Polygenic risk scores (PRS) for epithelial ovarian cancer (EOC) have the potential to improve risk stratification. Joint estimation of Single Nucleotide Polymorphism (SNP) effects in models could improve predictive performance over standard approaches of PRS construction. Here, we implemented computationally efficient, penalized, logistic regression models (lasso, elastic net, stepwise) to individual level genotype data and a Bayesian framework with continuous shrinkage, “select and shrink for summary statistics” (S4), to summary level data for epithelial non-mucinous ovarian cancer risk prediction. We developed the models in a dataset consisting of 23,564 non-mucinous EOC cases and 40,138 controls participating in the Ovarian Cancer Association Consortium (OCAC) and validated the best models in three populations of different ancestries: prospective data from 198,101 women of European ancestries; 7,669 women of East Asian ancestries; 1,072 women of African ancestries, and in 18,915 BRCA1 and 12,337 BRCA2 pathogenic variant carriers of European ancestries. In the external validation data, the model with the strongest association for non-mucinous EOC risk derived from the OCAC model development data was the S4 model (27,240 SNPs) with odds ratios (OR) of 1.38 (95% CI: 1.28–1.48, AUC: 0.588) per unit standard deviation, in women of European ancestries; 1.14 (95% CI: 1.08–1.19, AUC: 0.538) in women of East Asian ancestries; 1.38 (95% CI: 1.21–1.58, AUC: 0.593) in women of African ancestries; hazard ratios of 1.36 (95% CI: 1.29–1.43, AUC: 0.592) in BRCA1 pathogenic variant carriers and 1.49 (95% CI: 1.35–1.64, AUC: 0.624) in BRCA2 pathogenic variant carriers. Incorporation of the S4 PRS in risk prediction models for ovarian cancer may have clinical utility in ovarian cancer prevention programs.

European Journal of Human Genetics (2022) 30:349–362; https://doi.org/10.1038/s41431-021-00987-7

INTRODUCTION

Rare variants in known high and moderate penetrance susceptibility genes (BRCA1, BRCA2, BRIIP1, PALB2, RAD51C, RAD51D and the mis-match repair genes) account for about 40% of the inherited component of EOC disease risk [1, 2]. Common susceptibility variants, reviewed in Kar et al. and Jones et al., explain about 6% of the heritability of EOC [1, 3]. Polygenic risk scores (PRS) provide an opportunity for refined risk stratification in the general population and in carriers of rare moderate or high risk alleles.

A PRS is calculated as the weighted sum of the number of risk alleles carried for a specified set of variants. The best approach to identify the variant set and their weights to optimize the predictive power of a PRS is unknown. A common approach involves selecting a set of variants that reach a threshold for association based on the p-value for each variant with or without pruning to remove highly correlated variants [4, 5]. More complex machine learning approaches that do not assume variant independence have also been used [6, 7], but these methods have produced only modest gains in predictive power for highly polygenic phenotypes [6, 8]. Penalized regression approaches such as the lasso, elastic net and the adaptive lasso have also been used with individual level data [9], but a major drawback is the computational burden required to fit the models [9, 10].

We present novel, computationally efficient PRS models using two approaches: (1) penalized regression models including the lasso, elastic net and minimax concave penalty (MCP) for use with individual genotype data; and (2) a Bayesian regression model with continuous shrinkage priors for use where only summary statistics are available—referred to as the “select and shrink with summary statistics” (S4) method. We compare these models with two commonly used methods, stepwise regression with p-value thresholding and LDPred.

MATERIALS (SUBJECTS) AND METHODS

Model development study population

EOC is a highly heterogeneous phenotype with five major histotypes for invasive disease—high-grade serous, low-grade serous, endometrioid, clear cell, and mucinous histotype. The mucinous histotype is the least common and its origin is the most controversial with up to 60% of diagnosed cases of mucinous ovarian cancer often being misdiagnosed metastasis from non-ovarian sites [11]. Therefore, in this study, we performed PRS modeling and association testing for all cases of invasive, non-mucinous EOC. We used genotype data from 23,564 invasive non-mucinous EOC cases and 40,138 controls with >80% European ancestries from 63 case-control studies included in the Ovarian Cancer Association Consortium (OCAC) for model development. The distribution of cases by histotype was high-grade serous (13,669), low-grade serous (2,749), endometrioid (2,877), clear cell (1,427), and others (2,902). Sample collection, genotyping, and
quality control have been previously described [12]. Genotype data were imputed to the Haplotype Reference Consortium reference panel using 470,825 SNPs that passed quality control. Of the 32 million SNPs imputed, 10 million had imputation $r^2 > 0.3$ and were included in this analysis.

Model validation study populations
We validated the best-fitting PRS models developed in the OCAC data in 657 prevalent and incident cases of invasive, non-mucinous EOC and 198,101 female controls of European ancestries from the UK Biobank. Samples were genotyped using either the Affymetrix UK BILEVE Axiom Array or Affymetrix UK Biobank Axiom Array (which share 95% marker content), and then imputed to a combination of the Haplotype Reference Consortium, the 1000 Genomes phase 3 and the UK10K reference panels [13]. We restricted analysis to genetically confirmed females of European ancestries. We excluded individuals if they were outliers for heterozygosity, had low genotyping call rate <95%, had sex chromosome aneuploidy, or if they were duplicates (cryptic or intended) [12]. All SNPs selected in the model development phase were available in the UK Biobank.

We investigated transferability of the best-fitting PRS models to populations of non-European ancestries using genotype data from females of East Asian and African ancestries genotyped as part of the OCAC OncoArray Project [14, 15]. Women of East Asian ancestries—2,841 non-mucinous invasive EOC and 4,828 controls—were identified using a criterion of >80% Asian ancestries. This included samples collected from studies in China, Japan, Korea, and Malaysia as well as samples collected from women of Asian ancestry in studies conducted in the US, Europe and Australia [14]. Similarly, women of African ancestries—368 cases of non-mucinous invasive EOC and 704 controls—mainly from studies conducted in the US, were identified using a criterion of >80% African ancestries as described previously [15].

We also assessed the performance of the best-fitting PRS models in women of European ancestries (>80% European ancestries) with the pathogenic BRCA1 and BRCA2 variants from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). We used genotype data from 18,915 BRCA1 (2,053 invasive EOC cases) and 12,337 BRCA2 (717 invasive EOC cases) pathogenic variant carriers from 63 studies contributing to CIMBA [16]. Genotyping, data quality control measures, intercontinental ancestries assessment and imputation to the HRC reference panel are as described for the OCAC study population.

STATISTICAL ANALYSIS
Polygenic risk models
For all PRS models, we created scores as linear functions of the allele dosage in the general form $PRS_i = \sum_j x_j \beta_j$, where genotypes are denoted as $x$ (taking on the minor allele dosages of 0, 1, and 2), with $x_j$ representing the $i$th individual for the $j$th SNP (out of $p$ SNPs) on an additive log scale and $\beta_j$ represents the weight—the log of the odds ratio—for the $j$th SNP. We used different approaches to select and derive the optimal weights, $\beta_j$, in models as described below.

---

**Fig. 1** PRS model development using penalized regression and LDpred Bayesian approach. Shown in the left panel is the two-stage approach with five-fold cross validation used for individual level genotype data while the right panel shows the LDpred approach used for summary level data.
Penalized logistic regression models
A penalized logistic regression model for a set of SNPs aims to identify a set of regression coefficients that minimize the regularized loss function given by

\[ p(x; \lambda, \kappa) = \begin{cases} 
  x - \lambda \text{sign}(x)/(1 - \kappa) \frac{1}{1 + \kappa} & |x| < \lambda/\kappa \text{ and } |(x)| > \lambda \\
  x & |x| \geq \frac{\lambda}{\kappa} \\
  0 & \text{if } |x| < \lambda 
\end{cases} \]

where \( x \) is the effect estimate of a SNP, \( \lambda \) is the tuning parameter and \( \kappa \) is the threshold (penalty) for different regularization paths, \( \lambda \) and \( \kappa \) are parameters that need to be chosen during model development to optimize performance. The lasso, elastic net, MCP, and \( \kappa \)-value thresholds are instances of the function with different \( \kappa \) values. We minimized the winner’s curse effect on inflated effect estimates for rare SNPs by penalizing rarer SNPs more heavily than common SNPs. Details are provided in the Supplementary Methods.

We used a two-stage approach to reduce computational burden without a corresponding loss in predictive power. The first stage was a SNP selection stage using a sliding windows approach, with 5.5 Mb data blocks and a 500 kb overlap between blocks. SNP selection was performed for each block and selected SNPs were collated. Single SNP association analyses were then run, and all SNPs with a \( \chi^2 \) test statistic of less than 2.25 were excluded. The 2.25 cutoff was arbitrary and selected to maximize computational efficiency without loss in predictive power. Penalized regression models were applied to the remaining SNPs using \( \lambda \) values of 3.0 and \( \kappa \) values of 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. SNPs selected in any of these models were included in subsequent analyses. In the second stage, we fit penalized regression models to the training dataset with \( \lambda \) values ranging from 3.0 to 5.5 in increments of 0.1 iterated over \( \kappa \) values from \(-3.0\) to \(1\) in increments of 0.1. The lasso model (\( \kappa = 0 \)) for each value of \( \lambda \) was fitted first, to obtain a unique maximum. From the fitted maximum the \( \kappa \) value was changed, and the model refitted.

We applied this two-stage approach with five-fold cross-validation (Fig. 1). In each iteration, the data set was split into five, with one part constituting the test data and the other four constituting the training data. The variants and their weights from the two-stage penalized logistic regression modeling in the training data were used to calculate the area under the receiver operating characteristic curve (AUC) in the test data in each iteration. AUC estimates for each \( \kappa \) were obtained. We repeated this process for each cross-validation iteration to obtain a mean AUC for each combination of \( \lambda \) and \( \kappa \). Finally, we selected the tuning and threshold parameters from the lasso, elastic net and MCP models with the maximum mean cross-validated AUC and fitted penalized logistic regression models with these parameters to the entire OCAC dataset to obtain SNP weights for PRS scores.

Stepwise logistic regression with variable \( p \)-value threshold
This model is a general PLR model with \( \kappa = 1 \). As with the other PLR models, we investigated various values for \( \lambda \) values (corresponding to a variable \( p \)-value threshold for including a SNP in the model). However, we observed that the implementation of this model on individual level data was more difficult than for other \( \kappa \) values because the model would sometimes converge to a local optimum rather than the global optimum. Therefore, we applied an approximate conditional and joint association analysis using summary level statistics correcting for estimated LD between SNPs, and utilizing a reference panel of 5,000 individual level genotype OCAC data as described in Yang et al. [17]. Details are provided in the Supplementary Methods.

LDPred
LDPred is a Bayesian approach that shrinks the posterior mean effect size of each marker based on a point-normal prior and LD information from an external reference panel. We derived seven candidate PRSs assuming the fractions of associated variants were 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, and 1.0 respectively using the default parameters as detailed in Vilhjálmsdóttir et al. [18] and an LD reference panel of 503 samples of European ancestries from the 1000 Genomes phase 3 release with effect estimates from the OCAC model development data.

Select and shrink using summary statistics (S4)
The S4 algorithm is similar to the PRS-CS algorithm [19]—a Bayesian method that uses summary statistics and between-SNP correlation data from a reference panel to generate the PRS scores by placing a continuous shrinkage prior on effect sizes. We adapted this method with penalization of rarer SNPs by correcting for the standard deviation resulting in the selection of fewer SNPs. We varied three parameters, \( a, b, \varphi \), which control the degree of shrinkage of effect estimates. \( \varphi \), the overall shrinkage parameter, is influenced by values of \( a \) which controls shrinkage of effect estimates around 0 and \( b \) which control shrinkage of larger effect estimates. We generated summary statistics for each cross-validation training set and selected the parameters that gave the best results on average from the cross-validation and applied these to the set of summary statistics for the complete OCAC data set to obtain the final set of weights.

PRS based on meta-analysis of OCAC-CIMBA summary statistics
We conducted a meta-analysis of the EOC associations in \( BRCA1 \) variant carriers, \( BRCA2 \) variant carriers and the participants participating in OCAC (see Supplementary Methods) and constructed two PRS models. An S4 PRS was generated by applying the \( a, b \) and \( \varphi \) parameters from the S4 model described above. A stepwise PRS was generated by selecting all SNPs that were genome-wide significant (\( p < 5 \times 10^{-8} \)) in the meta-analysis, along with any independent signals in the same region with \( p < 10^{-5} \) from the histotype specific analyses for low-grade serous, high-grade serous, endometrioid, clear cell ovarian cancer and non-mucinous invasive EOC.

Polygenic risk score performance
The best lasso, elastic net, stepwise and S4 models from the model development stage were validated using two independent data sources: the UK Biobank data and \( BRCA1/BRCA2 \) pathogenic variant carriers from the CIMBA. In the UK Biobank data, we evaluated discriminatory performance of the models using the AUC and examined the association between standardized PRS and risk of non-mucinous EOC using logistic regression analysis. For the CIMBA data, we assessed associations for each version of the PRS and invasive non-mucinous EOC risk using weighted Cox regression methods [20]. PRSs in the CIMBA data were scaled to the same PRS standard deviations as the OCAC data, meaning that per standard deviation hazard ratios estimated on CIMBA data are comparable to PRS associations in the OCAC and UK Biobank data. The regression models were adjusted for birth cohort (<1920, 1920–1929, 1930–1939, 1940–1949, ≥1950) and the first four ancestries informative principal components (calculated separately by iCOGS/OncoArray genotyping array) and stratified by Ashkenazi Jewish ancestries and country. Absolute risks by PRS percentiles adjusting for competing risks of mortality from other causes were calculated as described in the Supplementary Material.

Transferability of PRS scores to non-European ancestries
We implemented two straightforward approaches to disentangle the role of ancestries on polygenic risk scoring. We selected homogeneous ancestral samples by using a high cut-off criterion of 80% ancestries and we standardized the PRSs by mean-centering within each population. These approaches led to a more uniform
distribution of PRSs within each ancestral population. Further adjustments using principal components of ancestries did not attenuate risk estimates.

RESULTS
Model development
The results for the models based on individual level genotype data are shown in Table 1. The elastic net model had the best predictive accuracy (AUC = 0.586). The optimal value of λ obtained from regularization paths for the MCP model was 3.3 meaning the best MCP model was equivalent to the lasso model. The best-fitting model based on summary statistics was the S4 (AUC = 0.593) and the LDPred model had the poorest performance of the methods tested (AUC = 0.552). Therefore, the LDPred model was not considered for further validation in other datasets. All SNPs selected and the associated weights for each model are provided in Supplementary Tables 1–6.

Model validation in women of European ancestries
Overall the PLR models performed slightly better in the UK Biobank than the model development data (Table 2). Of the models developed using the OCAC model development data, the association was strongest with the S4 PRS. In BRCA1 and BRCA2 variant carriers, prediction accuracy was generally higher among BRCA2 carriers than BRCA1 carriers. Consistent with results from the general population in the UK Biobank, the S4 PRS model also had the strongest association and predictive accuracy for invasive EOC risk in both BRCA1 and BRCA2 carriers. Sensitivity analyses were conducted in which the unadjusted models for BRCA1 and BRCA2 carriers were progressively adjusted for birth cohort and 6 principal components. There was little difference in HR estimates and association P-values going from the unadjusted model to the model adjusting for six principal components (Supplementary Table 7). The PRS models developed using the OCAC-CIMBA meta-analysis results had better discriminative ability in the UK Biobank than the PRS models developed using only OCAC data. Compared with the S4 PRS using only OCAC data, the S4 PRS model derived from the meta-analysis had fewer SNPs, a stronger association with invasive EOC risk and better predictive accuracy. Similarly, the stepwise model from the OCAC-CIMBA meta-analysis performed better than the stepwise model from only OCAC data, but included more SNPs.

The observed distribution of the OR estimates within centiles of the PRS distribution were consistent with ORs from predicted values under the assumption that all SNPs interact multiplicatively (Fig. 2), with all 95% confidence intervals intersecting with the theoretical estimates for women of European ancestries. Compared with women in the middle quintile, women of European ancestry (UK Biobank) in the top 95th percentile of the lasso derived PRS model had a 2.23-fold increased odds of non-mucinous EOC (95% CI: 1.64 - 3.02) (Table 3).

Absolute risk of developing ovarian cancer by PRS percentiles
We estimated cumulative risk of EOC within PRS percentiles for women in the general population (Fig. 3), by applying the odds

| Table 1. Performance of different PRS models in five-fold cross-validation of OCAC data. |
|-----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Model                             | Number of SNPs<sup>a</sup> | Tuning parameter for best performance | AUC   | OR per 1 SD of PRS | 95% CI          |
|-----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| (a) Models based on individual level genotype data |          |          |          |          |          |
| Lasso                             | 1403            | λ = 3.3        | 0.583           | 1.35           | 1.30–1.39       |
| Elastic net                       | 10,797          | λ = 3.3, κ = 2.2 | 0.586           | 1.36           | 1.31–1.40       |
| MCP                               | 1403            | λ = 3.3        | 0.583           | 1.35           | 1.30–1.39       |
| (b) Models based on summary statistics |        |          |          |          |          |
| LDPred                            | 5,291,719       | ρ = 0.001      | 0.552           | 1.21           | 1.13–1.29       |
| Stepwise                          | 22              | λ = 5.4        | 0.572           | 1.30           | 1.26–1.34       |
| Select and Shrink (OCAC)          | 27,240          | a = 2.75, b = 2, φ = 3e–6 | 0.593           | 1.39           | 1.34–1.44       |

AUC area under the receiver operating characteristic (ROC) curve, OR odds ratio, SD standard deviation, PRS polygenic risk score, CI confidence interval, NA not applicable.

<sup>a</sup>Number of SNPs in PRS model run on full OCAC data set after selection of model parameters.

| Table 2. External validation of PRS models in European populations using data from UK Biobank and CIMBA. |
|-----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Model (data set)                  | SNPs            | UK Biobank AUC | CI              | CIMBA BRCA1 carriers<sup>a</sup> |                |
|-----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| (a) PRS models based on OCAC data |          |          |          |          |          |
| Lasso (OCAC)                      | 1403            | 0.587           | 1.37           | 1.27–1.48       | 0.573           | 1.27           | 1.21–1.34       | 0.627           | 1.48           | 1.33–1.63       |
| Elastic net (OCAC)                | 10,797          | 0.588           | 1.36           | 1.26–1.47       | 0.583           | 1.32           | 1.26–1.39       | 0.617           | 1.47           | 1.33–1.63       |
| Stepwise (OCAC)                   | 22              | 0.588           | 1.35           | 1.26–1.46       | 0.563           | 1.21           | 1.16–1.26       | 0.605           | 1.39           | 1.26–1.54       |
| Select and shrink (OCAC)          | 27,240          | 0.588           | 1.38           | 1.28–1.48       | 0.592           | 1.36           | 1.29–1.43       | 0.624           | 1.49           | 1.35–1.64       |
| (b) PRS models based on meta-analysis of OCAC and CIMBA data |          |          |          |          |          |
| Stepwise (OCAC-CIMBA)<sup>b</sup> | 36              | 0.595           | 1.39           | 1.29–1.50       | NA              | NA              | NA              | NA              | NA              | NA              |
| Select and shrink (OCAC-CIMBA)    | 18,007          | 0.596           | 1.42           | 1.32–1.54       | NA              | NA              | NA              | NA              | NA              | NA              |

AUC area under the receiver operating characteristic curve, OR odds ratio, HR hazards ratio.
<sup>a</sup>Estimates are from unadjusted models.
<sup>b</sup>Results in CIMBA are overfitting.

E.O. Dareng et al.
European Journal of Human Genetics (2022) 30:349 – 362
ratio from the PRS models to age-specific population incidence and mortality data for England in 2016. For \textit{BRCA1} and \textit{BRCA2} pathogenic variant carriers, we applied the estimated hazard ratios from PRS models to age-specific incidence rates obtained from Kuchenbaecker et al. [21]. For women in the general population, the estimated cumulative risks of EOC by age 80 for women at the 99th centile of the PRS distribution were 2.24%, 2.18%, 2.54%, and 2.81% for the lasso, elastic net, stepwise and S4 models, respectively. In comparison, the absolute risks of EOC by age 80 for women at the 1st centile were 0.76%, 0.78%, 0.64%, and 0.56% for the lasso, elastic net, stepwise and S4 models, respectively.

The absolute risks of developing EOC in \textit{BRCA1} and \textit{BRCA2} pathogenic variant carriers were considerably higher than for women in the general population (Figs. S1 and S2). The estimated absolute risk of developing ovarian cancer by age 80 for \textit{BRCA1} carriers at the 99th PRS centiles were 63.2%, 66.3%, 59.0%, and 68.4% for the lasso, elastic net, stepwise and S4 models, respectively. The corresponding absolute risks for women at the 1st PRS centile were 27.7%, 25.6%, 30.8%, and 24.2%. For \textit{BRCA2} carriers the absolute risks for women at the 99th centile were 36.3%, 36.3%, 33.0%, and 36.9%; and 7.10%, 7.12%, 8.24%, and 6.92% at the 1st centile for the lasso, elastic net, stepwise and S4 models, respectively.

**PRS distribution and ancestries**

To investigate the transferability of the PRS to other populations, we applied the scores to women of African (\(N = 1,072\)) and Asian (\(N = 7,669\)) ancestries genotyped as part of the OncoArray project. In general, the distributions of the raw PRS were dependent on both the statistical methods used in SNP selection and ancestral group. PRS models that included more variants had less dispersion, such that the elastic net models had the least between individual variation in all ancestral groups (standard deviation = 0.15, 0.19, and 0.22 for individuals of Asian, African and European ancestries respectively), while the distributions from the stepwise models were the most dispersed (standard deviation = 0.23, 0.27, and 0.30 for individuals of Asian, African and European ancestries respectively). As expected, given the variation in variant frequencies by population, the distribution of polygenic scores was significantly different across the three ancestral groups, with the least dispersion among women of Asian ancestries and the most...
Table 3. Association between polygenic risk scores and non-mucinous EOC by PRS percentiles and ancestry.

| Percentile | UK Biobank (Controls (n) | Cases (n) | OR (95% CI) | East Asian (Controls (n) | Cases (n) | OR (95% CI) | African (Controls (n) | Cases (n) | OR (95% CI) |
|------------|--------------------------|-----------|-------------|-------------------------|-----------|-------------|---------------------|-----------|-------------|
| (a) Lasso   | 0-5 9880 12 0.42 (0.22–0.72) 278 106 0.65 (0.51–0.83) 35 19 0.89 (0.47–1.65) | 5-10 9870 24 0.83 (0.52–1.27) 271 112 0.71 (0.55–0.90) 41 13 0.52 (0.25–1.01) | 10-20 19,733 53 0.92 (0.66–1.27) 487 280 0.98 (0.62–1.51) 81 26 0.53 (0.31–0.94) | 20-40 39,468 104 0.90 (0.69–1.18) 993 541 0.93 (0.80–1.08) 154 60 0.64 (0.42–0.99) | 40-60 39,457 115 1.00 967 566 1.00 133 81 1.08 (0.85–1.35) 941 593 1.08 (0.85–1.35) 41 13 0.52 (0.25–1.01) |
|            | 60-80 39,425 147 1.28 (1.00–1.64) 941 593 1.08 (0.93–1.25) 136 78 0.94 (0.64–1.39) | 80-90 19,699 87 1.52 (1.14–2.00) 466 301 1.10 (0.92–1.32) 63 44 1.15 (0.71–1.84) | 90-95 9842 51 1.78 (1.27–2.46) 214 169 1.35 (1.07–1.69) 34 20 0.97 (0.51–1.78) | 95-100 9830 64 2.23 (1.64–3.02) 211 173 1.40 (1.12–1.76) 27 27 1.64 (0.90–3.00) | 0-5 9876 17 0.67 (0.39–1.09) 277 107 0.72 (0.56–0.92) 35 19 0.90 (0.47–1.64) | 5-10 9876 17 0.67 (0.39–1.09) 271 112 0.78 (0.61–0.99) 41 13 0.52 (0.25–1.01) | 10-20 19,740 45 0.89 (0.62–1.26) 497 270 1.02 (0.85–1.22) 81 26 0.53 (0.31–0.88) | 20-40 39,453 120 1.19 (0.91–1.55) 967 567 1.10 (0.95–1.28) 154 60 0.64 (0.42–0.96) | 40-60 39,471 101 1.00 1000 533 1.00 133 81 1.08 (0.85–1.35) 941 593 1.08 (0.85–1.35) 41 13 0.52 (0.25–1.01) |
|            | 60-80 39,413 159 1.58 (1.23–2.03) 926 608 1.23 (1.06–1.43) 136 78 0.94 (0.64–1.39) | 80-90 19,695 91 1.80 (1.36–2.40) 457 310 1.27 (1.06–1.52) 63 44 1.15 (0.71–1.84) | 90-95 9841 52 2.07 (1.47–2.87) 226 157 1.30 (1.04–1.64) 34 20 0.97 (0.51–1.78) | 95-100 9839 55 2.18 (1.56–3.02) 207 177 1.60 (1.28–2.01) 27 27 1.64 (0.90–3.00) | 0-5 9880 13 0.39 (0.21–0.76) 254 130 0.90 (0.71–1.14) 40 14 0.75 (0.37–1.44) | 5-10 9874 19 0.57 (0.34–0.91) 268 115 0.76 (0.59–0.96) 43 11 0.55 (0.26–1.01) | 10-20 19,742 44 0.67 (0.47–0.93) 494 273 0.98 (0.81–1.17) 80 27 0.72 (0.42–1.21) | 20-40 39,470 102 0.77 (0.60–1.00) 970 564 1.03 (0.89–1.19) 142 72 1.09 (0.73–1.63) | 40-60 39,440 132 1.00 979 564 1.00 146 68 1 |
|            | 60-80 39,414 158 1.20 (0.95–1.51) 951 583 1.08 (0.94–1.25) 130 84 1.39 (0.93–2.07) | 80-90 19,697 88 1.33 (1.02–1.75) 456 311 1.21 (1.01–1.44) 61 46 1.62 (1.00–2.61) | 90-95 9853 41 1.24 (0.86–1.75) 236 147 1.10 (0.87–1.38) 35 19 1.17 (0.61–2.17) | 95-100 9834 60 1.82 (1.33–2.46) 220 164 1.32 (1.04–1.65) 27 27 2.15 (1.17–3.95) | 0-5 9957 16 0.54 (0.31–0.89) 279 105 0.63 (0.49–0.81) 38 16 0.71 (0.36–1.33) | 5-10 9888 15 0.51 (0.29–0.85) 254 129 0.85 (0.67–1.08) 41 13 0.53 (0.26–1.03) | 10-20 19,812 51 0.87 (0.62–1.20) 489 278 0.96 (0.80–1.14) 81 26 0.54 (0.32–0.90) | 20-40 39,435 113 0.97 (0.75–1.25) 1013 521 0.86 (0.75–1.00) 156 58 0.62 (0.41–0.94) | 40-60 39,512 117 1.00 961 572 1.00 134 80 1 |
|            | 60-80 39,316 158 1.36 (1.07–1.73) 950 584 1.03 (0.89–1.20) 137 77 0.94 (0.63–1.40) | 10-20 20-40 40-60 60-80 |
variation in women of European ancestries. The difference in PRS distribution was minimized after correction for ancestry by standardizing the PRS to have unit standard deviation using the control subjects for each ancestral group.

High PRSs were significantly associated with risk of non-mucinous EOC in both Asian and African ancestries (Table 4), although the effects were weaker than in women of European ancestries. For example, with the lasso model, the odds ratio per unit standard deviation increment in polygenic score was 1.16 (95% CI: 1.11–1.22) in women of East Asian ancestries, 1.28 (95% CI: 1.13–1.45) in women of African ancestries and 1.37 (95% CI: 1.27–1.48) in women of European ancestries (p for heterogeneity <0.0001). Variability in effect sizes among ancestral groups was highest for the stepwise model (p2 = 92%) versus 84% and 83% for elastic net and lasso derived polygenic scores respectively. The best discriminative model among women of East Asian and African ancestries were the elastic net PRS (AUC = 0.543) and the S4 PRS derived from OCAC-CIMBA meta-analysis (AUC = 0.596) respectively. Women of African ancestries in the top 5% of the PRS had about two-fold increased risk compared to women in the middle quintile (lasso OR: 1.64, 95% CI: 0.90–3.00; elastic net OR: 1.64, 95% CI: 0.90–3.00; stepwise OR: 2.15, 95% CI: 1.17–3.95; S4 OR: 1.80, 95% CI: 0.99–3.31) (Table 3). Effect estimates were smaller in women of East Asian ancestries with women in the top 5% of the PRS, having about a 1.5 fold increased risk compared to women in the middle quintile (lasso OR: 1.40, 95% CI: 1.12–1.76; elastic net OR: 1.60, 95% CI: 1.28–2.01; stepwise OR: 1.32, 95% CI: 1.04–1.65; S4 OR: 1.32, 95% CI: 1.05–1.66) (Table 3).

**DISCUSSION**

Genetic risk profiling with PRSs has led to actionable outcomes for cancers such as breast and prostate [22, 23]. Previous PRS scores for invasive EOC risk in the general population and BRCA1/BRCA2 pathogenic variant carriers have been based on genetic variants for which an association with EOC risk had been established at nominal genome-wide significance [20, 24, 25]. Here, we explored the predictive performance of computationally efficient, penalized, regression methods in modeling joint SNP effects for EOC risk prediction in diverse populations and compared them with common approaches. By leveraging the correlation between SNPs which do not reach nominal genome-wide thresholds and including them in PRS models, the PRSs derived from penalized regression models provide stronger evidence of association with risk of non-mucinous EOC than previously published PRSs in both the general population and BRCA1/BRCA2 pathogenic variant carriers.

Recently, Barnes et al. derived a PRS score using 22 SNPs that were significantly associated with high-grade serous EOC risk (PRSHGS) to predict EOC risk in BRCA1/BRCA2 pathogenic variant carriers [20]. To make effect estimates obtained in this analysis comparable to the effect estimates obtained from the PRSHGS, we standardized all PRSs using the standard deviation from unaffected BRCA1/BRCA2 carriers and provide estimates which are directly comparable to the PRSHGS in Supplementary Table 9. All PRS models in this analysis except the Stepwise (OCAC only) had higher effect estimates [20]. The AUC estimates from the adjusted PLR methods implemented in this analysis, are higher than the corresponding PRSHGS estimates for BRCA1 carriers (0.604). In BRCA2 carriers, the AUC estimates for the lasso and S4 models did slightly better than the PRSHGS AUC estimate (0.667), while the stepwise did slightly worse and the elastic net estimate was comparable. The AUC estimates for women in the general population, as estimated from the UK Biobank, are slightly higher than estimates from previously published PRS models for overall EOC risk by Jia et al. (AUC = 0.57) and Yang et al. (AUC = 0.58) [25, 26].

The level of risk for women above the 95th percentile of the PRS is similar to that conferred by pathogenic variants in moderate penetrance genes such as FANCM (RR = 2.1, 95% CI = 1.1–3.9) and
The inclusion of other risk factors such as family history of ovarian cancer, presence of rare pathogenic variants, age at menarche, oral contraceptive use, hormone replacement therapy, parity, and endometriosis in combination with the PRS could potentially improve risk stratification as implemented in the CanRisk tool (www.canrisk.org), which currently uses a 36-SNP PRS with the potential to use other PRS models [29, 30].

We found that the discrimination of the PRS varied by ancestry with greater discrimination in women of European ancestries than in women of African and East Asian ancestries. The better performance in African than East Asian populations is in contrast to what one would expect given human demographic history, and the performance of PRS for other phenotypes in African populations. This may simply be the play of chance given the small number of samples of African ancestries. Alternatively it reflects the fact that the allele frequencies of the PRS SNPs were more similar between the African and European populations than they were with the East Asian population (Supplementary Tables 10–14).

Further optimization of the models could be achieved by varying the penalty function based on prior knowledge. For example, varying the penalty function to select more SNPs from genomic regions with known susceptibility variants given that susceptibility variants tend to cluster together. Alternatively, the penalty functions could be modified to incorporate information about functionally active regions of the genome such a promoters, enhancers, and transcription factor binding sites. However, incorporating functional annotation has resulted in limited gains in prediction accuracy for complex traits such as breast cancer, celiac disease, type 2 diabetes, and rheumatoid arthritis [31].

Machine/deep learning approaches are alternative ways to constructing PRS, but methods such as the neural net, support vector machine, and random forest have been shown to be computationally prohibitive or produce inferior results to other approaches [32, 33]. Other machine learning methods, such as those based on gradient boosting do not perform well in genomic regions where strong genetic interactions are present, for which alternative approaches such as the LDpred may perform better [18]. Our approach has several benefits over alternative machine learning methods, including its simplicity, and intrinsic robustness to minor misspecification of LD or association strength.
In conclusion, our results indicate that using the lasso model for individual level genotype data and the S4 model for summary level data in PRS construction provide an improvement in risk prediction for non-mucinous EOC over more common approaches. Our approach overcomes the computational limitations in the use of penalized methods for large-scale genetic data, particularly in the presence of highly correlated SNPs and when the use of cross-validation for parameter estimation is preferred. In practical terms, the PRS provides sufficient discrimination, particularly for women of European ancestries, to be considered for inclusion in risk prediction and prevention approaches for EOC in the future. Further studies are required to optimize these PRSs in ancestrally diverse populations and to validate their performance with the inclusion of other genetic and lifestyle risk factors.

DATA AVAILABILITY
OncoArray germline genotype data for the OCAC studies have been deposited at the European Genome-Phenome Archive (EGA; https://ega-archive.org/), which is hosted by the EBI and the CRG, under accession EGAS00001002305. Summary statistics for the Ovarian Cancer Association Consortium are available in the NHGRI-EBI GWAS catalogue (https://www.ebi.ac.uk/gwas/home) under the accession number, GCST90016665. A subset of the OncoArray germline genotype data for the CIMBA studies are publicly available through the database of Genotypes and Phenotypes (dbGaP) under accession phs001321.v1.p1. The complete data set will not be made publically available because of restraints imposed by the ethics committees of individual studies; requests for further data can be made to the Data Access Coordination Committee (http://cimba.ccge.medschl.cam.ac.uk/).

REFERENCES
1. Jones MR, Kamara D, Karlan BY, Pharoah PDP, Gayther SA. Genetic epidemiology of ovarian cancer and prospects for polygenic risk prediction. Gynecol Oncol. 2017;147:705–13.
2. Lyra PCM, Rangel LB, Monteiro ANA. Functional landscape of common variants associated with susceptibility to epithelial ovarian cancer. Curr Epidemiol Rep. 2020;7:49–57.
3. Kar SP, Berchuck A, Gayther SA, Goode EL, Mabjeesh NJ, Vastine C, et al. Common genetic variation and susceptibility to ovarian cancer: current insights and future directions. Cancer Epidemiol Biomar Prev Am Assoc Cancer Res Consop- responded Am Soc Prev Oncol. 2018;27:395–404.
4. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to dis- ease from genome-wide association studies. Genome Res. 2007;17:1520–8.
5. Purcell SM, Wray NR, Stone JL, Visscher PM, O’Donovan MC, International Schi- zophrenia Consortium, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460:748–52.
6. Abrahams BS, Kowalczuk Z, Zobel J, Inouye M. Performance and robustness of penalized and unpenalized methods for genetic prediction of complex human disease. Genet Epidemiol. 2011;35:608–15.
7. Haber D, Fernando RL, Kulkarni K, Gairing DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinform. 2011;12:186.
8. Szymczak S, Biernacka JM, Cordell HJ, Gonzalez-Rocío O, König IR, Zhang H, et al. Machine learning in genome-wide association studies. Genet Epidemiol. 2009;33: 551–57.
9. Privé F, Ashcha H, Blum MGB. Efficient implementation of penalized regression for genetic risk prediction. Genetika. 2019;212:65–74.
10. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized and unpenalized methods for genetic prediction of complex human traits. Genet Epidemiol. 2013;37:184–95.
11. Haber D, Fernando RL, Kulkarni K, Gairing DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinform. 2011;12:186.
12. Szymczak S, Biernacka JM, Cordell HJ, Gonzalez-Rocío O, König IR, Zhang H, et al. Machine learning in genome-wide association studies. Genet Epidemiol. 2009;33: 551–57.
13. Privé F, Ashcha H, Blum MGB. Efficient implementation of penalized regression for genetic risk prediction. Genetika. 2019;212:65–74.
14. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized and unpenalized methods for genetic prediction of complex human traits. Genet Epidemiol. 2013;37:184–95.
15. Haber D, Fernando RL, Kulkarni K, Gairing DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinform. 2011;12:186.
16. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet. 2017;49:680–91.
17. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropo- metric Traits (GIANT) Consortium, DiABetes Genetik Replication And Meta- analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44:369–75. 51-3
18. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am J Hum Genet. 2015;97:576–92.
19. Ge T, Chen CY, Ni Y, Feng Y-C, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10:1776.
20. Barnes DR, Rookus MA, McGuffog L, Leslie G, Mooij TM, Dennis J, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. Genet Med Off J Am Coll Med Genet. 2020;25:1557–62.
21. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017;317:2402–16.
22. Mavaddat N, Michaailidou K, Dennis J, Lush M, Facha L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet. 2019;104:21–34.
23. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50:928–36.
24. Kuchenbaecker KB, McGuffog L, Barrett-Connor E, Dite GS, Dite JS, Dite JS, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. J Natl Cancer Inst. 2017;109.
25. Yang X, Leslie G, Gentzy-Maharaj A, Ryan A, Internnaggio M, Lee A, et al. Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. J Med Genet. 2018;55:546–54.
26. Jia G, Lu Y, Wen W, Long L, Liu Y, Tao R, et al. Evaluating the utility of polygenic risk scores in identifying high-risk individuals for eight common cancers. JNCI Cancer Spectr. 2020;4:klk9a021.
27. Song H, Dicks EM, Tyrer J, Internnaggio M, Chenex-Nivens H, Bowtell DD, et al. Population-based targetted sequencing of 54 candidate genes identifies PALB2 as a susceptibility gene for high-grade serous ovarian cancer. J Med Genet. 2021;58:305–13.
28. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, et al. Cancer risks associated with germline PALB2 pathogenic variants: an international study of 524 families. J Clin Oncol Off J Am Soc Clin Oncol. 2020;38:674–85.
29. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Earley T, Hartley S, et al. BOA- DICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med Off J Am Coll Med Genet. 2019;21:1708–18.
30. Welcome to CanRisk. [cited 2021 Aug 15]. https://canrisk.org/
31. Hu Y, Lu Q, Powles R, Yao X, Yang C, Fang F, et al. Leveraging functional annota- tions in genetic risk prediction for human complex diseases. PLoS Comput Biol. 2017;13:e1005589.
32. Gola D, Erdmann J, Müller-Myshok B, Schunkert H, König IR. Polygenic risk scores outperform machine learning methods in predicting coronary artery disease status. Genet Epidemiol. 2020;44:125–38.
33. Paré G, Mao S, Deng WQ. A machine-learning heuristic to improve gene score prediction of polygenic traits. Sci Rep. 2017;7:12665.

ACKNOWLEDGEMENTS
See Supplementary Material.

AUTHOR CONTRIBUTIONS
EO Dareng, JP Tyrer, DR Barnes, MR Jones, X Yang, RK Aben, MA Adank, S Agata, IL Andrulis, H Anton-Culver, NN Antonsson, G Arvanitakis, BK Arun, A Augustsson, J Balm资, RB Barkardottir, D Barrowdale, MW Beckmann, A Beegly-Fadiel, J Benitez, M Bermisheva, MQ Bernardini, L Bjorge, NV Bogdanova, B Bonanni, A Bong, JG Brenton, A Budzilowska, R Butzow, SS Buys, H Cai, MA Caligo, I Campbell, R Cannito, H Cassingh, J Chang-Claude, SJ Channock, K Chen, Y Chiew, WK Chung, KB Claes, S Colonna, LS Cook, FJ Couch, MB Daly, F Dao, E Davies, M de la Hoye, R de Potter, J Dennis, A DePersia, P Devilee, O Dierk, Y Ding, JA Doherty, SM Domchek, T Dörr, A du Bois, M Durst, DM Eccles, HA Elsasser, C Engel, D Evans, PA Fasching, JM Flanagan, RT Fortner, E Machackova, E Friedman, PA Ganz, J Garber, F Gensini, GG Giles, G Glendon, AK Godwin, MT Goodman, MH Greene, J Gronwald, E Hahnen, CA Haiman, N Håkansson, U Hamann, TV Hansen, HR Harris, M Hartman, F Heitz, MA Hildrebrandt,
COMPETING INTERESTS

AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P. AOD has received a research grant from AstraZeneca, not directly related to the content of this manuscript. 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. AKW reports research funding to her institution from Myriad Genetics for an unrelated project. UM owns stocks in Abcodia and requests for materials should be addressed to Paul D. P.
Human Genetics and Oncology, Montréal, QC, Canada. 212University of Cambridge, Department of Medical Genetics, Cambridge, UK. 213Dartmouth College, Geisel School of Medicine, Hanover, NH, USA. 214The Ohio State University, Department of Cancer Biology and Genetics, Columbus, OH, USA. 215Pontificia Universidad Javeriana, Institute of Human Genetics, Bogota, Colombia. 216University of Oxford, Cancer Epidemiology Unit, Oxford, UK. 217Beth Israel Deaconess Medical Center, Department of Medical Oncology, Boston, MA, USA. 218University Medical Center Groningen, University Groningen, Department of Genetics, Groningen, The Netherlands. 219University of Pretoria, Department of Genetics, Arcadia, South Africa. 220University of Southern California Norris Comprehensive Cancer Center, Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA. 221University of Malaya, Department of Obstetrics and Gynaecology, Department of Biomedical Sciences, Women’s Health Research, Montréal, QC, Canada. 222University of Pretoria, Department of Genetics and Population Sciences, Durham, NC, USA. 223Duke University Hospital, Department of Community and Family Medicine, Durham, NC, USA. 224National Institute of Environmental Health Sciences, NIH, Biostatistics and Computational Biology Branch, Research Triangle Park, NC, USA. 225City of Hope, Clinical Cancer Genomics, Duarte, CA, USA. 226Fred Hutchinson Cancer Research Center, Seattle, WA, USA. 227Stanford University School of Medicine, Department of Biomedical Data Science, Stanford, CA, USA. 228University of Malaya, Department of Obstetrics and Gynaecology, University of Malaya Medical Centre, Kuala Lumpur, Malaysia. 229Hebei Medical University, Fourth Hospital, Department of Molecular Biology, Shijiazhuang, China. 230National Centre for Scientific Research ‘Demokritos’, Molecular Diagnostics Laboratory, INRATES, Athens, Greece. 231University of Uppsala, Department of Surgical Sciences, Uppsala, Sweden. 232University of Malaya, Department of Obstetrics and Gynaecology, University of Malaya Medical Centre, Kuala Lumpur, Malaysia. 233National Institute for Environmental Health Sciences, NICHD, National Institutes of Health, Bethesda, MD, USA. 234University of Southern California Norris Comprehensive Cancer Center, Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA. 235Women’s Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Centre, Department of Obstetrics and Gynecology, Los Angeles, CA, USA. 236Royal Pain Road, Tampa, FL, USA. 237University of NSW Sydney, School of Women’s and Children’s Health, Faculty of Medicine, Sydney, NSW, Australia. 238University of Michigan School of Public Health, Department of Epidemiology, Ann Arbor, MI, USA. 239University of Southern California Norris Comprehensive Cancer Center, Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA. 240Mayo Clinic, Department of Health Science Research, Division of Epidemiology, Rochester, MN, USA. 241Emory University, Department of Epidemiology, Rollins School of Public Health, Atlanta, GA, USA. 242These authors contributed equally: Eileen O. Dareng, Jonathan P. Tyrer.

GEMO Study Collaborators
Fabienne Lesueur and Nora Mebirouk

GC-HBOC Study Collaborators
Christoph Engel and Rita K. Schmutzler

EMBRACE Collaborators
Daniel Barrowdale, Eleanor Davies, Diana M. Eccles and D. Gareth Evans

KConFab Investigators
Georgia Chenevix-Trench

HEBON Investigators
Muriel A. Adank, Peter Devilee and Annemieke H. van der Hout

The OCAC Consortium
Eileen O. Dareng, Jonathan P. Tyrer, Michelle R. Jones, Katja K. H. Aben, Hoda Anton-Culver, Natalia N. Antonenkova, Gerasimos Aravantinos, Matthias W. Beckmann, Alicia Beeghly-Fadiel, Javier Benitez, Marina Bermishева, Marcus Q. Bernardini, Line Bjorge, Natalia V. Bogdanova, James D. Brenton, Agnieszka Budzilowska, Ralf Butzow, Hui Cai, Ian Campbell, Rikki Cannioto, Jenny Chang-Cluade, Stephen J. Chankoc, Ke Xin Chen, Yoke-Eng Chiew, Linda S. Cook, Fanny Dao, Joe Dennis, Jennifer A. Doherty, Thilo Dörk, Andreas du Bois, Matthias Dürst, Diana M. Eccles, Heather A. Eliassen, Peter A. Fasching, James M. Flanagan, Renée T. Fortner, Graham G. Giles, Marc T. Goodman, Jacek Gronwald, Christopher A. Haiman, Nicolas Håkansson, Holly R. Harris, Florian Heitz, Michelle A. Hildebrandt, Estrid Högstad, Claus K. Högdall, Rueba-Yea Huang, Chad Huff, David G. Huntsman, Anna Jakubowska, Allan Jensen, Michael E. Jones, Dahee Kang, Beth Y. Karlan, Anthony Karnezis, Linda E. Kelemen, Elza Khushnutdinova, Lamberts A. Kiemeyer, Byoung-Gie Kim, Susanne K. Kjaer, Jolanta Kupryjanczyk, Diether Lambrechts, Melissa C. Larson, Nhu D. Le, Jenny Lester, Douglas A. Levine, Karen H. Lu, Jan Lubiński, Jeffrey R. Marks, Rayna Kim Matsuno, Keitaro Matsuo, Taymra May, John R. McLaughlin, Iain A. McNeish, Roger L. Milne, Albina Minlikeeva, Francescmary Modugno, Kirsten B. Moysich, Elizabeth Munro, Helé Nevanlinna, Kunde Ondus, Siel Olbrecht, Sara H. Olson, Håkan Olsson, Ana Osoiro, Sue K. Park, Tanja Pejovic, Jennifer B. Permuith, Anna Piskorcz, Darya Prokofyeva, Marjorie J. Riggan, Harvey A. Risch, Cristina Rodriguez-Antonia, Mary Anne Rossing, Ingo Runnebaum, Dale P. Sandler, V. Wendy Setiawan, Kang Shan, Weiva Sieh, Honglin Song, Melissa C. Southey, Helen Steed, Rebecca Supthnen, Anthony J. Swerdlow, Soo Hwang Teo, Kathryn L. Terry, Pamela J. Thompson, Liv Cecilia Vesthrum Thomsen, Linda Titus, Britton Trabert, Ruth Travis, Shelley S. Tworoger, Ellen Valen, Anne M. van Altena, Els Van Nieuwenhuysen, Digna Velez Edwards, Robert A. Vierkant, Frances Wang, Penelope M. Webb, Clarice R. Weinberg, Nicolas Wentzensen, Emily White, Alice S. Whitemore, Stacey J. Winham, Alicia Wolk, Yin-Ling Woo, Anna H. Wu, Li Yan, Drakoulis Yannoukakos, Wei Zheng, Argyrios Zioas, Kate Lawsonron, Anna deFazio, Susan J. Ramus, Celeste L. Pearce, Alvaro N. Monteiro, Julie M. Cunningham, Ellen L. Goode, Joellen M. Schildkraut, Andrew Berchuck, Simon A. Gayther and Paul D. P. Pharoah.
The CIMBA Consortium
Daniel R. Barnes, Xiu Yang, Muriel A. Adank, Simona Agata, Irene L. Andrusis, Banu K. Arun, Annelie Augustinsson, Judith Balmaña, Rosa B. Barkardottir, Daniel Barrowdale, Bernardo Bonanni, Ake Borg, Saundra S. Buys, María A. Caligo, Hayley Cassingham, Wendy K. Chung, Kathleen B. M. Claes, Sarah Colonna, Fergus J. Couch, Mary B. Daly, Eleanor Davies, Miguel de la Hoya, Robin de Putter, Allison DePersia, Peter Devilee, Orland Diez, Yuan Chun Ding, Susan M. Domchek, Diana M. Eccles, Christoph Engel, D. Gareth Evans, Eva Machackova, Eitan Friedman, Patricia A. Ganz, Judy Garber, Francesca Gensini, Gord Glendon, Andrew K. Godwin, Mark H. Greene, Eric Hahnen, Ute Hamann, Thomas V. O. Hansen, Mikael Hartman, John L. Hopper, Peter J. Hulick, Evgeny N. Imyanitov, Claudine Isaacs, Paul A. James, Ramunas Janavicius, Óskar Th. Johannsson, Esther M. John, Ian Komenaka, Allison W. Kurian, Ava Kwong, Conxi Lazaro, Goska Leslie, Fabienne Lesueur, Jingmei Li, Jennifer T. Loud, Phuong L. Mai, Siranoush Manoukian, Lesley McGuffog, Noura Mebirouk, Austin Miller, Marco Montagna, Katherine L. Nathanson, Susan L. Neuhausen, Joanne Ngeow Yuen Yie, Henriette Roed Nielsen, Liene Nikitina-Zake, Kenneth Offit, Edith Olah, Olufunmilayo I. Olopade, Laura Papi, Michael T. Parsons, Harsha Pathak, Inge Søkilde Pedersen, Ana Peixoto, Pedro Perez-Segura, Beth Peshkin, Paolo Peterlongo, Paolo Radice, Johanna Rantala, Eric Ross, Marta Santamariña, Penny Soucy, Rita K. Schmutzler, Jacques Simard, Christian F. Singer, Anna P. Sokolenko, Dominique Stoppa-Lyonnet, Yen Yen Tan, Manuel R. Teixeira, Mary Beth Terry, Mads Thomassen, Darcy L. Thull, Marc Tischkowitz, Amanda E. Toland, Diana Torres, Nadine Tung, Annemieke H. van der Hout, Elizabeth J. van Rensburg, Ana Vega, Barbara Wappenschmidt, Jeffrey N. Weitzel, Katia M. Zavaglia, Kristin K. Zorn, Thomas A. Sellers, Georgia Chenevix-Trench and Antonis C. Antoniou