Published in final edited form as:

Biomark Med. 2014 ; 8(4): 589–603. doi:10.2217/bmm.13.143.

Growing recognition of the role for rare missense substitutions in breast cancer susceptibility

Sean V Tavtigian^{*,1} and Georgia Chenevix-Trench²

¹Huntsman Cancer Institute and Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, UT 84112, USA

²QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia

Abstract

Most cancer susceptibility genes function as tumor suppressors; accordingly, the focus of mutation screening in breast cancer families has been to identify protein-truncating mutations. However, it is now clear that, for some breast cancer susceptibility genes, a significant proportion of the burden of disease comes from rare missense substitutions. Among genes that have been extensively evaluated, *BRCA1*, *BRCA2*, *PALB2* and *BRIP1* stand as examples where the majority of mutations lead to protein truncation; *TP53* provides a counter example, where the majority of pathogenic variants are missense substitutions. In *ATM* and *CHEK2*, missense substitutions are probably equally or more important in terms of their frequency and attributable risk. Therefore, ongoing efforts to identify new susceptibility genes should not ignore missense variation.

Keywords

ATM; *BRCA1*; *BRCA2*; breast cancer; *CHEK2*; genetic susceptibility; homologous recombination repair; missense substitution; rare variant; unclassified variant

Evidence for a genetic component to breast cancer risk dates back at least to pedigree studies carried out in the 1860s [1]. Since that time, pedigree analyses have been complemented with linked genealogy and cancer registry studies, twin studies, and segregation analyses to estimate aspects of breast cancer risk, as well as a panoply of molecular genetic approaches to gene identification. In a descriptive sense, these studies have delivered two important pieces of information about breast cancer susceptibility: first, approximately 25% of the difference in risk between individuals is attributable to genetic susceptibility [2]; this

© 2014 Future Medicine Ltd

For reprint orders, please contact: reprints@futuremedicine.com

*Author for correspondence: sean.tavtigian@hci.utah.edu.

Disclaimer

The content of this review does not necessarily reflect the views or policies of the National Cancer Institute (NCI), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

measurement places a lower limit on the genetic population attributable fraction of breast cancer, but does not provide an upper limit. Second, the ratio of risk experienced by first-degree relatives of breast cancer cases compared with unselected cases (familial relative risk [FRR]) is approximately 2.0:1 to 2.5:1. Considerable evidence, such as the persistence of elevated FRRs to second-, third- and fourth-degree relatives of index cases, and particularly high FRRs in the close relatives of early-onset cases, is consistent with the idea that the majority of this FRR is due to inherited susceptibility [3–5].

One aim of breast cancer gene discovery and gene characterization studies is to explain the FRR of breast cancer in terms of genes and pathogenic alleles (with associated odds ratios and frequencies) of those genes. Along these lines, we generally recognize three classes of susceptibility genes. The pathogenic alleles of high-risk genes confer odds ratios (ORs) of 5–10 or higher, but the summed frequency of pathogenic alleles in any such gene is not usually more than approximately 0.2%. The pathogenic alleles of intermediate-risk genes confer ORs of 2.0–5.0, and for some of these genes the summed frequency of the pathogenic alleles approaches 1.0%. Common alleles of modest-risk single nucleotide polymorphisms confer ORs of 1.1–2.0 (mostly 1.1–1.3), and their frequencies can be as high as 50% [6]. In the case of breast cancer, *BRCA1*, *CHEK2* and *FGFR2* are archetypical examples of high-risk, intermediate-risk and modest-risk genes, respectively.

BRCA1 & BRCA1

In 1990, Mary-Claire King's laboratory at the University of California, Berkeley (CA, USA) described genetic linkage of susceptibility to early-onset breast cancer at chromosome 17q [7]. That publication then set off a world-wide scramble to find the underlying susceptibility gene, culminating with the discovery of *BRCA1* in 1994 [8]. From the initial wave of *BRCA1* validation papers [9–11], it became obvious that the majority of the pathogenic variants in this gene cause protein truncation. It is therefore now widely accepted that in *BRCA1* protein-truncating variants (except those in the last coding exon) are pathogenic unless there is contrary evidence. In addition, mutations that delete the gene's promoter result in little or no transcript from the affected allele, and are also unambiguously classified as pathogenic.

Missense substitutions, however, are different: changing just one amino acid in a protein's sequence, they may result in a variant protein that functions entirely normally, is severely defective or anything in-between. Although there are now clinically applicable methods for classifying missense substitutions [12], they are too complex to execute for each new missense substitution as it is identified. Consequently, in clinical mutation screening, newly observed missense substitutions are generally reported as unclassified variants (UVs) or variants of unknown significance (VUSs) where they pose a considerable clinical dilemma.

A second high-risk breast cancer susceptibility gene, *BRCA2*, was identified in late 1995 and its complete sequence published in early 1996 [13,14]. It was immediately clear that the spectrum of pathogenic variants in *BRCA2* is also dominated by protein truncating variants. In fact, it was not until 2004 that a *BRCA2* missense substitution (p.Asp2723His) was shown to be pathogenic [15].

BRCA1 and *BRCA2* both encode relatively large proteins (1863 and 3418 amino acids, respectively; Figure 1) that play roles in DNA double-strand break repair by homologous recombination (HRR; Figure 2) and function as tumor suppressors. Owing to their size, complete mutation screening of these genes using 1990s technologies was a nontrivial exercise and initially most of the methods used were severely biased toward detecting protein-truncating mutations. However, in a relatively early review of methods used to mutation screen these genes, Neuhausen and Ostrander noted that many missense substitutions reported in the genes were considered UVs, and that these would be difficult to reclassify into clinically actionable categories [16].

At the American Society of Human Genetics annual meeting held in Philadelphia (PA, USA) in October 2000, the Breast Cancer Information Core (BIC [24], which maintains a repository for information regarding mutations and polymorphisms in *BRCA1* and *BRCA2*) held a satellite meeting focused on development of methods to classify these missense UVs. The main outcome of this meeting was recognition that an effective method of classification would have to be able to combine several different kinds of data into a single statistical framework. A practical way to do this was to take a Bayesian statistical approach to the problem, which implied that the analysis of each individual type of data to be included in the framework would have to output either a probability in favor of pathogenicity or else a likelihood ratio in favor of pathogenicity.

The current incarnation of the integrated evaluation, which has been reviewed recently [25–27], uses a Bayesian framework to combine a sequence analysis-based prior probability in favor of pathogenicity with analyses of four types of observational data: cosegregation of UVs with cancer phenotype in pedigree; summary personal and family cancer history; co-occurrence with known pathogenic mutations; and tumor immunohistochemistry and histological grade. The output of the integrated evaluation is a posterior probability in favor of pathogenicity, which is a continuous variable that can take any value between 0.00 and 1.00. The posterior probability is supplemented by a five-category classification table that provides a clinically useful translation from the continuous variable output of the integrated evaluation to a qualitative system meant for use by clinical cancer geneticists and genetic counselors (Table 1) [12].

Two elements of the integrated evaluation are particularly relevant to this review: the summary personal and family cancer history because it played an important part in calibrating the sequence analysis based prior probability, and the sequence analysis itself because it has been important in assessing the contribution of rare missense substitutions to the fraction of breast cancers attributable to several susceptibility genes.

The summary personal and family cancer history likelihood ratio was first described by Goldgar *et al.* in 2004 and then included in the integrated evaluation by Easton *et al.* in 2007 [15,28]. The method was made possible because Myriad Genetic Laboratories had collected simple personal and family histories—basically age at diagnosis and ethnicity of the index cases, in addition to the number of affected first- and second-degree relatives—from thousands of the patients that they had tested. Setting aside details of the logistic regressions used to establish the method, the basic idea was that the summary family histories in the

database fall in a distribution from innocuous to severe. The family histories from subjects who were found to carry a clearly pathogenic mutation in *BRCA1* or *BRCA2* fall towards the severe end of the spectrum, whereas those from subjects who had an innocuous mutation screening result fell towards the innocuous end of the spectrum. Thus, given the family histories from several families that carried a specific UV (but no clearly pathogenic variant) it became possible to calculate a likelihood ratio in favor of pathogenicity for that UV. Perhaps more importantly, it also became possible to calculate a likelihood ratio in favor of pathogenicity for any defined set of sequence variants; for example, sets of missense substitutions grouped by *in silico* missense substitution analysis algorithms.

With respect to *in silico* analyses of missense substitutions, it has been known since the 1960s that conservation of specific sequences during evolution is indicative of functional constraint and that pathogenic missense substitutions tend to fall at evolutionarily-conserved residues [29,30]. In 1996, Szabo *et al.* determined the sequences of mouse and dog *BRCA1* and used them to make the first full-length *BRCA1* protein multiple sequence alignments [31]. Their analysis demonstrated that neutral missense substitutions in this gene tend to fall at positions that are variable across species, whereas then-UV and pathogenic substitutions were over-represented at conserved *BRCA1* residues. An important next step in this area came when Abkevich *et al.* used a deeper *BRCA1* alignment to quantitatively measure the degree to which *BRCA1* missense substitutions carried by breast cancer cases are over-represented at positions that are strongly conserved or invariant during vertebrate evolution [32]. This analysis indicated that approximately 7% of the load of pathogenic sequence variants in *BRCA1* is attributable to missense substitutions, a number that is entirely compatible with more recent analyses of larger data sets [28,33].

In a further elaboration of *in silico* sequence analysis methods, Tavtigian *et al.* merged measurement of the evolutionarily tolerated range of sequence variation at the site of a missense substitution and the physicochemical severity of a missense substitution with respect to that range of variation into a single graded variable, output by the program Align-GVGD [33,34,35], that provides a ranking of missense substitutions from evolutionarily most likely (C0) to least likely (C65). They then used the summary personal and family history method to calibrate the Align-GVGD output by measuring a probability in favor of pathogenicity for the grades of that variable, noting that the probability increased monotonically across the increasingly severe grades of the variable [33]. This analysis now provides a prior probability in favor of pathogenicity for rare *BRCA1* and *BRCA2* missense substitutions (it should be noted that other popular missense substitution analysis programs such as SIFT [36] and PolyPhen-2.1 [37] [38,39] output a variable that could be used similarly if properly calibrated).

Classification of missense variants in *BRCA1* and *BRCA2* has benefited greatly from a data set that was based on full sequence *BRCA1* and *BRCA2* testing of initially 20,000, and then 68,000, patients conducted at Myriad Genetics Laboratories [28,32–33]. This data set included 596 and 1131 distinct missense substitutions in *BRCA1* and *BRCA2*, respectively, that individually had frequencies of <1% in the tested population. From heterogeneity estimates, it appears that 9–12% of these substitutions are pathogenic [28], with an estimate that these substitutions are responsible for approximately 7% of the risk of breast cancer

attributable to *BRCA1* [32] (there is as yet no corresponding estimate for *BRCA2*). Despite the availability of this large data set, one of the biggest challenges in clinical cancer genetics remains the interpretation of UVs in *BRCA1* and *BRCA2*, for which an international consortium – the Evidence-based Network for the Interpretation of Germline Mutant Alleles – has been established [40]. As testing of other breast cancer susceptibility genes becomes more common, in both research and clinical settings, the efforts of Evidence-based Network for the Interpretation of Germline Mutant Alleles and similar initiatives should help inform the classification of the missense variants that are found, but this will not obviate the need for large sequencing and genotyping data sets.

TP53

The p53 tumor suppressor protein is encoded by the *TP53* gene. Germline mutations in *TP53* cause Li–Fraumeni syndrome, a severe multiorgan familial cancer syndrome characterized by early-onset breast cancer, soft tissue sarcoma, osteosarcoma, brain and other cancers [41]. In 1990, two groups used a candidate gene mutation screening strategy to demonstrate that *TP53* is the principal Li–Fraumeni syndrome predisposition gene [42,43] and found that in all six Li–Fraumeni families the pathogenic mutations were missense substitutions.

The p53 protein had originally been identified by coimmunoprecipitation with SV40 tumor antigens from SV40-transformed mouse tumor cells [44]. Results from early ras-plus-p53 cotransformation experiments were interpreted as evidence that *TP53* encodes an oncogene [45–47]. However, shortly before the germline mutations were found in Li–Fraumeni syndrome, it became apparent that wild-type p53 functions as a tumor suppressor with the unusual feature that specific missense substitutions and in-frame insertion–deletion mutations in the gene result in proteins with transdominant or dominantnegative characteristics – that is, the mutant protein can block function of the wild-type protein, reducing or eliminating its tumor suppressor activity [48,49]. Classification of missense variants in *TP53* has been greatly assisted by the fact that it is the most frequently somatically mutated gene in tumors, functional assays are well established [50] and a large database of genetic variation has been established at the International Agency for Research into Cancer [51]. A 2012 summary of this database reported that, between 1990 and 2009, 531 independent families or individuals with germline mutations in *TP53* were described in peer-reviewed literature, and approximately 73% of the pathogenic germline *TP53* mutations reported from those families or individuals are missense substitutions [52]. *TP53* therefore stands as a counter example to *BRCA1* and *BRCA2* in which missense substitutions stand as a much more important component of genetic risk than protein-truncating mutations.

Intermediate-risk genes

BRCA1, *BRCA2* and *TP53* are all high-risk susceptibility genes. The lifetime risks of breast and other cancers conferred by inherited pathogenic alleles of these genes are high enough to create clear patterns of segregation with cancer in a substantial fraction of pedigrees that harbor such alleles. The magnitude of risk conferred by the alleles made it possible to assign

susceptibility gene status to these genes based on results from a relatively limited number of pedigrees, and also made it possible to detect evidence of risk in defined sets of sequence variants, such as missense substitutions falling at invariant positions in a *BRCA1* protein multiple sequence alignment, using relatively crude analytic tools. By contrast, the combination of risk and frequency that characterizes intermediate-risk susceptibility genes, such as *ATM*, *CHEK2*, *BRIP1* and *PALB2*, creates a more severe analytic challenge. For the most part, pedigree-based studies have given way to population-based studies to meet this challenge. In particular, case-control mutation screening (CCMS), a study design in which all of the coding exons and proximal splice junction regions of genes of interest are mutation screened in both cases and controls, has become a tool of choice [53]. Moreover, we would argue from *post hoc* power calculations based on our CCMS of *CHEK2* that studies examining fewer than 1000 cases and 1000 controls are underpowered [54]; accordingly, we will limit most of this discussion to genes for which at least one such study is available.

ATM

Ataxia telangiectasia (A-T) is an autosomal recessive disorder characterized by a variable combination of neurological findings, immunodeficiency, radiosensitivity and cancer susceptibility. Almost a decade before the underlying susceptibility gene was discovered, studies of the family histories of A-T patients revealed that heterozygote carriers are at increased risk of a number of types of cancer, most notably breast cancer in female carriers [55]. After *ATM* was cloned, elucidating the molecular genetic basis for breast cancer sensitivity in heterozygote carriers became an important research topic. In the first large-scale study reported, FitzGerald *et al.* used the protein truncation test to mutation screen the gene in 401 breast cancer cases and 201 controls [56]. Identifying protein-truncating mutations in two cases and two controls, their results provided molecular evidence against a role for heterozygous *ATM* mutations in breast cancer susceptibility. Notably, because this study relied on the protein truncation test for mutation detection, it did not evaluate missense substitutions in the gene.

Between 1997 and 2006, a number of other groups reported results from case-only, case-control and control-only mutation screening studies of *ATM* (all are included in the Tavtigian *et al.* 2009 meta-analysis [57]). Some found evidence that heterozygous protein-truncating mutations confer increased risk of breast cancer, but others found that missense substitutions confer increased risk, and others found little evidence of risk. In 2006, Renwick *et al.* reported the largest *ATM* CCMS study completed to that date. Their study, which included 443 cases and 521 controls, had three important features: the cases were all *BRCA1*- and *BRCA2*-negative familial index cases, which increased statistical power to detect association as compared with a study based on population sampling of cases; mutation screening relied on methods that had reasonable sensitivity for both protein-truncating and missense substitutions; and their supplementary data included case and control counts for all of the individual rare missense substitutions observed during the study. In their analysis, Renwick *et al.* pooled protein-truncating variants (under the assumption that these would be pathogenic if the gene was indeed a breast cancer susceptibility gene) with missense substitutions that were already known to be pathogenic for A-T; in all, they observed 12 variants in the cases compared with two in the controls, resulting in a p-value of

0.0047 against the null hypothesis of no association [58]. This provided solid molecular genetic evidence that *ATM* is a breast cancer susceptibility gene.

In 2009, Tavtigian *et al.* reported a meta-analysis of all published *ATM* breast cancer CCMS studies, as well as including new data from mutation screening of 987 cases and 1021 controls, bringing the total number of subjects to 2531 cases and 2245 controls; the resulting study was large enough to support analysis of protein-truncating variants separately from rare missense substitutions. Rare missense substitutions were included in the study's statistical analyses by defining a graded variable that can be regarded as missense substitution severity and which was based on the same Align-GVGD analysis that is applied to unclassified missense substitutions in *BRCA1* and *BRCA2*. Three key results of the study were: protein-truncating variants were found to be pathogenic; independently, rare missense substitutions were also found to be pathogenic, with higher odds ratios estimated for the more severe grades of missense substitutions; and almost all of the evidence that some of the rare missense substitutions in the gene are pathogenic could be attributed to rare missense substitutions that mapped to *ATM*'s key FAT, protein kinase, and FATC functional domains, all of which fall in the carboxy-terminal third of the protein [57]. Finally, there was evidence that the most severe missense substitutions falling in the FAT, kinase, and FATC domains – those scored as C55 or C65 by Align-GVGD – actually confer greater risk of breast cancer than the protein-truncating variants do.

Shortly thereafter, Bernstein *et al.* reported an *ATM* mutation screening analysis that compared rare variant profiles in 1397 unilateral breast cancer cases to 708 bilateral breast cancer cases, and assessed radiation exposure data from their radiotherapy as a potential effects modifier [59]. Their study did not detect significant main effects from either protein-truncating variants or rare missense substitutions. However, when they stratified the mutation screening data on the basis of radiotherapy exposure, they found that the patients who carried a rare missense substitution and received radio-therapy had greater risk of a developing a contralateral tumor than did those who did not receive radiotherapy. Moreover, when the rare missense substitutions were stratified for predicted severity using the missense substitution analysis program SIFT [38], the carriers of likely pathogenic rare missense substitutions who received radiotherapy were found to be at elevated risk of a contralateral tumor. By contrast, there was no significant evidence for increased risk of contralateral tumors among carriers or protein-truncating variants in *ATM*, with or without stratification on radiotherapy dose.

To better define the main effect risks of breast cancer conferred by various classes of rare *ATM* sequence variants, Goldgar *et al.* used genotyping with a selected set of *ATM* protein-truncating variants and Align-GVGD C55–C65 missense substitutions from the gene's key FAT, kinase and FATC domains to identify 27 carriers from breast cancer case and case–control collections that had germline DNA from additional family members, both affected and unaffected [60]. The additional family members were then genotyped for the relevant *ATM* variant; their age, breast cancer status and genotype were then used to calculate relative risk and age-specific risk of breast cancer (penetrance) as a function of genotype. Grouped together, the truncating variants and key domain C55–C65 missense substitutions appeared to increase risk by 6.9-fold ($p = 0.00008$). When this relative risk was converted to

penetrance, cumulative risk of female breast cancer by the age of 80 years was estimated to be 60%, which was 75% of the risk conferred by *BRCA2* mutations. For the 11 families with a protein-truncating variant, estimated penetrance was 30% of the risk conferred by *BRCA2* mutations, whereas for the 16 families with a key domain C55–C65 missense substitution, estimated penetrance was 110% of the risk conferred by *BRCA2* mutations. One caveat to the analysis of missense substitutions in this study was that 15 of the 16 families with a key domain C55–C65 missense substitution actually carried the same substitution – p.Val2424Gly – which was originally linked to breast cancer risk in a Scottish pedigree [61] and has since been shown to have dominant-negative activity [62]; consequently, it is not clear how well these results will generalize to other likely pathogenic missense substitutions falling in the key functional domains of ATM.

Nonetheless, taken together, these last three studies paint a reasonably consistent picture. *ATM* is clearly a breast cancer susceptibility gene [57,59–60]. Protein-truncating variants and a subset of rare missense substitutions in the gene are probably both pathogenic for breast cancer. There is also a reasonable probability that the risk conferred by a subset of the rare missense substitutions is actually higher than risk conferred by truncating variants. Moreover, each of these three studies observed a numerically higher frequency of their likely pathogenic class of rare missense substitutions than protein-truncating variants among their breast cancer cases, raising the possibility that rare *ATM* missense substitutions are responsible for at least as large fraction, if not a larger, of the breast cancer risk attributable to this gene than are protein-truncating mutations.

CHEK2

CHEK2 was identified as the human ortholog of *S. cerevisiae* Rad53 in 1998 [63,64]. The gene encodes a protein kinase that sits downstream of the MRE11A/RAD50/NBN complex and ATM in the signaling cascade and transduces information about DNA damage to cell cycle and apoptosis regulators. By the time that the mammalian *CHEK2* gene was identified, it was clear from yeast studies that loss of Rad53 function results in a notable defect in cell cycle arrest in response to DNA damage [65,66], and so Bell *et al.* asked whether some *TP53* wild-type Li–Fraumeni syndrome and Li–Fraumeni-like pedigrees might be explained by germline mutations in either *CHEK1* or *CHEK2* [67]. Mutation screening of four unexplained Li–Fraumeni syndrome cases identified one that harbored the *CHEK2* protein-truncating variant c.1100delC, and screening of 14 Li–Fraumeni-like pedigrees revealed an additional protein-truncating variant, c.1422delT, as well as a rare missense substitution in the functionally important Forkhead domain, p.Ile157Thr.

While subsequent work has not confirmed *CHEK2* as a Li–Fraumeni syndrome susceptibility gene, germline variants in the gene certainly play a role in susceptibility to several cancers, most notably breast cancer. Two large case–control studies of the c.1100delC variant demonstrated beyond any reasonable doubt that *CHEK2* is an intermediate-risk breast cancer gene and that this particular protein-truncating variant confers about a 2.3-fold increased risk of breast cancer [68,69]. Similarly, case–control genotyping studies that either included or focused on the missense substitution p.Ile157 Thr [70,71] concluded that that variant confers about a 1.4-fold increased risk of breast cancer.

Demonstrating that *CHEK2* is an intermediate-risk breast cancer susceptibility gene was greatly eased, compared with the corresponding work with *ATM*, by the fact that c.1100delC is a founder mutation with carrier frequencies of 0.5–1.0% in several northern European countries, and that p.Ile157Thr is a founder mutation with a carrier frequency of >1.0% among individuals of Slavic ancestry [72,73]. Therefore, in populations where one or more founder mutations have drifted to frequencies of above 1%, these mutations dominate the spectrum of pathogenic *CHEK2* alleles. However, what about in more mixed populations, or populations that are not dominated by one or two founder mutations? Le Calvez-Kelm *et al.* mutation screened *CHEK2* in an ethnically diverse series of 1303 early-onset breast cancer cases and 1109 controls [54]. They observed seven distinct protein-truncating variants and 34 distinct rare missense substitutions, and found that both groups of variants made a significant contribution to the risk of breast cancer attributable to the gene. When the rare missense substitutions were stratified by Align-GVGD grade, they found that 18 cases carried a missense substitution from a grade with an estimated odds ratio of 3.0 or higher (grades C25–C65), against 17 cases who carried a protein-truncating variant. Thus, the number of cases accounted for by these two types of variants was about equal. At about the same time, Desrichard *et al.* mutation screened the gene in a series of 507 French familial breast cancer cases and 513 controls [74]. They observed five distinct protein-truncating variants and 11 distinct missense substitutions, and a significantly higher frequency of the combined set of rare variants in the cases compared with the controls (OR: 4.2; $p = 0.0065$). From a combination of *in silico* assessment and *in vitro* kinase assays, Desrichard *et al.* hypothesized that eight of the missense substitutions were likely pathogenic. With six of their cases carrying a protein-truncating variant against nine carrying a hypothetically pathogenic missense substitution, the truncation/missense balance of the Desrichard *et al.* study is quite congruent with that of the Le Calvez-Kelm *et al.* study. Therefore, combining the two studies, we estimate that pathogenic variants in *CHEK2* account for 2.5–3.0% of early-onset or familial breast cancer cases, with approximately half of the risk attributable to protein-truncating variants and the remainder to rare missense substitutions.

PALB2

PALB2 (also known as *FANCN*) was identified as a key *BRCA2* interacting protein in 2006 [75]. The protein serves as a bridge between *BRCA1* and *BRCA2* and is required for the localization of *BRCA2* to nuclear DNA double-strand break repair foci [75–77]. In 2007, biallelic protein-truncating variants in the gene were shown to cause Fanconi Anemia [78, 79]. At the same time, Rahman *et al.* demonstrated through a CCMS study of 923 familial breast cancer cases and 1084 controls that heterozygous protein-truncating mutations in the gene confer an approximate 2.3-fold increased risk of breast cancer [79]. In this study, both the direct case–control comparison and tests of segregation conducted in the families of *PALB2* mutation carriers indicated that *PALB2* is a susceptibility gene ($p = 0.0004$ and $p = 0.0025$, respectively). A spate of *PALB2* mutation screening studies then confirmed the gene's status as an intermediate-risk breast cancer susceptibility gene [80,81]. Almost all of the evidence in these studies was attributable to protein-truncating variants, and they provided little data through which to judge the balance of risk due to protein-truncating variants versus rare missense substitutions in the gene.

More recently, Tischkowitz *et al.* reported on mutation screening of the gene in 565 unilateral breast cancer cases versus 559 bilateral breast cancer cases [82]. They found five protein-truncating variants in the bilateral cases versus none in the unilateral cases, and also found a higher frequency of affected first-degree relatives among *PALB2* mutation carriers than among mutation non-carriers. As in the Rahman *et al.* study, both the direct unilateral–bilateral case comparison and the analysis of breast cancers in relatives supported the identification of *PALB2* as a breast cancer susceptibility gene.

Tischkowitz *et al.* went on to examine the pattern of rare missense substitutions observed during their *PALB2* mutation screening [82]. Overall, there was no evidence for a higher frequency of rare missense substitutions among the bilateral cases than among the unilateral cases. Severity grading of the rare missense substitutions with SIFT, Polyphen and Align-GVGD, followed by statistical analyses, showed at best marginal evidence that a subset of the rare missense substitution might be pathogenic. Similarly, SIFT- and Align-GVGD-based analyses of the rare missense substitutions reported in the Rahman *et al.* study's supplementary data did not reveal any consistent evidence that rare missense substitutions in *PALB2* contribute to breast cancer risk [Tavtigian SV, Unpublished Data]. Thus, in contrast to *TP53*, *ATM* and *CHEK2*, it appears unlikely that *PALB2* carries a high proportion of pathogenic missense variants. However, determining whether a small fraction of risk attributable to *PALB2* is due to rare missense substitutions will require analyses of many thousands of cases and controls.

BRIP1

BRIP1, also called *FANCI* and *BACH1*, is a member of the DEAH helicase family. Seal *et al.* screened 1212 individuals with breast cancer from *BRCA1/BRCA2* mutation-negative families and 2081 controls by conformation-sensitive gel electrophoresis and estimated that *BRIP1* protein-truncating mutations confer a relative risk of breast cancer of 2.0 [83]. They also identified 24 missense variants but found no evidence that, individually or collectively, these substitutions conferred any risk of breast cancer.

Genes on the radar: other high-moderate risk genes

There are many other known and suspected breast cancer susceptibility genes for which there are not enough data to examine the ratio of protein-truncating variants to rare missense substitutions. These include *BLM* [84,85], *CDH1* [86], *FANCC* [84], *STK11* [87], *MRE11A* [72, 88], *RAD50* [72,89–90] and *NBN* [72,91,7292] for which rare protein-truncating mutations have been identified in breast cancer families, and in some cases (*PTEN*, *RAD51C* [also known as *FANCO*] and *XRCC2*) also a small number of missense variants [93–96]. In the case of *BARD1*, no protein-truncating variant has been reported to be associated with breast cancer risk, but there is a missense variant, Cys557Ser, that confers an intermediate risk [97].

Conclusion

As massively parallel sequencing moves from the research arena into clinical genetics, testing formats will change from analyses of one or two genes to analyses of panels of genes

– or perhaps to whole-exome sequencing. For breast and ovarian cancer, testing is likely to expand from *BRCA1* and *BRCA2* to include at least *TP53*, *ATM*, *CHEK2*, *PALB2* and *PTEN*. Indeed, these seven genes plus *RAD51C* (which is better attested as ovarian cancer susceptibility genes than breast cancer susceptibility genes), are included in the combination of Myriad's plus Ambry Genetic's testing services (Myriad Genetics, Inc., UT, USA; Ambry Genetics, Inc., CA, USA) and are included on a testing panel developed by Mary-Claire King's laboratory at the University of Washington (WA, USA).

Our concern lies with the problems of test interpretability and the burden of unclassified variants. For three of the susceptibility genes that we have reviewed, *TP53*, *ATM* and *CHEK2*, the fraction of breast cancer risk attributable to rare missense substitutions in the genes is as high, if not higher, than the fraction of risk attributable to protein-truncating variants. Since currently unclassified missense substitutions plus newly observed rare missense substitutions are reported by testing services as UVs, this means that a large fraction of the *bona fide* genetic risk that will be detected by clinical mutation screening of these genes will be reported to clinical geneticists and their patients as UVs. Thus, a question: how large a fraction of the *bona fide* risk can be reported out as UVs before a particular gene test does not meet the American Society of Clinical Oncologists criterion of test interpretability [98,99]? A second question is, how large will the burden of UVs actually be? There are 12 HRR pathway genes that we know of that we or others have mutation screened in a large enough number of subjects to estimate the summed frequency of rare missense substitutions (Table 2). We estimate that mutation screening all 12 would reveal one or more rare missense substitutions in 36% of patients tested. There would also be a substantial frequency of sequence variants falling in and around the splice junction consensus regions that would be reported as UVs, pushing the total fraction of problematic sequence variants higher.

Just as accumulation of *BRCA1* and *BRCA2* UVs led to development of the integrated evaluation for UVs in those genes, accumulation of UVs in the expanded set of HRR genes will lead to development of an expanded system for UV assessment. The main challenge will be that, because intermediate-risk susceptibility genes do not create as strong patterns of segregation or as strong summary family histories as do the high-risk genes, and pathogenic variants in them are individually rare, these aspects of the integrated evaluation will be less powerful. In all likelihood, the difference in power will be made up by functional assays. Indeed, some functional assays have been developed for *ATM*, *CHEK2* and *TP53* missense substitutions [101–103]. Once the functional assays are sufficiently accurate and efficient to contribute to UV classification, there will still be two significant hurdles to cross: recalibration of the assay outputs into odds ratios in favor of pathogenicity so that they can be included in the integrated evaluation; and finding a way to fund the cost of the assays, either as an integral component of the testing process or else as a post-testing service.

BRCA1, *BRCA2* and *PALB2* seem to form a group in which the vast majority of genetic risk attributable to them is due to protein-truncating variants. On the other hand, *ATM*, *CHEK2* and *TP53* form a group in which half or more of the genetic risk attributable to them is due to rare missense substitutions. But why? We can point at three factors, but they make an incomplete explanation. First, looking at their protein multiple sequence alignments

(available at [20]) *BRCA1* and *BRCA2* share the characteristic that only a small fraction of the protein is evolutionarily well conserved; by contrast, a large fraction of the length of *CHEK2* and *TP53* is well conserved. Since missense substitutions that fall at conserved residues are more likely to be pathogenic than those falling at highly variable residues, it follows that a higher fraction of *CHEK2* and *TP53* missense substitutions should be pathogenic than *BRCA1* and *BRCA2* missense substitutions. However, this does not explain *ATM*, which also includes long stretches of poorly conserved residues. A second factor could be multiple repeated sequence elements. For example, *BRCA2* contains eight RAD51-binding BRC domains. Since facilitating the loading of RAD51 at sites of HRR appears to be a core function of *BRCA2*, one might expect that a severe BRC domain missense substitution would be pathogenic. In fact, however, there is no human genetic evidence that this is the case. Since a missense substitution would only knockout the function of one BRC domain, it could well be that this can only cause a small decrement in *BRCA2* functionality, perhaps causing a modestly increased risk of cancer rather than a severely increased risk. Thirdly, there is the more ineffable issue of dominant-negative missense substitutions. There is good evidence that both *TP53* and *ATM* can harbor dominant-negative missense substitutions, which have the potential to confer greater risk of cancer than do protein-truncating variants. But, as all of these proteins participate in large multiprotein functional complexes, it is not obvious that one can predict which proteins (or which missense substitutions) are most likely to harbor dominant-negative, rather than simple loss-of-function, substitutions.

Future perspective

Driven by technologic advance, the paradigm for clinical genetic predisposition diagnostic testing is shifting from screening patients for mutations in one or two susceptibility genes to panel tests in which patients are screened for mutations in many genes in parallel. While panel tests are more informative than their gene-by-gene predecessor, testing more genes complicates the challenge inherent in interpreting patients' test results. This challenge is encapsulated in a key difference between two pairs of breast cancer susceptibility genes: *BRCA1* and *BRCA2* – genes in which a large majority of pathogenic alleles are relatively easily recognized protein-truncating variants, versus *ATM* and *CHEK2* – genes in which a large fraction of the pathogenic alleles are more difficult to interpret missense substitutions. Since testing laboratories generally report potentially pathogenic missense substitutions as unclassified variants, a large fraction of the *bona fide* genetic risk observed in these latter two genes may initially be clinically uninterpretable.

Acknowledgments

Financial disclosure

G Chenevix-Trench is supported by grants from the Australian National Health and Medical Research Council and the Susan G. Komen Foundation. SV Tavtigian is supported by grants from the US NIH and also receives royalties from the NIH that originate from Myriad Genetics' BRACAnalysis testing. Work described here was also supported in part by NIH NCI R01CA121245.

References

Papers of special note have been highlighted as:

- of interest;
- of considerable interest

1. Broca, P. *Traite des Tumeurs*. P. Asselin; Paris, France: 1866.
2. Lichtenstein P, Holm NV, Verkasalo PK, et al. 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(2):78–85. [PubMed: 10891514]
3. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst*. 1994; 86(21):1600–1608. [PubMed: 7932824]
4. Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, et al. Cancer as a complex phenotype: pattern of cancer distribution within and beyond the nuclear family. *PLoS Med*. 2004; 1(3):e65. [PubMed: 15630470]
5. Kerber RA, O'Brien E. A cohort study of cancer risk in relation to family histories of cancer in the Utah population database. *Cancer*. 2005; 103(9):1906–1915. [PubMed: 15779016]
6. Boyle, P.; Levin, B., editors. *World Cancer Report 2008*. IARC Press; Lyon, France: 2008. Genetic susceptibility; p. 183-184.
7. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*. 1990; 250(4988):1684–1689. [PubMed: 2270482]
8. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science*. 1994; 266(5182):66–71. [PubMed: 7545954]
9. Castilla LH, Couch FJ, Erdos MR, et al. Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer. *Nat Genet*. 1994; 8(4):387–391. [PubMed: 7894491]
10. Simard J, Tonin P, Durocher F, et al. Common origins of *BRCA1* mutations in Canadian breast and ovarian cancer families. *Nat Genet*. 1994; 8(4):392–398. [PubMed: 7894492]
11. Friedman LS, Ostermeyer EA, Szabo CI, et al. Confirmation of *BRCA1* by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nat Genet*. 1994; 8(4):399–404. [PubMed: 7894493]
- 12•. Plon SE, Eccles DM, Easton D, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*. 2008; 29(11):1282–1291. Developed key rationale and guidelines for clinical evaluation and reporting of sequence variants observed in cancer susceptibility genes, with a focus on variants that would be reported as unclassified by the initial testing laboratory but can be classified into a clinically actionable category by an integrated evaluation. [PubMed: 18951446]
13. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature*. 1995; 378(6559):789–792. [PubMed: 8524414]
14. Tavtigian SV, Simard J, Rommens J, et al. The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds. *Nat Genet*. 1996; 12(3):333–337. [PubMed: 8589730]
- 15•. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to *BRCA1* and *BRCA2*. *Am J Hum Genet*. 2004; 75(4):535–544. Developed the first quantitative system for evaluation of unclassified variants in *BRCA1* and *BRCA2*. [PubMed: 15290653]
16. Neuhausen SL, Ostrander EA. Mutation testing of early-onset breast cancer genes *BRCA1* and *BRCA2*. *Genet Test*. 1997; 1(2):75–83. [PubMed: 10464630]
17. InterPro: protein sequence analysis & classification. www.ebi.ac.uk/interpro
18. UniProt. www.uniprot.org
19. NCBI. Conserved Domains and Protein Classification. www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

20. Alignments – Align GVG. <http://agvgd.iarc.fr/alignments.php>
21. O'Donovan PJ, Livingston DM. BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis*. 2010; 31:961–967. [PubMed: 20400477]
22. Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol*. 2010; 11:196–207. [PubMed: 20177395]
23. Chun J, Buechelmaier ES, Powell SN. Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. *Mol Cell Biol*. 2013; 33:387–395. [PubMed: 23149936]
24. NHGRI. Breast Cancer Information Core. <http://research.nhgri.nih.gov/bic/>
25. Spurdle AB. Clinical relevance of rare germline sequence variants in cancer genes: evolution and application of classification models. *Curr Opin Genet Dev*. 2010; 20(3):315–323. [PubMed: 20456937]
26. Lindor NM, Guidugli L, Wang X, et al. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum Mutat*. 2012; 33(1):8–21. [PubMed: 21990134]
27. Vallee MP, Francy TC, Judkins MK, et al. Classification of missense substitutions in the BRCA genes: a database dedicated to Ex-UVs. *Hum Mutat*. 2012; 33(1):22–28. [PubMed: 21990165]
28. Easton DF, Deffenbaugh AM, Pruss D, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the *BRCA1* and *BRCA2* breast cancer-predisposition genes. *Am J Hum Genet*. 2007; 81(5):873–883. [PubMed: 17924331]
29. Zuckerkandl E, Pauling L. Molecules as documents of evolutionary history. *J Theor Biol*. 1965; 8(2):357–366. [PubMed: 5876245]
30. Jukes TH, King JL. Deleterious mutations and neutral substitutions. *Nature*. 1971; 231(5298):114–115. [PubMed: 4930087]
31. Szabo CI, Wagner LA, Francisco LV, et al. Human, canine and murine *BRCA1* genes: sequence comparison among species. *Hum Mol Genet*. 1996; 5(9):1289–1298. [PubMed: 8872468]
32. Abkevich V, Zharkikh A, Deffenbaugh AM, et al. Analysis of missense variation in human *BRCA1* in the context of interspecific sequence variation. *J Med Genet*. 2004; 41(7):492–507. Demonstrates a clear excess of clinically observed missense substitutions at evolutionarily conserved amino acid positions in *BRCA1* and estimated the fraction of all pathogenic alleles in the gene that are missense substitutions. [PubMed: 15235020]
33. 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(11):1342–1354. [PubMed: 18951461]
34. Tavtigian SV, Samollow PB, de Silva D, Thomas A. An analysis of unclassified missense substitutions in human *BRCA1*. *Fam Cancer*. 2006; 5(1):77–88. [PubMed: 16528611]
35. Align GVG. <http://agvgd.iarc.fr/alignments.php>
36. SIFT. <http://sift.jcvi.org>
37. PolyPhen-2: prediction of functional effects of human nsSNPs. <http://genetics.bwh.harvard.edu/pph2>
38. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*. 2001; 11(5):863–874. [PubMed: 11337480]
39. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7(4):248–249. [PubMed: 20354512]
40. Spurdle AB, Healey S, Devereau A, et al. 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(1):2–7. [PubMed: 21990146]
41. Tinat J, Bougeard G, Baert-Desurmont S, et al. version of the Chompret criteria for Li Fraumeni syndrome. *J Clin Oncol*. 2009; 27(26):e108–e109. author reply e110 (2009). [PubMed: 19652052]

42. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science*. 1990; 250(4985):1233–1238. [PubMed: 1978757]
43. Srivastava S, Zou ZQ, Pirollo K, Blattner W, Chang EH. Germ-line transmission of a mutated *p53* gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature*. 1990; 348(6303):747–749. [PubMed: 2259385]
44. Linzer DI, Levine AJ. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell*. 1979; 17(1):43–52. [PubMed: 222475]
45. Parada LF, Land H, Weinberg RA, Wolf D, Rotter V. Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation. *Nature*. 1984; 312(5995):649–651. [PubMed: 6390217]
46. Jenkins JR, Rudge K, Currie GA. Cellular immortalization by a cDNA clone encoding the transformation-associated phosphoprotein p53. *Nature*. 1984; 312(5995):651–654. [PubMed: 6095117]
47. Eliyahu D, Raz A, Gruss P, Givol D, Oren M. Participation of p53 cellular tumour antigen in transformation of normal embryonic cells. *Nature*. 1984; 312(5995):646–649. [PubMed: 6095116]
48. Finlay CA, Hinds PW, Levine AJ. The p53 proto-oncogene can act as a suppressor of transformation. *Cell*. 1989; 57(7):1083–1093. [PubMed: 2525423]
49. Levine AJ. The p53 tumor suppressor gene and gene product. *Princess Takamatsu Symp*. 1989; 20:221–30. [PubMed: 2488233]
50. Soussi T, Kato S, Levy PP, Ishioka C. Reassessment of the TP53 mutation database in human disease by data mining with a library of TP53 missense mutations. *Hum Mutat*. 2005; 25(1):6–17. [PubMed: 15580553]
51. Hainaut P, Soussi T, Shomer B, et al. Database of *p53* gene somatic mutations in human tumors and cell lines: updated compilation and future prospects. *Nucleic Acids Res*. 1997; 25(1):151–157. [PubMed: 9016527]
- 52••. Ognjanovic S, Olivier M, Bergemann TL, Hainaut P. Sarcomas in *TP53* germline mutation carriers: a review of the IARC TP53 database. *Cancer*. 2012; 118(5):1387–1396. Documents that a large fraction of pathogenic alleles in *TP53* are missense substitutions. [PubMed: 21837677]
53. Tavtigian SV, Hashibe M, Thomas A. Tests of association for rare variants: case control mutation screening. *Nat Rev Genet*. 2011; 12(3):224. [PubMed: 21283087]
- 54••. Le Calvez-Kelm F, Lesueur F, Damiola F, et al. 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(1):R6. Case–control mutation screening study of *CHEK2* demonstrating that individually rare missense substitutions are responsible for a large fraction of the risk of breast cancer attributable to rare sequence variants in *CHEK2*. [PubMed: 21244692]
55. Swift M, Reitnauer PJ, Morrell D, Chase CL. Breast and other cancers in families with ataxia-telangiectasia. *N Engl J Med*. 1987; 316(21):1289–1294. [PubMed: 3574400]
56. FitzGerald MG, Bean JM, Hegde SR, et al. Heterozygous *ATM* mutations do not contribute to early onset of breast cancer. *Nat Genet*. 1997; 15(3):307–310. [PubMed: 9054948]
- 57•. Tavtigian SV, Oefner PJ, Babikyan D, et al. Rare, evolutionarily unlikely missense substitutions in *ATM* confer increased risk of breast cancer. *Am J Hum Genet*. 2009; 85(4):427–446. Meta-analysis of *ATM* case–control mutation screening data demonstrating that individually rare missense substitutions are responsible for a large fraction of the risk of breast cancer attributable to rare sequence variants in *ATM*. [PubMed: 19781682]
- 58••. Renwick A, Thompson D, Seal S, et al. *ATM* mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet*. 2006; 38(8):873–875. First convincing demonstration, in molecular genetic terms, that *ATM* is an intermediate-risk breast cancer susceptibility gene. [PubMed: 16832357]
59. Bernstein JL, Haile RW, Stovall M, et al. Radiation exposure, the *ATM* gene, and contralateral breast cancer in the women’s environmental cancer and radiation epidemiology study. *J Natl Cancer Inst*. 2010; 102(7):475–483. [PubMed: 20305132]

60. Goldgar DE, Healey S, Dowty JG, et al. Rare variants in the *ATM* gene and risk of breast cancer. *Breast Cancer Res.* 2011; 13(4):R73. [PubMed: 21787400]
61. Stankovic T, Kidd AM, Sutcliffe A, et al. 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(2):334–345. [PubMed: 9463314]
62. Waddell N, Jonnalagadda J, Marsh A, et al. Characterization of the breast cancer associated *ATM* 7271T>G (V2424G) mutation by gene expression profiling. *Genes Chromosomes Cancer.* 2006; 45(12):1169–1181. [PubMed: 17001622]
63. Matsuoka S, Huang M, Elledge SJ. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science.* 1998; 282(5395):1893–1897. [PubMed: 9836640]
64. Chaturvedi P, Eng WK, Zhu Y, et al. Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway. *Oncogene.* 1999; 18(28):4047–4054. [PubMed: 10435585]
65. Weinert TA, Kiser GL, Hartwell LH. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. *Genes Dev.* 1994; 8(6):652–665. [PubMed: 7926756]
66. Allen JB, Zhou Z, Siede W, Friedberg EC, Elledge SJ. The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes Dev.* 1994; 8(20):2401–2415. [PubMed: 7958905]
67. Bell DW, Varley JM, Szydlo TE, et al. Heterozygous germ line *hCHK2* mutations in Li-Fraumeni syndrome. *Science.* 1999; 286(5449):2528–2531. [PubMed: 10617473]
68. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to *CHEK2*(*)1100delC in noncarriers of *BRCA1* or *BRCA2* mutations. *Nat Genet.* 2002; 31(1):55–59. [PubMed: 11967536]
69. CHEK2 Breast Cancer Case-Control Consortium. *CHEK2**1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet.* 2004; 74(6):1175–1182. [PubMed: 15122511]
70. Kilpivaara O, Vahteristo P, Falck J, et al. *CHEK2* variant I157T may be associated with increased breast cancer risk. *Int J Cancer.* 2004; 111(4):543–547. [PubMed: 15239132]
71. Cybulski C, Gorski B, Huzarski T, et al. *CHEK2* is a multiorgan cancer susceptibility gene. *Am J Hum Genet.* 2004; 75(6):1131–1135. [PubMed: 15492928]
72. Hollestelle A, Wasielewski M, Martens JW, Schutte M. Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev.* 2010; 20(3):268–276. [PubMed: 20346647]
73. Shuen AY, Foulkes WD. Inherited mutations in breast cancer genes—risk and response. *J Mammary Gland Biol Neoplasia.* 2011; 16(1):3–15. [PubMed: 21461995]
74. Desrichard A, Bidet Y, Uhrhammer N, Bignon YJ. CHEK2 contribution to hereditary breast cancer in non-BRCA families. *Breast Cancer Res.* 2011; 13(6):R119. [PubMed: 22114986]
75. Xia B, Sheng Q, Nakanishi K, et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell.* 2006; 22(6):719–729. [PubMed: 16793542]
76. Oliver AW, Swift S, Lord CJ, Ashworth A, Pearl LH. Structural basis for recruitment of BRCA2 by PALB2. *EMBO Rep.* 2009; 10(9):990–996. [PubMed: 19609323]
77. Zhang F, Bick G, Park JY, Andreassen PR. MDC1 and RNF8 function in a pathway that directs BRCA1-dependent localization of PALB2 required for homologous recombination. *J Cell Sci.* 2012; 125(Pt 24):6049–6057. [PubMed: 23038782]
78. Xia B, Dorsman JC, Ameziane N, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nat Genet.* 2007; 39(2):159–161. [PubMed: 17200672]
79. Rahman N, Seal S, Thompson D, et al. *PALB2*, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet.* 2007; 39(2):165–167. [PubMed: 17200668]
80. Erkkö H, Xia B, Nikkila J, et al. A recurrent mutation in *PALB2* in Finnish cancer families. *Nature.* 2007; 446(7133):316–319. [PubMed: 17287723]
81. Tischkowitz M, Xia B, Sabbaghian N, et al. Analysis of PALB2/FANCN-associated breast cancer families. *Proc Natl Acad Sci USA.* 2007; 104(16):6788–6793. [PubMed: 17420451]

82. Tischkowitz M, Capanu M, Sabbaghian N, et al. Rare germline mutations in *PALB2* and breast cancer risk: a population-based study. *Hum Mutat.* 2012; 33(4):674–680. [PubMed: 22241545]
83. Seal S, Thompson D, Renwick A, et al. Truncating mutations in the Fanconi anemia J gene *BRIP1* are low-penetrance breast cancer susceptibility alleles. *Nat Genet.* 2006; 38(11):1239–1241. [PubMed: 17033622]
84. Thompson ER, Doyle MA, Ryland GL, et al. Exome sequencing identifies rare deleterious mutations in DNA repair genes *FANCC* and *BLM* as potential breast cancer susceptibility alleles. *PLoS Genet.* 2012; 8(9):e1002894. [PubMed: 23028338]
85. Sokolenko AP, Iyevleva AG, Preobrazhenskaya EV, et al. High prevalence and breast cancer predisposing role of the *BLM* c. 1642 C>T (Q548X) mutation in Russia. *Int J Cancer.* 2012; 130(12):2867–2873. [PubMed: 21815139]
86. Guilford PJ, Hopkins JB, Grady WM, et al. E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. *Hum Mutat.* 1999; 14(3):249–255. [PubMed: 10477433]
87. Schumacher V, Vogel T, Leube B, et al. *STK11* genotyping and cancer risk in Peutz-Jeghers syndrome. *J Med Genet.* 2005; 42(5):428–435. [PubMed: 15863673]
88. Bartkova J, Tommiska J, Oplustilova L, et al. Aberrations of the *MRE11*-*RAD50*-*NBS1* DNA damage sensor complex in human breast cancer: *MRE11* as a candidate familial cancer-predisposing gene. *Mol Oncol.* 2008; 2:296–316. [PubMed: 19383352]
89. Heikkinen K, Karppinen SM, Soini Y, et al. Mutation screening of *Mre11* complex genes: indication of *RAD50* involvement in breast and ovarian cancer susceptibility. *J Med Genet.* 2003; 40:e131. [PubMed: 14684699]
90. Heikkinen K, Rapakko K, Karppinen SM, et al. *RAD50* and *NBS1* are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis.* 2006; 27:1593–1599. [PubMed: 16474176]
91. di Masi A, Antoccia A. *NBS1* heterozygosity and cancer risk. *Curr Genomics.* 2008; 9:275–281. [PubMed: 19452044]
92. Zhang ZH, Yang LS, Huang F, et al. Current evidence on the relationship between two polymorphisms in the *NBS1* gene and breast cancer risk: a meta-analysis. *Asian Pac J Cancer Prev.* 2012; 13:5375–5379. [PubMed: 23317186]
93. Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime cancer risks in individuals with germline *PTEN* mutations. *Clin Cancer Res.* 2012; 18(2):400–407. [PubMed: 22252256]
94. Meindl A, Hellebrand H, Wiek C, et al. Germline mutations in breast and ovarian cancer pedigrees establish *RAD51C* as a human cancer susceptibility gene. *Nat Genet.* 2010; 42(5):410–414. [PubMed: 20400964]
95. Osorio A, Endt D, Fernandez F, et al. Predominance of pathogenic missense variants in the *RAD51C* gene occurring in breast and ovarian cancer families. *Hum Mol Genet.* 2012; 21(13):2889–2898. [PubMed: 22451500]
96. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in *XRCC2* increase the risk of breast cancer. *Am J Hum Genet.* 2012; 90(4):734–739. [PubMed: 22464251]
97. Karppinen SM, Barkardottir RB, Backenhorn K, et al. Nordic collaborative study of the *BARD1* Cys557Ser allele in 3956 patients with cancer: enrichment in familial *BRCA1/BRCA2* mutation-negative breast cancer but not in other malignancies. *J Med Genet.* 2006; 43(11):856–862. [PubMed: 16825437]
98. American Society of Clinical Oncology. American Society of Clinical Oncology policy statement update: genetic testing for cancer susceptibility. *J Clin Oncol.* 2003; 21(12):2397–2406. [PubMed: 12692171]
99. Robson ME, Storm CD, Weitzel J, Wollins DS, Offit K. American Society of Clinical Oncology policy statement update: genetic and genomic testing for cancer susceptibility. *J Clin Oncol.* 2010; 28(5):893–901. [PubMed: 20065170]
100. Le Calvez-Kelm F, Oliver J, Damiola F, et al. *RAD51* and Breast Cancer Susceptibility: No Evidence for Rare Variant Association in the Breast Cancer Family Registry Study. *PLoS ONE.* 2012; 7(12):e52374. [PubMed: 23300655]

101. Mitui M, Nahas SA, Du LT, et al. Functional and computational assessment of missense variants in the ataxia-telangiectasia mutated (*ATM*) gene: mutations with increased cancer risk. *Hum Mutat.* 2009; 30(1):12–21. [PubMed: 18634022]
102. Roeb W, Higgins J, King MC. Response to DNA damage of *CHEK2* missense mutations in familial breast cancer. *Hum Mol Genet.* 2012; 21(12):2738–2744. Functional analyses of clinically observed *CHEK2* missense substitutions revealed that a large fraction of these are probably pathogenic. [PubMed: 22419737]
103. Kato S, Han SY, Liu W, et al. Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci USA.* 2003; 100(14):8424–8429. [PubMed: 12826609]

Executive summary

BRCA1 & BRCA2

- High-risk breast and ovarian cancer susceptibility genes were discovered in the mid-1990s.
- Inherited pathogenic alleles confer 5–10-fold increased risk of breast cancer.
- Their spectra of pathogenic sequence variants is dominated by protein-truncating variants.
- By 1997, it was already clear that some rare missense substitutions in these genes are pathogenic.

The integrated evaluation of unclassified variants

- Efforts towards clinical evaluation and classification of *BRCA1* and *BRCA2* missense substitutions began in about 2000.
- The first quantitative classification model was described in 2004.
- The contemporary (circa 2013) version integrates five types of data to achieve classification.

TP53

- Inherited pathogenic alleles of *TP53* cause a high-risk multiorgan familial cancer syndrome, Li-Fraumeni syndrome.
- Almost three-quarters of the germline pathothogenic variants in this gene are missense substitutions.

Intermediate-risk genes

- Inherited pathogenic alleles confer a 2–5-fold increased risk of breast cancer.
- Intermediate-risk genes are more easily evaluated by case-control mutation screening than by segregation analysis.

ATM

- *ATM* is the susceptibility gene for the autosomal-recessive disorder ataxia telangiectasia.
- Heterozygous protein-truncating variants in *ATM* confer intermediate risk of breast cancer.
- Some key functional domain missense substitutions *in ATM* may confer greater risk of breast cancer than do truncating variants.
- Key functional domain missense substitutions may account for more breast cancer cases than truncating variants.
- *ATM* is rarely mutation screened in older gene-by-gene clinical tests, but is now a component of panel-based tests.

CHEK2

- A protein-truncating variant of this gene, c.1100delC, has a frequency of approximately 1% in northern European populations.
- Case-control genotyping revealed that c.1100delC confers a 2.3-fold increased risk of breast cancer.
- Case-control mutation screening revealed that *CHEK2* missense substitutions account for a similar number of breast cancer cases as protein-truncating variants.
- Like *ATM*, *CHEK2* is rarely screened in older gene-by-gene clinical mutation tests, but is now a component of panel-based tests.

PALB2 & BRIP1

- *PALB2* and *BRIP1* are also intermediate-risk breast cancer susceptibility genes.
- Their mutation spectra appear to be dominated by protein-truncating variants.
- Insufficient data are available to determine what fraction (if any) of their pathogenic variants are missense substitutions.

Genes on the radar

- There are more than ten additional known or suspected breast cancer susceptibility genes.
- There are insufficient data to examine the ratio of protein-truncating variants to rare missense substitutions for most of these.

Future perspective

- The transition from gene-by-gene cancer predisposition genetic testing to panel based tests increases test informativeness.
- The mutation spectra of some of these genes will feature a large fraction of pathogenic missense substitutions.
- Most of the pathogenic missense substitutions will be reported from testing laboratories as unclassified variants.
- Absent robust missense substitution evaluation models, a large fraction of the *bona fide* genetic risk detected will reside in the unclassified variants.

Figure 1. Functional domain organizations of BRCA1, BRCA2, p53, ATM and CHEK2, the homologous recombination pathway proteins known to harbor significant numbers of pathogenic missense substitutions

The protein domain diagrams are based on information from the InterPro protein sequence analysis and classification database [17], the Uniprot Protein Knowledgebase [18], and the NCBI Conserved Domains Database [19] and curated protein multiple sequence alignments available at the Align-GVGD website [20]. Note that the five protein diagrams are each on a different length scale, but that the length of each protein is given at the right side of each diagram.

aa: Amino acid.

Figure 2. Core genes in the pathway for DNA double-strand break repair by homologous recombination

Following detection of DNA double-strand breaks, the MRN complex (MRE11A–RAD50–NBN) activates the ATM kinase. ATM in turn phosphorylates downstream effectors including BRCA1, CHEK2 and p53. CHEK2 and p53 activate cell cycle arrest to provide time for DNA damage repair, and p53 also plays a key role in apoptotic signaling. PALB2 interacts with BRCA1 and BRCA2 and is important for the localization of BRCA2 to sites of DNA repair. BARD1 and BRCA1 are structural homologs and heterodimerize through their amino terminal RING domains. Phosphorylated CtIP (RBBP8) and BRIP1 interact with BRCA1's carboxy-terminal BRCT repeats; these BRCA1-containing complexes stimulate homologous recombination repair and inhibit the relatively mutagenic process of double-strand break repair by nonhomologous end joining. The central section of BRCA2 (encoded by exon 11) has eight RAD51 binding sites that play a role in concentrating RAD51 at sites requiring repair by homologous recombination. RAD51 and its paralogs, including RAD51C and XRCC2, then mediate strand invasion during repair by homologous recombination. Solid arrows represent a direct interaction between the proteins indicated.

Dotted arrows represent the impact of a proteins activity on the cellular process indicated [21–23].

Table 1

Classification system for initially unclassified sequence variants identified in cancer susceptibility genes by genetic testing.

Class	Description	Posterior probability in favor of pathogenicity
5	Definitely pathogenic	>0.99
4	Likely pathogenic	0.95–0.99
3	Uncertain	0.05–0.949
2	Likely not pathogenic or of little clinical significance	0.001–0.049
1	Not pathogenic or of no clinical significance	<0.001

Data taken from [12].

Table 2
Summed carrier rates of rare missense substitutions in homologous recombination pathway genes.

Gene	Subjects mutations screened (n)	Rare missense substitutions			Protein length	Ref.
		Distinct (n)	Carriers (n)	Carrier rate		
BRCA1	68,000	596	4525	0.067	1863	[33]
BRCA2	68,000	1131	10,031	0.148	3418	[33]
ATM	6115	167	395	0.065	3056	[57]
CHEK2	2619	34	61	0.023	543	[54]
PALB2	2007	31	73	0.036	1186	[79]
BRIP1	3293	22	80	0.024	1249	[83]
XRCC2	2420	6	6	0.002	280	[96]
RAD51	2453	3	4	0.002	339	[100]
BARD1	2422	28	54	0.022	777	†
MRE11A	~2430	21	25	0.010	708	†
RAD50	~2430	28	45	0.019	1312	†
NBN	~2430	11	18	0.007	754	†

Summed rare missense substitution carrier rate: 0.425.

Probability to carry one or more rare missense: 0.358. Assuming conditional independence of rare missense substitution carrier rates between homologous recombination pathway genes.

†Tavtigian laboratory unpublished case-control mutation screening results.