ORIGINAL ARTICLE

Rare, protein-truncating variants in *ATM*, *CHEK2* and *PALB2*, but not *XRCC2*, are associated with increased breast cancer risks

Brennan Decker,[1,2] Jamie Allen,[1] Craig Luccarini,[3] Karen A Pooley,[1] Mitul Shah,[3] Manjeet K Bolla,[1] Qin Wang,[1] Shahana Ahmed,[3] Caroline Baynes,[1] Don M Conroy,[1] Judith Brown,[1] Robert Luben,[1] Elaine A Ostrander,[2] Paul DP Pharoah,[1,3] Alison M Dunning,[3] Douglas F Easton[1,3]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2017-104588).

[1]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK
[2]Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA
[3]Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK

Correspondence to
Professor Douglas F Easton, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; dfe20@medschl.cam.ac.uk

Received 9 February 2017
Revised 9 May 2017
Accepted 22 May 2017
Published Online First
4 August 2017

To cite: Decker B, Allen J, Luccarini C, *et al*. *J Med Genet* 2017;**54**:732–741.

ABSTRACT

Background Breast cancer (BC) is the most common malignancy in women and has a major heritable component. The risks associated with most rare susceptibility variants are not well estimated. To better characterise the contribution of variants in *ATM*, *CHEK2*, *PALB2* and *XRCC2*, we sequenced their coding regions in 13 087 BC cases and 5488 controls from East Anglia, UK.

Methods Gene coding regions were enriched via PCR, sequenced, variant called and filtered for quality. ORs for BC risk were estimated separately for carriers of truncating variants and of rare missense variants, which were further subdivided by functional domain and pathogenicity as predicted by four *in silico* algorithms.

Results Truncating variants in *PALB2* (OR=4.69, 95% CI 2.27 to 9.68), *ATM* (OR=3.26; 95% CI 1.82 to 6.46) and *CHEK2* (OR=3.11; 95% CI 2.15 to 4.69), but not *XRCC2* (OR=0.94; 95% CI 0.26 to 4.19) were associated with increased BC risk. Truncating variants in *ATM* and *CHEK2* were more strongly associated with risk of oestrogen receptor (ER)-positive than ER-negative disease, while those in *PALB2* were associated with similar risks for both subtypes. There was also some evidence that missense variants in *ATM*, *CHEK2* and *PALB2* may contribute to BC risk, but larger studies are necessary to quantify the magnitude of this effect.

Conclusions Truncating variants in *PALB2* are associated with a higher risk of BC than those in *ATM* or *CHEK2*. A substantial risk of BC due to truncating *XRCC2* variants can be excluded.

INTRODUCTION

Breast cancer (BC) is the most common malignancy and the second leading cause of cancer deaths in women worldwide.[1] Both twin and population-based family studies indicate a substantial heritable component to BC risk.[2] To date, approximately half of familial risk of BC has been explained by a combination of common variants with small effect sizes,[3] together with rarer, protein coding variants that confer higher risks.[4] Truncating variants in some genes, including *BRCA1* and *BRCA2*, are associated with very high absolute risks, while deleterious variants in other genes, including *ATM*, *CHEK2* and *PALB2*, have been reported to confer more moderate risks.[4] However, since susceptibility alleles in these genes are rare, the risks associated with these have not yet been well estimated. Variants in several other genes, including *XRCC2*, have been suggested to contribute to risk, but evidence is more equivocal.[4]

ATM, *CHEK2*, *PALB2* and *XRCC2* play important roles in DNA repair. The *ATM* gene encodes a protein kinase that recognises double stranded DNA breaks and initiates multiple aspects of the damage response cascade. A recent meta-analysis estimated that truncating *ATM* variants were associated with a relative risk of 2.8,[4] and several studies have reported that certain subgroups of missense variants also contribute to BC risk.[5–9] *CHEK2* encodes a checkpoint kinase that interacts with cell cycle regulators and DNA repair proteins. The most common protein-truncating genetic variant in Western European populations is the frame shift c.1100delC (p.Thr367MetfsTer15, rs555607708); a recent analysis by the Breast Cancer Association Consortium (BCAC) estimated a relative risk of 2.26 for this variant.[10] Truncations of *PALB2*, the partner and localiser of *BRCA2*, have been associated with a fivefold relative risk,[4] though studies in different populations have produced divergent estimates.[11–14] *XRCC2* encodes a protein in the Rad51 family that participates in homologous recombination repair of double-stranded DNA breaks. Protein-truncating mutations in this gene are uncommon, but an association with BC risk has been suggested for rare missense variants.[15]

These four genes are now included in most multigene BC risk sequencing panels,[4] and it is therefore critical for genetic counselling to have accurate estimates of the BC risk associated with variants in these genes. To provide such estimates, we undertook a large, population-based study in which we sequenced the protein coding exons and intron-exon boundaries of *ATM*, *CHEK2*, *PALB2* and *XRCC2* in 13 087 BC cases and 5488 controls. We evaluated sets of rare variants, defined by predicted functional effect, and more common individual variants (population frequency >0.1%), for association with BC risk.

MATERIALS AND METHODS

Study population

Cases were drawn from SEARCH, a population-based study of BC in the region of East Anglia (UK) covered by the Eastern Cancer Registration and Information Centre (ECRIC).[16] The study enrolled subjects diagnosed before age 55 years with invasive BC from 1991 onwards and who were still alive at the start of the study in 1996 (prevalent cases, n=1087; median age=48 years), together with all patients diagnosed before age 70 years between 1996 and the present (n=12 000). Data on oestrogen-receptor and progesterone-receptor status and bilaterality were obtained through ECRIC and abstraction of medical records. Controls were drawn from three sources: (1) general practices participating in SEARCH who were frequency matched by age to the cases; (2) the European Prospective Investigation of Cancer (EPIC)-Norfolk study, a population-based cohort study of diet and health in Norfolk, East Anglia;[17] and (3) women undergoing breast screening as part of the National Health Service Breast Screening Programme in screening centres in Cambridgeshire, who participated in the Sisters in Breast Screening study.[18] Sequence analysis was conducted on samples from 13 824 BC cases and 5952 controls (pooled from the three studies above) of which 13 087 cases and 5488 controls passed all QC filters (see online supplementary figure s1 and supplementary methods) and were used in the analysis. Ethics approval was provided by the Cambridgeshire Research Ethics Committee, and written consent was obtained at the time of sample collection.

Amplicon design, enrichment, sequencing and variant calling

The Fluidigm Access Array 48.48 system was used for library preparation (see online supplementary methods). We designed 211 amplicons (see online supplementary table S1) to cover 98.1% of the bases within Consensus Coding Sequence (CCDS) exons of the four genes (see online supplementary table S2). Each library of 211 amplicons for 1536 samples was sequenced in 100-base paired-end mode on a single lane of an Illumina Hi-Seq2000. Raw sequence data were demultiplexed using the Illumina CASAVA 1.8 pipeline and aligned to the hg19 human reference sequence with BWA-MEM V.0.7.[19] GATK UnifiedGenotyper was used to perform SNP and indel discovery and variant calling across all samples simultaneously (see online supplementary methods).[20–22] After filtering samples and variants with >5% missing calls, 5488/5952 controls (92%) and 13 087/13 824 cases (95%) were retained for further analysis. Final variant calls were reproducible within this study and concordant with orthogonal methods (see online supplementary table s3 and supplementary methods).

Functional prediction and variant frequency classification

The Ensembl Variant Effect Predictor[23] was used to assign the canonical transcript-level and protein-level consequence for each genetic variant. Frameshift, stop/gain and canonical splice variants were grouped as protein truncating. Missense variants were further annotated with effect predictions from CADD,[24] PolyPhen2,[25] SIFT,[26] and AlignGVGD.[27] The consequences of the putative splice site variant *CHEK2* c.320–5T>A were evaluated using the *in silico* prediction tools SpliceSiteFinder-like,[28] MaxEntScan,[29] NNSPLICE,[30] GeneSplicer[31] and Human Splicing Finder.[32]

Variants detected in these four genes were annotated with allele frequencies observed in populations catalogued in the Exome Aggregation Consortium (ExAC) variation database (http://www.exac.broadinstitute.org).[33] All coding variants with a carrier frequency >0.1% in the 33 370 non-Finnish, European ExAC subjects were classified as common, while all others (including those not reported by ExAC) were classified as rare. The combined East Anglian case and control frequency of <0.1% was used to define rare non-coding variants.

Statistical analysis

For each variant with a carrier frequency >0.1% in ExAC European subjects (coding variants) or study cases and controls combined (non-coding variants), per-allele ORs and 95% CIs were computed, and Cochran-Armitage trend tests were carried out using the ‘prop.trend.test’ module of R.[34]

Since most variants were too rare to derive individual risk estimates, variants were grouped into classes (truncating, missense, synonymous and non-canonical splice) based on their predicted effect, and estimates were derived for carriers of any variant in each class. Rare missense variants were further subdivided based on domain, functional prediction scores and protein position. ORs were estimated by unconditional logistic regression, using the glm package in R. Profile likelihood-based confidence limits were derived using the confint routine. ORs were estimated using both the study controls (n=5488) and non-Finnish European ExAC subjects (n=33 370).

For truncating variants, separate analyses were conducted for oestrogen receptor (ER)-negative and ER-positive BC. Differences in ORs by subtype were assessed using case-only analyses.

RESULTS

Spectrum of variation

We identified 1273 variants in the four genes: 785 in *ATM*, 165 in *CHEK2*, 255 in *PALB2* and 68 in *XRCC2* (online supplementary tables S4 and S5). Among these, 72 were common variants (for coding variants: carrier frequency >0.1% in ExAC non-Finnish, European ancestry subjects; for non-coding variants: carrier frequency >0.1% in cases and controls combined) (see online supplementary table S4). However, the majority of variants (731/1273, 57.4%) were identified in a single subject (figure 1A). Most variants encoded a missense substitution (figure 1B), and 29.1% of the coding variants identified in this study had not been previously reported in the dbSNP, ExAC or COSMIC databases (figure 1C). All four genes had similar variant rates after adjusting for gene length, with approximately 60 variants per kilobase of coding sequence (figure 1C).

BC risks associated with *ATM*, *CHEK2* and *PALB2* truncating variants

Of the four genes, truncating variants in *PALB2* were associated with the highest BC risk, with an estimated OR=4.69 (95% CI 2.27 to 9.68, $p=6.9\times10^{-6}$)(table 1 and figure 2). Most truncating variants were rare, with 19/35 (54%) observed as singletons. Three variants, all in the final 280 codons of the gene, were more common: *PALB2* c.2718G>A (p.Trp906Ter, rs180177122; seven cases, one control), c.3113G>A (p.Trp1038Ter, rs180177132; 20 cases, one control) and c.3116delA (p.Asn1039IlefsTer2, rs180177133; eight cases, one control) (see online supplementary table 5). The combined OR for these three more common, C-terminal truncations was 3.57 (95% CI 1.27 to 10.1, p=0.016), somewhat, but not significantly, lower than the point estimate for all other truncating variants in the gene (OR=5.79; 95% CI 2.10 to 16.0, $p=2.2\times10^{-4}$; $P_{diff}=0.78$).

The risk estimate for *CHEK2*-truncating variants was OR=3.11 (95% CI 2.15 to 4.69, $p=5.6\times10^{-11}$; table 1 and

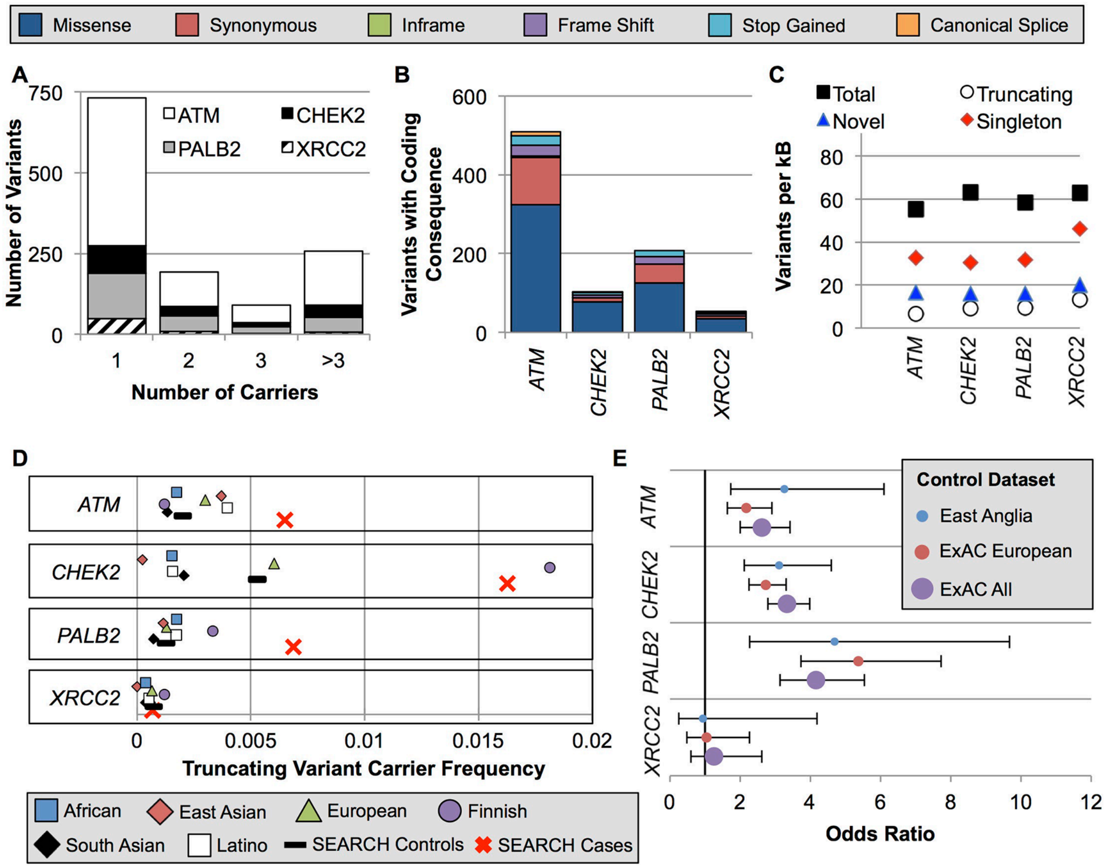

Figure 1 Spectrum of *ATM*, *CHEK2*, *PALB2* and *XRCC2* variation in ExAC populations and East Anglian cases and controls. (A) Nearly all variants were very rare or private. (B) The absolute number of variants per gene was highly divergent due to the difference in gene size. (C) However, when corrected for gene size, the genes had similar per-kB rates of total, truncating, singleton and novel variants not previously reported in the dbSNP, ExAC or COSMIC. (D) The carrier frequency for truncating variants in the study controls was near the mean of the ExAC populations. Notably, the ExAC Finnish population was an outlier for *CHEK2* and *PALB2* due to well-studied founder effects. (E) OR point estimates were similar using the study controls versus ExAC populations as controls. ExAC, Exome Aggregation Consortium.

figure 2). The most common truncating variant, *CHEK2* c.1100delC (p.Thr367MetfsTer15, rs555607708) accounted for 81% (196/242) of truncating variant carriers (online supplementary table S5) and was associated with an OR=3.18 (95% CI 2.01 to 4.92, p=6.1×10$^{−8}$). The risk estimate for the aggregate of the remaining 14 rare truncating variants was OR=2.83 (95% CI 1.20 to 6.69, p=0.020), consistent with the estimate for *CHEK2* c.1100delC.

Protein-truncating variants in *ATM* were associated with an elevated risk of BC similar to *CHEK2* (OR=3.26; 95% CI 1.82 to 6.46, p=2.1×10$^{−5}$; table 1 and figure 2).

Only 13 carriers of truncating variants in *XRCC2* were found in our study (table 1 and figure 2). There was no evidence for association with BC risk (OR=0.94; 95% CI 0.26 to 4.19, p=0.92).

For each of the four genes, the truncating variant carrier frequencies in the ExAC non-Finnish European population were similar to those in the study control group. Consequently, similar OR estimates were obtained when this dataset was substituted for the study controls (figure 1D–E and see online supplementary table S6).

BC associations for protein truncating variants disease subtype, age and family history

Truncating variants in *CHEK2* were associated with a higher relative risk for ER-positive (OR=3.42; 95% CI 2.33 to 5.21; table 2), and lower, non-significant risk for ER-negative BC (OR=1.59; 95% CI 0.80 to 3.00; P_{diff}=0.0032). A similar pattern was observed for progesterone receptor status, though

Table 1 Overall BC risk estimates for truncating mutations in each gene. Variant category tallies (online supplementary table S4) and carrier counts (online supplementary table S5) are available in the supplementary materials

	Cases			Controls				
	Carriers	Non-Carriers	Carriers Frequency	Carriers	Non-Carriers	Carrier Frequency	OR (95% CI)	P-Value
ATM	85	13,002	0.00649	11	5,477	0.00200	**3.26 (1.82–6.46)**	**2.1×10^{-5}**
CHEK2	213	12,874	0.01628	29	5,459	0.00528	**3.11 (2.15–4.69)**	**5.6×10^{-11}**
PALB2	89	12,998	0.00685	8	5,480	0.00146	**4.69 (2.27–9.68)**	**6.9×10^{-6}**
XRCC2	9	13,078	0.00069	4	5,484	0.00073	0.94 (0.26–4.19)	0.92

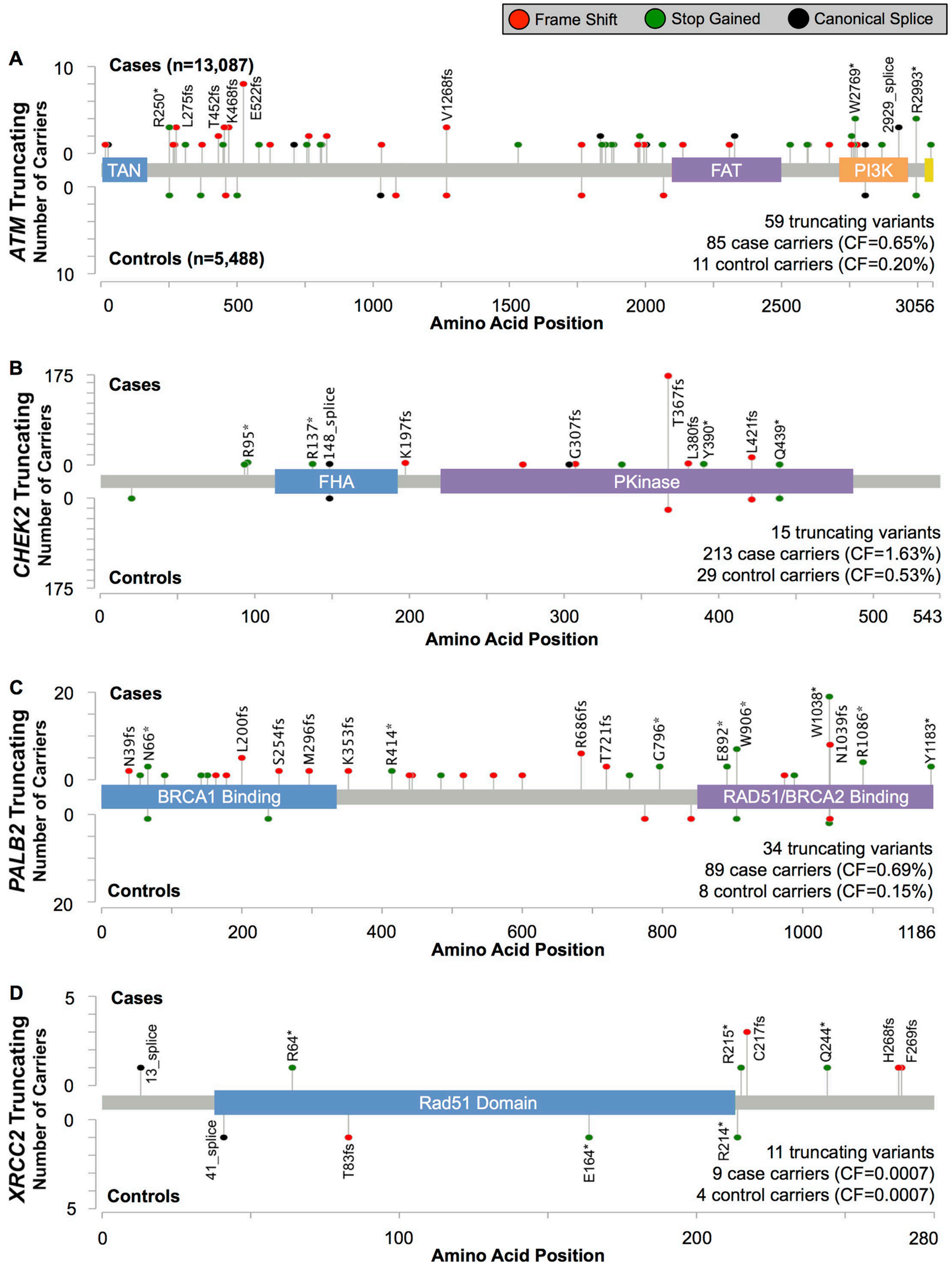

Figure 2 Position and frequency of protein-truncating variants in (A) *ATM*, (B) *CHEK2*, (C) *PALB2* and (D) *XRCC2*.

the difference was not significant (P_{diff}=0.18). Truncating variants in *ATM* were also associated with a higher risk for ER-positive disease (OR=3.42; 95% CI 2.33 to 5.21) than ER-negative disease (OR=1.59; 95% CI 0.80 to 3.00), though not significantly so (P_{diff}=0.11). There was no evidence of a difference in the estimated OR by ER-status for *PALB2* variants (ER-positive OR=4.32; 95% CI 2.07 to 10.5 vs ER-negative OR=5.58; 95% CI 2.19 to 15.2; P_{diff}=0.55).

Truncating *ATM* variants were more common in BC cases with a family history of BC among first-degree relatives (OR=2.06; 95% CI 1.12 to 3.64, p=0.022; table 2). There was some evidence of an association with a positive family history for carriers of truncating *CHEK2* variants (OR=1.51; 95% CI 0.97 to 2.28, p=0.070), but not for *PALB2* truncation carriers (OR=0.74; 95% CI 0.60 to 1.54, p=0.44). Bilateral BC was more common than unilateral disease in women with *CHEK2*-truncating variants (OR=3.27; 95% CI 1.66 to 5.83, p=0.0014; table 2). There was some evidence of an association with bilaterality for *PALB2* variant carriers (OR=2.85, 95% CI 0.86 to 6.91, p=0.080). No *ATM* truncations were found among bilateral BC cases.

Table 2 Associations by family history, synchronous or metachronous bilateral disease, age at diagnosis and hormone receptor subtype.

		ATM			*CHEK2*			*PALB2*		
	All Cases	Carriers	OR (95%CI)	P-Value	Carriers	OR (95%CI)	P-Value	Carriers	OR (95%CI)	P-Value
Family History*										
Positive	1301 (17.0%)	**16 (29.6%)**			28 (23.5%)			7 (13.2%)		
Negative	6330 (83.0%)	**38 (70.4%)**	**2.06 (1.12–3.65)**	**0.022**	91 (76.5%)	1.51 (0.97–2.28)	0.070	46 (86.8%)	0.74 (0.60–1.54)	0.44
Bilateral Disease*										
Positive	220 (1.7%)	0 (0.0%)			11 (5.3%)			4 (4.7%)		
Negative	12 526 (98.3%)	83 (100.0%)	NA	0.089	198 (94.7%)	**3.27 (1.66–5.83)**	**0.0014**	81 (95.3%)	2.85 (0.86–6.91)	0.080
Age at Diagnosis*										
Before 50	3674 (28.4%)	25 (29.8%)	3.41 (1.72–7.23)		**76 (36.0%)**	**3.98 (2.62–6.21)**		27 (31.4%)	5.80 (2.67–14.5)	
50s	4952 (38.3%)	34 (40.5%)	3.44 (1.80–7.13)		**87 (41.2%)**	**3.37 (2.24–5.22)**		41 (47.7%)	6.54 (3.12–16.0)	
After 60	4305 (33.3%)	25 (29.8%)	2.91 (1.47–6.17)	0.66	**48 (22.7%)**	**2.12 (1.35–3.41)**	**1.2×10^{-5}**	18 (20.9%)	3.29 (1.43–8.47)	0.22
ER Status†										
Positive	7845 (83.6%)	**50 (92.6%)**	**3.19 (1.73–6.47)**	**1.0×10^{-4}**	**140 (91.5%)**	**3.42 (2.33–5.21)**	**1.5×10^{-11}**	**43 (79.6%)**	**4.32 (2.07–10.5)**	**2.7×10^{-5}**
Negative	1544 (16.4%)	4 (7.4%)	1.28 (0.36–3.76)	0.67	13 (8.5%)	1.59 (0.80–3.00)	0.18	**11 (20.4%)**	**5.58 (2.19–15.2)**	**3.6×10^{-4}**
PR Status†										
Positive	3026 (66.5%)	**16 (64%)**	**2.65 (1.24–5.87)**	**0.012**	**61 (81.3%)**	**3.87 (2.51–6.12)**	**4.3E-10**	**16 (57.1%)**	**4.16 (1.77–10.8)**	**9.0×10^{-4}**
Negative	1524 (33.5%)	**9 (36%)**	**2.96 (1.19–7.16)**	**0.021**	14 (18.7%)	1.75 (0.89–3.25)	0.100	**12 (42.9%)**	**6.21 (2.5–16.7)**	**1.0×10^{-4}**

*Case only analysis.
†Case-control analysis.

The relative risk associated with *CHEK2*-truncating variants declined with increasing age, with estimated ORs of 3.98 for diagnosis before age 50, 3.37 between ages 50 and 59 and 2.12 after age 60 ($P_{trend}=1.2\times10^{-5}$; table 2). For *ATM* and *PALB2* variant carriers, there was no clear evidence for an OR trend by age ($P_{trend}=0.66$ and 0.22, respectively).

Risk associations for subsets of missense variants

Thirty-eight missense variants had an ExAC carrier frequency >0.1% (25/299 in *ATM*, 4/77 in *CHEK2*, 7/125 in *PALB2* and 2/34 in *XRCC2*; see online supplementary table S4), and two were significantly associated with BC risk: *ATM* c.7390T>C (p.Cys2464Arg, rs55801750, OR(Arg/Cys)=0.37; 95% CI 0.19 to 0.73, $P_{trend}=0.0028$) and *XRCC2* c.563G>A (p.Arg188His, rs3218536, OR(His/Arg)=0.90; 95% CI 0.83 to 0.97, $P_{trend}=0.0080$) (see online supplementary table S7). However, neither of these associations was significant after adjusting for multiple testing. Of note, *CHEK2* c.470T>C (p.Ile157Thr, rs17879961) was found in 0.13% of subjects (20 cases, four controls; see online supplementary table S5) in this study. The relative risk estimate (OR(Thr/Ile)=2.10; 95% CI 0.72 to 6.14, p=0.17), although non-significant, is compatible with the 1.4-fold increased risk previously reported for the same variant in Finnish and Eastern European populations.[3 35]

We tested for risks associated with the aggregate of all rare missense variants in each gene, irrespective of position or predicted deleteriousness (figure 3 and table 3). We found some evidence of increased BC risk associated with the combined rare missense substitutions in *ATM* (OR=1.18; 95% CI 0.99 to 1.40, p=0.073), *CHEK2* (OR=1.36; 95% CI 0.99 to 1.87, p=0.066) and *PALB2* (OR=1.28; 95% CI 0.95 to 1.73, p=0.12), but not in *XRCC2*. Considered together, rare missense variants in *ATM*, *CHEK2* and *PALB2* were associated with an estimated OR=1.24; 95% CI 1.08 to 1.43, p=0.0025 (table 3). Variants localised within protein domains (as defined by UniProt or Pfam) of these three genes had a somewhat higher risk estimate (OR=1.45; 95% CI 1.17 to 1.80; figure 3 and table 3) than those outside of the annotated domains (OR=1.09; 95% CI 0.92 to 1.31; $P_{diff}=0.060$).

There was no evidence that risk was higher among variants predicted to be deleterious by CADD, PolyPhen2, SIFT or AlignGVGD, for any of the four genes, and no subset of variants stratified by these annotations was significantly associated with risk (see online supplementary table S8). Similarly, the risk estimate for the aggregate of rare variants in *ATM*, *CHEK2* and *PALB2* with deleterious functional predictions was not significantly higher than for predicted benign variants (three genes combined $P_{diff}=0.91$, 0.74, 0.71 and 0.76 for CADD, PolyPhen, SIFT and AlignGVGD, respectively; see online supplementary table S8).

Previous analyses have indicated that *ATM* missense variants within the FRAP-ATM-TRRAP (FAT) and phosphatidylinositol 3-kinase (PI3K) domains were specifically associated with increased BC risk.[5 6] We found evidence for increased risk for variants in both these domains (combined OR=1.71; 95% CI 1.12 to 2.61, p=0.015), but estimates did not differ significantly from those for the aggregate of all rare missense variants ($P_{diff}=0.31$). Of note, c.7271T>G (p.Val2424Gly, rs28904921), which has been implicated in a milder Ataxia-Telangiectasia disease phenotype and has previously been associated with a substantial BC risk,[5 7 8] occurred in eight cases and no controls in our study (figure 3 and see online supplementary table S5). After excluding this variant, the remaining rare missense substitutions in the FAT and PI3K domains in aggregate were still associated with BC risk (OR=1.59; 95% CI 1.04 to 2.43, p=0.040).

In *PALB2*, missense variants within the N-terminal BRCA1 binding domain were most strongly associated with risk (OR=1.76; 95% CI 1.03 to 2.98, p=0.047). This signal was driven by rare missense variants (n=29) between amino acids 70 and 300, and few of these were predicted by CADD, PolyPhen2, SIFT or AlignGVGD to have a deleterious effect on the protein (see online supplementary table S5).

Non-canonical splice variants and BC risk

We also examined associations for common variants in non-coding regions (see online supplementary table S7). Among these, only *CHEK2* c.320–5T>A (rs121908700; OR=13.9; 95% CI 1.89 to 101, $P_{trend}=6.7\times10^{-4}$) was significantly associated with risk after correction for multiple testing. This variant, in a non-canonical splice site, was predicted to reduce recognition of the normal splice acceptor site of exon 3 and introduce a new acceptor site three nucleotides upstream. At the protein level, this change would preserve the reading frame and cause the insertion of a valine residue. There was some suggestion of an association for the aggregate of other non-canonical splice variants in *CHEK2*, which were found in 12 cases and two controls (OR=2.52, 95% CI 0.56 to 11.3, p=0.26; online supplementary table S5).

DISCUSSION

This study, the largest experiment to date to systematically sequence the coding and exon-flanking regions of these genes in a population-based series of BC cases and controls, provides additional confirmation that protein-truncating mutations in *ATM*, *CHEK2* and *PALB2* are associated with increased BC risks. For *ATM* and *CHEK2*, the relative risks were higher for ER-positive than ER-negative disease, but we observed no differential effect by ER-status for *PALB2*. In contrast, *XRCC2*-truncating variants were not significantly associated with risk, but a twofold increased risk could not be excluded because these variants were very rare (13/18 575 samples; upper 95% confidence limit 4.19). These findings underscore the fact that, despite the large size of the study, the data are too sparse to accurately estimate risks for very rare variant classes and less common BC subtypes (eg, triple negative disease).

The BC risk estimates for all three of the associated genes were similar to estimates from smaller case–control studies and studies based on family-based designs (in the combined analysis of previous studies reported by Easton *et al*[4]: *PALB2*: meta analysis OR=5.3; 95% CI 3.0 to 9.4 vs table 1 OR=4.69; 95% CI 2.27 to 9.68; *ATM*: meta analysis OR=2.8; 95% CI 2.2 to 3.7 vs table 1 OR=3.26; 95% CI 1.82 to 6.46; and *CHEK2:* meta analysis OR=3.0; 95% CI 2.6 to 3.5 vs table 1 OR=3.11; 95% CI 2.15 to 4.69). Based on the estimated population frequencies and relative risks from this study, truncating variants in *ATM*, *CHEK2* and *PALB2* would explain approximately 4% of the twofold familial relative risk of BC and approximately 2% of all BC cases. While these estimates were derived from a study in the UK, the comparability of the combined frequency of truncating variants in our study with those from ExAC suggests that these estimates are likely to be broadly applicable to other European populations. Somewhat surprisingly, we observed no association between carrying a truncating *PALB2* variant and a BC family history, but this may reflect lack of power: there were only 53 carriers for whom family history data were available.

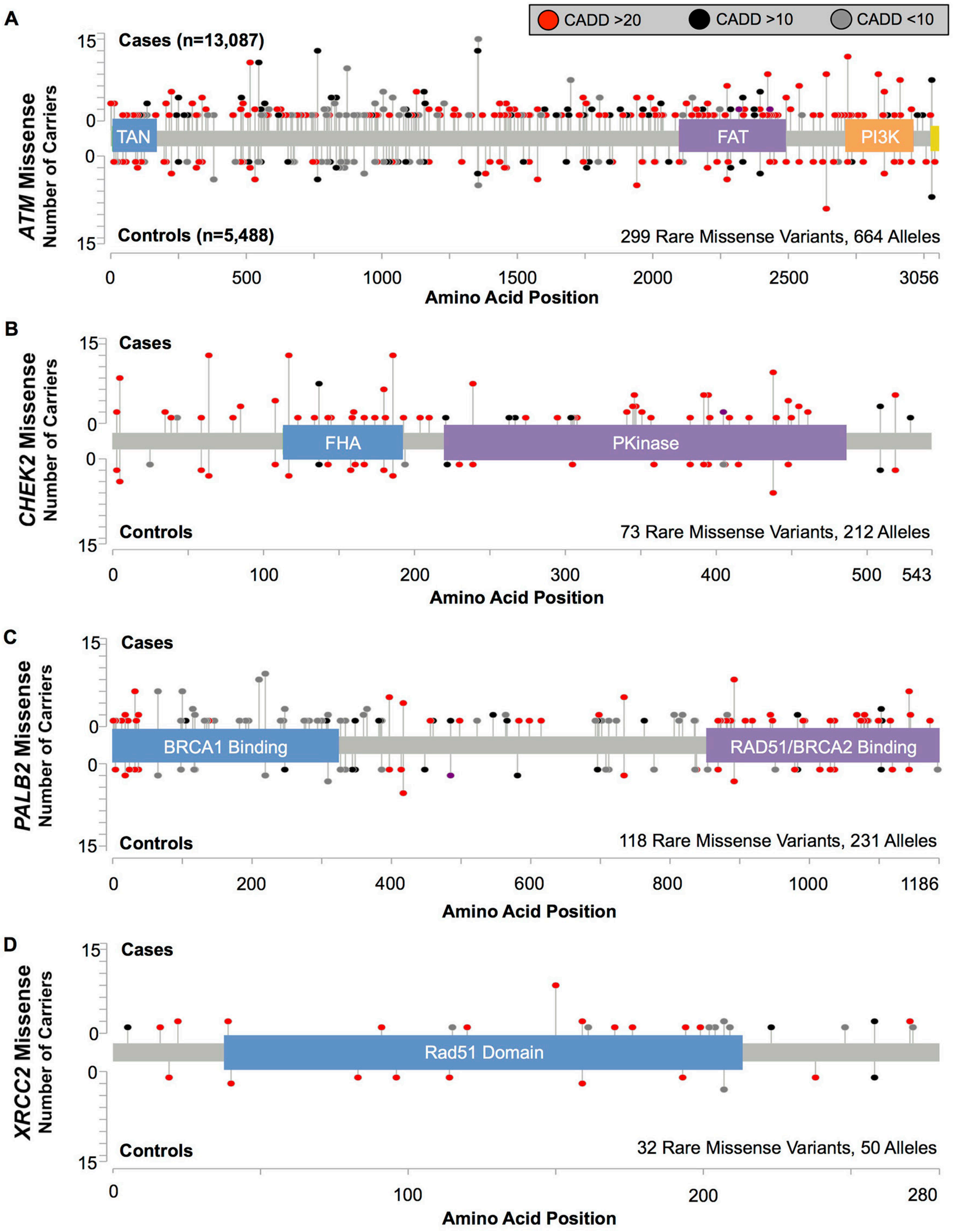

Figure 3 Position and frequency for rare missense variants in (A) *ATM* (B) *CHEK2* and (C) *PALB2*. ORs were calculated for all Pfam (ATM and CHEK2) and UniProt (PALB2) domains (table 2).

The vast majority of the truncating variants in this study were very rare: 117/119 were found in <0.1% of samples. The most notable exception was *CHEK2* c.1100delC, which was identified in approximately 1.1% of subjects and accounted for 81% of truncation carriers in this gene. Our risk estimate for this variant (OR=3.18; 95% CI to 2.01 to 4.92) was somewhat higher than two recent analyses (BCAC: OR=2.26; 95% CI 1.90 to 2.69[10]; Danish cohort: OR=2.08; 95% CI 1.51 to 2.85).[36] These differences might be explained by, for example, differences in the age distribution of the study subjects. The risk estimate for aggregated non-c.1100delC truncating variants in *CHEK2* was similar to that for c.1100delC, suggesting that results for this founder variant can reasonably be extrapolated to other truncating variants.

No individual missense variants showed evidence of association with BC risk at $p<0.001$, nor did we find strong evidence for the aggregate of rare missense variants in a single gene. There was, however, an association with BC risk for all rare,

Table 3 Rare missense variant associations with BC by domain

	Domain (AA Start-AA End)	Case carriers	Control carriers	OR (95% CI)	p Value	P_{diff}
ATM	All rare missense	489	175	1.18 (0.99 to 1.40)	0.073	
	In any domain	127	37	1.44 (1.00 to 2.08)	0.059	0.24
	Not in any domain	362	138	1.10 (0.90 to 1.35)	0.36	
	TAN (7-165)	17	10	0.71 (0.33 to 1.56)	0.52	
	FAT (2097–2488)	67	18	1.56 (0.93 to 2.63)	0.12	
	PI3K (2714–2961)	43	9	2.01 (0.98 to 4.12)	0.074	
	FAT + PI3K domains	**110**	**27**	**1.71 (1.12 to 2.61)**	**0.015**	
CHEK2	All rare missense	162	50	1.36 (0.99 to 1.87)	0.066	
	In any domain	**115**	**32**	**1.51 (1.02 to 2.24)**	**0.047**	0.45
	Not in any domain	47	18	1.10 (0.64 to 1.89)	0.85	
	FHA (113-191)	47	14	1.41 (0.78 to 2.56)	0.32	
	Pkinase (220-486)	68	18	1.59 (0.94 to 2.67)	0.10	
PALB2	All rare missense	174	57	1.28 (0.95 to 1.73)	0.12	
	In any domain	132	40	1.39 (0.97 to 1.98)	0.083	0.50
	Not in any domain	42	17	1.04 (0.59 to 1.82)	1.00	
	BRCA1 binding (1-319)	**71**	**17**	**1.76 (1.03 to 2.98)**	**0.047**	
	ChAM (395-466)	11	8	0.58 (0.23 to 1.43)	0.34	
	RAD51/BRCA2 binding (853–1186)	50	15	1.40 (0.79 to 2.49)	0.31	
	Both binding domains	**121**	**32**	**1.59 (1.08 to 2.35)**	**0.024**	
Combined	**All rare missense**	**825**	**282**	**1.24 (1.08 to 1.43)**	**0.0025**	
	In any domain	**374**	**109**	**1.45 (1.17 to 1.80)**	**7.3x10^{-4}**	0.060
	Not in any domain	451	173	1.09 (0.92 to 1.31)	0.33	

Domain boundaries were determined from Pfam (*ATM* and *CHEK2*) or UniProt (*PALB2*).
BC, breast cancer.

non-synonymous substitutions combined across *ATM*, *CHEK2* and *PALB2*. This risk could be mediated by a small subset of variants conferring a high risk, or a larger subset of variants associated with a lower risk. We observed little evidence of association by predicted effect severity, but there was, however, some suggestion that rare missense variants within functional domains may contribute to BC risk.

CONCLUSIONS

This report, based on a large population-based study, provides relative risk estimates associated with truncating variants in *ATM*, *CHEK2* and *PALB2*. Our results confirm that risk estimates for *ATM* and *CHEK2* gene variants are similar and firmly within the twofold to fourfold range. *PALB2* protein-truncating variants conferred a somewhat higher risk, supporting previous suggestions that specific management may be justified in *PALB2* carriers.[11] The absolute risks and age-specific penetrance in carriers will depend on additional influences, including common susceptibility variants, lifestyle risk factors and family history—considerations that can be built into more comprehensive risk prediction models.[37] Clinically useful risk estimates for rarer disease subtypes and for missense variants will require studies that are substantially larger than the current experiment; these are becoming possible through large consortia and technological advances.

Acknowledgements BD is a student in the NIH-Cambridge Scholars Program. BD and EAO were supported by the Intramural Research Program of the National Human Genome Research Institute. SEARCH is funded by a programme grant from Cancer Research UK (C490/A10124) and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. Targeted sequencing in SEARCH was supported by Cancer Research UK grants C1287/A16563 to DFE and C8197/A16565 to AMD. We thank Maaike Vreeswijk for her help with the *in silico* predictions for the *CHEK2* splice variant rs121908700.

Collaborators Maaike Vreeswijk.

Contributors BD designed the targeted sequencing panel, performed bioinformatics and statistical analyses, and drafted the manuscript. JA performed bioinformatics analysis. CL and AMD coordinated the targeted sequencing and AMD also assisted with drafting the manuscript. KAP and SA contributed Sanger validation of variant calls. MS, MKB and QW provided data management support. CB, DMC and JB prepared targeted sequencing libraries. RL provided samples and phenotypic data from EPIC. EAO assisted with drafting the manuscript. PDPP coordinated SEARCH. DFE conceived and designed the study, performed statistical analysis and assisted with drafting the manuscript. All authors read and approved the final manuscript.

Funding Cancer Research UK (C490/A10124, C1287/A16563, C8197/A16565), NIH, NHGRI

Competing interests No, there are no competing interests.

Patient consent Obtained.

Ethics approval Cambridgeshire Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and Major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.

2 Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.

3 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 LF, 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 MR, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken Z, Aaltonen K, Heikkinen T, Broeks A, Veer LJ, van der Schoot CE, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Zamora MP, Perez JI, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MW, 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 AM, 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 VM, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dörk T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo KY, Noh DY, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng CC, Van Den Berg D, Stram DO, Shu XO, Lu W, Gao YT, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng JH, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Shen CY, Hsiung CN, Wu PE, Ding SL, 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 PD, Chenevix-Trench G, Dunning AM, Benitez J, Easton DF; Breast and Ovarian Cancer Susceptibility CollaborationHereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON)kConFab InvestigatorsAustralian Ovarian Cancer Study GroupGENICA (Gene Environment Interaction and Breast Cancer in Germany) Network. Large-scale genotyping identifies 41 new loci associated with breast Cancer risk. *Nat Genet* 2013;45:353–61.

4 Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DG, Chenevix-Trench G, 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.

5 Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, Lesueur F, Byrnes GB, Chuang SC, 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 C-TG. Australian Cancer StudyBreast Cancer Family Registries (BCFR)Kathleen Cuningham Foundation Consortium for Research into Familial Aspects of Breast Cancer (kConFab). Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast Cancer. *Am J Hum Genet 2009*;85:427–46.

6 Young EL, Feng BJ, Stark AW, Damiola F, Durand G, Forey N, Francy TC, Gammon A, Kohlmann WK, Kaphingst KA, McKay-Chopin S, Nguyen-Dumont T, Oliver J, Paquette AM, Pertesi M, Robinot N, Rosenthal JS, Vallee M, Voegele C, Hopper JL, Southey MC, Andrulis IL, John EM, Hashibe M, Gertz J, Le Calvez-Kelm F, Lesueur F, Goldgar DE, Tavtigian SV. Breast Cancer Family Registry. Multigene testing of moderate-risk genes: be mindful of the missense. *J Med Genet* 2016;53.

7 Stankovic T, Kidd AM, Sutcliffe A, McGuire GM, Robinson P, Weber P, Bedenham T, Bradwell AR, Easton DF, Lennox GG, Haites N, Byrd PJ, Taylor AM. ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of Mutant ATM and the risk of leukemia, lymphoma, and breast Cancer. *Am J Hum Genet* 1998;62:334–45.

8 Goldgar DE, Healey S, Dowty JG, Da Silva L, Chen X, Spurdle AB, Terry MB, Daly MJ, Buys SM, Southey MC, Andrulis I, John EM, Khanna KK, Hopper JL, Oefner PJ, Lakhani S, Chenevix-Trench G; BCFRkConFab. Rare variants in the ATM gene and risk of breast Cancer. *Breast Cancer Res* 2011;13:R73.

9 Bernstein JL, Teraoka S, Southey MC, Jenkins MA, Andrulis IL, Knight JA, John EM, Lapinski R, Wolitzer AL, Whittemore AS, West D, Seminara D, Olson ER, Spurdle AB, Chenevix-Trench G, Giles GG, Hopper JL, Concannon P. Population-based estimates of breast Cancer risks associated with ATM gene variants c.7271T>G and c.1066-6T>G (IVS10-6T>G) from the breast Cancer family Registry. *Hum Mutat* 2006;27:1122–8.

10 Schmidt MK, Hogervorst F, van Hien R, Cornelissen S, Broeks A, Adank MA, Meijers H, Waisfisz Q, Hollestelle A, Schutte M, van den Ouweland A, Hooning M, Andrulis IL, Anton-Culver H, Antonenkova NN, Antoniou AC, Arndt V, Bermisheva M, Bogdanova NV, Bolla MK, Brauch H, Brenner H, Brüning T, Burwinkel B, Chang-Claude J, Chenevix-Trench G, Couch FJ, Cox A, Cross SS, Czene K, Dunning AM, Fasching PA, Figueroa J, Fletcher O, Flyger H, Galle E, García-Closas M, Giles GG, Haeberle L, Hall P, Hillemanns P, Hopper JL, Jakubowska A, John EM, Jones M, Khusnutdinova E, Knight JA, Kosma VM, Kristensen V, Lee A, Lindblom A, Lubinski J, Mannermaa A, Margolin S, Meindl A, Milne RL, Muranen TA, Newcomb PA, Offit K, Park-Simon TW, Peto J, Pharoah PD, Robson M, Rudolph A, Sawyer EJ, Schmutzler RK, Seynaeve C, Soens J, Southey MC, Spurdle AB, Surowy H, Swerdlow A, Tollenaar RA, Tomlinson I, Trentham-Dietz A, Vachon C, Wang Q, Whittemore AS, Ziogas A, van der Kolk L, Nevanlinna H, Dörk T, Bojesen S, Easton DF. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *J Clin Oncol* 2016;34:2750–60.

11 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 KB, Yannoukakos D, Concannon P, Bernstein JL, James PA, Easton DF, Goldgar DE, Hopper JL, Rahman N, Peterlongo P, Nevanlinna H, King MC, 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.

12 Erkko H, Xia B, Nikkilä J, Schleutker J, Syrjäkoski K, Mannermaa A, Kallioniemi A, Pylkäs K, Karppinen SM, Rapakko K, Miron A, Sheng Q, Li G, Mattila H, Bell DW, Haber DA, Grip M, Reiman M, Jukkola-Vuorinen A, Mustonen A, Kere J, Aaltonen LA, Kosma VM, Kataja V, Soini Y, Drapkin RI, Livingston DM, Winqvist R. A recurrent mutation in PALB2 in finnish Cancer families. *Nature* 2007;446:316–9.

13 Erkko H, Dowty JG, Nikkilä J, Syrjäkoski K, Mannermaa A, Pylkäs K, Southey MC, Holli K, Kallioniemi A, Jukkola-Vuorinen A, Kataja V, Kosma VM, Xia B, Livingston DM, Winqvist R, Hopper JL. Penetrance analysis of the PALB2 c.1592delt founder mutation. *Clin Cancer Res* 2008;14:4667–71.

14 Southey MC, Teo ZL, Dowty JG, Odefrey FA, Park DJ, Tischkowitz M, Sabbaghian N, Apicella C, Byrnes GB, Winship I, Baglietto L, Giles GG, Goldgar DE, Foulkes WD, Hopper JL; kConFab for the Beast Cancer Family Registry. A PALB2 mutation associated with high risk of breast Cancer. *Breast Cancer Res* 2010;12:R109.

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. Breast Cancer Family RegistryKathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer. Rare mutations in XRCC2 increase the risk of breast Cancer. *Am J Hum Genet* 2012;90:734–9.

16 Dunning AM, Healey CS, Pharoah PD, Teare MD, Ponder BA, Easton DF. A systematic review of genetic polymorphisms and breast Cancer risk. *Cancer Epidemiol Biomarkers Prev* 1999;8:843–54.

17 Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, Wareham N. EPIC-Norfolk: study design and characteristics of the cohort. european prospective investigation of Cancer. *Br J Cancer* 1999;80 Suppl 1(Suppl 1):95–103.

18 Kataoka M, Antoniou A, Warren R, Leyland J, Brown J, Audley T, Easton D. Genetic models for the familial aggregation of mammographic breast density. *Cancer Epidemiol Biomarkers Prev* 2009;18:1277–84.

19 Li H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* 2012;28:1838–44.

20 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.

21 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.

22 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K V, Altshuler D, Gabriel S, DePristo MA. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current protocols in bioinformatics*: John & Wiley Sons,Inc, 2013. 43:11.10.1-11.10.33.

23 McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics* 2010;26:2069–70.

24 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.

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 Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001;11:863–74.

27 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.

28 Zhang MQ. Statistical features of human exons and their flanking regions. *Hum Mol Genet* 1998;7:919–32.

29 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–94.

30 Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311–23.

31 Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001;29:1185–90.

32 Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37:e67.

33 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan

LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.

34 Team R. R: a language and environment for Statistical Computing. R Found Stat. Comput. 2016.

35 Kilpivaara O, Vahteristo P, Falck J, Syrjäkoski K, Eerola H, Easton D, Bartkova J, Lukas J, Heikkilä P, Aittomäki K, Holli K, Blomqvist C, Kallioniemi OP, Bartek J, Nevanlinna H. CHEK2 variant I157T may be associated with increased breast Cancer risk. *Int J Cancer* 2004;111:543–7.

36 Näslund-Koch C, Nordestgaard BG, Bojesen SE. Increased risk for other cancers in addition to breast Cancer for CHEK2*1100delc heterozygotes estimated from the Copenhagen General Population Study. *J Clin Oncol* 2016;34:1208–16.

37 Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC; Consortium of Investigators of Modifiers of BRCA1/2Breast Cancer Association Consortium. BOADICEA breast Cancer risk prediction model: updates to Cancer incidences, tumour pathology and web interface. *Br J Cancer* 2014;110:535–45.