



OPEN ACCESS

ORIGINAL ARTICLE

# Multigene testing of moderate-risk genes: be mindful of the missense

E L Young,<sup>1</sup> B J Feng,<sup>2</sup> A W Stark,<sup>1</sup> F Damiola,<sup>3</sup> G Durand,<sup>4</sup> N Forey,<sup>4</sup> T C Francy,<sup>1</sup> A Gammon,<sup>5</sup> W K Kohlmann,<sup>5</sup> K A Kaphingst,<sup>6</sup> S McKay-Chopin,<sup>4</sup> T Nguyen-Dumont,<sup>7</sup> J Oliver,<sup>8</sup> A M Paquette,<sup>1</sup> M Pertesi,<sup>4</sup> N Robinot,<sup>4</sup> J S Rosenthal,<sup>1</sup> M Vallee,<sup>9</sup> C Voegelé,<sup>4</sup> J L Hopper,<sup>10,11</sup> M C Southey,<sup>6</sup> I L Andrusis,<sup>12</sup> E M John,<sup>13,14</sup> M Hashibe,<sup>15</sup> J Gertz,<sup>1</sup> Breast Cancer Family Registry,<sup>13</sup> F Le Calvez-Kelm,<sup>4</sup> F Lesueur,<sup>16</sup> D E Goldgar,<sup>2</sup> S V Tavtigian<sup>1</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2015-103398>).

For numbered affiliations see end of article.

## Correspondence to

Professor Sean V Tavtigian, Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA; sean.tavtigian@hci.utah.edu

Received 17 July 2015

Revised 17 December 2015

Accepted 18 December 2015

Published Online First

19 January 2016

## ABSTRACT

**Background** Moderate-risk genes have not been extensively studied, and missense substitutions in them are generally returned to patients as variants of uncertain significance lacking clearly defined risk estimates. The fraction of early-onset breast cancer cases carrying moderate-risk genotypes and quantitative methods for flagging variants for further analysis have not been established.

**Methods** We evaluated rare missense substitutions identified from a mutation screen of *ATM*, *CHEK2*, *MRE11A*, *RAD50*, *NBN*, *RAD51*, *RINT1*, *XRCC2* and *BARD1* in 1297 cases of early-onset breast cancer and 1121 controls via scores from Align-Grantham Variation Grantham Deviation (GVGD), combined annotation dependent depletion (CADD), multivariate analysis of protein polymorphism (MAPP) and PolyPhen-2. We also evaluated subjects by polygenotype from 18 breast cancer risk SNPs. From these analyses, we estimated the fraction of cases and controls that reach a breast cancer  $OR \geq 2.5$  threshold.

**Results** Analysis of mutation screening data from the nine genes revealed that 7.5% of cases and 2.4% of controls were carriers of at least one rare variant with an average  $OR \geq 2.5$ . 2.1% of cases and 1.2% of controls had a polygenotype with an average  $OR \geq 2.5$ .

**Conclusions** Among early-onset breast cancer cases, 9.6% had a genotype associated with an increased risk sufficient to affect clinical management recommendations. Over two-thirds of variants conferring this level of risk were rare missense substitutions in moderate-risk genes. Placement in the estimated  $OR \geq 2.5$  group by at least two of these missense analysis programs should be used to prioritise variants for further study. Panel testing often creates more heat than light; quantitative approaches to variant prioritisation and classification may facilitate more efficient clinical classification of variants.

## INTRODUCTION

For the last 15+ years, most clinical cancer genetics involving predisposition to breast (and ovarian) cancer have been driven by mutation screening of *BRCA1* and *BRCA2*. For these two genes, the ratio of truncating and splice junction variants (T+SJV) to pathogenic rare missense substitutions (rMS) is

about 10:1,<sup>1,2</sup> fostering a view among clinical cancer geneticists that rMS are only a minor part of the spectrum of breast cancer predisposing variants.

Following the discoveries of *BRCA1* and *BRCA2*, many additional genes have been identified as breast cancer susceptibility genes. A prominent group of these are referred to as moderate-risk susceptibility genes because protein truncating variants and severely dysfunctional missense substitutions in them appear to confer, on average, twofold to fivefold increased risk of breast cancer. This magnitude of increased risk is less dramatic than that conferred by most pathogenic alleles in the high-risk genes *BRCA1*, *BRCA2* and *PALB2*, but potentially high enough to influence the medical management of carriers.<sup>3–6</sup> Beyond the moderate-risk genes, many common SNPs have been identified as markers for slightly increased breast cancer risk.<sup>7–9</sup> A challenge posed by these modest-risk SNPs is that, individually, they do not confer enough risk to influence the medical management of a carrier, but considered as an ensemble they may.

In our previously published case-control mutation screening studies of *ATM*, *CHEK2*, *MRE11A*, *NBN*, *RAD50*, *RAD51*, *RINT1* and *XRCC2*,<sup>10–15</sup> we repeatedly found, albeit with some variations in the methodology, that the summed frequency of predicted deleterious missense substitutions exceeded that of protein truncating variants. This study used the same ethnically diverse sample of 1297 breast cancer cases and 1121 controls, negative for pathogenic variants in *BRCA1*, *BRCA2* or *PALB2* (table 1). Here, we added *BARD1* mutation screening data to the original gene-by-gene analyses and applied consistent analytic models across the rare variants from all nine genes. Setting an  $OR \geq 2.5$  as a threshold for clinical significance, we estimated the scores from the missense substitution analysis programs Align-GVGD,<sup>16</sup> CADD,<sup>17</sup> MAPP<sup>18</sup> and PolyPhen-2<sup>19</sup> required to identify a group of missense substitutions that reach an average  $OR \geq 2.5$ . These results were used to determine the proportion of cases and controls carrying a potential risk-conferring rMS. We also explored a combined evaluation of 18 Breast Cancer Association Consortium (BCAC)-confirmed



Open Access  
Scan to access more  
free content



CrossMark

**To cite:** Young EL, Feng BJ, Stark AW, et al. *J Med Genet* 2016;**53**:366–376.

**Table 1** Distribution of cases and controls by age, race/ethnicity and study centre

Distributions	Case-control mutation screening for rare variants in nine moderate-risk genes				Case-control SNP genotyping for 18 BCAC SNPs			
	Case	%	Control	%	Case	%	Control	%
Age range, years								
≤30	106	8.2	67	6.0	97	7.8	61	5.8
31–35	319	24.6	171	15.3	300	24.3	157	14.9
36–40	433	33.4	238	21.2	409	33.1	220	20.8
41–45	439	33.9	203	18.1	430	34.8	183	17.3
46–50	0	0	230	20.5	0	0	225	21.3
51–55	0	0	212	18.9	0	0	211	20.0
Race/ethnicity								
Caucasian	840	64.8	967	86.3	788	63.8	904	85.5
East Asian	202	15.6	71	6.3	193	15.6	70	6.6
Latina	158	12.2	47	4.2	158	12.8	47	4.5
Recent African Ancestry	97	7.5	36	3.2	97	7.9	36	3.4
Study centre								
BCFR-Australia	588	45.4	522	46.6	551	44.6	472	44.7
BCFR-Canada	299	23.1	463	41.3	284	23.0	499	42.5
BCFR-Northern California	410	31.6	136	12.1	401	32.4	136	12.9
Total	1297		1121		1236		1057	

Subjects were excluded from mutation-screening if performance was poor; percentage data are the total number of cases on control DNA in the category indicated that met the mutation-screening quality control standards.  
BCAC, Breast Cancer Association Consortium; BCFR, Breast Cancer Family Registry.

modest-risk SNPs as a polygene, compared predicted to empirically observed ORs and estimated the prevalence of genotype combinations across these 18 SNPs with an average  $OR \geq 2.5$ . Our data set is unique in that the moderate-risk gene mutation screening and SNP genotyping were performed on the same subjects, giving us the opportunity to compare prevalence of the  $OR \geq 2.5$  threshold across T+SJV and  $OR \geq 2.5$  groupings of rMS or normalised polygene score (NPS).

## MATERIALS AND METHODS

### Subjects

Patients were selected from women systematically recruited by population-based sampling by the Australian, Northern Californian and Ontarian sites of the Breast Cancer Family Registry (BCFR). Patients were recruited between 1995 and 2005. The selection criteria for cases ( $N=1297$ ) were diagnosis at or before age 45 years and self-reported race/ethnicity plus grandparents' country of origin consistent with Caucasian, East Asian, Hispanic/Latino or Recent African racial or ethnic heritage.<sup>20</sup> The controls ( $N=1121$ ) were frequency matched to cases within each centre on racial or ethnic group, with age at selection not more than  $\pm 10$  years difference from the age range at diagnosis of the patients systematically recruited from the same centre.

### Mutation screening and SNP genotyping

Mutation screening was as described previously<sup>10–15</sup> and is included in online supplementary methods, as is SNP genotyping. The following methods focus on the analysis of missense substitutions and of the ensemble of 18 modest-risk SNPs.

### Allele frequency threshold

Following our allele frequency analysis of *ATM*, *BRCA1*, *BRCA2* and *CHEK2* from Damiola *et al*,<sup>12</sup> we applied a minor allele frequency (q) threshold of  $\leq 0.1\%$ , based on exome variant server and 1000 genomes project allele frequency data

that are independent of this study's mutation screening, for all variants of the eight genes in which biallelic truncating variants are often either embryonic lethal or else cause a highly deleterious phenotype from the ataxia telangiectasia/Fanconi anaemia spectrum. Biallelic *CHEK2* carriers are superficially healthy, and our analysis suggested a cut-off of  $q < 0.32\%$  for that gene.<sup>12</sup>

### In silico missense substitution scoring

Align-GVGD ([agvgd.iarc.fr/agvgd\\_input.php](http://agvgd.iarc.fr/agvgd_input.php)) and MAPP ([mendel.stanford.edu/SidowLab/downloads/MAPP/index.html](http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html)) require user-supplied protein multiple sequence alignments (pMSAs) to score missense substitutions; both compare the physicochemical features of the missense residue to the physicochemical range of variation at the relevant position in the pMSA to calculate their scores. Align-GVGD produces a score with seven discrete grades from C0 (most likely neutral) to C65 (most likely deleterious). MAPP, which additionally requires a phylogenetic tree detailing the evolutionary relationships and distances between the organisms with sequences represented in the pMSA, outputs a continuous variable, the MAPP score.

For programs that require a user-generated pMSA, it has been suggested that the pMSA for each gene needs enough variation to average at least three amino acid substitutions per position (3S/P).<sup>21</sup> For each gene, we created an initial pMSA containing the human sequence and 13 additional orthologs. To maintain harmony across the pMSAs, orthologs were sampled from a phylogenetically similar set of organisms ranging from a non-human primate (*Macaca mulatta*) to the non-chordate deuterostomate *Strongylocentrotus purpuratus* (see details in online supplementary methods).

Ortholog sequences downloaded from GenBank were aligned using the *expresso* extension of T-Coffee to create the initial pMSA.<sup>22–23</sup> The initial alignment was checked by hand in Geneious V7.1.4 (<http://www.geneious.com>) for anomalies that might be attributed to gene model errors rather than actual

sequence divergence. Potential anomalies were corrected by reference to, and gene reprediction from, genomic DNA sequence available on the UCSC genome browser (<http://genome.ucsc.edu>).

Routines from the PHYLIP package (V3.69),<sup>24</sup> constrained using the known phylogeny of the species included in our alignments, were used to estimate substitutions per position within each alignment and to calculate the distance matrices required by MAPP. The complete alignments, phylogenetic trees and distances are available upon request.

PolyPhen-2 ([genetics.bwh.harvard.edu/pph2/](http://genetics.bwh.harvard.edu/pph2/))<sup>25</sup> and CADD ([cadd.gs.washington.edu/](http://cadd.gs.washington.edu/))<sup>17</sup> operate without user-supplied alignments. PolyPhen-2 uses a combination of internally generated pMSAs, functional annotations and structural information to evaluate missense substitutions;<sup>19</sup> we used its output variable 'pph2\_prob' as a continuous variable score. CADD uses a series of 63 gene annotations, combined through a support vector machine linear kernel, to define a PHRED-like score (their 'scaled C-score') for all possible single-nucleotide substitutions and small insertion–deletion mutations to the human genome.<sup>17</sup>

Although CADD has a built-in method for short indels, the other missense analysis programs do not. For Align-GVGD, MAPP and PolyPhen-2, nonsense substitutions in the final exon and non-frameshift indels received the score of the most severe missense substitution possible in the affected interval. Variant pathogenicity scores are summarised in online supplementary table S1.

### Statistics

To assess evidence of risk from the case–control frequency distribution of T+SJV and rMS, we constructed a table with one entry per subject; the variants per subject; and annotations for whether the variant was in a key functional domain (see online supplementary table S2), its frequency, as well as study centre, case–control status, race/ethnicity and age for the subject. In-frame deletions (IFDs) were treated as rMS. For the subjects who carried more than one rare variant of interest, only the most deleterious score was considered. We then divided the subjects into groups: a reference group of non-carriers and carriers of common variants (only), carriers of rMS not in a functional domain, carriers of rMS in a key functional domain divided into two groups via score and carriers of T+SJVs. For each of the four rMS analysis programs, we toggled the program's severity score from a very relaxed to a very stringent value. We repeatedly estimated two ORs as the stringency increased: the OR for subjects that carried one or more rMS at or above the score (and no T+SJV), and the OR for subjects who carried an rMS that was below the score (and carried neither a T+SJV nor a higher scoring rMS). From this analysis, we determined a threshold severity score for each program at which subjects carrying an rMS at or above the threshold had an average  $OR \geq 2.5$ .

For the 18 SNP polygene, we created a polygenic risk score (PRS) by multiplying together the appropriate published OR estimate from each individual SNP genotype.<sup>7, 9, 26, 27</sup> The geometric mean of the PRS of the controls was used to normalise the PRS into an NPS. Because risk estimates from Caucasian populations may not be applicable to women from other race/ethnicities, we gave each non-Caucasian subject a race/ethnicity-specific risk estimate derived from the population of the subject,<sup>26, 28–30</sup> and normalised each race/ethnicity separately. Risk estimates of rs1045485 could not be found for the non-Caucasian subjects in this study, so the risk estimate based

on Caucasian populations was used for all race/ethnicities. In instances where a Latina-specific risk estimate could not be found, we used the average between Caucasian and East Asian race/ethnicities. To determine the correlation between NPS and the observed OR, we grouped the subjects into a series of 10 contiguous bins based on percentile, using the central quintile (40–60 percentile) as the reference group. We treated groups outside of the reference as categorical variables for OR calculations. For the threshold analysis, we used the same reference group and adjusted the NPS threshold until the group containing scores above the set threshold had an  $OR \geq 2.5$ .

For the regressions, NPS was treated as the independent variable and the resultant OR of each group as the dependent variable weighted to the number of individuals in each group, excluding subjects in the middle quintile. p Values were found by testing the regression coefficient equal to 0 or 1. To combine the risk estimates from the rMS and NPS, we multiplied the NPS and OR from the rMS.

All analyses were performed using multivariable unconditional logistic regression using Stata V12.1 software (StataCorp, College Station, Texas, USA). Adjustments were made for race/ethnicity and study centre, unless otherwise noted.

## RESULTS

### Initial evaluation of rare variants

From mutation screening of 1297 cases and 1121 controls, we observed 22 T+SJV, 9 IFDs and 196 rMS with minor allele frequencies  $<0.32\%$  for *CHEK2* and  $<0.1\%$  for the remaining genes. T+SJVs falling before the final exon were considered pathogenic and were associated with an OR of 3.32 ( $p=0.0023$ , table 2). Nonsense mutations located in the final exon were considered as IFDs.

The National Comprehensive Cancer Network (NCCN) and the American College of Radiology (ACR) recommend screening beginning at age 30 years and offering breast MRI in addition to mammograms for women with a  $\geq 20\%$  lifetime breast cancer risk.<sup>31, 32</sup> The American Cancer Society (ACS) recommends breast MRI for women with a 20–25% or greater lifetime risk.<sup>33</sup> In the USA, the lifetime risk of a woman to develop breast cancer is estimated to be 12.3%;<sup>34</sup> however, this figure is an overestimate for our purposes because it includes women who are at high risk because of inherited mutations in genes such as *BRCA1* or *BRCA2*, or very strong family history. For a woman with minimal risk factors, for example, age at menarche  $\geq 14$  years, first childbirth at age  $\leq 20$  years and no family history, the Gail model<sup>35</sup> and Tyrer–Cuzick model<sup>36</sup> suggest a lifetime risk of 6.9% and 11% for developing breast cancer, respectively. If we assume that the average of these two estimates (9%) is approximately correct, carriage of a genotype conferring a 2.5-fold increase of risk, even in this low-risk population, would result in a lifetime risk estimate exceeding the NCCN, ACR and American Cancer Society (ACS) medically actionable threshold of a 20% lifetime risk. Subject to formal variant classification, carriers may then qualify, under current recommendations, for early mammography and/or enhanced screening with breast MRI. We note that threshold for intensified screening may be higher in other countries.

Considering all rMS as a group, we obtained a risk estimate that was elevated but that did not reach an  $OR \geq 2.5$  threshold ( $OR=1.42$ ,  $p=0.0091$ , table 2). We focused our analyses of rMSs to those that are relatively likely to impact key functions. This grouping included all of the rMS from the relatively small proteins encoded by *CHEK2*, *RAD51*, *RINT1* and *XRCC2*. Noting the structural similarity between *BARD1* and *BRCA1*,

**Table 2** Combined OR estimates from case-control mutation screening of nine moderate-risk genes

Analysis	Distinct variants	Control	%	Case	%	Adjusted OR*	CI	p Value
Non-carrier	148	998	89.0	1094	84.4	Reference		
Carrier of truncating or splice junction variant	22	9	0.8	27	2.1	3.31	1.53 to 7.16	2.36×10 <sup>-3</sup>
<i>Rare missense substitution analyses</i>								
Carrier of rare missense substitution	205	114	10.2	176	13.6	1.42	1.09 to 1.84	8.70×10 <sup>-3</sup>
Carrier of key domain rare missense substitution	140	65	5.8	136	10.5	1.94	1.41 to 2.67	5.11×10 <sup>-5</sup>
Carrier of non-key domain rare missense substitution	65	49	4.4	40	3.1	0.74	0.48 to 1.16	0.1926
Key domain rMS: MAPP								
rMS<11	63	39	3.5	58	4.5	1.47	0.95 to 2.26	0.0818
rMS≥11	77	26	2.3	78	6.0	2.63	1.65 to 4.21	5.32×10 <sup>-5</sup>
Key domain rMS : Align-GVGD								
rMS<C35	93	54	4.8	87	6.7	1.58	1.09 to 2.27	0.0146
rMS≥C35	47	11	1.0	49	3.8	3.62	1.84 to 7.13	2.03×10 <sup>-4</sup>
Key domain rMS : CADD								
rMS<23	97	50	4.5	88	6.8	1.66	1.14 to 2.41	0.0082
rMS≥23	43	15	1.3	48	3.7	2.87	1.57 to 5.26	6.40×10 <sup>-4</sup>
Key domain rMS : PolyPhen-2								
rMS<0.9	61	37	3.3	55	4.2	1.50	0.96 to 2.34	0.0752
rMS≥0.9	79	28	2.5	81	6.3	2.49	1.58 to 3.92	7.88×10 <sup>-5</sup>
Overlap of missense analysis programs								
One or more	89	34	3.0	93	7.2	2.37	1.57 to 3.60	4.56×10 <sup>-5</sup>
Two or more	65	18	1.6	70	5.4	3.18	1.85 to 5.46	2.68×10 <sup>-5</sup>
Three or more	45	14	1.2	52	4.0	3.27	1.77 to 6.04	1.51×10 <sup>-4</sup>
All four	19	2	0.2	20	1.5	8.61	1.96 to 37.81	4.35×10 <sup>-3</sup>
Total	375	1121		1297				

\*Adjusted for race/ethnicity and study centre.  
rMS, rare missense substitutions.

and that *BRCA1* pathogenic rMS are so far only known from the RING and BRCT domains,<sup>37–40</sup> we limited analyses of *BARD1* rMS to RING and BRCT domain substitutions. For the relatively large proteins encoded by *ATM*, *MRE11A*, *NBN* and *RAD50*, we focused analyses of rMS on the same key functional domains specified in our prior publications (see online supplementary table S2).<sup>10–12</sup> We observed 140 rMS and IFDs in key functional domains (OR 1.94,  $p=5.1\times 10^{-05}$ , table 2), which still did not reach an  $OR\geq 2.5$  threshold.

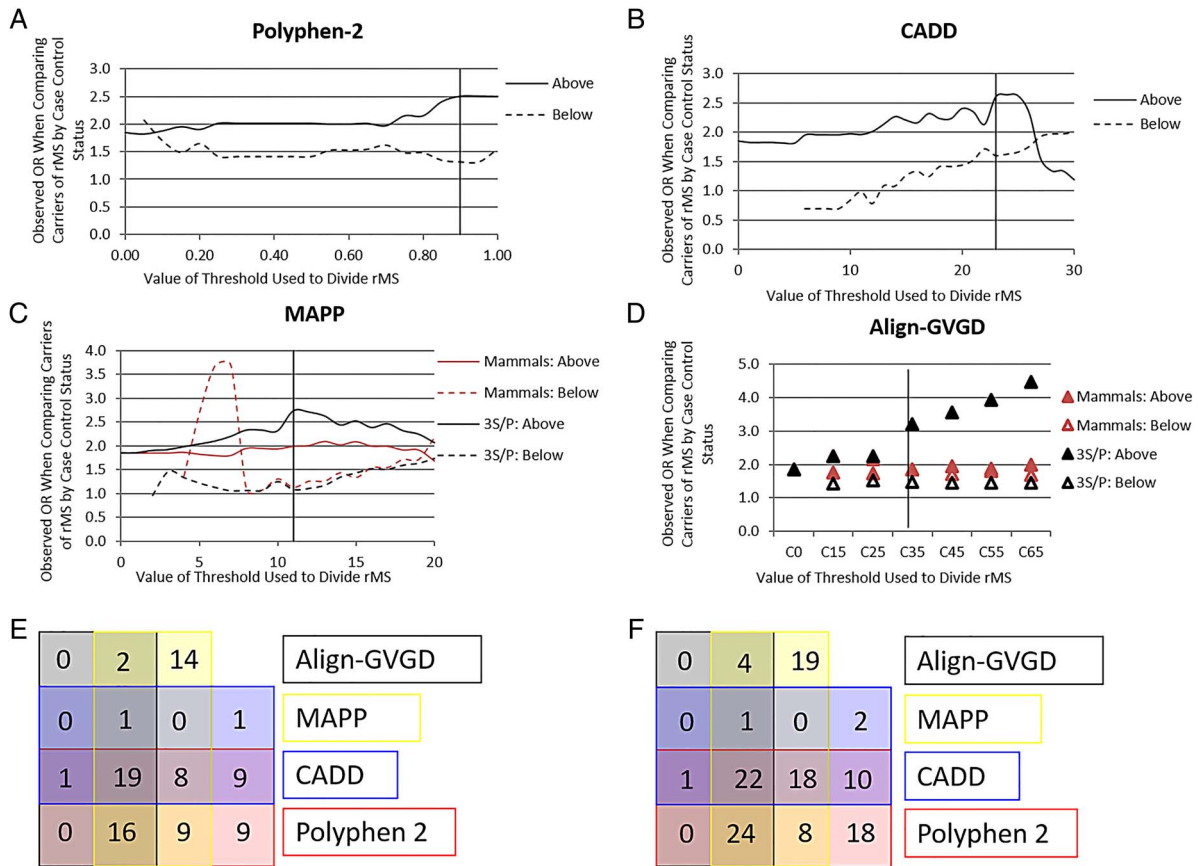
### Grouping rMS to estimate risk and carrier rates

To define a higher-risk subset of rMS, we focused the next analyses on the three established moderate-risk genes: *ATM*, *CHEK2* and *NBN*.<sup>41</sup> There is no fully accepted method for rMS analysis. Instead of introducing a new method for variant classification, we used four existing missense analysis programs, Align-GVGD, CADD, MAPP and PolyPhen-2,<sup>16–18, 25</sup> to assign severity scores to the key domain rMS from these genes. Align-GVGD was selected because its scores contribute to determination of prior probabilities of pathogenicity for key domain missense substitutions in *BRCA1* and *BRCA2*,<sup>2</sup> MAPP and PolyPhen-2 because of their strong performance in our recent analyses of mismatch repair protein missense substitutions,<sup>42</sup> and CADD because of its reported ability to prioritise variants across functional categories and effect sizes.<sup>17</sup> For variant evaluation, we adjusted our pMSAs to two depths: human through platypus (mammals only), and human through the organism required for 3S/P for each individual gene. Receiver operating characteristic (ROC) curves were generated for each method and depth (if applicable). Area under the curve (AUCs) were similar for all methods (see online supplementary table S3 and supplementary figure S1). The correlations between the missense analysis programs were highest between Polyphen-2 and

CADD ( $R^2=0.56$ ), but the  $R^2$  for any combination of missense analysis programs was never  $>0.8$ , so none of the missense analysis programs were dropped from further analysis (see online supplementary table S4).

We then toggled the severity score for each of the four programs to find the lowest score where the OR for key domain rMS above the score reached at least 2.5 (figure 1A–D). The thresholds at which each of the four rMS analysis programs reached  $OR\geq 2.5$  for key domain rMS were above a score of 11 for MAPP when using pMSAs consisting of organisms from human through 3S/P, C35 for Align-GVGD when using pMSAs consisting of organisms from human through 3S/P, 23 for CADD and 0.9 for PolyPhen-2 (see online supplementary table S5). It was interesting that neither Align-GVGD nor MAPP was able to achieve an  $OR\geq 2.5$  with a pMSA that consisted only of mammals (human through platypus). It appears that these sequences, although generally more complete than those from more distant organisms, do not offer adequate variation to stratify variants. Examining the variants that were placed in the  $OR\geq 2.5$  category by multiple missense analysis programs, we found that an overlap of at least two of the missense analysis programs resulted in a classification of variants with an  $OR\geq 2.5$  (OR 2.59,  $p=0.0044$ ; online supplementary table S6).

Applying the score thresholds determined from the *ATM-CHEK2-NBN* group to the key domain rMS observed in the remaining six less established moderate-risk genes, rMS ORs for the *BARD1-MRE11A-RAD50-RAD51-RINT1-XRCC2* group ranged from 2.41 ( $p=0.0078$ ) using PolyPhen-2 to 4.86 ( $p=0.0129$ ) using Align-GVGD (data not shown). We also found that concordance between at least two of the missense analysis programs resulted in a grouping of rMS with an  $OR\geq 2.5$  (OR 4.90,  $p=0.0012$ ; online supplementary table S6).



**Figure 1** Observed OR and thresholds for each missense substitution analysis program. The observed OR for the carriers of key domain rare missense substitutions (rMS) of both the ‘above’ and ‘below’ groups for each severity threshold tested with (A) PolyPhen-2, (B) CADD, (C) MAPP and (D) Align-GVGD, adjusting for race/ethnicity and study centre for just the combined ensemble of ATM, CHEK2 and NBN. Vertical lines are indicative of the threshold for which the observed  $OR \geq 2.5$ . (E) A four-way Venn diagram detailing the number of rMS for which CADD, PolyPhen-2, Align-GVGD and MAPP placed in the ‘above threshold’ group for key domain rMS observed in all nine moderate-risk genes. (F) A four-way Venn diagram detailing the number of individuals for which CADD, PolyPhen-2, Align-GVGD and MAPP were placed in the ‘above threshold’ group for key domain rMS observed in all nine moderate-risk genes. Mammals=score obtained from using protein multiple sequence alignments (pMSA) containing sequences from human through platypus; 3S/P=scores are from gene-specific 3S/P depth pMSAs.

Having established that the thresholds identified with the *ATM-CHEK2-NBN* group were able to extract  $OR \geq 2.5$  groupings from the remaining six genes, we used these thresholds to evaluate the proportions of cases and controls with above-threshold variants across the nine-gene ensemble (table 2). We found that 3.7–6.3% of cases and 1.0–2.5% of controls carried an above-threshold key domain rMS. Considering only the key domain rMS that were placed in an  $OR \geq 2.5$  grouping by more than one of the missense analysis programs (figure 1E,F), concordance between two or more missense analysis programs was associated with an  $OR \geq 2.5$  ( $OR\ 3.18, p=2.68 \times 10^{-5}$ ), affecting 5.4% of cases and 1.6% of controls (table 2). These results are comparable to other studies, but include data from controls.<sup>43–45</sup> Comparing the proportion of above-threshold key domain rMS carriers to T+SJV carriers, rMS carriers appear to outnumber T+SJV carriers by a ratio of about 2.5:1.

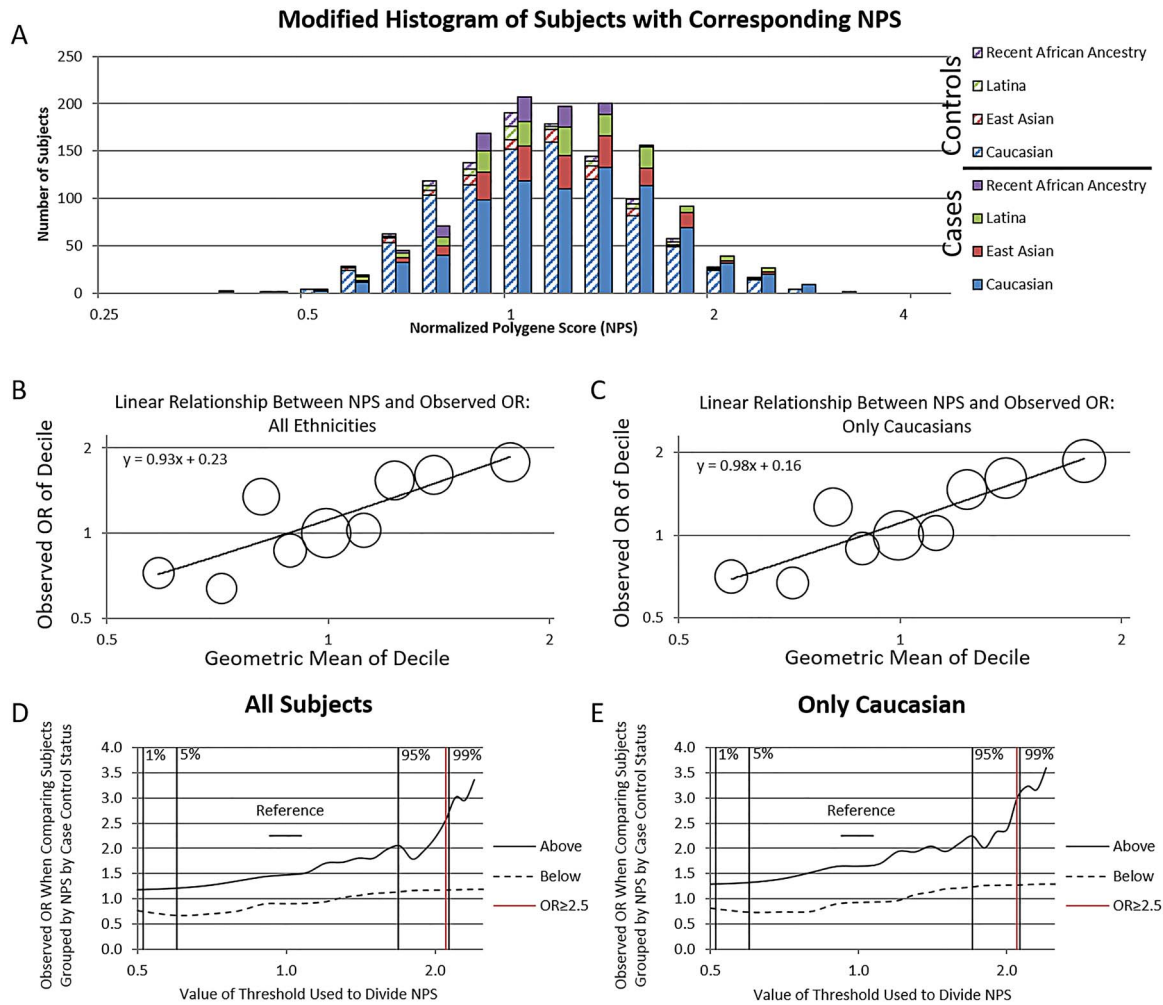
Asking whether the results reported here are robust to the loss of any one gene from the less established moderate-risk gene set, we performed a series of analyses in which the genotype information of one of the genes was dropped and then the OR, rMS to T+SJV ratio, and carrier percentage was re-determined for the rMS from the remaining eight genes. We observed that, in each subset of eight genes, 3.0–5.9% of cases and 0.8–2.7% of controls were carriers of a variant from the above-threshold grouping, with a ratio of rMS to T+SJVs

consistently over 1.6:1 for cases (see online supplementary table S7).

### Common SNP-based polygene scores and above-threshold carrier rates

Generally, individual modest-risk SNPs do not confer enough risk to impact clinical practice. An attractive method for using SNPs in a clinical setting is to combine the risk estimates from multiple SNPs. Indeed, a recent large study combined risk estimates from 77 SNPs and found  $ORs \geq 2.5$  at and above the 99th percentile of the combined scores.<sup>9</sup> We genotyped 18 BCAC-confirmed SNPs on the same subjects from the case-control mutation screening phase of this study (table 1). Using per-allele ORs from recent large studies,<sup>7, 26, 28–30</sup> we treated the SNPs as a polygene and created an NPS for each subject (see online supplementary table S8 and figure 2A).

To determine how closely the NPS predicted OR, we grouped the NPS scores into deciles and compared the mean NPS of each decile to its observed OR. With all subjects grouped together, the NPS correlated highly with the observed OR (coeff. = 0.9232,  $R^2=0.70, p=0.0060$ ) (figure 2B). Evaluating each race/ethnicity individually, Caucasians were the only group to achieve significance (coeff.=0.9835,  $R^2=0.81, p=0.0014$ ) (figure 2C, data not shown), likely due to small sample sizes of the non-Caucasian groups. We also tested the alternate



**Figure 2** The normalised polygene score (NPS): distribution, NPS-OR correlation and threshold for medically actionable. (A) NPS distribution for all subjects. Comparison of observed OR and NPS for each decile for (B) all subjects and (C) only Caucasians, with the corresponding equations derived from linear regressions, excluding the central quintile. The observed OR when dividing subjects based on NPS, using the middle quintile as reference for (D) all subjects and (E) only Caucasians, adjusting for race/ethnicity and study centre. Vertical lines are indicative of the threshold for which the observed  $OR \geq 2.5$ , as well as the 1, 5, 95 and 99 percentiles. Bubble sizes are proportional to the number of subjects in each decile.

hypothesis,  $NPS = \text{observed OR}$ , and did not observe a significant difference ( $p = 0.75$  and  $0.93$  for all subjects and Caucasians, respectively).

Using the NPS, how many women are at a medically actionable risk? Using an approach analogous to that applied to the key domain rMS, we toggled the NPS to find the lowest score where the observed OR for the group of subjects with NPS above the score exceeded 2.5 (figure 2D–E). From the data, 2.1% of cases and 1.2% of controls carry a combination of SNPs such that they have a polygene score associated with an average  $OR \geq 2.5$  (table 3). Limiting the analysis to Caucasians, we found that 3.2% of cases and 1.3% of controls have an NPS associated with an average  $OR \geq 2.5$ .

To explore the possibility of integrating gene mutation screening with SNP genotyping, we tested for interactions between carriage of an  $OR \geq 2.5$  rare variant (combining T+SJV and above-threshold rMS into a single group) and the NPS. In these tests, the interaction term never approached significance ( $p = 0.52, 0.82, 0.51$  and  $0.96$  for analyses with rMS scored by Align-GVGD, CADD, MAPP and PolyPhen-2, respectively). Accordingly, as the multiplicative OR model does appear to apply to combinations of rare variants from these nine genes with the NPS, we isolated the subjects who carried a rare

variant in the  $OR \geq 2.5$  category and multiplied the OR estimated from their rMS or T+SJV with their NPS. These combined ORs varied from  $\sim 1.0$  to  $> 5.0$  (figure 3).

## DISCUSSION

Neither pathogenic alleles in moderate-risk breast cancer susceptibility genes nor individual modest-risk breast cancer-associated SNPs confer the magnitude of risk of early-onset breast cancer conferred by pathogenic alleles in high-risk genes such as *BRCA1* and *BRCA2*. Nonetheless, under a generalised understanding of NCCN, ACR and ACS guidelines, a  $\geq 2.5$ -fold increased risk of breast cancer is high enough to impact the medical management of otherwise healthy carriers. Across the nine moderate-risk susceptibility genes examined here, two classes of sequence variants meet or exceed this 2.5-fold risk threshold. 2.1% of cases carried a T+SJV, and these were associated with an OR of 3.32 ( $p = 0.0023$ ). Each of the four missense substitution analysis programs that we evaluated was able to define a set of key functional domain rMS that reached the  $OR \geq 2.5$  threshold. 5.4% carried an rMS that two or more of the programs agreed was above the threshold, and this group of rMS was associated with an OR of 3.18 ( $p = 2.68 \times 10^{-5}$ ). In

**Table 3** OR estimates and p-values for case–control screening of 18 Breast Cancer Association Consortium SNPs

	Control	%	Case	%	Adjusted OR	CI*	p> z
Utilizing NPS as a continuous score							
All	1057	100	1236	100			5.76×10 <sup>-10</sup>
Only Caucasians	904	85.5	788	63.8			3.47×10 <sup>-10</sup>
Excluding Caucasian	153	14.5	448	36.3			0.32
Number of subjects at risk indicated by NPS score above Threshold group							
All†	13	1.2	26	2.1	2.56	1.26 to 5.19	0.009
Only Caucasian†	12	1.3	25	3.2	3.02	1.45 to 6.29	0.003
Excluding Caucasian					Never	≥2.5	
OR of top and bottom percentiles using middle quintile as reference							
All							
0–1%	11	1.0	5	0.4	0.50	0.16 to 1.56	0.232
1–5%	42	4.0	28	2.3	0.72	0.42 to 1.24	0.239
95–99%	41	3.9	73	5.9	1.99	1.27 to 3.12	0.003
99–100%	11	1.0	22	1.8	2.74	1.27 to 5.90	0.010
Only Caucasian							
0–1%	10	1.1	3	0.4	0.48	0.13 to 1.81	0.278
1–5%	36	4.0	19	2.4	0.80	0.44 to 1.48	0.482
95–99%	36	4.0	51	6.5	2.00	1.22 to 3.26	0.006
99–100%	10	1.1	22	2.8	3.20	1.45 to 7.08	0.004

\*Adjusted for study centre and race/ethnicity.

†NPS≥2.1. NPS, normalised polygene score.

addition, 2.1% of cases carried an above-threshold SNP polygene genotype.

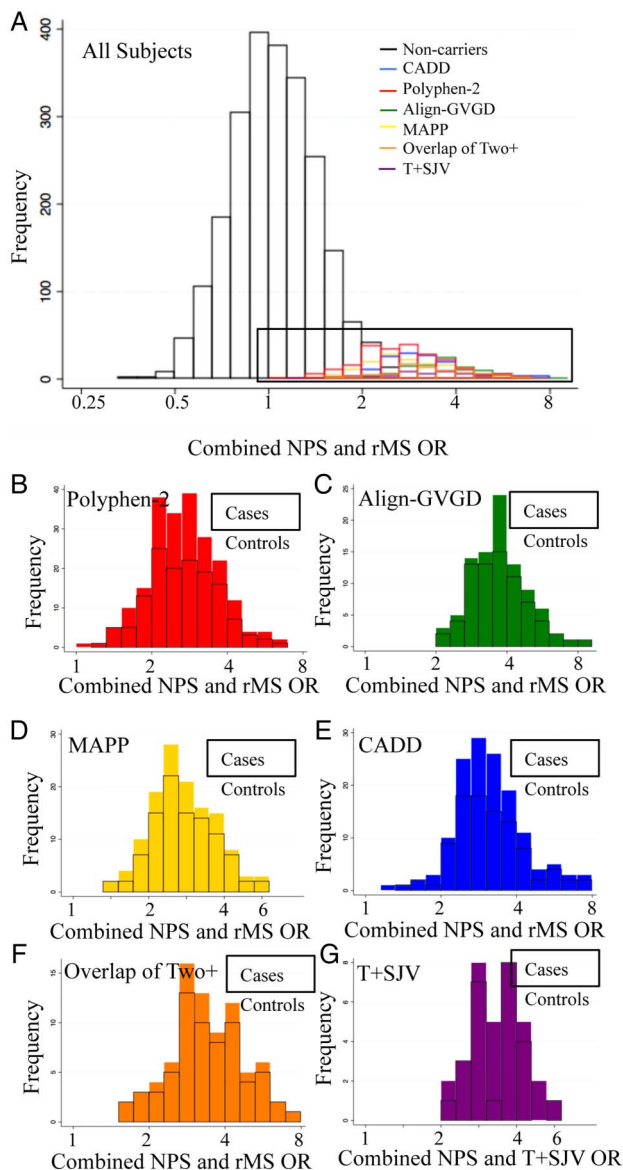
Whether focusing on the confirmed moderate-risk genes *ATM*, *CHEK2* and *NBN* or looking at all nine genes, the ratio of carriers of T+SJV to above-threshold key domain rMS was in the range of 1:2–1:3. This finding is different than the ~10:1 ratio observed in *BRCA1/2*; the preponderance of above-threshold rMS in these moderate-risk genes is much more reminiscent of the situation with *TP53*.<sup>5</sup> Because most rMS observed during clinical testing of these moderate-risk genes would be returned as variants of uncertain significance (VUS) in test reports, the relatively high proportion of above-threshold rMS reported here creates a challenge for test interpretation. During clinical counselling, to alleviate patient distress, observations of VUS rMS, especially in moderate-risk genes, are often downplayed as of minimal significance—‘normalised’. However, at least for the nine genes that we examined, normalising the rMS amounts to disregarding approximately 2/3 of sequence variants with OR≥2.5 detectable by the genetic tests.

Within the logical structure of the analyses presented here, the OR≥2.5 threshold applied to rMS and SNP polygene groupings was a device used to align the analyses with current patient management standards. A consequence is that the OR point estimates reported for those groupings in tables 2 and 3 are circularly dependent on the threshold selected. Nonetheless, the following four key results are independent to the circular logic underlying those OR point estimates: (i) the a priori existence of groupings with OR≥2.5; (ii) the p values associated with those groupings; (iii) the ratios of subjects with T+SJVs, rMSs with OR≥2.5 and SNP polygene with OR≥2.5; and (iv) the frequencies among controls and early-onset cases of individuals with a genotype falling into one of these groupings. These findings all correspond to open, medically relevant questions.

Although we did not accompany this study with functional assays, a yeast complementation assay applied to 25 *CHEK2* missense substitutions included applicable Align-GVGD and

PolyPhen-2 scores.<sup>46</sup> Among the six rMS with Align-GVGD and PolyPhen-2 scores meeting our severity criterion, the average activity was –0.062 (SD=0.027) in an assay where the internal wild-type and dysfunctional variants were given scores of 1.00 and 0.00, respectively. In contrast, the average score among the 19 rMS not meeting our concordant severity criterion was +0.472 (SD=0.388), resulting in a p value of 1.12×10<sup>-5</sup> against the hypothesis that the two groups have the same mean activity. Moving forward, it will become important to develop methods that combine patient observational data with in silico and functional assay results towards clinical classification of these rMS. Such methods may leverage the Bayesian classification framework already developed for rMS in *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, etc.<sup>37 47 48</sup>

One weakness of this study is that it focuses on early-onset cases using a data set that already contributed either to association of rMS in these genes with breast cancer susceptibility (*ATM*, *CHEK2*),<sup>10 11</sup> or susceptibility to breast cancer in general (*MRE11A*, *NBN*, *RAD50*, *RINT1*, *XRCC2*).<sup>12 14 15</sup> While the impact on the overall results of a possible false association for one or another of the genes is addressed by the leave-one-out analysis, the possibility remains that the ORs that we report are systematically inflated either because this was a study of early-onset cases or because of winner’s curse.<sup>41</sup> These issues were partly ameliorated in two ways: (i) an OR≥2.5 grouping of rMS could be isolated by each of the four rMS analysis programs that we used, and (ii) the group of women that can benefit most from early or intensified breast cancer screening is primarily those at risk of early-onset breast cancer—largely, the group of women from which the cases used in this study are drawn. Looking forward, the ratio of the above-threshold rMS to T+SJV can be re-evaluated in case–control studies, but accurate assessment of risk will have to come from prospective cohort studies. A second weakness in our analytic strategy is that the rMS analyses in five of the genes included here—*ATM*, *BARD1*, *MRE11A*, *RAD50* and *NBN*—are somewhat dependent on our definitions of key protein functional domains. This



**Figure 3** Combined normalised polygene score (NPS) and rare variant OR. (A) Distribution of subjects by NPS when carriers of rare missense substitutions (rMS) in the 'above threshold' group had their NPS increased by the factor of their OR. Stacked histograms of the expansion of the combined contribution of NPS and rMS for carriers with a variant conferring an average  $OR \geq 2.5$  with (B) PolyPhen-2, (C) Align-GVGD, (D) MAPP, (E) CADD as well as (F) carriers of variants that were placed in the 'above threshold' group for two or more missense analysis programs and (G) truncation and splice junction variant (T+SJV) carriers.

analytic element is in need of independent evaluation and refinement.

Our analysis raises additional questions regarding standard clinical genetic testing practices using panel tests. For the established moderate-risk genes *ATM*, *CHEK2* and *NBN*, the majority of the pathogenic variants that the test can actually detect are rMS, likely to be reported to patients as VUS, and likely to be normalised during counselling. In this circumstance, how does one answer the clinical validity question, "Are the variants the test is intended to identify associated with disease risk, and are these risks well quantified?"<sup>41</sup> What is the impact on studies intended to explore the penetrance and tumour spectrum of pathogenic variants in these genes if the studies focus

on T+SJVs even though these may represent a minority of the pathogenic variants? One path forward lies in a more nuanced use of the IARC 5-class system for variant classification and reporting to incorporate more data from ongoing research on missense substitution evaluation.<sup>49</sup> From work that defined the sequence analysis-based prior probabilities of pathogenicity for rMS in *BRCA1*, *BRCA2* and the mismatch repair genes, one can clearly define subsets of rMS that have relatively high probabilities of pathogenicity.<sup>2, 42</sup> A straightforward approach for clinicians could be to make systematic efforts to enrol carriers of high probability of pathogenicity rMS in research studies, such as those coordinated through the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium,<sup>50</sup> while still describing these findings to patients as VUS. For *BRCA1*, *BRCA2* and the mismatch repair genes, these could be defined as rMS with prior probabilities of pathogenicity of  $\geq 0.66$  as defined at the calibrated prior probability of pathogenicity websites (priors.hci.utah.edu/PRIORS/index.php and hci-lovd.hci.utah.edu/home.php, respectively). rMS from the nine genes examined here that are placed in an  $OR \geq 2.5$  grouping by two or more of the missense analysis programs similarly fall into a relatively high probability of pathogenicity subset. VUS with lower probabilities of pathogenicity could reasonably be normalised since future reclassification to a clearly pathogenic variant is rather unlikely. Such an approach would better prioritise those missense substitutions with high probabilities of pathogenicity, leading to better understanding of these VUS by clinicians and patients. This approach should empower research towards gene validation, penetrance and tumour spectrum and thereby address the question of clinical validity in the future.

#### Author affiliations

- <sup>1</sup>Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
- <sup>2</sup>Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
- <sup>3</sup>Breast Cancer Genetics Group, Cancer Research Centre of Lyon, Centre Léon Bérard, Lyon, France
- <sup>4</sup>Genetic Cancer Susceptibility group, International Agency for Research on Cancer, Lyon, France
- <sup>5</sup>Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
- <sup>6</sup>Department of Communication and Huntsman Cancer Institute, University of Utah
- <sup>7</sup>Genetic Epidemiology Laboratory, The University of Melbourne, Melbourne, Victoria, Australia
- <sup>8</sup>Instituto de Ciencias Básicas y Medicina Experimental del Hospital Italiano de Buenos Aires, Buenos Aires, Argentina
- <sup>9</sup>Cancer Genomics Laboratory, CHUQ Research Center, Quebec City, Canada
- <sup>10</sup>Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
- <sup>11</sup>Department of Epidemiology (Genome Epidemiology Lab), Seoul National University School of Public Health, Seoul, Korea
- <sup>12</sup>Department of Molecular Genetics, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada
- <sup>13</sup>Cancer Prevention Institute of California, Fremont, California, USA
- <sup>14</sup>Department of Health Research and Policy, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA
- <sup>15</sup>Department of Family and Preventive Medicine, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
- <sup>16</sup>Genetic Epidemiology of Cancer Team, Inserm, U900, Institut Curie, Paris, France

**Twitter** Follow Javier Oliver at @javiom

**Acknowledgements** The authors wish to thank all participants in the BCFR for their contribution to the study. The authors also appreciate the support of J. McKay and the Genetic Cancer Susceptibility group at the International Agency for Research on Cancer (IARC).

**Contributors** ELY and BJF were involved in data analysis, critical revision of manuscript and final approval. AWS, FD, GD, NF, TCF, AG, WKK, SMcK-C, TN-D, JO, AMP, MP, NR, JSR, MH, JG, FLC-K, FL, DEG, MV and CV were involved in data acquisition, critical revision of manuscript and final approval. KAK was involved



critical revision of manuscript and final approval. J.L.H., M.C.S., I.L.A. and E.M.J. were involved in the acquisition of subjects, critical revision of manuscript and final approval. S.V.T. was involved in the acquisition of data, critical revision of manuscript, final approval and is the corresponding author.

**Funding** This work was supported by the United States National Institutes of Health (NIH) National Cancer Institute (NCI) grants R01 CA121245 and R01 CA155767, by the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer programme, by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, and the Ministère de l'enseignement supérieur, de la recherche, de la science, et de la technologie du Québec through Génome Québec. The BCFR is supported by grant UM1 CA164920 from the USA National Cancer Institute.

**Disclaimer** The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the BCFR, nor does mention of trade names, commercial products or organisations imply endorsement by the US government or the BCFR.

**Competing interests** None declared.

**Patient consent** Obtained.

**Ethics approval** The studies described here were approved by the institutional review board (IRB) of the International Agency for Research on Cancer (IARC), the University of Utah IRB and the local IRBs of the BCFR centres from which we received samples.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** A table of the data used for this analysis including case-control status and variant calling information is available upon contacting the corresponding author. Additionally, protein multiple sequence alignments for the nine genes are also available upon request.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Abkevich V, Zharkikh A, Deffenbaugh AM, Frank D, Chen Y, Shattuck D, Skolnick MH, Gutin A, Tavtigian SV. Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. *J Med Genet* 2004;41:492–507.
- Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* 2008;29:1342–54.
- Kean S. The 'other' breast cancer genes. *Science* 2014;343:1457–9.
- Stadler ZK, Schrader KA, Vijai J, Robson ME, Offit K. Cancer genomics and inherited risk. *J Clin Oncol* 2014;32:687–98.
- Tavtigian SV, Chenevix-Trench G. Growing recognition of the role for rare missense substitutions in breast cancer susceptibility. *Biomark Med* 2014;8:589–603.
- Byrnes GB, Southey MC, Hopper JL. Are the so-called low penetrance breast cancer genes, ATM, BRIP1, PALB2 and CHEK2, high risk for women with strong family histories?. *Breast Cancer Res* 2008;10:208.
- Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Fletcher O, Peto J, Gibson L, Dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Czene K, Iwanto A, Liu J, Waisfisz Q, Meijers-Heijboer H, Adank M, van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Müller-Miyhok B, Lichtner P, Hopper JL, Southey MC, Makalic E, Schmidt DF, Uitterlinden AG, Hofman A, Hunter DJ, Chanock SJ, Vincent D, Bacot F, Tessier DC, Canisius S, Wessels LFA, Haiman CA, Shah M, Luben R, Brown J, Luccarini C, Schoof N, Humphreys K, Li J, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang X, Vachon C, Stevens KN, Lambrechts D, Moisse M, Paridaens R, Christiaens M-R, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken T, Aaltonen K, Heikkinen T, Broeks A, Veer LJV, van der Schoot CE, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Zamora MP, Perez JIA, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MWR, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Henderson BE, Schumacher F, Le Marchand L, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Lindblom A, Margolin S, Hoening MJ, Hollestelle A, van den Ouweland AMW, Jager A, Bui QM, Stone J, Dite GS, Apicella C, Tsimiklis H, Giles GG, Severi G, Baglietto L, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Brenner H, Müller H, Arndt V, Stegmaier C, Swerdlow A, Ashworth A, Orr N, Jones M, Figueroa J, Lissowska J, Brinton L, Goldberg MS, Labreche F, Dumont M, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Brüning T, Radice P, Peterlongo P, Manoukian S, Bonanni B, Devilee P, Tollenaar RA, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dörk T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo K-Y, Noh D-Y, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng C-C, Van Den Berg D, Stram DO, Shu X-O, Lu W, Gao Y-T, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng J-H, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsang P, Shen C-Y, Hsiung C-N, Wu P-E, Ding S-L, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Blot WJ, Signorello LB, Cai Q, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Simard J, Garcia-Closas M, Pharoah PDP, Chenevix-Trench G, Dunning AM, Benitez J, Easton DF. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;45:353–61, 361e1–2.
- Maxwell KN, Nathanson KL. Common breast cancer risk variants in the post-COGS era: a comprehensive review. *Breast Cancer Res* 2013;15:212.
- Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Czene K, Darabi H, Eriksson M, Peto J, Dos-Santos-Silva I, Dudbridge F, Johnson N, Schmidt MK, Broeks A, Verhoef S, Rutgers EJ, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Figueroa J, Chanock SJ, Brinton L, Lissowska J, Couch FJ, Olson JE, Vachon C, Pankratz VS, Lambrechts D, Wildiers H, Van Ongeval C, van Limbergen E, Kristensen V, Grenaker Alnaes G, Nord S, Borresen-Dale A-L, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Trentham-Dietz A, Newcomb P, Titus L, Egan KM, Hunter DJ, Lindstrom S, Tamimi RM, Kraft P, Rahman N, Turnbull C, Renwick A, Seal S, Li J, Liu J, Humphreys K, Benitez J, Pilar Zamora M, Arias Perez JJ, Menendez P, Jakubowska A, Lubinski J, Jaworska-Katana B, Durda K, Bogdanova NV, Antonenkova NN, Dörk T, Anton-Culver H, Neuhausen SL, Ziogas A, Bernstein L, Devilee P, Tollenaar RAEM, Seynaeve C, van Asperen CJ, Cox A, Cross SS, Reed MWR, Khusnutdinova E, Bermisheva M, Prokofyeva D, Takhirova Z, Meindl A, Schmutzler RK, Sutter C, Yang R, Schurmann P, Bremer M, Christiansen H, Park-Simon T-W, Hillemanns P, Guénel P, Truong T, Menegaux F, Sanchez M, Radice P, Peterlongo P, Manoukian S, Pensotti V, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Brauch H, Brüning T, Ko Y-D, Sigurdson AJ, Doody MM, Hamann U, Torres D, Ulmer H-U, Forsti A, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Andrulis IL, Knight JA, Glendon G, Marie Mulligan A, Chenevix-Trench G, Balleine R, Giles GG, Milne RL, McLean C, Lindblom A, Margolin S, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Eilber U, Wang-Gohrke S, Hoening MJ, Hollestelle A, van den Ouweland AMW, Koppert LB, Carpenter J, Clarke C, Scott R, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Brenner H, Arndt V, Stegmaier C, Karina Dieffenbach A, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Offit K, Vijai J, Robson M, Rau-Murthy R, Dwek M, Swann R, Annie Perkins K, Goldberg MS, Labreche F, Dumont M, Eccles DM, Tapper WJ, Rafiq S, John EM, Whittemore AS, Slager S, Yannoukakos D, Toland AE, Yao S, Zheng W, Halverson SL, Gonzalez-Neira A, Pita G, Rosario Alonso M, Alvarez N, Herrero D, Tessier DC, Vincent D, Bacot F, Luccarini C, Baynes C, Ahmed S, Maranian M, Healey CS, Simard J, Hall P, Easton DF, Garcia-Closas M. Prediction of Breast Cancer Risk Based on Profiling With Common Genetic Variants. *J Natl Cancer Inst* 2015;107:djv036–djv036.
- Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, Lesueur F, Byrnes GB, Chuang S-C, Forey N, Feuchtinger C, Gioia L, Hall J, Hashibe M, Herte B, McKay-Chopin S, Thomas A, Vallée MP, Voegelé C, Webb PM, Whiteman DC, Sangrajrang S, Hopper JL, Southey MC, Andrulis IL, John EM, Chenevix-Trench G. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet* 2009;85:427–46.
- Le Calvez-Kelm F, Lesueur F, Damiola F, Vallée M, Voegelé C, Babikyan D, Durand G, Forey N, McKay-Chopin S, Robinot N, Nguyen-Dumont T, Thomas A, Byrnes GB, Hopper JL, Southey MC, Andrulis IL, John EM, Tavtigian SV. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res* 2011;13:R6.
- Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegelé C, Young EL, Robinot N, Forey N, Durand G, Vallée MP, Tao K, Roane TC, Williams GJ, Hopper JL, Southey MC, Andrulis IL, John EM, Goldgar DE, Lesueur F, Tavtigian SV. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res* 2014;16:R58.
- Le Calvez-Kelm F, Oliver J, Damiola F, Forey N, Robinot N, Durand G, Voegelé C, Vallée MP, Byrnes GB, Registry BCF, Hopper JL, Southey MC, Andrulis IL, John EM, Tavtigian SV, Lesueur F. RAD51 and Breast Cancer Susceptibility: No Evidence for Rare Variant Association in the Breast Cancer Family Registry Study. *PLoS One* 2012;7:e52374.
- Park DJ, Tao K, Le Calvez-Kelm F, Nguyen-Dumont T, Robinot N, Hammet F, Odefrey F, Tsimiklis H, Teo ZL, Thingholm LB, Young EL, Voegelé C, Lonie A, Pope BJ, Roane TC, Bell R, Hu H, Shankaracharya, Huff CD, Ellis J, Li J, Makunin IV, John EM, Andrulis IL, Terry MB, Daly M, Buys SS, Snyder C, Lynch HT,

- Devilee P, Giles GG, Hopper JL, Feng B-J, Lesueur F, Tavtigian S V, Southey MC, Goldgar DE. Rare Mutations in RINT1 Predispose Carriers to Breast and Lynch Syndrome-Spectrum Cancers. *Cancer Discov* 2014;4:804–15.
- 15 Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, Neuhausen SL, John EM, Andrusis IL, Terry MB, Daly M, Buys S, Le Calvez-Kelm F, Lonie A, Pope BJ, Tsimiklis H, Voegelé C, Hillers FM, Hoogerbrugge N, Barroso A, Osorio A, Giles GG, Devilee P, Benitez J, Hopper JL, Tavtigian SV, Goldgar DE, Southey MC. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet* 2012;90:734–9.
- 16 Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de Silva D, Zharkikh A, Thomas A. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 2006;43:295–305.
- 17 Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
- 18 Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res* 2005;15:978–86.
- 19 Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet* 2013;Chapter 7: Unit7.20.
- 20 John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, Ziogas A, Andrusis IL, Anton-Culver H, Boyd N, Buys SS, Daly MB, O’Malley FP, Santella RM, Southey MC, Venne VL, Venter DJ, West DW, Whittemore AS, Seminara D. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res* 2004;6:R375–89.
- 21 Greenblatt MS, Beaudet JG, Gump JR, Godin KS, Trombley L, Koh J, Bond JP. Detailed computational study of p53 and p16: using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene* 2003;22:1150–63.
- 22 Armougoum F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res* 2006;34:W604–8.
- 23 Taly JF, Magis C, Bussotti G, Chang J-M, Di Tommaso P, Erb I, Espinosa-Carrasco J, Kemena C, Notredame C. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat Protoc* 2011;6:1669–82.
- 24 Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. *Distrib by author Dep Genome Sci Univ Washington, Seattle* 2005.
- 25 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
- 26 Meyer KB, O’Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, Prathalingham R, Dennis J, Bolla MK, Wang Q, de Santiago I, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Schmidt MK, Broeks A, Van ’t Veer LJ, Hogervorst FB, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsang P, Fasching PA, Lux MP, Ekici AB, Beckmann MW, Peto J, Dos Santos Silva I, Fletcher O, Johnson N, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Milne RL, Zamora MP, Arias JI, Benitez J, Neuhausen S, Anton-Culver H, Ziogas A, Dur CC, Brenner H, Müller H, Arndt V, Stegmaier C, Meindl A, Schmutzler RK, Engel C, Ditsch N, Brauch H, Brüning T, Ko Y-D, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Matsuo K, Ito H, Iwata H, Yatabe Y, Dörk T, Helbig S, Bogdanova NV, Lindblom A, Margolin S, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Chenevix-Trench G, Wu AH, Tseng C-C, Van Den Berg D, Stram DO, Lambrechts D, Thienpont B, Christiaens M-R, Smeets A, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Radice P, Peterlongo P, Bonanni B, Bernard L, Couch FJ, Olson JE, Wang X, Purrington K, Giles GG, Severi G, Baglietto L, McLean C, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Simard J, Goldberg MS, Labrèche F, Dumont M, Teo S-H, Yip C-H, Phuah S-Y, Kristensen V, Grenaker Alnæs G, Børresen-Dale A-L, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Kauppila S, Andrusis IL, Knight JA, Glendon G, Tchatchou S, Devilee P, Tollenaar RAEM, Seynaeve CM, García-Closas M, Figueroa J, Chanock SJ, Lissowska J, Czene K, Darabi H, Eriksson K, Hoening MJ, Martens JWM, van den Ouweland AMW, van Deurzen CHM, Hall P, Li J, Liu J, Humphreys K, Shu X-O, Lu W, Gao Y-T, Cai H, Cox A, Reed MWR, Blot W, Signorello LB, Cai Q, Pharoah PDP, Ghousaini M, Harrington P, Tyrer J, Kang D, Choi J-Y, Park SK, Noh D-Y, Hartman M, Hui M, Lim W-Y, Buhari SA, Hamann U, Försti A, Rüdiger T, Ulmer H-U, Jakubowska A, Lubinski J, Jaworska K, Durda K, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Vachon C, Slager S, Fostira F, Pilarski R, Shen C-Y, Hsiung C-N, Wu P-E, Hou M-F, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Ponder BAJ, Dunning AM, Easton DF. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet* 2013;93:1046–60.
- 27 Song M, Kraft P, Joshi AD, Barndahl M, Chatterjee N. Testing calibration of risk models at extremes of disease risk. *Biostatistics* 2015;16:143–54.
- 28 Huo D, Zheng Y, Ogundiran TO, Adebamowo C, Nathanson KL, Domchek SM, Rebbeck TR, Simon MS, John EM, Hennis A, Nemesure B, Wu S-Y, Leske MC, Ambis S, Niu Q, Zhang J, Cox NJ, Olopade OI. Evaluation of 19 susceptibility loci for breast cancer in women of African ancestry. *Carcinogenesis* 2012;33:835–40.
- 29 Fejerman L, Stern MC, Ziv E, John EM, Torres-Lmejia G, Hines LM, Wolff R, Wang W, Baumgartner KB, Giuliano AR, Slattery ML. Genetic ancestry modifies the association between genetic risk variants and breast cancer risk among Hispanic and non-Hispanic white women. *Carcinogenesis* 2013;34:1787–93.
- 30 Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, Choi J-Y, Long J, Dennis J, Humphreys MK, Wang Q, Lu W, Gao Y-T, Li C, Cai H, Park SK, Yoo K-Y, Noh D-Y, Han W, Dunning AM, Benitez J, Vincent D, Bacot F, Tessier D, Kim S-W, Lee MH, Lee JW, Lee J-Y, Xiang Y-B, Zheng Y, Wang W, Ji B-T, Matsuo K, Ito H, Iwata H, Tanaka H, Wu AH, Tseng C, Van Den Berg D, Stram DO, Teo SH, Yip CH, Kang IN, Wong TY, Shen C-Y, Yu J-C, Huang C-S, Hou M-F, Hartman M, Miao H, Lee SC, Putti TC, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsang P, Sangrajrang S, Shen H, Chen K, Wu P-E, Ren Z, Haiman CA, Sueta A, Kim MK, Khoo US, Iwasaki M, Pharoah PDP, Wen W, Hall P, Shu X-O, Easton DF, Kang D. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum Mol Genet* 2013;22:2539–50.
- 31 Lee CH, Dershaw DD, Kopsan D, Evans P, Monsees B, Monticciolo D, Brenner RJ, Bassett L, Berg W, Feig S, Hendrick E, Mendelson E, D’Orsi C, Sicles E, Burhenne LW. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol* 2010;7:18–27.
- 32 Daly M, Pilarski R, Axilbund J, Buys S, Crawford B, Friedman S, Garber J, Horton C, Kaklamani V, Klein C, Kohlmann W, Kurian A, Litton J, Madlensky L, Marcom P, Merajver S, Offit K, Pal T, Pasche B, Reiser G, Shannon K, Swisher E, Voian N, Weitzel J, Whelan A, Wiesner G, Dwyer M, Kumar R. National comprehensive cancer network. Genetic/familial high-risk assessment: breast and ovarian, version 1.2014. *J Natl Compr Canc Netw* 2014;12:1326–38.
- 33 Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, Morris E, Pisano E, Schnall M, Sener S, Smith RA, Warner E, Yaffe M, Andrews KS, Russell CA. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57:75–89.
- 34 SEER Cancer Statistics Factsheets: Breast Cancer. *Natl Cancer Institute Bethesda, MD*. 2014. <http://seer.cancer.gov/statfacts/html/breast.html>
- 35 Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.
- 36 Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004;23:1111–30.
- 37 Vallée MP, Francy TC, Judkins MK, Babikian D, Lesueur F, Gammon A, Goldgar DE, Couch FJ, Tavtigian SV. Classification of missense substitutions in the BRCA genes: a database dedicated to Ex-UVs. *Hum Mutat* 2012;33:22–8.
- 38 Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;266:66–71.
- 39 Castilla LH, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE, Boyd J, Lubin MB, Deshano ML, Brody LC. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat Genet* 1994;8:387–91.
- 40 Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, Bennett LM, Haugen-Strano A, Swensen J, Miki Y. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* 1994;266:120–2.
- 41 Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DGR, Trench GC, Rahman N, Robson M, Domchek SM, Foulkes WD. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 2015;372:2243–57.
- 42 Thompson BA, Greenblatt MS, Vallee MP, Herkert JC, Tessereau C, Young EL, Adzhubei IA, Li B, Bell R, Feng B, Mooney SD, Radivojac P, Sunyaev SR, Frebourg T, Hofstra RMW, Sijmons RH, Boucher K, Thomas A, Goldgar DE, Spurdle AB, Tavtigian SV. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat* 2013;34:255–65.
- 43 Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, Horick N, Yang S, Shannon KM, Tung N, Ford JM, Lincoln SE, Ellisen LW. Clinical Actionability of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Risk Assessment. *JAMA Oncol* 2015;02114:1–9.
- 44 Maxwell KN, Wubbenhorst B, D’Andrea K, Garman B, Long JM, Powers J, Rathbun K, Stopfer JE, Zhu J, Bradbury AR, Simon MS, DeMichele A, Domchek SM, Nathanson KL. Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. *Genet Med* 2015;17:630–8.
- 45 Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, Timms K, Garber JE, Herold C, Ellisen L, Krejdovsky J, DeLeonardis K, Sedgwick K, Soltis K, Roa B,

- Wenstrup RJ, Hartman A-R. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer* 2015;121:25–33.
- 46 Roeb W, Higgins J, King MC. Response to DNA Damage of CHEK2 Missense Mutations in Familial Breast Cancer. *Hum Mol Genet* 2012;21:2738–44.
- 47 Thompson BA, Goldgar DE, Paterson C, Clendenning M, Walters R, Arnold S, Parsons MT, Walsh MD, Gallinger S, Haile RW, Hopper JL, Jenkins MA, Lemarchand L, Lindor NM, Newcomb PA, Thibodeau SN, Young JP, Buchanan DD, Tavtigian SV, Spurdle AB, Michael DW, Michael DW, Gallinger S, Haile RW, Hopper JL, Jenkins MA, Lemarchand L, Lindor NM, Newcomb PA, Thibodeau SN, Young JP, Buchanan DD, Tavtigian SV, Spurdle AB. A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the Colon Cancer Family Registry. *Hum Mutat* 2013;34:200–9.
- 48 Lindor NM, Guidugli L, Wang X, Vallée MP, Monteiro ANA, Tavtigian S, Goldgar DE, Couch FJ. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum Mutat* 2012;33:8–21.
- 49 Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FBL, Hoogerbrugge N, Spurdle AB, Tavtigian SV. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 2008;29:1282–91.
- 50 Spurdle AB, Healey S, Devereau A, Hogervorst FBL, Monteiro ANA, Nathanson KL, Radice P, Stoppa-Lyonnet D, Tavtigian S, Wappenschmidt B, Couch FJ, Goldgar DE. ENIGMA-evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat* 2012;33:2–7.