OPEN ACCESS

ORIGINAL ARTICLE

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

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

▸ 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

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

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≥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≥2.5. 2.1% of cases and 1.2% of controls had a polygenotype with an average OR≥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≥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,[1 2] 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.[3–6] Beyond the moderate-risk genes, many common SNPs have been identified as markers for slightly increased breast cancer risk.[7–9] 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*,[10–15] 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≥2.5 as a threshold for clinical significance, we estimated the scores from the missense substitution analysis programs Align-GVGD,[16] CADD,[17] MAPP[18] and PolyPhen-2[19] required to identify a group of missense substitutions that reach an average OR≥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

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

	Case–control mutation screening for rare variants in nine moderate-risk genes				Case–control SNP genotyping for 18 BCAC SNPs			
Distributions	**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
36–10	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≥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≥2.5 threshold across T+SJV and OR≥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.[20] The controls (N=1121) were frequency matched to cases within each centre on racial or ethnic group, with age at selection not more than ±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[10–15] 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*,[12] we applied a minor allele frequency (q) threshold of ≤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.[12]

In silico missense substitution scoring

Align-GVGD (agvgd.iarc.fr/agvgd_input.php) and MAPP (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).[21] 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.[22 23] The initial alignment was checked by hand in Geneious V.7.1.4 (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 (V.3.69),[24] 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/)[25] and CADD (cadd.gs.washington.edu/)[17] operate without user-supplied alignments. PolyPhen-2 uses a combination of internally generated pMSAs, functional annotations and structural information to evaluate missense substitutions;[19] 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.[17]

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≥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.[7 9 26 27] 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,[26 28–30] 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≥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 V.12.1 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 ≥20% lifetime breast cancer risk.[31 32] The American Cancer Society (ACS) recommends breast MRI for women with a 20–25% or greater lifetime risk.[33] In the USA, the lifetime risk of a woman to develop breast cancer is estimated to be 12.3%;[34] 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 ≥14 years, first childbirth at age ≤20 years and no family history, the Gail model[35] and Tyrer–Cuzick model[36] 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≥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^{-3}
Rare missense substitution analyses								
Carrier of rare missense substitution	205	114	10.2	176	13.6	1.42	1.09 to 1.84	8.70×10^{-3}
Carrier of key domain rare missense substitution	140	65	5.8	136	10.5	1.94	1.41 to 2.67	5.11×10^{-5}
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^{-5}
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^{-4}
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^{-4}
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^{-5}
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^{-5}
Two or more	65	18	1.6	70	5.4	3.18	1.85 to 5.46	2.68×10^{-5}
Three or more	45	14	1.2	52	4.0	3.27	1.77 to 6.04	1.51×10^{-4}
All four	19	2	0.2	20	1.5	8.61	1.96 to 37.81	4.35×10^{-3}
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,[37–40] 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).[10 12] We observed 140 rMS and IFDs in key functional domains (OR 1.94, $p=5.1\times10^{-05}$, table 2), which still did not reach an OR≥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*.[41] 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,[16–18 25] 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*,[2] MAPP and PolyPhen-2 because of their strong performance in our recent analyses of mismatch repair protein missense substitutions,[42] and CADD because of its reported ability to prioritise variants across functional categories and effect sizes.[17] 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≥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≥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≥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≥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≥2.5 (OR 4.90, p=0.0012; online supplementary table S6).

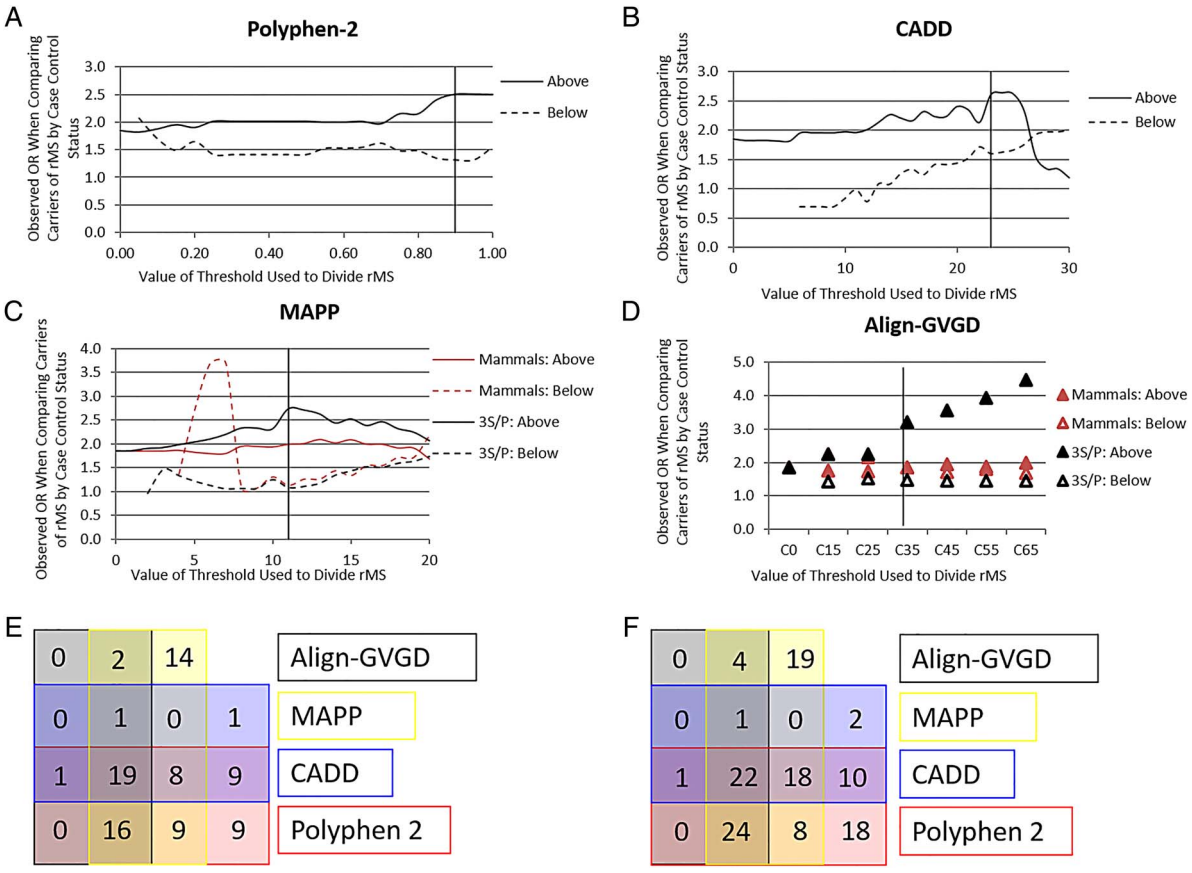


Figure 1 Observed OR and thresholds for each missense substitution analysis program. The observed OR for the carriers of key domain rare missense substitutions (rMS) of both the 'above' and 'below' groups for each severity threshold tested with (A) PolyPhen-2, (B) CADD, (C) MAPP and (D) Align-GVGD, adjusting 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≥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≥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≥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≥2.5 (OR 3.18, $p=2.68\times10^{-5}$), affecting 5.4% of cases and 1.6% of controls (table 2). These results are comparable to other studies, but include data from controls.[43–45] 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≥2.5 at and above the 99th percentile of the combined scores.[9] 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,[7 26 28–30] 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

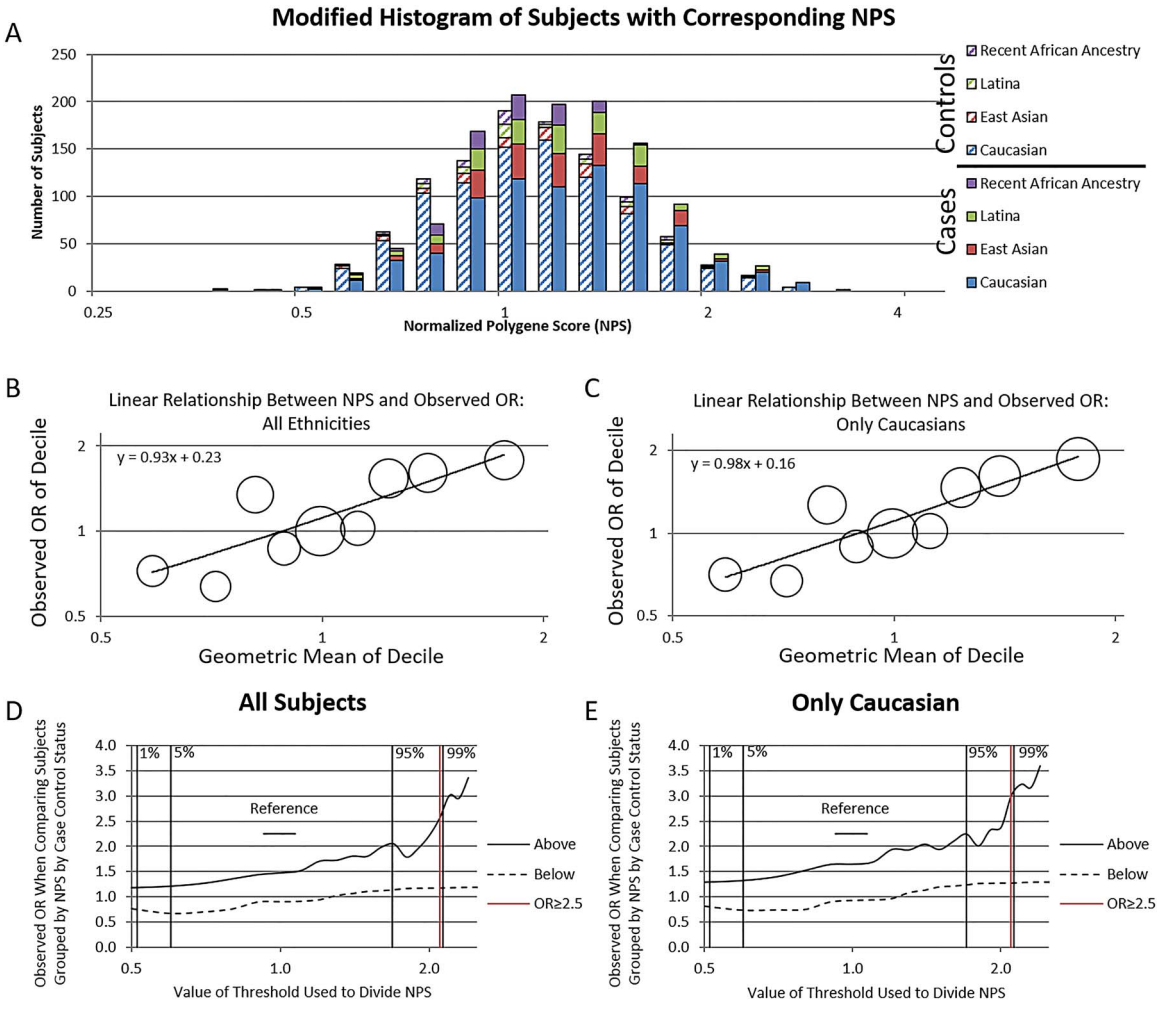


Figure 2 The normalised polygene score (NPS): distribution, NPS-OR correlation and threshold for medically actionable. (A) NPS distribution for all subjects. Comparison of observed OR and NPS for each decile for (B) all subjects and (C) only Caucasians, with the corresponding equations derived from linear 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≥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=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≥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≥2.5.

To explore the possibility of integrating gene mutation screening with SNP genotyping, we tested for interactions between carriage of an OR≥2.5 rare variant (combining T+SJVs 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≥2.5 category and multiplied the OR estimated from their rMS or T+SJV with their NPS. These combined ORs varied from ~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 ≥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≥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 ($p=2.68\times10^{-5}$). 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^{-10}
Only Caucasians	904	85.5	788	63.8			3.47×10^{-10}
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*.[5] 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.[46] 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^{-5} 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.[37 47 48]

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*),[10 11] or susceptibility to breast cancer in general (*MRE11A*, *NBN*, *RAD50*, *RINT1*, *XRCC2*).[12 14 15] 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.[41] 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

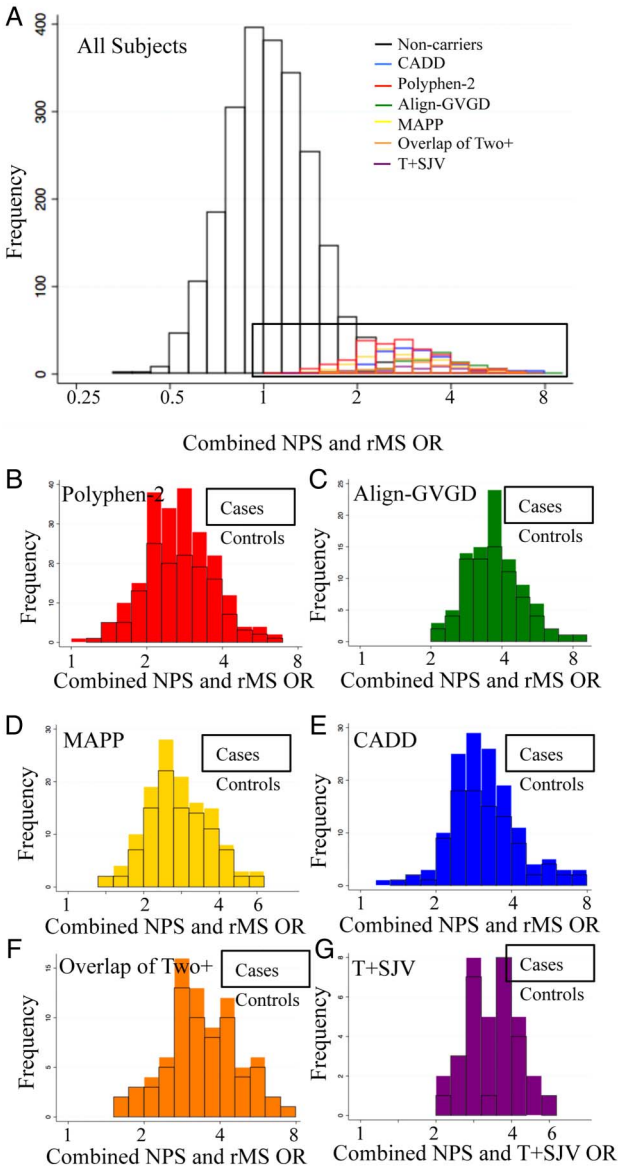


Figure 3 Combined normalised polygene score (NPS) and rare variant OR. (A) Distribution of subjects by NPS when carriers of rare missense substitutions (rMS) in the 'above threshold' group had their NPS increased by the factor of their OR. Stacked histograms of the expansion of the combined contribution of NPS and rMS for carriers with a variant conferring an average OR≥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?"[41] 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.[49] 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.[2 42] 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,[50] 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 ≥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≥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
[1]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[2]Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[3]Breast Cancer Genetics Group, Cancer Research Centre of Lyon, Centre Léon Bérard, Lyon, France
[4]Genetic Cancer Susceptibility group, International Agency for Research on Cancer, Lyon, France
[5]Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[6]Department of Communication and Huntsman Cancer Institute, University of Utah
[7]Genetic Epidemiology Laboratory, The University of Melbourne, Melbourne, Victoria, Australia
[8]Instituto de Ciencias Básicas y Medicina Experimental del Hospital Italiano de Buenos Aires, Buenos Aires, Argentina
[9]Cancer Genomics Laboratory, CHUQ Research Center, Quebec City, Canada
[10]Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
[11]Department of Epidemiology (Genome Epidemiology Lab), Seoul National University School of Public Health, Seoul, Korea
[12]Department of Molecular Genetics, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada
[13]Cancer Prevention Institute of California, Fremont, California, USA
[14]Department of Health Research and Policy, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA
[15]Department of Family and Preventive Medicine, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA
[16]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. JLH, MCS, ILA and EMJ were involved in the acquisition of subjects, critical revision of manuscript and final approval. SVT 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.

REFERENCES

1 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.
2 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.
3 Kean S. The 'other' breast cancer genes. *Science* 2014;343:1457–9.
4 Stadler ZK, Schrader KA, Vijai J, Robson ME, Offit K. Cancer genomics and inherited risk. *J Clin Oncol* 2014;32:687–98.
5 Tavtigian SV, Chenevix-Trench G. Growing recognition of the role for rare missense substitutions in breast cancer susceptibility. *Biomark Med* 2014;8:589–603.
6 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.
7 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, Irwanto 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-Myhsok 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 Z, 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, Hooning 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, Labrèche 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, Siriwanarangsan 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.
8 Maxwell KN, Nathanson KL. Common breast cancer risk variants in the post-COGS era: a comprehensive review. *Breast Cancer Res* 2013;15:212.
9 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, Aittomaki 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 JI, Menendez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Bogdanova NV, Antonenkova NN, Dork 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, Guenel 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, Bruning 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, Hooning 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, Pylkas 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.
10 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, Voegele 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.
11 Le Calvez-Kelm F, Lesueur F, Damiola F, Vallée M, Voegele 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.
12 Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegele 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.
13 Le Calvez-Kelm F, Oliver J, Damiola F, Forey N, Robinot N, Durand G, Voegele C, Vallée MP, Byrnes G, 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.
14 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, Voegele 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, Andrulis IL, Terry MB, Daly M, Buys S, Le Calvez-Kelm F, Lonie A, Pope BJ, Tsimiklis H, Voegele C, Hilbers 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, Andrulis 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 Armougom 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, Siriwanarangsan 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, Andrulis 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, Hooning 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, Ghoussaini 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, Barrdahl 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, Ambs S, Niu Q, Zhang J, Cox NJ, Olopade OI. Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis* 2012;33:835–40.
29 Fejerman L, Stern MC, Ziv E, John EM, Torres-Mejia 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, Siriwanarangsan 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, Kopans D, Evans P, Monsees B, Monticciolo D, Brenner RJ, Bassett L, Berg W, Feig S, Hendrick E, Mendelson E, D'Orsi C, Sickles 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, Babikyan 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, Adzhubey 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.

Supplemental Methods

Ethics Statement

The mutation screening and analyses described here were approved by the Ethics committee of the International Agency for Research on Cancer (IARC), the University of Utah Institutional Review Board (IRB), and the local IRBs of the Breast Cancer Family Registry (BCFR) centers from which we received samples. These local IRBs were the Health Sciences Human Ethics Subcommittee of the University of Melbourne, Australia; the Institutional Review Board of the Northern California Cancer Center (now the Cancer Prevention Institute of California); and the Research Ethics Board of Mount Sinai Hospital, Ontario, Canada. All participants gave written informed consent.

Participants

The design of this study has been described in detail previously[1–6]. Briefly, eligible participants included women ascertained by population-based sampling by the Australian, Northern California, and Ontario sites of the Breast Cancer Family Registry (BCFR)[7]. Participants were recruited between 1995 and 2005. Selection criteria for cases (N = 1,297) were diagnosis of breast cancer at or before age 45 years and self-reported race/ethnicity plus grandparents' country of origin information consistent with Caucasian, East Asian, Hispanic/Latino, or African American racial/ethnic heritage. The controls (N = 1,121) were frequency matched to the cases within each center on racial/ethnic group, with age at selection not more than ± 10 years from the age range at diagnosis of the cases gathered from the same center (Table 1). Because of the shortage

of available controls in some racial/ethnic and age groups, the frequency matching was not one-to-one in all subgroups. Known carriers of pathogenic variants in *BRCA1* and *BRCA2* were excluded. Known *PALB2* carriers were also excluded due to recent results upgrading that gene to high-risk status[8]. *PTEN*, *TP53*, and *CDH1* carrier status was unknown, but would only affect a small number of individuals in the study. In the analysis of the sequence variants, we excluded participants with a PCR failure rate higher than 20% of the coding sequence from the rare variants analyses, and a SNP genotyping failure rate of more than 4 of 18 SNPs from the SNP polygene analyses.

Mutation Screening for Variants in the 9 Moderate-Risk Genes

For mutation screening of the coding exons and proximal splice junction regions of *ATM* (NM_000051.3), *BARD1* (NM_000465.3), *CHEK2* (NM_007194.3), *MRE11A* (NM_005591.3), *NBN* (NM_002485.4), *RAD50* (NM_005732.3), *RAD51* (NM_002875.4), *RINT1* (NM_021930.4), and *XRCC2* (NM_005431.1), we used 30 ng of whole-genome amplified (WGA) DNA obtained by mixing 15 ng of amplified DNA from each of two independent WGA reactions. The laboratory process was as described in detail in the prior studies of the 9 moderate-risk genes used in this study[1–6]. Our semi-automated approach, handled by a Laboratory Information Management System (LIMS)[9], relies on mutation scanning by high-resolution melt curve (HRM) analysis followed by direct Sanger sequencing of the individual samples for which an aberrant melting curve profile is indicative of the presence of a sequence variant.

All exonic sequence variants, plus intronic sequence variants that fell within 20 bp of a splice acceptor or eight bp of a splice donor, and were either unreported or had an

allele frequency of <1% in the large scale reference groups “Caucasian Americans”, “African Americans” and “East Asians” based on exome variant server (EVS) and 1,000 genomes project (1000G) data (http://evs.gs.washington.edu/EVS; http://browser.1000genomes.org/index.html), were confirmed either by independent re-amplification and sequencing from each of the two independent WGA reaction products and concordant variant calls, or, for five variants, by re-amplification and sequencing from genomic DNA.

All samples that failed either at the primary PCR, secondary PCR, or sequencing reaction stage were re-amplified from WGA DNAs or genomic DNAs. Samples that still did not provide satisfactory mutation screening results for at least 80% of the concatenated MRN coding sequence were excluded from further analysis. Primer and probe sequences are available from the authors upon request.

Genotyping for 18 SNPs

Genotyping for 18 of the first breast cancer modest-risk SNPs confirmed by the Breast Cancer Association Consortium(BCAC).[10, 11] Genotyping for rs13387042, rs13281615, rs2981578, rs4973768, rs11249433, rs2046210, rs704010, rs10995190, rs10941679, rs2380205, rs6504950, rs614367, rs1011970, and rs999737) involved a nested and multiplexed polymerase chain reaction (PCR) of whole-genome amplified (WGA) DNA followed by high-resolution melting (HRM) curve analysis to identify major and minor alleles. Our HRM analysis consisted of two assays, either unlabeled probe or small amplicon-based genotyping. Genotyping for rs1045485, rs3803662, rs889312, rs3817198 were genotyped by Taqman. Genotyping began at the World Health

Organization's (WHO) International Agency for Research on Cancer (IARC) before moving to the University of Utah's Huntsman Cancer Institute (HCI). Primer and probe sequences are available from the authors upon request.

Protein Multiple Sequence Alignment Organisms

Orthologs from Human (*Homo sapiens*), either mouse (*Mus musculus*) or rat (*Rattus norvegicus*) from clade Murinae, either pig (*Sus scrofa*), cow (*Bos taurus*), dog (*Canis lupus*) or panda (*Ailuropoda melanoleuca*) from clade Laurasiatheria, elephant (*Loxodonta africana*), armadillo (*Dasypus novemcinctus*), either opossum (*Monodelphis domestica*) or tasmanian devil (*Sarcophilus harrisii*) from clade Metatheria, platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*) and lizard (*Anolis carolinensis*) or painted turtle(*Chrysemys picta bellii*) from clade Sauria, clawed frog (*Xenopus laevis* or *Xenopus tropicalis*), coelacanth (*Latimeria chalumnae*), either zebrafish (*Danio rerio*), Tetraodon (*Tetraodon nigroviridis*), or Fugu (*Takifugu rubripes*) from clade Clupeocephala, lancelet (*Branchiostoma floridae*), and sea urchin (*Strongylocentrotus purpuratus*) were included in our initial alignments. RAD50 and RAD51 did not reach three substitutions per position for their individual pMSA using the preceding organisms, so orthologous protein sequences from the model organisms *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Arabidopsis thaliana* were added until three substitutions per position was achieved.

Ortholog sequences downloaded from Genbank were aligned using the expresso extension of T-Coffee to create the initial pMSA[12, 13]. These initial alignments were 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 re-prediction from, genomic DNA sequence available on the UCSC genome browser (http://genome.ucsc.edu).

Protein Multiple Sequence Alignment Depth Determination

We estimated substitutions in each alignment by using Protpars from PHYLIP v3.69[14] with a constrained phylogeny to make a maximum parsimony estimate of the number of substitutions that occurred for each organism in the underlying phylogeny. In order to create a distance table for MAPP, we used Protdist to establish a distance matrix for protein sequences using maximum likelihood estimates with a constrained phylogeny followed by Fitch to generate a phylogenetic tree under the additive tree model. The complete alignments, phylogenetic trees, and distances are available upon request.

Statistics

To assess evidence of risk from the case-control frequency distribution of protein-truncating variants (T), known or very likely spliceogenic splice-junction variants (SJV), and rare missense substitutions (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, its frequency, as well as study center, case-control status, race/ethnicity, and age for the subject. For the subjects who carried more than one rare variant of interest, only the most deleterious score was considered. Data from key functional domain rare silent substitutions, key functional domain rare missense substitutions (including in-frame deletions), and protein truncating variants were used in a ROC analysis. The silent

substitutions were scored using CADD and MAPP, and the missense substitutions were scored using Align-GVGD, CADD, MAPP, and PolyPhen-2[15–18]. ROC curves were drawn (Supplemental Figure 1) and AUC was determined (Supplemental Table 3) using the ROCR library (http://rocr.bioinf.mpi-sb.mpg.de). Correlations between the missense analysis programs were performed using the continuous variables of each missense analysis program for linear regressions via Stata version 12.1 software (StataCorp, College Station, TX, USA). We also created a synthetic variable "consensus" that counted the number of missense analysis programs that considered a variant "above threshold" and preformed similar linear regressions between the variable output of the missense analysis program with our synthetic consensus variable (Supplemental Table 4).

For MAPP, Align-GVGD, and Polyphen2, logistic regression trend tests were formatted such that participants who did not carry any rare variant were given a score of 0, carriers of rMS outside of a domain were given a 1, carriers of a "below threshold" rMS were given a 2, carriers of an "above threshold" rMS were given a 3, carriers of an in-frame deletion were given a 4 and T+SJVs were given a 5. CADD gives severity scores for more than rMS and was treated differently as follows: carriers of a non-rMS variant that would have been in the "above threshold" group were given a 2, carriers of a "below threshold" rMS were given a 3, carriers of an "above threshold" rMS were given a 4, carriers of an in-frame deletion were given a 5 and T+SJVs were given a 6.

These row labels were then used as a categorical variable in the logistic regressions. The reference non-carrier group (assigned logistic regression row label 0 for each classifier) comprised the participants who were not reported to carry an rMS, an in-frame deletion, or a T+SJV variant in a domain of one of the 9 genes. The same reference

group of non-carriers was used for MAPP, Align-GVGD and Polyphen2 analyses. The participants in CADD that carried a non-rMS variant that would have been in the "above threshold" group would have been classified as non-carriers in the alternative methods, but were not considered as part of the reference group for the analyses involving CADD. Differences in the case–control ratio between racial/ethnic groups and study center were accounted for by including categorical variables for each racial/ethnic group and each study center. Adjustment for racial/ethnic group should also capture confounding of genetic and social factors with interaction terms, allowing that this confounding effect may be different for the broadly labeled racial/ethnic groups in different centers.

To determine the correlation between NPS and the observed OR, we divided our subjects into deciles based on the percentiles observed in the controls. The middle quintile (40-60) was used as reference group and given a value of 0. The remaining deciles were given values 1-8 respectively, which were used as these categorical variables in logistic regressions. For the threshold analysis, we used the middle quintile as the reference group and gave this group a value of 0. The "below threshold, outside of reference" group was given a value of 1 and the "above threshold" group was given a value of 2. To create the histogram for the moderate-risk gene combined with the SNP polygene, the NPS and OR from the moderate-risk gene were multiplied together. The middle quintile was excluded for the linear regression equation, and the NPS was treated as the independent variable and the resultant OR of each group as the dependent variable. P-values were found by testing the regression coefficient equal to 0 or 1.

Supplementary Methods Sources

1 Tavtigian S V, 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, Voegele 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.

2 Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegele 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 S V. 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.

3 Le Calvez-Kelm F, Oliver J, Damiola F, Forey N, Robinot N, Durand G, Voegele C, Vallée MP, Byrnes G, Registry BCF, Hopper JL, Southey MC, Andrulis IL, John EM, Tavtigian S V, 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.

4 Le Calvez-Kelm F, Lesueur F, Damiola F, Vallée M, Voegele 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 S V. 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.

5 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, Voegele C, Lonie A, Pope BJ, Roane TC, Bell R, Hu H, Shankaracharya, Huff CD, Ellis J, Li J, Makunin I V., 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.

6 Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, Neuhausen SL, John EM, Andrulis IL, Terry MB, Daly M, Buys S, Le Calvez-Kelm F, Lonie A, Pope BJ, Tsimiklis H, Voegele C, Hilbers FM, Hoogerbrugge N, Barroso A, Osorio A, Giles GG, Devilee P, Benitez J, Hopper JL, Tavtigian S V, Goldgar DE, Southey MC. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet* 2012;90:734–9.

7 John EM, Hopper JL, Beck JC, Knight J a, Neuhausen SL, Senie RT, Ziogas A, Andrulis 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.

8 Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkäs K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomäki K, Moilanen JS, Turnbull C, Seal S, Mannermaa A, Kallioniemi A, Lindeman GJ, Buys SS, Andrulis IL, Radice P, Tondini C, Manoukian S, Toland AE, Miron P, Weitzel JN, Domchek SM, Poppe B, Claes KBM, Yannoukakos D, Concannon P, Bernstein JL, James P a., Easton DF, Goldgar DE, Hopper JL, Rahman N, Peterlongo P, Nevanlinna H, King M-C, Couch FJ, Southey MC, Winqvist R, Foulkes WD, Tischkowitz M. Breast-cancer risk in families with mutations in PALB2. *N Engl J Med* 2014;371:497–506.

9 Voegele C, Tavtigian S V, de Silva D, Cuber S, Thomas a, Le Calvez-Kelm F. A Laboratory Information Management System (LIMS) for a high throughput genetic platform aimed at candidate gene mutation screening. *Bioinformatics* 2007;23:2504–6.

10 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 T a, Aittomäki K, Blomqvist C, Czene K, Irwanto 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-Myhsok 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 LF a, Haiman C a, 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 Z, 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 J a, Glendon G, Mulligan AM, Lindblom A, Margolin S, Hooning 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 P a, 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, Labrèche 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 R a EM, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Bogdanova N V, 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, Siriwanarangsan 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.

11 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, Siriwanarangsan P, Fasching P a, 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 T a, Aittomäki K, Blomqvist C, Matsuo K, Ito H, Iwata H, Yatabe Y, Dörk T, Helbig S, Bogdanova N V, 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 C a, 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, Andrulis IL, Knight J a, Glendon G, Tchatchou S, Devilee P, Tollenaar R a EM, Seynaeve CM, García-Closas M, Figueroa J, Chanock SJ, Lissowska J, Czene K, Darabi H, Eriksson K, Hooning 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, Ghoussaini M, Harrington P, Tyrer J, Kang D, Choi J-Y, Park SK, Noh D-Y, Hartman M, Hui M, Lim W-Y, Buhari S a, 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 B a J, 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.

12 Armougom 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.

13 Taly J-F, 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.

14 Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. *Distrib by author Dep Genome Sci Univ Washington, Seattle* 2005.

15 Tavtigian S V, Deffenbaugh a M, 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.

16 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.

17 Stone E a, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res* 2005;15:978–86.

18 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.

Supplemental Table 1. The list of variants with their respective scores for each variant classifier

Gene	DNA	Protein	Align-GVGD Mammals[a]	Align-GVGD 3 S/P[b]	Polyphen-2	CADD	MAPP Mammals[a]	MAPP 3 S/P[b]	Frequency (ESP)	Domain
ATM	c.136-139 del CATT	frameshift	FS	FS	1	21.8	45.4	45.4	0	1
ATM	c.170 G>A	W57X	X	X	1	36	45.4	45.4	0	1
ATM	c.295 A>G	S99G	Class C55	Class C0	0.997	28.8	9.08	2.41	0.000154	0
ATM	c.363 delA	frameshift	FS	FS	1	36	45.4	45.4	0	1
ATM	c.544 G>C	V182L	Class C0	Class C0	0.002	12.03	10.22	4.77	0.008004	0
ATM	c.936 A>C	L312F	Class C15	Class C15	1	19.5	12.59	12.59	0	0
ATM	c.1229 T>C	V410A	Class C65	Class C25	0.434	15.88	27.75	19.81	0.00177	0
ATM	c.G1464T	W488C	Class C65	Class C65	0.653	14.15	29.05	29.05	0.000077	0
ATM	c.1810 C>T	P604S	Class C65	Class C0	0.635	15.46	17.5	4.5	0.004003	0
ATM	c.2021 A>G	H674R	Class C0	Class C0	0	4.044	6.96	3.51	0.000077	0
ATM	c.2158 C>T	R720C	Class C0	Class C0	0	10.15	3.6	2.92	0	0
ATM	c.2494 C>T	R832C	Class C0	Class C0	0	9.764	1.91	1.77	0.000154	0
ATM	c.3004 G>T	G1002C	Class C0	Class C0	0.002	8.811	3.42	6.66	0	0
ATM	c.3295 G>A	D1099N	Class C0	Class C0	0.002	9.962	2.37	1.39	0	0
ATM	c.3925 G>A	A1309T	Class C0	Class C0	0.009	15.47	8.56	4.95	0.001077	0
ATM	c.3963 G>A	M1321I	Class C0	Class C0	0.003	9.956	6.1	2.1	0	0
ATM	c.4388 T>G	F1463C	Class C65	Class C65	1	24.2	17.91	17.91	0.001154	0
ATM	c.4424 G>T	Y1475C	Class C25	Class C25	0.874	11.19	8.85	8.45	0	0
ATM	c.4724 G>A	R1575H	Class C25	Class C25	1	25.4	14.41	14.41	0	0
ATM	c.4949 A>G	N1650S	Class C0	Class C0	0	2.775	1.05	1.32	0.000077	0
ATM	c.5071 A>C	S1691R	Class C0	Class C0	0.001	9.876	3.72	4.19	0.002154	0
ATM	c.5089 A>G	T1697A	Class C0	Class C0	0.001	2.424	17.04	3.2	0.000154	0
ATM	c.5267 C>G	T1756R	Class C15	Class C0	0.054	8.393	33.46	2.63	0	0
ATM	c.5713 insT	frameshift	FS	FS	1	40	45.4	45.4	0	1
ATM	c.5882 A>G	Y1961C	Class C65	Class C65	1	23	23.02	23.02	0.000154	1
ATM	c.5890 A>G	K1964E	Class C15	Class C0	0.053	12.9	20.1	13.14	0	1
ATM	c.5906 A>T	D1969V	Class C35	Class C0	0.242	13.97	25.01	5.76	0	1
ATM	c.5932 G>T	E1978X	X	X	1	49	45.4	45.4	0	1
ATM	c.5975 A>C	K1992T	Class C65	Class C0	0.73	17.27	27.99	16.51	0.000231	1
ATM	c.6067 G>A	G2023R	Class C65	Class C25	1	30	36.52	45.39	0.002308	1
ATM	c.6088 A>G	I2030V	Class C15	Class C0	0	8.212	12.49	5.81	0.005001	1
ATM	c.6235 G>A	V2079I	Class C0	Class C0	0.002	9.681	4.05	2.45	0.006001	1
ATM	c.6437 G>C	S2146T	Class C0	Class C0	0	0.87	2.16	1.38	0.003925	1
ATM	c.6482 G>A	R2161H	Class C0	Class C0	0.503	19.29	7.16	6.29	0	1
ATM	c.6551 G>C	S2184T	Class C0	Class C0	0.109	14.82	5.08	2.65	0.000154	1
ATM	c.6574 T>G	S2192P	Class C65	Class C0	0.015	4.401	39.26	4.22	0	1
ATM	c.6820 G>A	A2274T	Class C55	Class C55	1	35	33.24	33.24	0	1

ATM	c.6860 G>C	G2287A	Class C0	Class C0	0.216	12.31	3.75	4.4	0.000077	1
ATM	c.6919 C>T	L2307F	Class C15	Class C0	0.999	22.4	12.59	7.04	0.001846	1
ATM	c.6995 T>C	L2332P	Class C25	Class C0	0.005	5.623	21.58	2.37	0.008463	1
ATM	c.6998 C>A	T2333K	Class C0	Class C0	0.005	11.5	30.09	31.23	0.000077	1
ATM	c.7004 C>T	T2335I	Class C0	Class C0	0.004	16.79	4.02	5.75	0.000231	1
ATM	c.7174 C>T	R2392W	Class C65	Class C65	1	25.1	22.31	22.31	0.000231	1
ATM	c.7187 C>G	T2396S	Class C55	Class C0	0.095	13.57	11.33	1.85	0.000231	1
ATM	c.7187C>G	T2396S	Class C55	Class C0	0.095	13.57	11.33	1.85	0.000231	1
ATM	c.7271 T>G	V2424G	Class C65	Class C65	1	26.1	23.22	23.22	0.000077	1
ATM	c.7271T>G	V2424G	Class C65	Class C65	1	26.1	23.22	23.22	0.000077	1
ATM	c.7390 T>C	C2464R	Class C65	Class C0	0.838	5.902	34.86	10.39	0.000462	1
ATM	c.7475 T>G	L2492R	Class C65	Class C45	1	24	33.02	31.81	0.000077	1
ATM	c.7636_7644del9	SRI2546_2548del3	IFD	IFD	1	19.45	40.7	34.66	0	1
ATM	c.7638-7646 del9	RIS2547-2549del	IFD	IFD	1	19.45	40.7	41.72	0	1
ATM	c.7740 A>C	R2580S	Class C0	Class C0	0.001	7.937	4.86	5.16	0	1
ATM	c.7775 C>G	S2592C	Class C65	Class C65	1	17.69	20.41	20.41	0	1
ATM	c.7867 C>A	L2623I	Class C0	Class C0	1	29.2	15.46	15.46	0	1
ATM	c.7870 T>G	C2624G	Class C65	Class C65	0.999	23.8	15.41	15.41	0	1
ATM	c.7912 T>G	W2638G	Class C65	Class C15	0.964	10.94	23.29	11.35	0	1
ATM	c.7999 A>G	M2667V	Class C15	Class C0	0.179	13.49	30.01	7.51	0	1
ATM	c.8006 T>C	I2669T	Class C65	Class C65	0.999	20.8	19.6	19.6	0	1
ATM	c.8011 -6 T>G	NA	intronic	intronic	0	2.559	0	0	0	0
ATM	c.8268_8268+4 del5	splice donor deletion	splice	splice	1	11.96	45.4	45.4	0	1
ATM	c.8269 -5 T>G	splice	splice	splice	1	6.733	45.4	45.4	0	1
ATM	c.8285 A>G	Q2762R	Class C35	Class C0	0.961	29	33.7	7.16	0	1
ATM	c.8481 T>G	F2827L	Class C15	Class C15	1	25.3	12.82	12.82	0	1
ATM	c.8494 C>T	R2832C	Class C65	Class C45	1	23.2	29.01	20.28	0	1
ATM	c.8732 C>G	T2911S	Class C0	Class C0	0.212	14.61	3.53	3.04	0	1
ATM	c.8734 A>G	R2912G	Class C65	Class C65	1	20.7	25.9	25.9	0.000231	1
ATM	c.8741 T>C	I2914T	Class C65	Class C65	1	25.1	19.6	19.6	0	1
ATM	c.8773 G>A	G2925S	Class C55	Class C55	1	35	7.5	7.5	0	1
ATM	c.8938 C>A	L2980I	Class C0	Class C0	0.148	9.526	15.46	2.88	0	1
ATM	c.9078_9079 insA	frameshift	IFD	IFD	1	43	44.95	43.98	0	1
ATM	c.9086 G>A	G3029D	Class C65	Class C15	0.011	12.26	35.86	37.66	0.000308	1
ATM	c.9139 C>T	R3047X	IFD	IFD	1	50	43.98	43.98	0	1
BARD1	c.33 G>T	Q11H	Class C0	Class C0	0.41	13.17	2.38	2.04	0.001608	0
BARD1	c.57 G>C	E19D	Class C0	Class C0	0.083	12.36	8.64	7.82	0	0
BARD1	c.73 G>C	A25P	Class C0	Class C0	0.876	15.92	10.89	10.94	0	0
BARD1	c.90 T>A	G30G	SYN	SYN	0	3.557	0.86	1.97	0.000805	1
BARD1	c.216 -21T>G	NA	intronic	intronic	0	4.3	0	0	0	0

BARD1	c.216 T>C	C53C	SYN	SYN	0	10.75	0.15	0.15	0	1
BARD1	c.221 G>T	C74F	Class C65	Class C65	1	19.7	21.32	21.32	0	1
BARD1	c.279 A>G	Q93Q	SYN	SYN	0	8.479	0.62	0.94	0.000154	1
BARD1	c.281 A>C	D94A	Class C65	Class C65	1	20	29.51	29.51	0	1
BARD1	c.326 G>A	S109N	Class C45	Class C0	0.988	17.82	14.09	3.67	0	1
BARD1	c.346C>T	H116Y	Class C35	Class C0	0.006	12.26	10.85	3.25	0.000231	1
BARD1	c.348 T>C	H116H	SYN	SYN	0	8.504	0.93	1.99	0.000308	1
BARD1	c.600 A>G	A200A	SYN	SYN	0	0.22	3.97	5.19	0	0
BARD1	c.609 A>C	G203G	SYN	SYN	0	0.36	3.1	2.25	0.008996	0
BARD1	c.632T>C	L211S	Class C35	Class C35	0.999	17.75	16.77	18.85	0	0
BARD1	c.690 C>G	D230E	Class C0	Class C0	0.028	9.815	5.24	2.86	0	0
BARD1	c.716 T>A	L239Q	Class C0	Class C0	0.014	2.114	2.22	2.87	0.000154	0
BARD1	c.722 C>G	S241C	Class C15	Class C15	0.999	13.17	1.3	2.3	0	0
BARD1	c.882 G>A	R294R	SYN	SYN	0	8.827	1.72	2.4	0	0
BARD1	c.1339 C>G	L447V	Class C25	Class C25	0.998	16.89	25.8	25.8	0.000154	0
BARD1	c.1347 A>G	Q449Q	SYN	SYN	0	7.22	0.11	2.4	0.000231	0
BARD1	c.1364 A>G	N455S	Class C45	Class C45	1	18.38	13.4	13.4	0	0
BARD1	c.1395 +50T>C	NA	intronic	intronic	0	5.779	0	0	0.001231	0
BARD1	c.1409 A>G	N470S	Class C0	Class C0	0.339	8.571	3.78	5.52	0	0
BARD1	c.1429 G>A	V477M	Class C15	Class C15	1	24.5	24.43	24.43	0	0
BARD1	c.1491 A>G	P497P	SYN	SYN	0	8.669	0.14	0.14	0	0
BARD1	c.1586 G>A	R529Q	Class C35	Class C35	0.998	10.28	14.43	14.43	0	0
BARD1	c.1613 G>A	S538N	Class C0	Class C0	0	0.007	1.25	1.43	0.000077	0
BARD1	c.1678A>G	M560V	Class C0	Class C0	0	1.319	6.31	4.65	0	0
BARD1	c.1738 G>A	E580K	Class C0	Class C0	0.041	9.977	5.4	2.55	0.004767	1
BARD1	c.1835 A>T	D612V	Class C35	Class C0	0.015	6.447	24.1	2.69	0.000077	1
BARD1	c.1904 -10T>C	NA	intronic	intronic	0	13.96	0	0	0	1
BARD1	c.1933 T>C	C645R	Class C0	Class C0	0.001	2.457	5.85	2.53	0.022913	1
BARD1	c.1955 A>G	E652G	Class C65	Class C65	1	24	28.23	28.23	0	1
BARD1	c.1957 A>G	I653V	Class C0	Class C0	0.001	9.207	5.48	6.67	0	1
BARD1	c.1972 C>T	R658C	Class C35	Class C35	0.995	24	23.61	20.58	0.00592	1
BARD1	c.1977 A>G	R659R	SYN	SYN	0	8.834	0.59	0.99	0.002384	1
BARD1	c.1989 C>T	N663N	SYN	SYN	0	7.497	0.09	0.09	0	1
BARD1	c.2002 -11C>T	NA	intronic	intronic	0	11.61	0	0	0	1
BARD1	c.2161 G>A	A721T	Class C0	Class C0	0.999	17.3	13.1	15.95	0	1
BARD1	c.2191 C>G	R731G	Class C65	Class C0	0.992	19.7	25.9	3.78	0.002076	1
BARD1	c.2212 A>G	I738V	Class C0	Class C0	0.027	9.469	5.48	6.67	0.006689	1
BARD1	c.2235 T>C	Y745Y	SYN	SYN	0	4.321	1.16	1.36	0	1
BARD1	c.2279 C>T	S760L	Class C65	Class C65	1	24.8	28.64	28.64	0	0
BARD1	c.2282 G>A	S761N	Class C0	Class C0	0.007	14.35	2.86	2.25	0.001461	0

CHEK2	c.14C>T	S5L	Class C0	Class C0	0.011	12.92	9.98	14.03	0.000077	1
CHEK2	c.74T>C	V25A	Class C0	Class C0	0	0.272	4.86	2.91	0	1
CHEK2	c.254C>T	P85L	Class C0	Class C0	0.728	12.03	26.92	2.37	0.002384	1
CHEK2	c.283C>T	R95X	X	X	1	37	45.4	45.4	0	1
CHEK2	c.320 -5T>A	NA	intronic	intronic	0	11.93	0	0	0.000384	1
CHEK2	c.349A>G	R117G	Class C65	Class C65	1	19.93	25.9	25.9	0.000154	1
CHEK2	c.381A>G	E127E	SYN	SYN	0	4.791	1.49	2.42	0	1
CHEK2	c.405delA	K135delA	FS	FS	1	24.3	45.4	45.4	0	1
CHEK2	c.410G>A	R137Q	Class C0	Class C0	0.027	12.54	11.22	2.45	0.000077	1
CHEK2	c.444+24c>T	NA	intronic	intronic	0	5.15	0	0	0.001538	1
CHEK2	c.470T>C	I157T	Class C65	Class C25	0.514	21.1	19.6	14.55	0.001615	1
CHEK2	c.474A>C	A158A	SYN	SYN	0	11.2	1.69	1.5	0	1
CHEK2	c.474A>G	A158A	SYN	SYN	0	11.18	1.69	1.5	0	1
CHEK2	c.538C>T	R180C	Class C65	Class C25	0.64	17.36	29.01	9.34	0.001077	1
CHEK2	c.539G>A	R180H	Class C25	Class C0	0.125	12.53	14.41	2.99	0	1
CHEK2	c.575C>T	S192L	Class C65	Class C15	0.899	16.82	28.64	13.99	0	1
CHEK2	c.592+50a>T	NA	intronic	intronic	0	2.498	0	0	0.002846	1
CHEK2	c.593 -14C>T	NA	intronic	intronic	0	8.971	0	0	0.000233	1
CHEK2	c.593 -20delTCT	NA	intronic	intronic	0	0	0	0	0	1
CHEK2	c.593 -45T>A	NA	intronic	intronic	0	5.903	0	0	0	1
CHEK2	c.593-6 InsCCTT	NA	intronic	intronic	0	0	0	0	0	1
CHEK2	c.651A>G	R217R	SYN	SYN	0	9.542	0.14	1.36	0	1
CHEK2	c.663C>G	I221M	Class C0	Class C0	0.953	17.5	15.7	5.14	0.000155	1
CHEK2	c.688G>T	A230S	Class C65	Class C15	0.076	19.88	23.39	6.45	0	1
CHEK2	c.715G>A	E239K	Class C55	Class C15	0.985	24	19.98	9.44	0.000154	1
CHEK2	c.727T>C	C243R	Class C65	Class C0	0.887	19.81	34.86	27.77	0.000077	1
CHEK2	c.751A>T	I251F	Class C15	Class C15	1	25.4	10.65	10.65	0	1
CHEK2	c.792+62 delAA	NA	intronic	intronic	0	0	0	0	0	1
CHEK2	c.793 -17T>C	NA	intronic	intronic	0	11.37	0	0	0	1
CHEK2	c.823G>T	E275X	X	X	1	49	45.4	45.4	0	1
CHEK2	c.847 -10C>G	NA	intronic	intronic	0	11.2	0	0	0	1
CHEK2	c.911T>C	M304T	Class C65	Class C55	0.999	19.47	25.97	18.59	0	1
CHEK2	c.917G>C	G306A	Class C55	Class C55	1	27.3	22.05	22.05	0	1
CHEK2	c.931G>A	D311N	Class C0	Class C0	0.005	14.33	13.96	3.26	0	1
CHEK2	c.967A>C	T323P	Class C35	Class C0	0.995	21.3	43.72	13.2	0	1
CHEK2	c.1036C>T	R346C	Class C65	Class C65	1	21.1	29.01	29.01	0	1
CHEK2	c.1037G>A	R346H	Class C25	Class C25	1	34	14.41	14.41	0	1
CHEK2	c.1054A>T	N352Y	Class C65	Class C65	1	25.8	20.2	20.2	0	1
CHEK2	c.1065G>A	L355L	SYN	SYN	0	4.288	0.13	0.13	0	1
CHEK2	c.1100delC	T367delC	FS	FS	1	42	45.4	45.4	0	1

CHEK2	c.1111C>T	H371Y	Class C35	Class C35	0.011	16.74	12.32	14.91	0	1
CHEK2	c.1138delCT	L380delCT	FS	FS	1	42	45.4	45.4	0	1
CHEK2	c.1182A>T	E394D	Class C35	Class C35	1	19.09	13.31	13.31	0	1
CHEK2	c.1216C>T	R406C	Class C15	Class C15	0.745	15.7	19.52	15.39	0	1
CHEK2	c.1253T>G	F418C	Class C65	Class C65	1	23.3	17.91	17.91	0	1
CHEK2	c.1263delT	L421delT	FS	FS	1	41	45.4	45.4	0	1
CHEK2	c.1276C>T	P426S	Class C65	Class C65	1	27.3	17.5	17.5	0	1
CHEK2	c.1312G>T	D438Y	Class C65	Class C25	0.999	21.2	21.06	9.42	0.000308	1
CHEK2	c.1313A>G	D438G	Class C65	Class C15	0.661	20.1	20.85	10.68	0	1
CHEK2	c.1336A>G	N446D	Class C15	Class C0	0.003	9.55	29.27	3.69	0	1
CHEK2	c.1343T>G	I448S	Class C65	Class C65	0.124	11.73	20.83	20.83	0.006766	1
CHEK2	c.1427C>T	T476M	Class C65	Class C15	1	19.54	29.5	23.02	0.000401	1
CHEK2	c.1451C>T	P484L	Class C65	Class C65	0.999	25.8	26.92	26.92	0	1
CHEK2	c.1491T>C	D497D	SYN	SYN	0	5.179	1.88	2.95	0	1
CHEK2	c.1528C>T	Q510X	X	X	1	47	45.4	45.4	0	1
CHEK2	c.1534C>G	L512V	Class C0	Class C0	0	3.416	3.1	3.33	0	1
CHEK2	c.1542 +11T>A	NA	intronic	intronic	0	2.948	0	0	0.022796	1
CHEK2	c.1542 +92A>G	NA	intronic	intronic	0	3.726	0	0	0	1
CHEK2	c.1556G>T	R519L	Class C25	Class C25	0.61	16.9	21.16	20.41	0	1
MRE11A	c.1-42G>A	NA	intronic	intronic	0	14.46	0	0	0.000077	0
MRE11A	c.18A>T	A6A	SYN	SYN	0	0.694	2.55	3.32	0	0
MRE11A	c.19C>G	L7V	Class C25	Class C0	0.004	8.845	25.8	6.38	0	0
MRE11A	c.21-12T>C	NA	intronic	intronic	0	17.76	0	0	0	0
MRE11A	c.21-17C>T	NA	intronic	intronic	0	15.68	0	0	0	0
MRE11A	c.37T>C	F13L	Class C15	Class C0	0.402	12.79	12.82	7.26	0	1
MRE11A	c.120C>T	L40L	SYN	SYN	0	5.058	1.09	3.58	0.000539	1
MRE11A	c.162T>C	F54F	SYN	SYN	0	8.823	0.11	0.11	0	1
MRE11A	c.259C>T	R87W	Class C65	Class C15	1	17.06	22.31	11.19	0	1
MRE11A	c.463C>T	R155C	Class C65	Class C35	0.89	22.4	29.01	18.41	0	1
MRE11A	c.529G>C	A177P	Class C25	Class C25	1	24.7	38.15	38.15	0	1
MRE11A	c.660-7G>T	NA	intronic	intronic	0	13.4	0	0	0	1
MRE11A	c.704A>G	D235G	Class C65	Class C65	1	25.4	20.85	20.85	0	1
MRE11A	c.826C>T	P276S	Class C0	Class C0	0.998	26.8	6.38	6.04	0	1
MRE11A	c.846-12T>C	NA	intronic	intronic	0	6.496	0	0	0	1
MRE11A	c.846-20T>G	NA	intronic	intronic	0	6.152	0	0	0	1
MRE11A	c.940C>T	L314L	SYN	SYN	0	9.282	0.13	0.13	0.000154	1
MRE11A	c.1074A>G	P358P	SYN	SYN	0	10.64	0.14	0.14	0	1
MRE11A	c.1084C>G	L362V	Class C25	Class C25	0.856	18.34	25.8	25.8	0	1
MRE11A	c.1096C>A	R366R	SYN	SYN	0	13.01	0.14	0.69	0	1
MRE11A	c.1238A>G	N413S	Class C45	Class C0	0.947	19.01	13.4	4.35	0	1

MRE11A	c.1281A>T	L427F	Class C0	Class C0	0.87	20.6	9.32	3.53	0	1
MRE11A	c.1442C>T	T481I	Class C65	Class C0	0.999	29.2	33.48	10.24	0	1
MRE11A	c.1443A>G	T481T	SYN	SYN	0	9.867	0.19	1.56	0	1
MRE11A	c.1462C>T	R488C	Class C65	Class C65	1	19.65	29.01	29.01	0.000077	0
MRE11A	c.1504C>T	R502C	Class C65	Class C35	0.028	16.07	29.01	17.61	0.000077	0
MRE11A	c.1508G>A	R503H	Class C0	Class C0	0.007	14.93	5.75	4.35	0	0
MRE11A	c.1564-18delTCT	NA	intronic	intronic	0	0	0	0	0	0
MRE11A	c.1667A>G	N556S	Class C0	Class C0	0.001	7.308	3.28	2.3	0.000462	0
MRE11A	c.1811G>A	R604H	Class C0	Class C0	0.002	5.527	2.66	2.54	0.000231	0
MRE11A	c.1852A>G	M618V	Class C15	Class C0	0.025	12.99	30.01	6.51	0.000154	0
MRE11A	c.1897C>G	R633G	Class C0	Class C0	0.001	6.68	5.3	7.63	0	0
MRE11A	c.1927-9C>A	NA	intronic	intronic	0	4.056	0	0	0	0
MRE11A	c.1995-19delTGTTTTTGTTTT	NA	intronic	intronic	0	0	0	0	0	0
MRE11A	c.2049G>A	G683G	SYN	SYN	0	0.413	0.12	0.12	0	0
MRE11A	c.2054A>G	D685G	Class C35	Class C0	0.191	14.43	17.47	3.63	0	0
MRE11A	c.2070+2T>A	NA	intronic	intronic	0	10.44	0	0	0	0
MRE11A	c.2109delAAGAAGAAA	frameshift	IFD	IFD	0	13.73	37.49	31.21	0	0
MRE11A	c.2113A>G	R705G	Class C15	Class C0	0.1	13	12.68	3.53	0	0
NBN	c.1-2C>T	NA	intronic	intronic	0	3.515	0	0	0	0
NBN	c.37+3A>T	NA	intronic	intronic	0	7.684	0	0	0	0
NBN	c.104T>C	I35T	Class C65	Class C65	0.999	22	21.74	21.62	0	1
NBN	c.127C>T	R43*	X	X	1	25.3	45.4	45.4	0.000077	1
NBN	c.425A>G	N142S	Class C0	Class C0	0.129	11.05	7.02	4.26	0.000077	1
NBN	c.426T>C	N142N	SYN	SYN	0	9.874	0.54	1.4	0.000154	1
NBN	c.456G>A	M152I	Class C0	Class C0	0.999	24.9	21.99	21.99	0.000077	1
NBN	c.464T>C	V155A	Class C65	Class C65	0.982	20.7	27.75	27.75	0	1
NBN	c.591insTTGDel13	frameshift	FS	FS	1	28.9	45.4	45.4	0	1
NBN	c.628G>T	V210F	Class C0	Class C0	0.215	12.43	7.11	5.19	0.001001	0
NBN	c.671G>A	G224E	Class C65	Class C65	1	21.1	35.36	35.36	0	0
NBN	c.804delGT	frameshift	FS	FS	1	17.34	45.4	45.4	0	1
NBN	c.950T>A	M317K	Class C45	Class C0	0.814	14.7	31.58	25.87	0	1
NBN	c.954T>A	T318T	SYN	SYN	0	9.618	1.2	1.78	0	1
NBN	c.1065C>T	S355S	SYN	SYN	0	0.1	0.75	0.62	0	0
NBN	c.1176A>G	K392K	SYN	SYN	0	3.512	0.94	1.26	0	0
NBN	c.1202C>G	P401R	Class C0	Class C0	0.03	0.634	25.23	3.07	0	0
NBN	c.1222A>G	K408E	Class C0	Class C0	0.722	10.69	5.41	5.48	0.002768	0
NBN	c.1301C>A	T434N	Class C0	Class C0	0.698	13.87	7.15	7.28	0	0
NBN	1398-19delC	NA	intronic	intronic	0	0	0	0	0	0
NBN	c.1398-9insT	NA	intronic	intronic	0	0	0	0	0	0
NBN	c.1509T>G	N503K	Class C0	Class C0	0.835	13.76	2.24	1.41	0	0

NBN	c.1720 T>A	L574I	Class C0	Class C0	0.085	5.199	3.58	3.72	0.000615	0
NBN	c.2235-18C>T	NA	intronic	intronic	0	14.87	0	0	0.0014	0
NBN	c.2235-18C>T	NA	intronic	intronic	0	14.87	0	0	0.0014	0
RAD50	c.24C>T	S8S	SYN	SYN	0	12.06	0.12	0.12	0.000154	1
RAD50	c.113A>G	N38S	Class C45	Class C45	1	30	13.4	13.4	0	1
RAD50	c.129G>A	T43T	SYN	SYN	0	16.26	0.19	0.19	0	1
RAD50	c.250C>G	L84V	Class C25	Class C0	1	21.4	25.8	24.81	0	1
RAD50	c.260G>A	R87H	Class C25	Class C0	0.962	15.33	14.41	10.56	0.000154	1
RAD50	c.280A>C	I94L	Class C0	Class C0	0	15.56	7.01	8.94	0.003537	1
RAD50	c.379G>A	V127I	Class C25	Class C0	0.999	20.1	11.96	5.57	0.001768	1
RAD50	c.412C>T	R138X	X	X	1	36	45.4	45.4	0	1
RAD50	c.511G>T	A171S	Class C65	Class C0	0.098	13.15	23.39	2.22	0	0
RAD50	c.552-1G>A	NA	intronic	intronic	0	19.33	0	0	0	0
RAD50	c.572C>T	T191I	Class C65	Class C0	0.584	19.41	33.48	5.11	0.004536	1
RAD50	c.597A>C	Q199H	Class C15	Class C0	1	19.09	14.36	14.01	0	0
RAD50	c.671G>A	R224H	Class C0	Class C0	1	20.5	8.23	1.81	0.000615	0
RAD50	c.715G>C	E239Q	Class C0	Class C0	0.978	16.98	13.82	2.73	0	0
RAD50	c.741A>G	E247E	SYN	SYN	0	7.999	0.92	1.95	0	0
RAD50	c.756+5C>T	NA	intronic	intronic	0	1.684	0	0	0	0
RAD50	c.756+6delT	NA	intronic	intronic	0	0	0	0	0	0
RAD50	c.757-12C>A	NA	intronic	intronic	0	9.495	0	0	0	0
RAD50	c.785T>G	L262H	Class C65	Class C0	0.202	24	20.32	3.17	0.000077	0
RAD50	c.885+5G>A	NA	intronic	intronic	0	13.35	0	0	0	0
RAD50	c.921A>G	L307L	SYN	SYN	0	7.949	1.28	1.71	0.000538	0
RAD50	c.943G>T	V315L	Class C25	Class C0	0.777	21.8	20.81	4.73	0.001999	0
RAD50	c.980G>A	R327H	Class C25	Class C0	0.994	27.7	14.41	3.66	0.002384	0
RAD50	c.1094G>A	R365Q	Class C0	Class C0	0.013	15.29	5.73	2.03	0.000308	0
RAD50	c.1246-10A>G	NA	intronic	intronic	0	0.022	0	0	0	0
RAD50	c.1277A>G	Q426R	Class C35	Class C35	0.992	24.2	33.7	33.7	0.000308	0
RAD50	c.1336A>G	K446E	Class C55	Class C0	1	24.7	23.7	16.55	0.000154	0
RAD50	c.1456C>T	R486C	Class C25	Class C15	0.752	17.94	20.93	7.76	0.000077	0
RAD50	c.1670C>G	S557C	Class C15	Class C0	0.996	20	18.06	5.44	0	0
RAD50	c.1713C>T	P571P	SYN	SYN	0	9.889	0.14	0.78	0	0
RAD50	c.1794-59A>C	NA	intronic	intronic	0	12.18	0	0	0	0
RAD50	c.1875C>G	Y625X	X	X	1	37	45.4	45.4	0	1
RAD50	c.1878A>G	E626E	SYN	SYN	0	8.423	0.16	1.17	0	0
RAD50	c.1911T>A	D637E	Class C35	Class C0	0.972	21.3	23.98	5.93	0.000154	1
RAD50	c.2025C>T	D675D	SYN	SYN	0	7.498	0.15	2.97	0.028141	1
RAD50	c.2047G>A	V683I	Class C25	Class C0	0.635	15.23	11.96	1.48	0.000077	1
RAD50	c.2054delAG	Q685fs	FS	FS	1	38	45.4	45.4	0	1

RAD50	c.2091C>T	V697V	SYN	SYN	0	9.127	0.74	3.96	0.001153	1
RAD50	c.2163A>G	K721K	SYN	SYN	0	9.862	1.5	1.54	0	1
RAD50	c.2173C>T	R725W	Class C65	Class C35	0.981	23.3	22.31	18.98	0.000077	1
RAD50	c.2283A>G	I761M	Class C0	Class C0	0.212	11.68	15.7	9.86	0	0
RAD50	c.2288G>A	R763H	Class C0	Class C0	0.004	15.11	5.06	2.36	0.000154	0
RAD50	c.2397G>C	Q799H	Class C15	Class C0	0.966	24.3	14.36	5.57	0.000384	0
RAD50	c.2525T>C	V842A	Class C65	Class C0	0.048	13.26	27.75	5.55	0.000308	0
RAD50	c.2750C>T	T917I	Class C15	Class C0	0.001	12.85	6.59	2.73	0	0
RAD50	c.2837A>T	D946V	Class C35	Class C0	0.029	11.58	23.09	6.45	0	0
RAD50	c.2841T>C	I947I	SYN	SYN	0	8.619	0.1	1.42	0	0
RAD50	c.2910C>T	D970D	SYN	SYN	0	7.222	0.9	2.49	0.004204	0
RAD50	c.2938del5	L980fs	FS	FS	1	41	45.4	45.4	0	1
RAD50	c.3036+37T>C	NA	intronic	intronic	0	7.01	0	0	0	0
RAD50	c.3036+5G>A	NA	intronic	intronic	0	11.14	0	0	0.000616	0
RAD50	c.3037-3T>C	NA	intronic	intronic	0	9.988	0	0	0.006	0
RAD50	c.3153G>A	L1051L	SYN	SYN	0	7.995	0.13	1.65	0.004154	0
RAD50	c.3165-4A>T	NA	intronic	intronic	0	7.717	0	0	0.000923	0
RAD50	c.3165-8T>G	NA	intronic	intronic	0	2.824	0	0	0.002693	0
RAD50	c.3239G>A	G1080D	Class C65	Class C0	0.804	25.4	35.86	9.73	0	0
RAD50	c.3260A>G	H1087R	Class C0	Class C0	0	4.723	7.03	0.97	0	0
RAD50	c.3278G>A	R1093Q	Class C35	Class C0	0.265	17.12	14.43	2.45	0	0
RAD50	c.3311A>G	Y1104C	Class C25	Class C25	0.942	20.6	8.34	7.58	0	0
RAD50	c.3363G>T	L1121L	SYN	SYN	0	8.752	0.13	0.13	0	1
RAD50	c.3476-12delTTC	NA	intronic	intronic	0	0	0	0	0	0
RAD50	c.3496C>T	R1166W	Class C65	Class C0	0.999	24.4	22.31	4.79	0	1
RAD50	c.3790C>T	L1264F	Class C15	Class C15	1	32	12.59	12.59	0	1
RAD50	c.3836G>A	R1279H	Class C25	Class C0	0.982	33	14.41	3.6	0.000154	1
RAD50	c.3852delGAAA	E1284fs	IFD	IFD	1	38	44.34	40.7	0	1
RAD50	c.3879C>T	I1293I	SYN	SYN	0	7.962	1.86	1.91	0.003306	1
RAD50	c.3902A>G	K1301R	Class C0	Class C0	0.001	15.75	2.1	2.07	0	0
RAD51	c.-1 dupA	unknown	intronic	intronic	0	14.41	0	0	0	1
RAD51	c.-2 -19A>G	NA	intronic	intronic	0	5.169	0	0	0.000461	1
RAD51	c.108 C>T	N36N	SYN	SYN	0	9.624	2.26	1.34	0.000077	1
RAD51	c.414 T>C	H138H	SYN	SYN	0	16.07	0.12	0.54	0	1
RAD51	c.449 G>A	R150Q	Class C0	Class C0	0	9.975	2.33	1.54	0.001384	1
RAD51	c.531 -31C>T	NA	intronic	intronic	0	6.909	0	0	0	1
RAD51	c.645 G>A	R215R	SYN	SYN	0	14.49	2.4	1.06	0.000154	1
RAD51	c.671 C>G	A224G	Class C0	Class C0	0.927	35	9.35	5.58	0.000077	1
RAD51	c.720 C>G	A240A	SYN	SYN	0	10.47	2.04	1.83	0.000154	1
RAD51	c.895 +33G>A	NA	intronic	intronic	0	0	0	0	0	1

RAD51	c.895 +5delG	NA	intronic	intronic	0	0	0	0	0	1
RAD51	c.976 A>G	M326V	Class C0	Class C0	0	5.798	11.2	4.53	0	1
RINT1	c.43-12delTTC	NA	intronic	intronic	0	0	0	0	0.0092	1
RINT1	c.43-20C>G	NA	intronic	intronic	0	9.827	0	0	0	1
RINT1	c.177C>T	F59F	SYN	SYN	0	12.21	0.11	1.24	0.000154	1
RINT1	c.281C>G	T94R	Class C65	Class C65	1	20.7	37.76	37.76	0	1
RINT1	c.301A>G	K101E	Class C55	Class C0	0.935	16.27	23.7	5.97	0	1
RINT1	c.319T>G	L107V	Class C25	Class C0	0.985	16.48	25.8	7.33	0	1
RINT1	c.376C>T	H126Y	Class C0	Class C0	0.028	6.361	2.26	2.87	0.000308	1
RINT1	c.388G>A	A130T	Class C0	Class C0	0.01	9.408	8.8	3.32	0.000538	1
RINT1	c.399C>T	S133S	SYN	SYN	0	1.297	1.38	0.94	0.000077	1
RINT1	c.413C>T	A138V	Class C65	Class C0	0.234	13.92	35.49	6.61	0.000384	1
RINT1	c.501A>T	Q167H	Class C15	Class C0	0.014	13.52	14.36	3.51	0	1
RINT1	c.543C>T	T181T	SYN	SYN	0	9.692	0.19	1.51	0	1
RINT1	c.690-15A>G	NA	intronic	intronic	0	0.019	0	0	0.000077	1
RINT1	c.732C>T	I244I	SYN	SYN	0	6.717	1.15	0.68	0.006074	1
RINT1	c.736C>A	P246T	Class C0	Class C0	0.418	12.74	4.45	2.64	0	1
RINT1	c.778G>T	A260S	Class C0	Class C0	0	4.759	4.31	2.16	0	1
RINT1	c.782C>T	P261L	Class C65	Class C0	0.001	7.221	26.92	2.69	0.000154	1
RINT1	c.891T>C	P297P	SYN	SYN	0	8.597	0.14	1.81	0	1
RINT1	c.1025T>C	M342T	Class C45	Class C0	0.003	7.485	22.41	2.88	0.000308	1
RINT1	c.1121G>A	R374Q	Class C35	Class C35	1	36	14.43	14.43	0	1
RINT1	c.1256C>G	P419R	Class C0	Class C0	0.997	17.98	27.5	27.41	0.000077	1
RINT1	c.1270A>T	S424C	Class C15	Class C15	0.034	14.35	5.02	6.25	0	1
RINT1	c.1334-5delA, c.1334-1_1335delGTT	IFR	IFD	IFD	1	21.2	43.15	43.15	0	1
RINT1	c.1377C>T	A459A	SYN	SYN	0	10.56	0.23	0.23	0.000154	1
RINT1	c.1385C>T	S462L	Class C65	Class C15	0.997	18.39	28.64	16.77	0	1
RINT1	c.1428A>G	P476P	SYN	SYN	0	10.1	0.14	1.06	0	1
RINT1	c.1449G>T	M483I	Class C0	Class C0	0.105	14.59	21.99	21.99	0	1
RINT1	c.1453C>T	L485L	SYN	SYN	0	6.03	0.13	0.13	0	1
RINT1	c.1465A>G	I489V	Class C0	Class C0	0.212	16.67	9.59	9.87	0	1
RINT1	c.1519G>A	E507K	Class C55	Class C0	0.837	23.9	19.98	4.73	0.000077	1
RINT1	c.1562C>T	T521I	Class C65	Class C0	0.969	19.98	33.48	10.08	0	1
RINT1	c.1656T>C	D552D	SYN	SYN	0	9.369	0.15	1.38	0	1
RINT1	c.1672-3C>T	NA	intronic	intronic	0	12.25	0	0	0	1
RINT1	c.1949C>T	P650L	Class C65	Class C15	0.997	14.16	26.92	14.45	0	1
RINT1	c.1962G>A	T654T	SYN	SYN	0	0.01	0.91	2.3	0.001384	1
RINT1	c.1985T>C	L662S	Class C65	Class C55	0.997	23.9	22.76	19.82	0.000154	1
RINT1	c.2003T>C	F668S	Class C55	Class C0	0.015	10.68	16.4	2.6	0.003767	1
RINT1	c.2036T>C	V679A	Class C0	Class C0	0	5.292	6.42	1.83	0	1

RINT1	c.2067+6T>A	NA	intronic	intronic	0	10.85	0	0	0.008381	1
RINT1	c.2090A>G	N697S	Class C45	Class C45	0.966	21.6	13.4	13.4	0	1
RINT1	c.2128C>T	R710W	Class C65	Class C65	1	21.2	22.31	22.31	0	1
RINT1	c.2159G>T	C720F	Class C65	Class C0	0.999	23.3	21.32	17.68	0	1
RINT1	c.2176T>C	Y726H	Class C65	Class C35	0.866	24.3	16.52	17.35	0.000231	1
RINT1	c.2179T>C	F727L	Class C15	Class C15	0.998	32	12.82	12.82	0	1
RINT1	c.2184A>G	K728K	SYN	SYN	0	12.4	0.14	0.14	0.000077	1
RINT1	c.2193A>G	K731K	SYN	SYN	0	8.839	0.14	0.14	0	1
RINT1	c.2253G>T	L751L	SYN	SYN	0	6.883	0.13	1.03	0	1
RINT1	c.2276C>T	P759L	Class C25	Class C15	0.015	17.59	24.42	22.59	0.006766	1
RINT1	c.2361G>A	W787*	IFD	IFD	1	40	43.98	43.98	0	1
RINT1	c.2362C>T	P788S	Class C65	Class C0	0.999	27.8	17.5	3.71	0	1
XRCC2	c.1-1G>A	NA	Class C65	Class C65	1	13.74	36.3	36.3	0	1
XRCC2	c.40-16T>C	NA	intronic	intronic	0	14.8	0	0	0.001231	1
XRCC2	c.46G>T	A16S	Class C65	Class C0	0.584	15.35	23.39	8.62	0	1
XRCC2	c.49C>T	R17X	X	X	1	37	45.4	45.4	0	1
XRCC2	c.181C>A	L61I	Class C0	Class C0	0.604	17.26	15.46	15.46	0	1
XRCC2	c.271C>T	R91W	Class C65	Class C65	1	14.77	22.31	22.31	0	1
XRCC2	c.283A>G	I95V	Class C0	Class C0	0.013	0.61	3.96	5.81	0.000231	1
XRCC2	c.651_652delTG	C217X	IFD	IFD	1	11.76	43.98	43.98	0	1
XRCC2	c.693G>T	W231C	Class C65	Class C65	1	16.34	29.05	29.05	0	1
XRCC2	c.808T>G	F270V	Class C45	Class C45	1	19.47	17.61	17.61	0.002692	1

[a]PMSAs include protein sequences from Human through *Ornithorynchus anatinus*; [b]3 S/P=3 substitutions per position: Human through 3 substitutions per position pMSA depth, which was different for each gene: *ATM, Brachiostoma floridae*; *BARD1, Xenopus*; *CHEK2, Danio rerio; MRE11A, Branchiostoma floridae; NBN, Anolis carolinensis; RAD50, Drosophila melanogaster; RAD51, Arabidopsis thaliana; RINT1, Strongylocentrotus purpuratus; XRCC2, Danio rerio*;

Supplemental Table 2. A listing of the regions used for the domain boundaries rMS inclusion

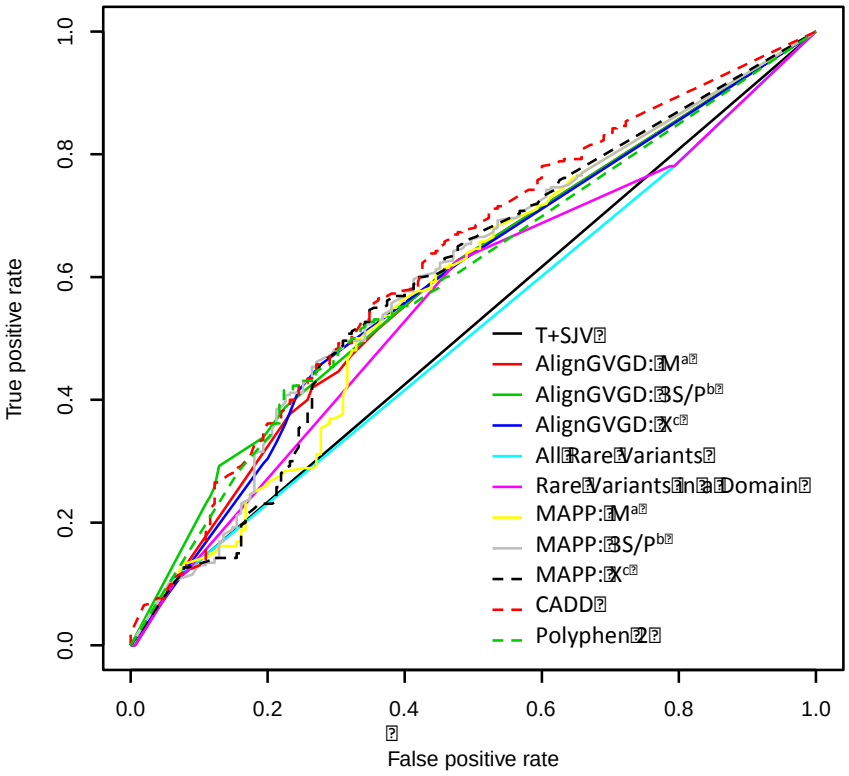
Genes	mRNA Assession	Protein Assession	Gene Length, a	Summed Length of Domain, a	Domain Region
ATM	NM_000051.3	NP_000042.3	3056	1,097 (36%)	1960+ (includes FAT and FATC)
BARD1	NM_000465.3	NP_000456.2	777	273 (35%)	26-119 (RING & BRCA1 interaction); 568-748 (BRCT)
CHEK2	NM_007194.3	NP_009125.1	543	543 (100%)	Entire Gene
RAD51	NM_002875.4	NP_002866.2	339	339 (100%)	Entire Gene
RINT1	NM_021930.4	NP_068749.3	792	792 (100%)	Entire Gene
XRCC2	NM_005431.1	NP_005422.1	280	280 (100%)	Entire Gene
MRE11A	NM_005591.3	NP_005582.1	708	497 (70%)	13-482 (Phosphoesterase; RAD50 interaction); 568-594 (GAR)
RAD50	NM_005732.3	NP_005723.2	1312	455 (35%)	3-166 (ATPase); 189-198 (MRE11A interaction); 635-734 (Zinc hook); 1117-1297 (ATPase)
NBN	NM_002485.4	NP_002476.2	754	369 (49%)	1-110 (FHA); 111-206 & 227-334 (BRCT); 640-662, 681-691 & 734-754 (MRE11A interaction)

Supplemental Table 3. ROC curve estimates for each variant classifier

Variant Classifier: pMSA depth	Observation	AUC
Align-GVGD: M[a]	416	0.5926
Align-GVGD: 3 S/P[b]	416	0.6025
MAPP: M[a]	416	0.5817
MAPP: 3 S/P[b]	416	0.6000
CADD	416	0.6235
Polyphen2	416	0.5941
All Rare Variants	416	0.5105
All Rare Variants in Domains	416	0.5596
Truncations and Splice Junction Variants	416	0.5193

[a]M=Mammals: Human through *Ornithorynchus anatinus* pMSA depth; [b]3 S/P=3 substitutions per position: Human through 3 substitutions per position pMSA depth, which was different for each gene: *ATM, Brachiostoma floridae*; *BARD1*, *Xenopus*; *CHEK2, Latimeria chalumnae; MRE11A, Branchiostoma floridae; NBN, Anolis carolinensis; RAD50, Drosophila melanogaster; RAD51, Arabidopsis thaliana; RINT1, Strongylocentrotus purpuratus; XRCC2, Danio rerio*; [c]X=*Xenopus*: Human through *Xenopus* pMSA depth.

Supplemental Figure 1. ROC curves for each variant classifier. [a]M=Mammals: Human through *Ornithorynchus anatinus* pMSA depth; [b]3S/P: Human through 3M pMSA depth, which was different for each gene: *ATM*, *Brachiostoma floridae*; *BARD1*, *Xenopus*; *CHEK2*, *Danio rerio; MRE11A*, *Branchiostoma floridae; NBN*, *Anolis carolinensis; RAD50*, *Drosophila melanogaster; RAD51*, *Arabidopsis thaliana; RINT1*, *Strongylocentrotus purpuratus; XRCC2*, *Danio rerio*; [c]X=*Xenopus*: Human through *Xenopus* pMSA depth. T+SJV: Truncations and Splice Junction Variants



Supplemental Table 4. Correlations between missense analysis programs for rare key-domain missense variants. Adjusted R^2 and (P-values).

	MAPP[a]	CADD	Polyphen-2	Synthetic Consensus
Align-GVGD[a]	0.43($8.84X10^{-18}$)	0.19($8.40X10^{-8}$)	0.27($1.07X10^{-10}$)	0.64($4.07X10^{-31}$)
MAPP[a]	X	0.15($3.18X10^{-6}$)	0.26($2.49X10^{-10}$)	0.47($2.81X10^{-20}$)
CADD	X	X	0.56($2.36X10^{-25}$)	0.55($8.62X10^{-25}$)
Polyphen-2	X	X	X	0.59($1.84X10^{-27}$)

[a]Protein Multiple Sequence Alignments were from Human through 3S/P

Supplemental Table 5. Combined Odds Ratio Estimates from Case-Control Mutation Screening of ATM, CHEK2, and NBN

Analysis	Distinct Variants	Control(%)	Case(%)	Adjusted OR[a](CI)	P-value
Noncarrier	43	1,054(94.02)	1,192(91.90)	Reference	
Carrier of Rare Missense Outside of Domain	19	22(1.96)	9(0.69)	0.33(0.15-0.74)	$7.28X10^{-3}$
Carrier of Truncation or Splice Junction Variant	17	6(0.54)	24(1.85)	4.20(1.69-10.44)	0.0020
Rare Missense Substitutions: MAPP					
rMS < 11	29	21(1.87)	22(1.70)	1.08(0.58-2.02)	0.8104
rMS ≥ 11	45	18(1.61)	50(3.86)	2.73(1.56-4.79)	$4.55X10^{-4}$
Rare Missense Substitutions: Align-GVGD					
rMS < C35	46	31(2.77)	42(3.24)	1.47(0.90-2.40)	0.1234
rMS ≥ C35	28	8(0.71)	30(2.31)	3.19(1.43-7.13)	$4.65X10^{-3}$
Rare Missense Substitutions: CADD					
rMS < 23	51	29(2.59)	48(3.70)	1.59(0.98-2.60)	0.0611
rMS ≥ 23	23	10(0.89)	24(1.85)	2.60(1.22-5.57)	0.0136
Rare Missense Substitutions: PolyPhen-2					
rMS < 0.9	35	22(1.96)	28(2.16)	1.32(0.73-2.38)	0.3556
rMS ≥ 0.9	39	17(1.52)	44(3.39)	2.51(1.40-4.49)	$1.96X10^{-3}$
Total	153	1,121	1,297		

[a]Adjusted for race/ethnicity and study center; CI=95% Confidence Interval

Supplemental Table 6. Missense Analysis Program Overlap

Variant Classifier	Control	(%)	rMS/T+SJV	Case	(%)	rMS/T+SJV	Adjusted OR[a]	(CI)	P-Value
Only NBN, CHEK2, and ATM									
Carrier of T+SJV	6	(0.5)		24	(1.9)		4.20	(1.69-10.44)	0.0020
One or More	23	(2.1)	3.83	52	(4.0)	2.17	2.22	(1.33-3.72)	0.0023
Two or More	13	(1.2)	2.17	37	(2.9)	1.54	2.59	(1.35-5.00)	0.0044
Three or More	11	(1.0)	1.83	31	(2.4)	1.29	2.77	(1.36-5.63)	0.0050
All Four	2	(0.2)	0.33	12	(0.9)	0.50	6.14	(1.34-28.14)	0.0194
BARD1-MRE11A-RAD50-RAD51-RINT1-XRCC2									
Carrier of T+SJV	3	(0.3)		3	(0.2)		1.07	(0.20-5.60)	0.9401
One or More	13	(1.2)	4.33	44	(3.4)	14.67	2.53	(1.33-4.83)	0.0047
Two or More	5	(0.4)	1.67	34	(2.6)	11.33	4.90	(1.87-12.84)	0.0012
Three or More	3	(0.3)	1.00	21	(1.6)	7.00	5.25	(1.52-18.14)	0.0088
All Four	0	(0.0)	0.00	8	(0.6)	2.67			<0.000[a]
All Nine Genes									
Carrier of T+SJV	9	(0.8)		27	(2.1)		3.31	(1.53-7.16)	$2.36X10^{-3}$
One or More	34	(3.0)	5.67	93	(7.2)	3.88	2.37	(1.57-3.60)	$4.56X10^{-5}$
Two or More	18	(1.6)	3.00	70	(5.4)	2.92	3.18	(1.85-5.46)	$2.68X10^{-5}$
Three or More	14	(1.2)	2.33	52	(4.0)	2.17	3.27	(1.77-6.04)	$1.51X10^{-4}$
All Four	2	(0.2)	0.33	20	(1.5)	0.83	8.61	(1.96-37.81)	$4.35X10^{-3}$

[a] P-value from Fisher's Exact Test

Supplemental Table 7. Excluding a single gene from the 9 gene ensemble

Variant Classifier	Control	(%)	rMS/T+SJV	Case	(%)	rMS/T+SJV	Adjusted OR[a]	(CI)	P-Value
Dropping BARD1									
Non-carrier	1,014	(90.45)		1,140	(87.90)		Reference		
rMS Outside of Domain	50	(4.46)		44	(3.39)		0.77	(0.50-1.18)	0.2298
Carrier of T+SJV	6	(0.54)		10	(0.77)		1.80	(0.63-5.15)	0.2718
rMS Inside of Domain									
Align-GVGD	9	(0.80)	1.50	35	(2.70)	3.50	3.25	(1.52-6.96)	0.0024
MAPP	20	(1.78)	3.33	56	(4.32)	5.60	2.42	(1.42-4.15)	0.0012
CADD	15	(1.34)	2.50	42	(3.24)	4.20	2.48	(1.34-4.58)	0.0038
Polyphen2	25	(2.23)	4.17	63	(4.86)	6.30	2.13	(1.31-3.47)	0.0024
Overlap of Two	16	(1.43)	2.67	53	(4.09)	5.30	2.82	(1.58-5.06)	4.84E-04
Dropping MRE11A									
Non-carrier	994	(88.67)		1,098	(84.66)		Reference		
rMS Outside of Domain	55	(4.91)		44	(3.39)		0.75	(0.49-1.15)	0.1868
Carrier of T+SJV	9	(0.80)		27	(2.08)		3.24	(1.49-7.01)	0.0029
rMS Inside of Domain									
Align-GVGD	11	(0.98)	1.22	48	(3.70)	1.78	3.58	(1.81-7.05)	2.37E-04
MAPP	25	(2.23)	2.78	74	(5.71)	2.74	2.69	(1.67-4.34)	4.82E-05
CADD	15	(1.34)	1.67	44	(3.39)	1.63	2.68	(1.45-4.93)	1.60E-03
Polyphen2	30	(2.68)	3.33	76	(5.86)	2.81	2.25	(1.44-3.52)	3.67E-04
Overlap of Two	18	(1.61)	2.00	62	(4.78)	2.30	2.96	(1.71-5.12)	9.96E-05
Dropping RAD50									
Non-carrier	1,016	(90.63)		1,125	(86.74)		Reference		
rMS Outside of Domain	37	(3.30)		22	(1.70)		0.54	(0.31-0.94)	0.0298
Carrier of T+SJV	6	(0.54)		25	(1.93)		4.53	(1.83-11.23)	0.0011
rMS Inside of Domain									
Align-GVGD	10	(0.89)	1.67	46	(3.55)	1.84	3.79	(1.87-7.69)	2.22E-04
MAPP	24	(2.14)	4.00	72	(5.55)	2.88	2.71	(1.67-4.41)	5.85E-05
CADD	13	(1.16)	2.17	40	(3.08)	1.60	2.97	(1.55-5.68)	0.0010
Polyphen2	25	(2.23)	4.17	71	(5.47)	2.84	2.55	(1.58-4.11)	1.34E-04
Overlap of Two	17	(1.52)	2.83	60	(4.63)	2.40	3.01	(1.72-5.28)	1.18E-04
Dropping RAD51									
Non-carrier	1,000	(89.21)		1,096	(84.50)		Reference		
rMS Outside of Domain	49	(4.37)		40	(3.08)		0.75	(0.48-1.16)	0.1940
Carrier of T+SJV	9	(0.80)		27	(2.08)		3.31	(1.53-7.16)	0.0024
rMS Inside of Domain									
Align-GVGD	11	(0.98)	1.22	49	(3.78)	1.81	3.62	(1.84-7.13)	2.02E-04
MAPP	26	(2.32)	2.89	78	(6.01)	2.89	2.63	(1.65-4.21)	5.28E-05

CADD	14	(1.25)	1.56	48	(3.70)	1.78	3.05	(1.64-5.67)	4.25E-04
Polyphen2	27	(2.41)	3.00	81	(6.25)	3.00	2.57	(1.63-4.07)	5.47E-05
Overlap of Two	17	(1.52)	1.89	70	(5.40)	2.59	3.38	(1.95-5.87)	1.54E-05
Dropping RINT1									
Non-carrier	1,009	(90.01)		1,121	(86.43)		Reference		
rMS Outside of Domain	50	(4.46)		41	(3.16)		0.72	(0.46-1.12)	0.1428
Carrier of T+SJV	9	(0.80)		27	(2.08)		3.26	(1.51-7.05)	0.0027
rMS Inside of Domain									
Align-GVGD	9	(0.80)	1.00	42	(3.24)	1.56	3.75	(1.78-7.90)	5.02E-04
MAPP	22	(1.96)	2.44	68	(5.24)	2.52	2.72	(1.64-4.52)	1.06E-04
CADD	13	(1.16)	1.44	39	(3.01)	1.44	2.85	(1.48-5.48)	0.0017
Polyphen2	24	(2.14)	2.67	69	(5.32)	2.56	2.48	(1.52-4.04)	2.69E-04
Overlap of Two	16	(1.43)	1.78	61	(4.70)	2.26	3.14	(1.77-5.58)	9.12E-05
Dropping XRCC2									
Non-carrier	1,010	(90.10)		1,106	(85.27)		Reference		
rMS Outside of Domain	45	(4.01)		40	(3.08)		0.81	(0.51-1.27)	0.3561
Carrier of T+SJV	9	(0.80)		26	(2.00)		3.14	(1.44-6.81)	0.0039
rMS Inside of Domain									
Align-GVGD	11	(0.98)	1.22	45	(3.47)	1.73	3.31	(1.67-6.56)	6.11E-04
MAPP	21	(1.87)	2.33	70	(5.40)	2.69	2.84	(1.70-4.75)	6.56E-05
CADD	10	(0.89)	1.11	44	(3.39)	1.69	3.66	(1.80-7.45)	3.49E-04
Polyphen2	21	(1.87)	2.33	72	(5.55)	2.77	2.82	(1.70-4.70)	6.50E-05
Overlap of Two	13	(1.16)	1.44	63	(4.86)	2.42	3.83	(2.06-7.12)	2.11E-05

[a]Adjusted for ethnicity and study center. OR=Odds Ratio; CI=95% Confidence Interval; rMS=rare missense substitution; T+SJV = Truncation or Splice Junction Variant.

Supplemental Table 8. Previously reported associated SNPs and corresponding breast cancer risk in the multiple papers used for reference in comparison of our dataset and risk estimates for non-Caucasians.

			Caucasian[h]			Recent African Ancestry[j]			Latina[k]				East Asian[l]		
Locus	SNP	Allele[a]	MAF[b]	OR[c]	Observed MAF	AF[d]	OR[c]	Observed MAF	AF[d]	Heterozygous OR	Homozygous OR	Observed MAF	AF[d]	OR[c]	Observed MAF
1p11.2	rs11249433	A/G	0.40	1.09	0.45	0.101	1.07	0.16		1.13[f]		0.27	0.03	1.16	0.06
CASP8	rs1045485	G/C	0.13	0.97	0.12		0.97[g]	0.05		0.97[g]		0.06		0.97[g]	0.00
2q35	rs13387042	A/G	0.49	0.88	0.46	0.251	1.01[e]	0.30	0.67	0.85[e]	0.67[e]	0.59	0.90	0.94	0.88
SLC4A7	rs4973768	C/T	0.47	1.10	0.49	0.35	1.06	0.39		1.11[f]		0.59	0.19	1.11	0.23
5p12	rs10941679	A/G	0.25	1.13	0.26	0.189	0.94	0.22		1.11[f]		0.36	0.48	1.08	0.53
MAP3K1	rs889312	A/C	0.28	1.12	0.28	0.34	0.93	0.33	0.43	1.04	1.09	0.41	0.52	1.05[e]	0.43
ESR1	rs2046210	G/A	0.41	1.08	0.38	0.627	1.02	0.58		1.18[f]		0.33	0.34	1.27	0.39
8q24	rs13281615	A/G	0.41	1.09	0.42	0.44	1.00	0.44		1.06[f]		0.60	0.52	1.03	0.51
CDKN2A/B	rs1011970	G/T	0.17	1.06	0.18	0.342	0.90	0.26		1.06[f]		0.33	0.08	1.06	0.12
ANKRD16	rs2380205	C/T	0.44	0.98	0.38	0.584	1.03[e]	0.57		0.98[f]		0.28	0.11	0.98	0.13
ZNF365	rs10995190	G/A	0.16	0.86	0.18	0.168	1.14[e]	0.17		0.96[f]		0.10	0.02	1.06	0.04
ZMIZ1	rs704010	C/T	0.38	1.08	0.40	0.076	1.04	0.08		1.07[f]		0.45	0.29	1.05[e]	0.40
FGFR2	rs2981578	T/C	0.50[i]	1.24[i]	0.48	0.871	1.24	0.86		1.22[f]		0.52		1.19[i]	0.62
LSP1	rs3817198	T/C	0.31	1.07	0.33	0.159	0.85	0.17	0.20	1.02	1.20	0.21	0.13	1.07	0.10
11q13	rs614367	C/T	0.15	1.21	0.15	0.128	1.09	0.15		1.25[f]		0.08		1.29	0.03
RAD51L1	rs999737	C/T	0.23	0.92	0.23	0.033	1.07[e]	0.06	0.17	1.01	1.17	0.18	0.00	1.08	0.00
TOX3	rs3803662	G/A	0.26	1.24	0.29	0.516	0.96	0.42	0.41	1.27	1.25	0.42	0.64	1.15[e]	0.54
COX11	rs6504950	G/A	0.28	0.94	0.28	0.346	1.11[e]	0.35		0.96[f]		0.21	0.07	0.98	0.15

[a]Major/Minor Allele; [b]MAF=Minor Allele Frequency in controls; [c]OR=Per allele OR; [d]AF=Risk Allele Frequency; [e]The alternative allele was determined as the risk allele, so the reciprocal of the published OR was used; [f]Latian OR = (Caucasian OR + East Asian OR)/2; [g]Minority OR could not be found, so Caucasian was used; [h]Michailidou, K. *et al.* (2013). N=10,052 cases and 12575 controls; [i]Meyer, K. B. et al. (2013). N=89,050 Caucasians, N=13,983 East Asians; [j]Huo, D. *et al.* (2012). N= 1509 cases and 1383 controls; [k]Fejerman, L. *et al.* (2013). N= 603 cases and 730 controls; [l]Zheng, W. *et al.* (2013). N= 6,000+ cases and 6,000+ controls