medical Center (HMC) in Israel.

Results In the BCAC-IL cohort, the highest OR per 1 SD was 1.56 (±0.09). The OR for AJ women at the top 10% of the PRS distribution compared with the middle quintile was 2.10 (±0.24). In the HMC cohort, the OR per 1 SD of the EUR-based PRS that performed best in the BCAC-IL cohort was 1.58±0.27. The OR per 1 SD of the commonly used 313-SNP BC PRS was 1.64 (±0.28).

Conclusions Extant EUR GWAS data can be used for generating PRSs that identify AJ women with markedly elevated risk of BC and therefore hold promise for improving BC risk assessment in AJ women.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Genome-wide association studies (GWASs) on breast cancer (BC) were, to date, mainly done on women of European (EUR) ancestry, and recent studies showed that polygenic risk score (PRS) based on these GWAS can effectively stratify EUR women according to their BC risk.

⇒ However, PRS performance declines with the increase of the genetic distance between the population used in the GWAS and the population on which the PRS is applied.

WHAT THIS STUDY ADDS

⇒ Here, we systematically evaluated the performance of EUR-based BC PRS on Ashkenazi Jewish (AJ) women from Israel. Our results demonstrate that extant EUR GWAS data can be used for generating PRSs that identify AJ women with markedly elevated risk of BC.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our study suggests the possibility of personalised BC screening programmes in Israel that could potentially improve early detection of BC while reducing overdiagnosis.

INTRODUCTION

Breast cancer (BC) is the most common cancer diagnosed among women in Western countries including Israel, where some 5500 BC cases are diagnosed annually.[1] An early diagnosis of BC leads to a higher cure rate and improved survival. Thus, it is essential to develop accurate risk prediction methods for identifying women at high risk of BC. An ongoing debate over the optimal approach to BC screening has led to discordant professional society recommendations.[2] Two fundamental questions—whether to screen annually or at a lower frequency and whether screening should start at the age of 40 or at a later point in life—have been debated for over 20 years.[2–4] In Israel, health providers generally recommend biennial mammography screening starting at age 50 for women, except for those with a family history of relevant cancer or carriers of pathogenic variants in BC-associated genes, who are recommended to start earlier and screen more frequently. This 'one size fits all' approach to nationwide BC screening might be suboptimal as it assumes an equal risk of developing BC to most women. A personalised screening strategy based on individual risk could enhance the early detection of BC, decrease the harm of overdiagnosis and unnecessary screens and improve the use of medical resources.[5]

Rare pathogenic variants in the *BRCA1* and *BRCA2* genes confer high risk of developing BC but account for only a small proportion (<10%) of BC cases in the general population.[6 7] In contrast, numerous common BC susceptibility variants have been discovered over the last decade through genome-wide association studies (GWASs).[8 9] Each of these variants confers only a small risk individually, but their combined effect, commonly estimated by a polygenic risk score (PRS), can be substantial.[10 11] Importantly, recent studies on women of European (EUR) ancestry demonstrated that PRS models can effectively stratify women according to their BC risk. In particular, women in the top 1% of an optimised PRS model, based on 313 BC risk SNPs, have >4-fold elevated risk of developing BC compared with those in the middle quintile (40%–60%).[12] This amounts to ~3.5% of BC incidence falling in this top percentile. In terms of absolute risk, women in the top 1% had a lifetime risk of 32.6%, similar to the risk conferred by pathogenic variants in some of the moderate-impact BC predisposition genes such as *ATM* and *CHEK2*.[13 14] These results show that PRS models can be powerful BC risk predictors and hold a promise for improving BC prevention programmes and assisting in early diagnosis of BC. These advances have led to the launching of clinical trials in which prevention programmes are guided by novel personalised risk prediction models that integrate PRS information.[5]

Unfortunately, PRS performance declines substantially as the genetic distance increases between the *discovery population* (used in the GWAS) and the *target population* (on which the PRS is used).[15] The decline in performance is due to differences in effect sizes, allele frequencies and linkage disequilibrium (LD) patterns between populations. Since the vast majority of the currently available GWAS was done on people of EUR ancestry, the clinical usefulness of PRS models in other populations is limited. The decline in PRS performance in non-EUR populations might aggravate disparities in clinical genetics care between ethnic groups.[15] Several studies showed that BC PRS generated from EUR GWAS summary statistics (EUR BC PRS) has lower performance on non-EUR women (eg, African–Americans).[16 17] Yet, some studies demonstrated that EUR BC PRS performance on Latin American women—a large group with variable levels of Indigenous American, EUR and African ancestries—was similar to its performance on women of EUR ancestry.[18]

The population in Israel is highly heterogeneous, with Ashkenazi Jews (AJ) being one of its largest ethnic group. Given the relatively low genetic distance between the EUR and AJ populations,[19 20] we hypothesised that EUR BC PRS could be used to develop clinically relevant PRS models for AJ women in Israel. To that end, we used the massive genetic resource generated by the multinational Breast Cancer Association Consortium (BCAC),[8] which also contains an Israeli cohort, to conduct a systematic evaluation of the predictive performance of EUR BC PRS models on Israeli AJ women. We demonstrate that an EUR BC PRS can be adjusted to the AJ population and identify women with markedly elevated BC risk (OR >2.0 for AJ women in the top 10% compared with the middle quintile). We substantiate these findings using an independent cohort of AJ Israeli women.

MATERIALS AND METHODS

BCAC dataset

We analysed 132 335 EUR women from the BCAC: 72 899 cases and 59 436 controls. In addition, the BCAC includes an Israeli cohort (BCINIS/BCAC cohort from Israel (BCAC-IL)) of 2161

women: 1437 cases and 724 controls. According to the 'ethnOt' field in the BCAC phenotype file, all the women in the BCAC-IL cohort are tagged as 'Jewish Ashkenazi'. In addition, there are 73 samples in the EUR cohort that are tagged as AJ.

All samples analysed were genotyped using the OncoArray chip. In our analysis, we used an imputed version of the data provided by BCAC. The imputation was done against the 1000 Genomes Project imputation panel. In BCAC-IL, 119 (5.5%) *BRCA1/2* mutation carriers were identified by a self-reporting field provided by the BCAC.

Hadassah Medical Center (HMC) cohort

The HMC dataset contains 181 Israeli AJ women, of whom 118 are BC cases under the age of 45 years and 63 are controls older than 75 years. We validated either by sequencing or genotyping that none of the women carried one of the three AJ founder mutations in *BRCA1/2*. Samples were genotyped using the Axiom PMDA chip. Likely pathogenic variants in selected genes are covered by this chip.Three women carried such variants in *BRCA1/2*, and none bore pathogenic variants in other BC susceptibility genes.

We phased the data using SHAPEIT2[21] and imputed it using IMPUTE2.[22] The imputation reference panel was generated using SHAPEIT2 from the EUR samples from the 1000 Genomes Project (n=503). Using PLINK, we filtered out SNPs with uncertainty greater than 0.1.

For the evaluation of the 313 PRS, we were able to map 304 SNPs, of which 248 were called (either by genotyping or imputation) in more than 90% of the samples.

Quality check (QC) of discovery sets

We performed QC on each discovery set using PLINK.[23 24] We kept only SNPs with minor alllele frequency (MAF) of ≥5%, HWE p value of ≥1e-6 and missing rate of ≤10%. In addition, we kept only samples where less than 10% of SNPs present in the set were missing. In addition, we filtered out ambiguous and duplicated alleles. A total of 4 617 515 SNPs remained in the BCAC-EUR cohort and 4 973 754 SNPs in the BCAC-EUR cohort after exclusion of the Polish samples.

Similarly, we used PLINK to perform QC on each target set. We kept the same HWE, missing rate and MAF thresholds as in the discovery set, filtered out duplicated alleles and kept samples where less than 10% of SNPs present in the set were missing. This process left 5 549 031 and 5 704 856 SNPs on the entire Israeli (BCAC-IL) and, used as a control, the entire Polish (BCAC cohort from Poland (BCAC-PL)) cohorts, respectively. Note that in cross-validation (CV) analyses (see further), to avoid information leakage, we performed QC on each fold separately, so the number of SNPs in each fold slightly varied, depending on the subset of individuals in the fold.

GWAS analysis

We ran GWAS analyses for two sets: EUR (n=132 335) and EUR without the PL cohort (n=128 153). Both sets did not contain the BCAC-IL women. For each analysis, we ran PCA and GWAS using PLINK2 (with the --glm command)[24] and generated GWAS summary statistics with the first five principal components as covariates.

Nested CV

We applied nested CV for optimising PRS models generated by four different methods (pruning and thresholding using European linkage disequilibrium (P+T EUR-LD), pruning and thresholding using linkage disequilibrium of the target population (P+T target set LD), LDpred2 and Lassosum; see futher). Specifically, for each PRS method, we split the BCAC-IL cohort into six sets (each of size 360). Next, we held out one set (red box in figure 1) and used the other five sets (green boxes in figure 1) to perform a standard 5-fold CV, in which four out of five parts (training set; light green) are used to derive PRS

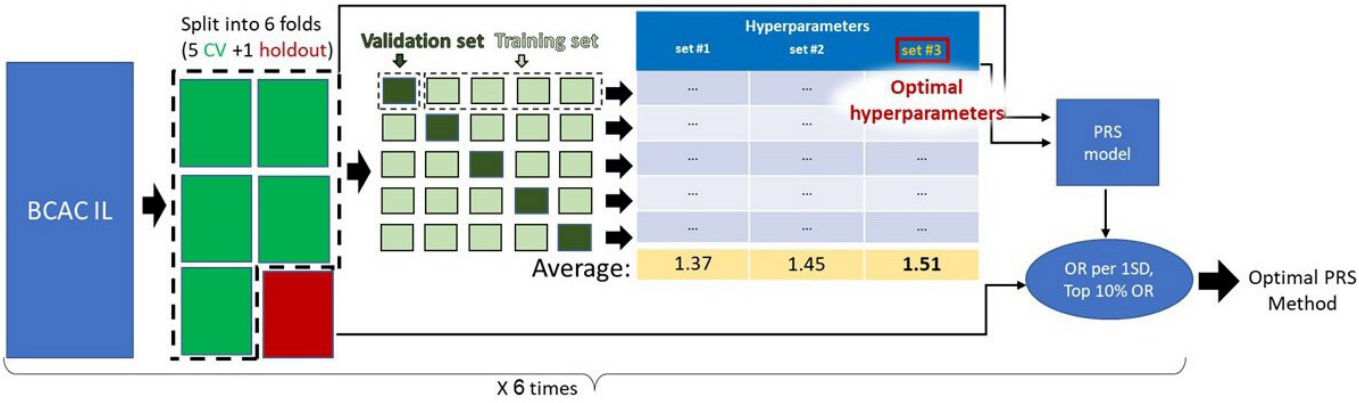

Figure 1 Outline of the CV scheme used to construct and evaluate the PRS models. We applied nested CV to optimise PRS models on the AJ cohort. Specifically, we split the BCAC-IL cohort into six sets (each of size 360). Next, we held out one set (red box) and used the other five sets (green boxes) to perform a standard fivefold CV in which four out of five parts (training set, light green) are used to derive PRS models with different predefined sets of hyperparameters (see the Materials and methods section), and then the resulting models are applied on the fifth part (validation set, dark green). For each PRS model, we measured the OR per 1 SD and the top 10% OR. After iterating over the five combinations of training and test sets, we chose the hyperparameter set with the highest average performance (see detailed ranking criteria in the Materials and methods section). Then, we retrained a PRS model on the five CV folds with the chosen hyperparameters. Finally, we applied the resulting PRS model on the holdout set and measured the OR per 1 SD and for the top 10% OR. We repeated this entire process six times, each with a different holdout set. We applied this scheme to each of the four PRS methods included in our analysis (P+T EUR-LD, P+T target LD, LDpred2 and Lassosum). The method that obtained the highest average performance on the six holdout sets is selected as the best one. AJ, Ashkenazi Jewish; BCAC, Breast Cancer Association Consortium; BCAC-IL, BCAC cohort from Israel; CV, cross validation; PRS, polygenic risk score; P+T EUR-LD, pruning and thresholding using European linkage disequilibrium; P+T target LD, pruning and thresholding using linkage disequilibrium of the target population.

models with different predefined sets of hyperparameters, and then the resulting models are applied on the fifth part (validation set, dark green). For each model, we measured the OR per 1SD (using logistic regression with the first six principal components as covariates) and OR of women at the top 10% of the PRS distribution compared with the middle quintile. After iterating over the five combinations of training and test sets, we chose the hyper-parameter set that performed the best on average (see detailed ranking criteria below). Then, using these optimal hyper-parameters, we retrained a PRS model on the entire five CV folds (green boxes). Finally, we applied the resulting PRS model on the holdout set and measured the OR per 1SD and top 10% OR. We repeated this entire process six times, each with a different holdout set. The method with the highest average on the six holdout sets is nominated as the best one.

In all analyses, PRS were standardised to the control samples of the respective target set.

Criteria for choosing an optimal PRS model

We tested the performance of each PRS method with a predefined set of hyper-parameters (see below). For each method, we ranked runs with different hyper-parameters using two metrics: (1) OR per 1SD and (2) top-10% OR, and combined these rankings by taking their sum. We broke ties using the model with the higher OR per 1SD, as this metric is less noisy.

Pruning and thresholding using European linkage disequilibrium

Using PLINK, we clumped the GWAS results according to LD in the EUR population derived from the EUR samples in the 1000 Genomes Project (n=503) with r^2 =0.2. Then, we filtered the remaining SNPs based on a significance threshold (T). We tested the following threshold values T:

$$5 \cdot 10^{-8},\ 10^{-7},\ 10^{-6},\ 10^{-5}, 10^{-4}, 10^{-3},$$
$$5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5$$

For each T, we calculated the PRS from the SNPs that passed the filtering.

Pruning and thresholding using linkage disequilibrium of the target population

Here, when applying LD clumping in PLINK, we used LD inferred from the training set. The training set comes from the same population as the target set. Namely, in each fold of the CV, LD was calculated using the genotype data of individuals in the training set. On the HMC cohort, we used the LD from the BCAC-IL cohort. The subsequent steps of the analysis are identical to the P+T EUR-LD method.

LDpred2

LDpred2 (grid mode) generates a PRS model using SNP correlations calculated from genotype data (ie, the training set). We supplied LDpred2 with a training set that comes from the same population as the target set, as for the P+T method previoously. We ran LDpred2 using the set of hyper-parameter values for the proportion of causal variants, heritability, and sparseness that were recommended by.[25] The rest of the hyper-parameters were left with their default values.

Lassosum

Lassosum generates a PRS model using a reference panel calculated from genotype data (ie, the training set). We supplied Lassosum with a training set that comes from the same population as the target set, as above. We ran Lassosum using LD blocks option 'EUR.hg19' and the values of the regularisation hyper-parameter *s* that were recommended by.[25] The rest of the hyper-parameters were left with their default values.

313-SNPs EUR BC PRS model

We downloaded the weights for the EUR PRS model from.[12] Originally, the model consisted of 313 SNPs. In the imputed data, we managed to retain all the 313 SNPs for the BCAC-IL cohort and 304 SNPs for the HMC cohort. Risk scores for each sample were calculated using PLINK.

RESULTS

We set to build and evaluate EUR-based BC PRS for AJ women from Israel. For this task, we used an Israeli cohort of 2161 AJ women (1437 BC cases and 724 controls) that is a part of the BCAC (Methods). We refer to the Israeli sub-cohort of the BCAC as *BCAC-IL*. In order to avoid inflation of the predictive performance, the target set should be independent of the discovery set. Therefore, we could not reliably assess how the commonly used EUR BC 313-SNP PRS[12] performs on the BCAC-IL cohort since this PRS was derived from BCAC GWAS, which included the BCAC-IL cohort. Therefore, we first removed the Israeli women from the EUR BCAC cohort and recomputed GWAS summary statistics using only data from the 132 335 non-Israeli EUR women (72 899 cases and 59 436 controls; Methods). A PCA on the BCAC genotype data confirmed the close genetic relatedness of AJ to the EUR population (figure 2).

Next, we set to adapt an EUR-based BC PRS for AJ women from Israel. We constructed PRS models from the GWAS we generated using four different methods: P+T[26 27] EUR-LD; P+T using LD of the target (AJ) population (P+T target LD), LDpred2[28] and Lassosum.[29] We used two metrics to evaluate the models produced by these algorithms: (1) the OR per 1 unit SD and (2) the OR of women in the top 10% of the PRS distribution relative to those in the middle quantile (top 10% OR). We constructed and evaluated the PRS models using a nested CV scheme (see the Materials and methods section). The outline of our evaluation procedure is depicted in figure 1.

Of the four methods we tested, Lassosum performed best, obtaining an OR per 1 SD of 1.56 (±0.09) and a top 10% OR of 2.1 (±0.24) (table 1 and online supplemental figure S1; see online supplemental table S1 for performance on the validation sets in the CV). We also examined the OR of other deciles of the PRS (compared with the middle quintile) and found that it increased nearly monotonically (figure 3). Further, women in the top 10% were estimated to have fourfold higher OR for BC compared with AJ women in the bottom 10% (figure 3, online supplemental figure S2). Notably, these top and bottom 10% OR estimates that we obtained for AJ women were comparable to those reported using EUR BC PRS on women of EUR ancestry.[12]

Next, to estimate the decline in the performance of EUR-based BC PRS when applied to AJ women relative to women of EUR ancestry, we compared the performance obtained on women from BCAC-IL and women from BCAC-PL. We compared BCAC-IL to the Polish cohort as the AJ population is mainly from Eastern Europe. Specifically, we now excluded the Polish and Israeli samples from the BCAC discovery set and reran a GWAS analysis (see the Materials and methods section). Then, we applied the same nested CV scheme to the BCAC-PL (4537 women: 2318 cases and 2219 controls) and BCAC-IL cohorts using the same four PRS methods as previously discussed. As expected, the results obtained on BCAC-PL were mostly higher

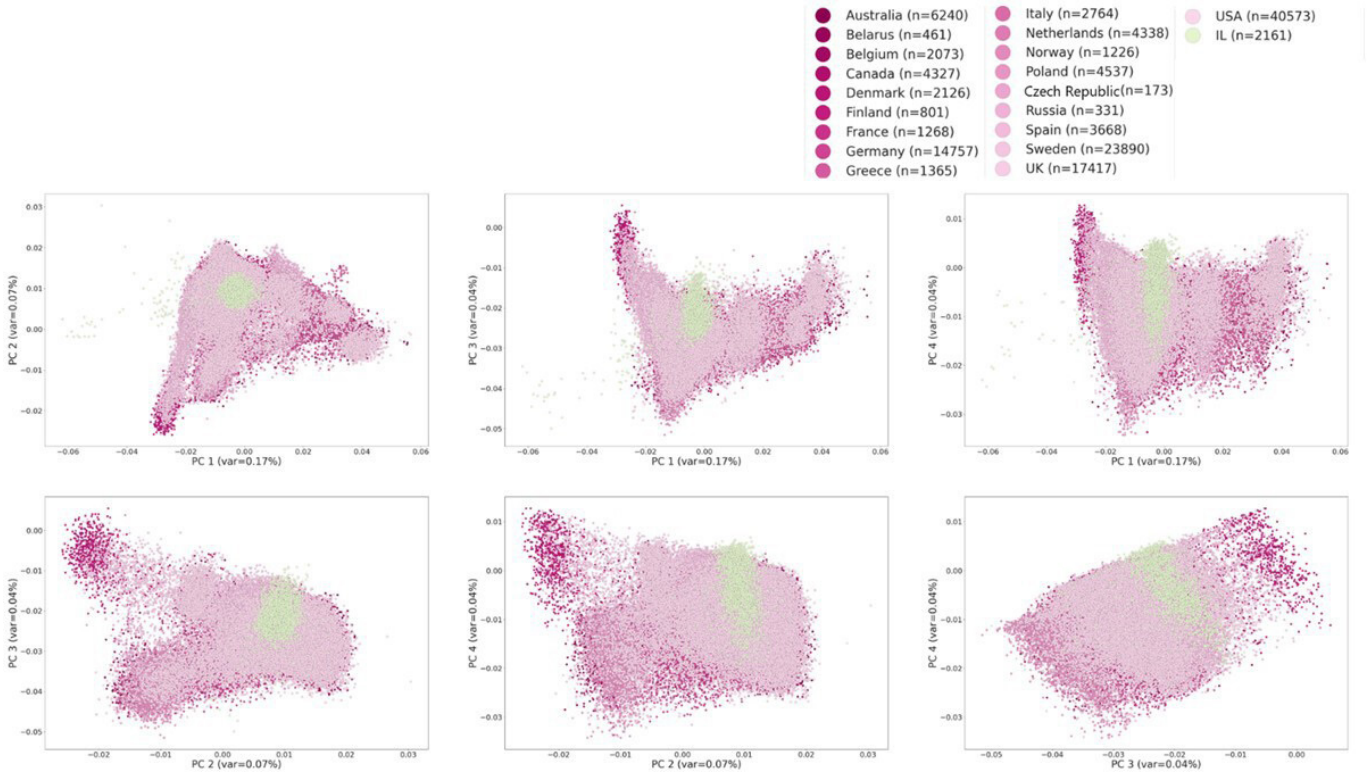

Figure 2 PCA on the EUR BCAC dataset. PCA was computed without BCAC-IL, which was later projected on it. Shown are two-dimesnional projections of PCs 1–4. The plot demonstrates high genetic similarity between the EUR and Israeli AJ populations. AJ, Ashkenazi Jewish; BCAC, Breast Cancer Association Consortium; EUR, European; BCAC-IL, BCAC cohort from Israel; PC, principal component; PCA, principal component analysis

than those on BCAC-IL, reflecting the greater genetic distance of the AJ population from the EUR population (table 2; see online supplemental table S2 for performance on the validation sets).

Pathogenic variants in *BRCA1/2* confer a very high risk of BC. In BCAC-IL, 119 women were flagged as carriers of the *BRCA1/2* mutation (106 cases and 13 controls). To test the impact of the inclusion of these *BRCA1/2* carriers on PRS performance, we measured the performance of the P+T EUR-LD PRS on the BCAC-IL cohort after excluding these 119 samples. As shown in online supplemental figure S3, there was no significant difference between the two runs in the estimates for the OR per 1 SD and the top 10% OR.

To further examine the performance of EUR-based BC PRS on AJ women in Israel, we genotyped an independent sample of 181 Israeli AJ women recruited at the HMC in Jerusalem. This cohort comprises 118 patients with BC and 63 healthy women as controls. All the patients in the HMC cohort were diagnosed with BC at an early age (<45 years old) and tested negative for the three AJ founder variants in *BRCA1/2*. The controls were women aged 75 years and over who were never diagnosed with

Table 1 Performance of different PRS methods on the BCAC-IL cohort

Method	OR per 1SD	Top 10% OR	SNPs (n)
P+T EUR-LD	1.43±0.08	1.43±0.27	1483±502
P+T target set LD	1.39±0.07	1.65±0.25	8591±5136
LDpred2	1.31±0.07	1.96±0.43	740 919±18
Lassosum	1.56±0.09	2.1±0.24	65 632±16 126

Performance of different PRS methods on the BCAC-IL cohort. ORs per 1 SD and top 10% OR were obtained using the nested CV outlined in figure 1. The last column is the average number of SNPs. Shown are means and SEMs over the six holdout sets. BCAC-IL, BCAC cohort from Israel; CV, cross validation; PRS, polygenic risk score; P+T EUR-LD, pruning and thresholding using European linkage disequilibrium; P+T target set LD, pruning and thresholding using linkage disequilibrium of the target population.

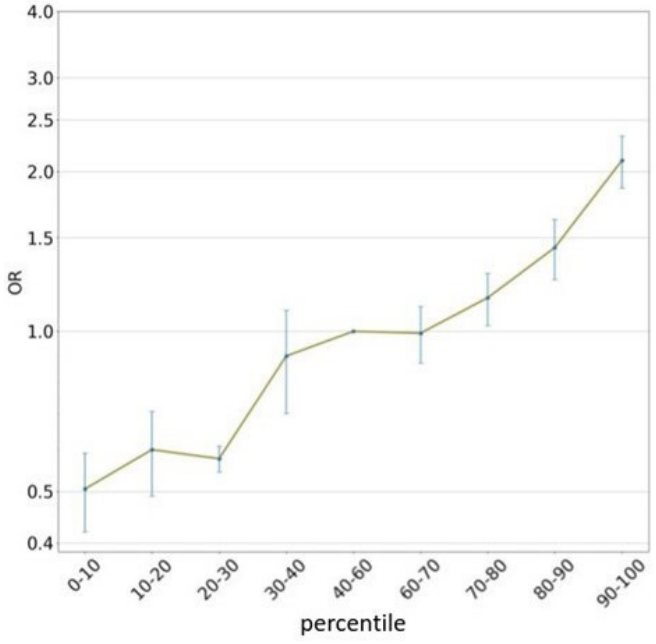

Figure 3 OR of BC risk as a function of BC PRS deciles. PRS was generated using Lassosum. OR is measured relative to scores in the middle PRS quintile (40%–60%). Shown are means and SEMs over the six holdout sets. BC, breast cancer; PRS, polygenic risk score.

Table 2 Performance of EUR PRS when excluding the Polish and Israeli cohorts from the discovery set and using these respective populations as the target cohorts

Method	Target set cohort	OR per 1 SD	Top 10% OR
P+T (EUR)	BCAC-IL	1.37±0.06	2.41±0.76
	BCAC-PL	1.46±0.06	2.2±0.15
P+T target LD	BCAC-IL	1.36±0.03	1.64±0.17
	BCAC-PL	1.5±0.07	1.85±0.16
LDPred	BCAC-IL	1.31±0.1	1.56±0.17
	BCAC-PL	1.17±0.06	1.82±0.33
Lassosum	BCAC-IL	1.50±0.06	1.82±0.3
	BCAC-PL	1.53±0.05	3.11±0.45

Shown are average ORs per 1 SD and top 10% ORs. Errors were measured using SEM for the six holdout sets.
BCAC-IL, BCAC cohort from Israel; BCAC-PL, BCAC cohort from Poland; P+T, pruning and thresholding; P+T target LD, pruning and thresholding using linkage disequilibrium of the target population.

cancer. We first evaluated how the EUR BC 313-SNP PRS (313 PRS)[12] performs on this cohort. Notably, the OR per 1 SD of the 313-PRS model was 1.64±0.28 on the HMC cohort, similar to the effect reported for this PRS model on EUR women (1.65 OR per 1 SD, 95% CI 1.59 to 1.79)[12]. For comparison, we also measured the performance of the 313 PRS on the BCAC-IL cohort and obtained OR per 1 SD of 1.77±0.09. This result is likely inflated due to the inclusion of the BCAC-IL in the discovery set used to infer the 313-PRS model. On the other hand, the OR estimate for the BCAC-IL cohort was less noisy than the one obtained in the HMC cohort due to its larger size (the BCAC-IL cohort is >10 times larger than the HMC).

Last, we evaluated Lassosum—the best performing method on BCAC-IL—on HMC. Using the EUR GWAS we generated, we trained the PRS model on the BCAC-IL cohort in fivefold CV (online supplemental figure S4). Applying this PRS to the HMC cohort yielded an OR of 1.58±0.27 per 1 SD (number of SNPs: 4540).

Overall, the results obtained on the HMC cohort reaffirm that EUR-based BC PRS has clinically relevant predictive capacity for Israeli AJ women.

DISCUSSION

PRS models have the potential to play an essential role in detecting women's risk of developing BC. Nevertheless, at present, clinically relevant BC PRS models have been constructed primarily for women of EUR ancestry, for whom large discovery sets are currently available.[15] Whether these models perform well on women of other ancestries and how they can be adapted for women of other ancestries are key open questions. Our study focuses on a major ethnic group in Israel, the Ashkenazi Jewish (AJ) population, which is genetically close to the EUR population. We tested whether a large number of available EUR genotypes of patients with BC and healthy women could be used to generate a clinically relevant BC PRS model for AJ women in Israel.

We evaluated four PRS methods on the Israeli cohort from BCAC (BCAC-IL) and found that Lassosum had the best prediction performance. Notably, there was a fourfold increased BC risk between women in the top and bottom 10% of the PRS distribution (figure 3 and online supplemental figure S2), suggesting that BC PRS models derived from EUR GWAS may help fit personalised recommendations for BC preventive screening for Israeli AJ women. The results obtained on the independent HMC cohort further support this conclusion. While the BCAC-IL cohort is too small to calculate reliable risk estimates for women in the top 5% and 1%, the monotonic increase of the OR with the deciles (figure 3) and results by similar BC PRS on EUR women[12] suggest that this model has the capacity to identify at its very top percentiles AJ women with even higher risk of developing BC. Follow-up studies with larger samples of AJ women are needed to substantiate this expectation.

Notably, the HMC cohort has extreme age differences between the case and control arms: healthy women are older than 75 and patients with BC are younger than 45. Thus, the high prediction performance of the BC PRS models on this cohort suggests that EUR-based PRS models may also be relevant for detecting early-onset cases of BC among Israeli AJ women. In addition, these results indicate that for AJ women, low-impact common genetic variants—and not only pathogenic variants with high and moderate impact—play an important role in predisposing women to early-onset BC.

One limitation of our study is that *BRCA1/2* carriers were identified in the BCAC-IL only by self-reporting. Thus, there might be additional women carrying *BRCA1/2* variants who were marked as non-carriers as identified by.[30] Still, our analysis indicates that inclusion of a limited group of patients who carry pathogenic variants in *BRCA1/2* genes does not have a significant impact on the PRS performance (online supplemental figure S3).

As the patients with BC at HMC were under 45, we could not directly generalise the prediction performance obtained on HMC for older AJ Israeli patients. However, online supplemental figure S5 indicates that there is no substantial difference in the PRSs between age groups of BCAC-IL patients, consistent with previous findings on EUR population.[12]

Our finding indicates that the currently available EUR BC GWAS data can be used to generate BC PRS models for Israeli AJ women. Nevertheless, this observation should not nullify the effort to genotype a higher number of individuals in Israel. First, an increased sample of AJ women would provide more accurate risk estimates for women at the top tail of the PRS distribution. Second, the Israeli population is highly heterogeneous, comprising many different ethnic groups, including North African and Middle Eastern Jews, as well as Palestinians, Druzes and Bedouins. Moreover, many of the younger generation in Israel are of mixed ethnicities. Therefore, to cover additional groups in nationwide BC prevention programmes, large-scale genotyping initiatives should include women from other ethnic groups in Israel, including admixed groups. Such data would allow a systematic evaluation of EUR-derived PRS BC models on non-AJ Israeli populations. We hope that this study will expedite the realisation of the potential for personalised BC risk stratification and encourage the development of screening protocols for high-risk women.

Author affiliations
[1]The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
[2]Department of Human Molecular Genetics and Biochemistry, Tel Aviv University, Tel Aviv, Israel
[3]Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel
[4]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel
[5]Institute of Oncology, Davidoff Cancer Center, Rabin Medical Center, Beilinson Hospital, Petah Tikva, Israel
[6]Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel
[7]Department of oncology, Hadassah Medical Center, Jerusalem, Israel
[8]Hebrew University of Jerusalem, Jerusalem, Israel
[9]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[10]Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus

[11]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[12]Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada
[13]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
[14]Department of Medicine, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA
[15]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[16]Oncology, Department of Clinical Sciences in Lund, Lund University, Lund, Sweden
[17]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland
[18]Institute of Clinical Medicine, Oncology, University of Eastern Finland, Kuopio, Finland
[19]Department of Oncology, Cancer Center, Kuopio University Hospital, Kuopio, Finland
[20]Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
[21]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[22]Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
[23]Department of Population Science, American Cancer Society, Atlanta, GA, USA
[24]Department of Radiation Oncology, Hannover Medical School, Hannover, Germany
[25]Gynaecology Research Unit, Hannover Medical School, Hamburg, Germany
[26]N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus
[27]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
[28]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
[29]Faculty of Health and Medical Sciences, Copenhagen, Denmark
[30]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[31]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
[32]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
[33]Department of Internal Medicine and Huntsman Cancer Institute, University of Utah, Salt lake city, UT, USA
[34]Oncology and Genetics Unit, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain
[35]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[36]CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain
[37]Departments of Pediatrics and Medicine, Columbia University, New York, NY, USA
[38]Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia
[39]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway
[40]Institute of Clinical Medicine, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
[41]Department of Research, Vestre Viken Hospital, Drammen, Norway
[42]Section for Breast- and Endocrine Surgery, Department of Cancer, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital, Oslo, Norway
[43]Department of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway
[44]Department of Pathology, Akershus University Hospital, Lorenskog, Norway
[45]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
[46]Department of Oncology, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Radiumhospitalet, Oslo, Norway
[47]National Advisory Unit on Late Effects after Cancer Treatment, Oslo University Hospital, Oslo, Norway
[48]Department of Oncology, Akershus University Hospital, Lorenskog, Norway
[49]Oslo Breast Cancer Research Consortium, Oslo University Hospital, Oslo, Norway
[50]Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway
[51]Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, Netherlands
[52]Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA, USA
[53]City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA, USA
[54]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
[55]Department of Oncology and Metabolism, Sheffield Institute for Nucleic Acids (SInFoNiA), University of Sheffield, Sheffield, UK
[56]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK
[57]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[58]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA
[59]Department of Pathology, Leiden University Medical Center, Leiden, Netherlands
[60]Department of Human Genetics, Leiden University Medical, Leiden, Netherlands
[61]Nutrition and Metabolism Section, International Agency for Research on Cancer (IARC-WHO), Lyon, France
[62]Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK
[63]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[64]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[65]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[66]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
[67]The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK
[68]Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
[69]School of Population Health, Curtin University, Perth, Western Australia, Australia
[70]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
[71]Medical Oncology, Hospital Clinico San Carlos, Madrid, Spain, USA
[72]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA
[73]Herbert Irving Comprehensive Cancer Center, New York, New York, USA
[74]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia
[75]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
[76]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
[77]Department of Medicine, McGill University, Montreal, Quebec, Canada
[78]Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montreal, QU, Canada
[79]Team 'Exposome and Heredity', CESP, Gustave Roussy, INSERM, University Paris-Saclay, UVSQ, Villejuif, France
[80]Department of Oncology, Södersjukhuset, Stockholm, Sweden
[81]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
[82]Medical Oncology, Erasmus Medical Center, Rotterdam, Netherlands
[83]Dr Margarete Fischer Bosch Institute of Clinical Pharmacology, Stuttgart, Germany
[84]University of Tübingen, Tubingen, Germany
[85]Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia
[86]Research Centre for Genetic Engineering and Biotechnology 'Georgi D. Efremov', Skopje, North Macedonia
[87]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
[88]Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland
[89]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA
[90]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA
[91]Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, UK
[92]Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[93]Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[94]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia
[95]Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
[96]Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium
[97]VIB Center for Cancer Biology, VIB, Leuven, Belgium
[98]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
[99]Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel
[100]Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

[101]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden
[102]German Breast Group, GmbH, Neu Isenburg, Germany
[103]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland
[104]Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland
[105]Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece
[106]MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College, London, UK
[107]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada
[108]Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada
[109]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[110]Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada
[111]Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium
[112]Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[113]Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA
[114]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
[115]Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA
[116]Dipertimento Di Medicina Clinca e Chirurgia, Federico II University, Naples, Italy
[117]Genome Diagnostics Program, IFOM ETS - the AIRC Institute of Molecular Oncology, Milan, Italy
[118]Department of Gynaecology and Obstetrics, University Hospital Ulm, Ulm, Germany
[119]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy
[120]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
[121]Laboratorio de Oncología Molecular, Hospital Clínico San Carlos, Madrid, Spain
[122]Department of Oncology, University Hospital of Larissa, Larissa, Greece
[123]Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, Netherlands
[124]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, Netherlands
[125]Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands
[126]Department of Internal Medicine, Division of Medical Oncology, University of Kansas Medical Center, Westwood, KS, USA
[127]Genomics Center, Molecular Medicine, Université Laval, Quebec, Quebec, Canada
[128]Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia
[129]Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, Western Australia, Australia
[130]Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA
[131]Department of Cancer Biology and Genetics, The Ohio State University, Columbus, OH, USA
[132]Family Cancer Clinic, The Netherlands Cancer Institute, Amsterdam, Netherlands
[133]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA
[134]Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden
[135]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA
[136]Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, CA, USA
[137]Clalit Research Institute, Clalit Health Services, Ramat Gan, Israel
[138]Department of Genetics, Hadassah Medical Center, Jerusalem, Israel

Twitter Kyriaki Michailidou @KyrMichailidou

Acknowledgements We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who enabled this work to be carried out. ABCFS thanks Maggie Angelakos, Judi Maskiell and 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 and Mythily Sachchithananthan. Samples are made available to researchers on a non-exclusive basis. BBCS thanks Eileen Williams, Elaine Ryder-Mills and Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie and Rachel Lloyd. The BCINIS study would not have been possible without the major contribution of Ms H Rennert and the contributions of Dr M Pinchev, Dr O Barnet, Dr N Gronich, Dr K Landsman, Dr A Flugelman, Dr WSaliba, Dr E Liani, Dr I. Cohen, Dr S Kalet and Dr V Friedman of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. BIGGS thanks Niall McInerney, Gabrielle Colleran, Andrew Rowan and 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. CBCS thanks the study participants, coinvestigators, 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 and Konstantinos Pompodakis. CGPS thanks staff and participants of the Copenhagen General Population Study. The authors thank the following for the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank and Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. The Danish Breast Cancer Cooperative Group is acknowledged for its provision of clinical case data. CNIO-BCS thanks Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez and 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. 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 programme. The authors thank the California Teachers Study (CTS) Steering Committee, which is responsible for the formation and maintenance of the study within which this research was conducted. A full list of 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 European Prospective Investigation into Cancer and Nutrition (EPIC). ESTHER thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier and 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 (Hiltrud Brauch, Reiner Hoppe and Wing-Yee Lo), Department of Internal Medicine; Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany (YDK, Christian Baisch); Institute of Pathology, University of Bonn, Germany (Hans-Peter Fischer); Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany (UH); Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum, Bochum, Germany (Thomas Brüning, Beate Pesch, Sylvia Rabstein and Anne Lotz); and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany (Volker Harth). GLACIER thanks Kelly Kohut, Patricia Gorman and Maria Troy. HABCS thanks Peter Schürmann, Peter Hillemanns, Natalia Bogdanova, Michael Bremer, Johann Karstens, Hans Christiansen and the Breast Cancer Network in Lower Saxony for continuous support. HMBCS thanks Peter Hillemanns, Hans Christiansen and Johann H Karstens. HUBCS thanks Darya Prokofyeva and Shamil Gantsev. ICICLE thanks Kelly Kohut, Michele Caneppele and Maria Troy. 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 Institutes of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. 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. Milan Breast Cancer Study Group: Siranoush Manoukian, Bernard Peissel, Jacopo Azzollini, Erica Rosina, 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. MSKCC thanks Marina Corines and Lauren Jacobs. MTLGEBCS thanks 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. JS is chair holder of the Canada Research Chair in Oncogenetics. The following are NBCS collaborators: Kristine K Sahlberg (PhD), Anne-Lise Børresen-Dale (Prof Emeritus), Lars Ottestad (MD), Rolf Kåresen (Prof Emeritus), 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 the study participants and research staff for their contributions and commitment to the studies. We 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 and WY. The authors assume full responsibility for analyses and interpretation of these data. OBCS thanks Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri Otsukka, Leena Keskitalo and Kari Mononen for their contributions 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 and Michael Stagner. We thank staff in the Experimental Cancer Medicine Centre, which supported the Faculty of Medicine Tissue Bank and the Faculty of Medicine DNA Banking resource. The authors 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. SBCS thanks Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and Malcolm W R Reed. 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. 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. UBCS thanks all study participants as well as the ascertainment, laboratory, analytics and informatics teams at Huntsman Cancer Institute and Intermountain Healthcare for their important contributions to this study. 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.

Collaborators Members of the Breast Cancer Association Consortium (BCAC) Consortium who have contributed the genotypoe data of the BCAC cohort analysed in our study: QW, MKB, JD, KM, ML, TUA, ILA, HAC, ACA, VA, AA, PA, LBF, MWB, SB, MB, CB, NVB, SEB, HBr, HBy, NJC, JEC, JCC, M-DC, WKC, CLC, JMC, SVC, FJC, AC, SSC, KC, MBD, PD, TD, LD, DME, AHE, ME, DGE, PAF, OF, HF, LF, MG, MGD, MGC, JGS, JG, GGG, MSG, PG, PH, UH, WH, PH, AH, RH, JLH, SJ, AJ, HJ, EMJ, NJ, MEJ, VJ, RK, EKK, CMK, SK, VNK, AWK, JVL, DL, LLM, FL, AL, SL, AL, JL, AM, MM, DM, UM, AMM, RAM, IN, WGN, NO, KOB, KO, AFO, JEO, SP, TWPS, AVP, PP, DPK, BR, PR, GD, VR, AR, ES, DPS, MKS, LS, MS, PS, JS, MCS, JS, WJT, JAT, LRT, AET, MAT, TT, LEK, CRW, CW, XRY, WZ, AZ, AMD, PDPP, DFE.

Contributors RE and RS conceived the project. HL designed, developed and performed the analysis under the supervision of RE and RS. SC, SR, NE and SBS were consulted on the analyses. NE, SBS, RY, AZ and TYP collected and provided the HMC data. The rest of the authors are part of Breast Cancer Association Consortium (BCAC), which contributed the BCAC dataset. All authors wrote the manuscript and approved the final manuscript. Guarantor: RE.

Funding This study was supported in part by grants from the Israeli Science Foundation (number 3165/19, within the Israel Precision Medicine Partnership program; number 2206/22 to RS, and number 407/17 to SC), from the Tel Aviv University Center for AI and Data Science (TAD) to RE and RS, by a joint program grant from the Cancer Biology Research Center, Djerassi Oncology Center, Edmond J. Safra Center for Bioinformatics and TAD to RE, and by the Koret-UC Berkeley-Tel Aviv University Initiative in Computational Biology and Bioinformatics to RE and RS. HL was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. The Breast Cancer Association Consortium (BCAC) is funded by the Confluence project, which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health. Additional funding for BCAC is provided by Cancer Research UK grant: PPRPGM-Nov20\100002, 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. 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 number 223175) (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692 and 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 BRIDGES panel sequencing was supported by the European Union Horizon 2020 research and innovation program BRIDGES (grant number, 634935) and the Wellcome Trust (v203477/Z/16/Z).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 organisations 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. JLH is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. MCS is an NHMRC senior research fellow. The ABCS study was supported by the Dutch Cancer Society (grants NKI 2007-3839 and 2009 4363) and an institutional grant of the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport. The Australian Breast Cancer Tissue Bank 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 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). 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. BOCS is supported by funds from Cancer Research UK (C8620/A8372/A15106) and the Institute of Cancer Research (UK). BOCS acknowledges NHS funding to the Royal Marsden / Institute of Cancer Research NIHR Specialist Cancer Biomedical Research Centre. The Breast Oncology GAlician Network 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 CIHR. CCGP is supported by funding from the University of Crete. The CECILE study was supported by Fondation de France, Institut National du Cancer, Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail and Agence Nationale de la Recherche. 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 DKFZ, Heidelberg, Germany. DT was in part supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation. The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. 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 numbers 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 CTS 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 programmes. 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 (France); German Cancer Aid, 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, Netherlands Cancer Registry, LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund, 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) (UK). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). FHRISK and PROCAS are funded from NIHR (grant PGfAR 0707-10031). DGE, AH and WGN were supported by the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007). The GC-HBOC (German Consortium of Hereditary Breast and Ovarian Cancer) was supported by the German Cancer Aid (grant numbers 110837 and 70114178, coordinator: Rita K Schmutzler, Cologne) and the Federal Ministry of Education and Research, Germany (grant number 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 and 14575/2470). The GENICA was funded by the BMBF Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, DKFZ, Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum, Bochum, as well as the Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany. Generation Scotland 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', reference 104036/Z/14/Z). Funding for identification of cases and contribution to BCAC was provided 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 GEPARSIXTO study was conducted by the German Breast Group GmbH. The GESBC was supported by the Deutsche Krebshilfe e. V. (70492) and the DKFZ. GLACIER was supported by Breast Cancer Now, CRUK and Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. The HABCS study was supported by German Research Foundation (DFG Do761/15-1), 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 and 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 the German Research Foundation (DFG Do761/15-1), a grant from the Friends of Hannover Medical School and by the Rudolf Bartling Foundation. The HUBCS was supported by German Research Foundation (DFG Do761/15-1), a grant from the German Federal Ministry of Research and Education (RUS08/017). BM was supported by the Russian Foundation for Basic Research (grants 17-44-020498 and 17-29-06014). DP was supported by the Russian Foundation for Basic Research (grant 18-29-09129). EK 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). ICICLE was supported by Breast Cancer Now, CRUK and Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. 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 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 NHMRC (400413, 400281 and 199600). GCT and PW were supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. LMBC was supported by the 'Stichting tegen Kanker'. DL was 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 and 110828), the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the BMBF Germany (01KH0402). MBCSG was supported by grants from the AIRC. The MCBCS was supported by the NIH (grants R35CA253187, R01CA192393, R01CA116167 and R01CA176785), an 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 the 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 was 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. 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' programme (grant number CRN-87521) and the Ministry of Economic Development, Innovation and Export Trade (grant number 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-LB-D and VNK, and grant 193387/H10 to A-LB-D and VNK), South Eastern Norway Health Authority (grant 39346 to A-LB-D) and the Norwegian Cancer Society (to A-LB-D and VNK). The NBHS was supported by NIH (grant R01CA100374). Biological sample preparation was conducted by the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California BCFR and Ontario

Familial Breast Cancer Registry were supported by grant U01CA164920 from the USA National Cancer Institute of the National Institutes of Health. The content of this article does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products or organisations 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 and U01 CA179715) and the North Carolina University Cancer Research Fund. The NGOBCS was supported by the National Cancer Center Research and Development Fund (Japan). 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). OBCS was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant numbers 250083 and 122715, and Center of Excellence grant number 251314), the Sigrid Juselius Foundation, the University of Oulu general as well as strategic funding, the University of Oulu Support Foundation and the special Governmental VTR funds towards Oulu University Hospital-based research activities. 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 and C1275/A15956) and Breast Cancer Campaign (2010PR62 and 2013PR044). The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124 and DDHK 2009-4318). The SBCGS was supported primarily by NIH (grants R01CA64277, R01CA148667, UMCA182910 and R37CA70867). Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The scientific development and funding of this project were, in part, supported by the Genetic Associations and Mechanisms in Oncology Network U19 CA148065. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank. The SCCS is supported by a grant from the National Institutes of Health (R01 CA092447). Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, Kentucky; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the CDC. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry. SEARCH is funded by Cancer Research UK (C490/A10124 and 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. SEBCS was supported by the BRL (Basic Research Laboratory) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2012-0000347). 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 EnvironmentalResearch (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 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. 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 Huntsman Cancer Foundation, HCF), the University of Utah, 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 number NU58DP006320), the University of Utah and HCF. The UCIBCS component of this research was supported by the NIH (CA58860 and 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. The US3SS study was supported by Massachusetts (KME, R01CA47305), Wisconsin (PAN, R01 CA47147) and New Hampshire (LT-E, R01CA69664) centres, and Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA.

Competing interests BCAC conflict of interest: MWB conducts research funded by Amgen, Novartis and Pfizer. PAF conducts research funded by Amgen, Novartis and Pfizer. He received Honoraria from Roche, Novartis and Pfizer. JV is one ofthe inventors of diagnosis and treatment of ERCC3-mutant cancer. AWK has a research funding for his institution from Myriad Genetics for an unrelated project (funding dates 2017–2019). UM has research collaborations with Mercy BioAnalytics, RNA Guardian, Dana Farber and iLOF (Intelligent Lab on Fiber). RAM is a consultant for Pharmavite.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval This study involves human participants and the research was carried out in accordance with principles stated in the Declaration of Helsinki. The research was approved by the Breast Cancer Association Consortium. This study on the Hadassah Medical Center cohort was approved by the institutional reciew board (IRB) (IRB-0346-12), and all subjects provided informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The Breast Cancer Association Consortium (BCAC) data are available upon request from Cambridge University (see the BCAC website: https://bcac.ccge.medschl.cam.ac.uk/bcacdata/).The Hadassah Medical Center data are available from the corresponding author upon a reasonable request.

ORCID iDs
Kyriaki Michailidou http://orcid.org/0000-0001-7065-1237
Antonis C Antoniou http://orcid.org/0000-0001-9223-3116
Peter Devilee http://orcid.org/0000-0002-8023-2009
Diana M Eccles http://orcid.org/0000-0002-9935-3169
William G Newman http://orcid.org/0000-0002-6382-4678
Paolo Radice http://orcid.org/0000-0001-6298-4111
Ran Elkon http://orcid.org/0000-0003-3440-1286

REFERENCES

1 Ministry of Health. Breast cancer in women in Israel, update of morbidity and mortality data. 2022. Available: https://www.gov.il/en/departments/news/23102022-01
2 Esserman LJ, WISDOM Study and Athena Investigators. The WISDOM study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer* 2017;3:34.
3 Román M, Sala M, Domingo L, *et al*. Personalized breast cancer screening strategies: a systematic review and quality assessment. *PLoS One* 2019;14:e0226352.
4 Oeffinger KC, Fontham ETH, Etzioni R, *et al*. Breast cancer screening for women at average risk: 2015 guideline update from the American cancer society. *JAMA* 2015;314:1599–614.
5 Shieh Y, Eklund M, Madlensky L, *et al*. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. *J Natl Cancer Inst* 2017;109.

6 Kurian AW, Ward KC, Howlader N, *et al*. Genetic testing and results in a population-based cohort of breast cancer patients and ovarian cancer patients. *J Clin Oncol* 2019;37:1305–15.
7 Armstrong N, Ryder S, Forbes C, *et al*. A systematic review of the International prevalence of *BRCA* Mutation in breast cancer. *Clin Epidemiol* 2019;11:543–61.
8 Michailidou K, Lindström S, Dennis J, *et al*. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;551:92–4.
9 Zhang H, Ahearn TU, Lecarpentier J, *et al*. Genome-wide Association study identifies 32 novel breast cancer susceptibility Loci from overall and subtype-specific analyses. *Nat Genet* 2020;52:572–81.
10 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of Polygenic risk scores. *Nat Rev Genet* 2018;19:581–90.
11 Burton H, Chowdhury S, Dent T, *et al*. Public health implications from COGS and potential for risk stratification and screening. *Nat Genet* 2013;45:349–51.
12 Mavaddat N, Michailidou K, Dennis J, *et al*. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet* 2019;104:21–34.
13 Willoughby A, Andreassen PR, Toland AE. Genetic testing to guide risk-stratified screens for breast cancer. *J Pers Med* 2019;9.
14 Apostolou P, Papasotiriou I. Current perspectives on Chek2 mutations in breast cancer. *Breast Cancer (Dove Med Press)* 2017;9:331–5.
15 Martin AR, Kanai M, Kamatani Y, *et al*. Clinical use of current Polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019;51:584–91.
16 Wang S, Qian F, Zheng Y, *et al*. Genetic variants demonstrating flip-flop phenomenon and breast cancer risk prediction among women of African ancestry. *Breast Cancer Res Treat* 2018;168:703–12.
17 Allman R, Dite GS, Hopper JL, *et al*. SNPs and breast cancer risk prediction for African American and Hispanic women. *Breast Cancer Res Treat* 2015;154:583–9.
18 Shieh Y, Fejerman L, Lott PC, *et al*. A Polygenic risk score for breast cancer in US Latinas and Latin American women. *J Natl Cancer Inst* 2020;112:590–8.
19 Behar DM, Yunusbayev B, Metspalu M, *et al*. The genome-wide structure of the Jewish people. *Nature* 2010;466:238–42.
20 Carmi S, Hui KY, Kochav E, *et al*. Sequencing an Ashkenazi reference panel supports population-targeted personal Genomics and Illuminates Jewish and European origins. *Nat Commun* 2014;5:4835.
21 Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of Genomes. *Nat Methods* 2011;9:179–81.
22 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
23 Purcell S, Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
24 Chen Z-L, Meng J-M, Cao Y, *et al*. A high-speed search engine pLink 2 with systematic evaluation for Proteome-scale identification of cross-linked peptides. *Nat Commun* 2019;10:3404.
25 Choi SW, Mak T-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020;15:2759–72.
26 Purcell SM, Wray NR, Stone JL, *et al*. Common Polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;460:748–52.
27 Privé F, Vilhjálmsson BJ, Aschard H, *et al*. Making the most of Clumping and Thresholding for Polygenic scores. *Am J Hum Genet* 2019;105:1213–21.
28 Privé F, Arbel J, Vilhjálmsson BJ. Ldpred2: better, faster, stronger. *Bioinformatics* 2021;36:5424–31.
29 Mak TSH, Porsch RM, Choi SW, *et al*. Polygenic scores via penalized regression on summary Statistics. *Genet Epidemiol* 2017;41:469–80.
30 Dorling L, Carvalho S, Allen J, *et al*. Breast cancer risk genes — association analysis in more than 113,000 women. *N Engl J Med* 2021;384:428–39.