Skol et al. Breast Cancer Research (2016) 18:99
DOI 10.1186/s13058-016-0759-4

Breast Cancer Research

REVIEW Open Access

The genetics of breast cancer risk in the post-genome era: thoughts on study design to move past *BRCA* and towards clinical relevance

Andrew D. Skol[1], Mark M. Sasaki[2] and Kenan Onel[1,3*]

Abstract

More than 12 % of women will be diagnosed with breast cancer in their lifetime. Although there have been tremendous advances in elucidating genetic risk factors underlying both familial and sporadic breast cancer, much of the genetic contribution to breast cancer etiology remains unknown. The discovery of *BRCA1* and *BRCA2* over 20 years ago remains the seminal event in the field and has paved the way for the discovery of other high-penetrance susceptibility genes by linkage analysis. The advent of genome-wide association studies made possible the next wave of discoveries, in which over 80 low-penetrance and moderate-penetrance variants were identified. Although these studies were highly successful at discovering variants associated with both familial and sporadic breast cancer, the variants identified to date explain only 50 % of the heritability of breast cancer. In this review, we look back at the investigative strategies that have led to our current understanding of breast cancer genetics, consider the challenges of performing association studies in heterogeneous complex diseases such as breast cancer, and look ahead toward the types of study designs that may lead to the identification of the genetic variation accounting for the remaining missing heritability.

Keywords: Genetics, Genome-wide association study, Genomics, Linkage analysis, Next-generation sequencing, Penetrance, Predisposition

Background

Among women, breast cancer accounts for over 25 % of cancer diagnoses and 15 % of cancer-related deaths [1]. Ten percent of women with breast cancer have a family history of the disease [2]. Compared with women without a family history, women with one premenopausal first-degree relative with breast cancer are at 3.3-fold greater risk, and women with two first-degree relatives with breast cancer are at 3.6-fold greater risk [3], demonstrating that germline genetics contributes significantly to risk.

To identify genetic factors associated with breast cancer predisposition, early studies used linkage analysis and positional cloning in families with multiple affected individuals to discover highly penetrant susceptibility genes such as *BRCA1* and *BRCA2* [4, 5]. Although these studies were successful and could explain about 20 % of the familial aggregation of breast cancer risk [6], they provided little insight into the role of genetics in nonfamilial breast cancer.

More recently, genome-wide association studies (GWAS) have identified over 80 loci significantly associated with sporadic breast cancer. Collectively, however, these variants explain only 16 % of breast cancer heritability [7]. The inability of GWAS to identify a greater proportion of the genetic risk stems from many factors, including genotyping platform limitations in interrogating rare variation. Consequently, attention has shifted recently to investigating the impact of rare variation on disease, motivated in part by the precipitous drop in next-generation sequencing (NGS) costs.

Here, we will review linkage studies, GWAS, and NGS studies that have led to our current knowledge of the

* Correspondence: konel@northwell.edu
[1]Department of Pediatrics, The University of Chicago, Chicago, IL 60637, USA
[3]Section of Hematology/Oncology, Department of Pediatrics, The University of Chicago, KCBD 5140, 900 East 57th Street, Chicago, IL 60637, USA
Full list of author information is available at the end of the article

genetics of breast cancer susceptibility. Our emphasis will be on the strengths and limitations of different study designs with the potential to yield clinically translatable discoveries.

Family linkage studies and rare high-penetrance and medium-penetrance risk variants

Clinically, the most important breast cancer susceptibility genes are *BRCA1* and *BRCA2*. The loci where these genes reside were first observed as linkage peaks on chromosomes 17q21 and 13q12, in studies of just 23 and 15 families [8, 9]. The genes were identified shortly thereafter by fine-mapping using linkage analysis of more families [10], followed by positional cloning [4, 11] and mutation screening [5]. All told, 5 % of all breast cancer cases and up to 25 % of familial breast cancer cases can be attributed to high-penetrance mutations in *BRCA1* and *BRCA2* [12].

The impact of mutations in either gene can be dramatic; 65 % and 45 % of women with deleterious mutations in *BRCA1* or *BRCA2*, respectively, will develop breast cancer by age 70 [13], and the risk increases to 85 % and 84 %, respectively, for women with a family history of breast cancer [14, 15]. Generally, *BRCA* susceptibility variants identified in breast cancer patients with a positive family history are unique to each family. However, founder mutations are observed within certain populations. For example, in Ashkenazi Jews, the *BRCA2* c.5946delT (previously 6174delT) mutation is found at an allele frequency (AF) of 0.009–0.015 [16].

There are other genes in which germline mutations have been identified that substantially increase the risk of breast cancer. Most of these were initially discovered because they cause a syndrome of which breast cancer is a component. Li–Fraumeni syndrome (LFS), for example, is a cancer predisposition syndrome due to germline mutations in *TP53*, in which the most common cancer type in women is breast cancer. Malkin et al. [17] initially investigated a link between *TP53* and LFS because somatic mutations in *TP53* were identified in cancer types commonly observed in LFS families. They sequenced *TP53* exons 5–8 in five LFS families because this region contains the highly conserved DNA binding domain and harbors most *TP53* somatic mutations. Affected members of all families were found to have segregating germline mutations in this region, with inheritance consistent with a dominant model.

A deletion in *CHEK2* was investigated by the same group in 1071 breast cancer patients from 718 families with a positive history of breast cancer and no *BRCA* mutation, a population-based set of 636 patients, and 1620 controls. They found the *CHEK2**1100delC variant at a frequency of 1.1 % in controls, 5.1 % in cases with a family history, and 13.5 % in cases with a family history of male breast cancer [18]. Intriguingly, the AF in nonfamilial breast cancer cases did not differ from that of controls (1.4 %).

A similar strategy of examining genes causing syndromes with a high incidence of breast cancer led to the discovery of *PALB2*. Biallelic *PALB2* mutations cause a Fanconi anemia (FA) phenotype similar to that caused by biallelic *BRCA2* mutations. Rahman et al. [19] investigated whether heterozygous *PALB2* mutation carriers, like *BRCA2* carriers, were at increased breast cancer risk. They sequenced *PALB2* in 1084 controls and 923 cases with a family history of breast cancer but no *BRCA* mutation. They found 10 *PALB2* truncating mutations among cases, but none among controls. More recently, Antoniou et al. [20] examined the breast cancer risk in 362 members of 154 families with loss of function mutations in *PALB2*. They found an age-dependent trend in breast cancer risk among *PALB2* mutation carriers relative to age-matched controls (age 40–44 years, RR = 8.02; age 50–54, RR = 6.55; age 60–64, RR = 5.45). Interestingly, women with *PALB2* mutations from families with a history of breast cancer had substantially greater breast cancer risk than women with *PALB2* mutations but no family history.

BRIP1 also causes FA when deleted biallelically. *BRIP1* was investigated as a breast cancer susceptibility gene in heterozygous carriers because it interacts with other breast cancer predisposing genes such as *BRCA1*. Seal et al. [21] sequenced the exons and exon–intron boundaries of *BRIP1* in 1212 breast cancer cases with a family history of disease and no *BRCA* mutation and in 2081 controls, and found mutations in nine cases (0.74 %) but only in two controls (0.10 %). Intriguingly, no *BRIP1*-mutated FA family had a family history of breast cancer. More recently, Easton et al. [22] sequenced the coding region of *BRIP1* in more than 13,000 population-based breast cancer cases and 8000 controls, and found no excess of truncating mutations in cases relative to controls (0.21 % vs 0.23 %, respectively). The apparently discrepant results between these two studies may be another example of the importance of family history in determining the penetrance of a risk variant. However, these results also illustrate the challenges inherent in drawing conclusions about rare variants of modest effect, even when analyzing tens of thousands of samples.

ATM, in which biallelic mutations cause ataxia-telangiectasia (AT), was also suspected to be a breast cancer susceptibility gene in carriers because of an increased breast cancer incidence among relatives of AT patients. Renwick et al. [23] sequenced *ATM* in 443 *BRCA*-negative cases from families with at least three breast cancer-affected members and in 521 controls. Nine truncating and exon-skipping mutations were identified in cases, while only two were found in controls.

All mutations found in cases were predicted to cause AT, and seven had been observed previously in AT cases. Another group performed a meta-analysis using *ATM* sequence data from 1544 breast cancer cases and 1224 controls [24]. They found only marginal evidence for an excess of truncating and splice site variants within cases relative to controls, but greater evidence when restricting attention to variants with the greatest evidence of evolutionary constraint. Bernstein et al. [25] performed an *ATM* mutation screen in 708 unilateral breast cancer survivors who developed contralateral breast cancer following radiotherapy and 1397 who did not. They found that women with AT-associated *ATM* mutations treated previously with radiation had significantly greater risk of contralateral breast cancer than unexposed women either with no mutation (Gy < 1.0, RR = 2.8; Gy ≥ 1.0, RR = 3.3) or unexposed women with the same mutation (Gy < 1.0, RR = 5.3; Gy ≥ 1.0, RR = 5.8). These studies suggest that *ATM* mutations causing AT but not other *ATM* variants are associated with increased breast cancer risk in heterozygous carriers and that this risk may be increased by radiation exposure; however, these results await replication, and current guidelines do not recommend that heterozygous *ATM* mutation carriers should avoid radiation.

Some genes are uniquely associated with risk for specific breast cancer subtypes. *CDH1*, for example, is a tumor suppressor mutated in invasive lobular carcinoma of the breast (ILCB) but not ductal breast cancer [26]. Because germline mutations in *CDH1* cause hereditary diffuse gastric cancer (HDGC) and HDGC patients have a high incidence of ILCB (50 % lifetime risk) [27], Pharoah et al. [28] investigated the penetrance of *CDH1* germline mutations by performing segregation analysis in 11 families with at least three HDGC cases and a confirmed *CDH1* mutation. They estimated the cumulative risk of HDGC and ILCB by age 80 among women in these families to be 83 % and 39 %, respectively.

In summary, high-penetrance and moderate-penetrance variants in these genes collectively explain approximately 20 % of the familial risk of breast cancer [29]. Undoubtedly, continued investigation of families with multiple cancer-affected members will lead to the identification of other variants in these genes that also predispose to breast cancer, and will also shed light on the penetrance of these variants. Additionally, as the true prevalence of other cancer-predisposing syndromes becomes clear, it is likely that new associations between these syndromes and increased breast cancer risk will be discovered. Importantly, two themes are emerging from family studies that have important clinical and research implications. First, there is growing recognition that some variants causing heritable cancer syndromes when mutated biallelically also increase cancer risk among heterozygous carriers. Second, it is becoming increasingly clear that the contribution of some variants to breast cancer risk can be significantly modified by family history. Thus, there are clearly many lessons remaining to be learned through the continued study of familial breast cancer.

Genome-wide association studies and common low-penetrance risk variants

Although rare high-penetrance mutations explain much of the genetic breast cancer risk in a small number of cases, they do not shed light on the role of genetics in nonfamilial breast cancer. There is, however, considerable evidence for a strong genetic contribution to risk even for sporadic breast cancer [30]. Most investigators believe that the genetic architecture of sporadic disease is polygenic, in which susceptibility results from the aggregate effect of many low-penetrance variants. GWAS are used to search for these variants by testing for AF differences in single nucleotide polymorphisms (SNPs) genotyped across the genome in a large sample of cases and healthy controls.

The first three breast cancer GWAS were published concurrently in 2007. In one of these studies, Stacey et al. [31] used 4554 cases and 17,577 controls of predominantly European ancestry (EA) to identify two common SNPs, rs13387042 and rs3803662, with odds ratios (ORs) of 1.2 and 1.28, respectively. In the second of these GWAS, Easton et al. [32] identified five independent susceptibility loci in EA individuals using 4398 breast cancer cases and 4316 controls in a discovery stage, and more than 20,000 cases and 20,000 controls in a confirmation stage. These loci contained SNPs in or near *FGFR2*, *TNRC9*, *MAP3K1*, *LSP1*, and *H19*. Finally, Hunter et al. [33] conducted a two-stage genome-wide association study using 2921 European postmenopausal breast cancer cases and 3214 controls, and identified four intronic *FGFR2* SNPs, thereby independently replicating Easton et al.'s finding.

To date, more than 60 breast cancer GWAS have been performed [34]. As this number grows, the advantage of meta-analysis—the combining of evidence across multiple studies—becomes obvious. The first large-scale meta-analysis, conducted by Michailidou et al. [35] in 2013, employed 55,342 EA cases and 54,455 controls from nine GWAS and identified 41 new susceptibility loci. Two years later, an even larger meta-analysis, comprising more than 120,000 individuals from 52 studies, found 15 more susceptibility loci [7], bringing the current number of susceptibility loci identified by GWAS to 84.

Many variants in GWAS show consistent associations across populations; apparent population-specific associations can often be explained by differences in AF among populations. For example, in 2016 African American (AA) breast cancer patients and 2745 controls, 36 of 47 (67 %) EA breast cancer risk SNPs had ORs in AA in

the same direction, and seven (15 %) had nominally significant *P* values [36]. In East Asian women (23,637 cases and 25,579 controls), 31 of 67 EA susceptibility loci were significantly associated with breast cancer. Thus, variants contributing to sporadic breast cancer risk are likely to be similar across ancestries.

Typically, a homogeneous disease model is assumed in genetic studies, and cases are lumped together because of the increase in power that comes with increased sample size. Splitting cases by subtype is an alternative study design, with the potential to increase power despite decreasing sample size.

There have been GWAS investigating specific breast cancer subtypes based on the presence or absence of estrogen receptor (ER), progesterone receptor (PR), and/or HER2 expression [37]. Broeks et al. [38] and Figueroa et al. [39] investigated 10 validated SNPs for heterogeneity of effect size between ER+ and ER− patients. They found that seven SNPs had significantly larger effects in ER+ patients than in ER− patients, and only two SNPs remained associated with ER− breast cancer after adjusting for multiple testing. Stevens et al. [37] studied 65 validated breast cancer variants and found that while 38 were associated with both ER+ and ER− disease, the rest were unique to only one subtype. Recently, three meta-analyses of ER− breast cancer were performed that identified seven risk loci specific to this disease subtype [38, 40, 41]. Although no subtype-specific association had a particularly large effect size, these results suggest that subsetting cases based upon clinical or molecular characteristics may be an important strategy for future investigations.

While the 80+ breast cancer-associated loci identified to date have greatly expanded our knowledge of the genetics of the disease, they also have the potential to be of clinical utility. Recent studies have assessed the clinical utility of variants in GWAS using the polygenic risk score (PRS), a crude estimate of a patient's OR for disease calculated by summing the ORs for each risk allele carried by the patient [42–45]. In one study, Mavaddat et al. [42] used the PRS in a logistic regression model to demonstrate that the OR for disease differed significantly between patients with a PRS in either the highest or lowest one percentile as compared with patients with an average PRS ($OR_{1\%} = 0.32$, $OR_{99\%} = 3.36$). The discriminative accuracy of the PRS as measured by a *C*-statistic, however, was modest ($C = 0.62$). The authors estimated that the lifetime risk of cancer for women below the first and above the 99th percentile of the PRS is 3.5 % and 29.0 %, respectively. In the UK, enhanced surveillance is recommended for women with both a family history of breast cancer and a lifetime risk of breast cancer above 17 %. Using the PRS, about 8 % of UK women at this risk level—accounting for about 17 % of breast cancer cases—can be identified. Thus, risk assessment can be marginally improved by incorporating susceptibility variants from GWAS. Although variants in GWAS currently have little impact on public health, this is likely to change in the future.

Next-generation sequencing and rare variation

Taken together, these results suggest that lumping and splitting strategies for GWAS are unlikely to identify much of the missing non-high- or non-moderate-penetrance genetic contribution to breast cancer risk. One explanation for this is that GWAS are designed to identify common variants (MAF > 0.01), and to only poorly interrogate rare variants (MAF < 0.01) [46]. Thus, much of the rare variation in the genome remains uninvestigated. That rare variants may contribute significantly to risk is an appealing hypothesis because variants strongly predisposing to disease should be associated with lower fitness and be maintained at low AFs due to purifying selection. NGS can directly interrogate every position in the genome, and therefore identifies both common and rare variation. Consequently, many investigators have turned to NGS to study rare variants in complex diseases.

NGS approaches can be divided into four broad experimental strategies: sequence large numbers of unrelated patients and healthy controls to identify rare variants with AFs differing significantly between cases and controls; perform a staged study in which unrelated individuals from high-risk families meeting certain criteria (e.g., no identified mutations in *BRCA1* or *BRCA2*) are sequenced in stage one, and identified candidate risk variants are genotyped in a much larger set of cases and controls in stage two; perform a staged study in which unrelated individuals sharing critical clinical or other characteristics, such as driver somatic mutations, are sequenced in stage one, and candidate variants are subsequently genotyped in stage two; and sequence multiple related affected individuals from families enriched for disease to identify novel candidate variants and/or genes, and then interrogate these variants and genes in large case–control sets.

In the first strategy, when comparing case–control AF differences, the comparisons can be at a specific chromosomal position, in aggregate within a single gene, or in aggregate across multiple genes within a molecular or functional pathway [47]. These studies are followed up in large numbers of cases and controls investigating only genes with evidence for association.

This study design is essentially that of GWAS, except that the number of variants tested is reduced and the AF spectrum is shifted from common to rare. There are, however, statistical issues with this approach that significantly reduce its power. First, even when restricting attention to rare variation, tens of thousands of variants

are tested for association. Second, the AF of the variants tested profoundly influences their power to be detected. Consider three SNPs with the same effect size but AFs of 0.10, 0.01, and 0.001. If a study using 700 cases and 700 controls has 80 % power to identify a risk allele with AF = 0.10, then 5910 cases and 5910 controls are required for the same power to identify the risk allele with AF = 0.01, and 58,130 cases and 58,130 controls are required for the risk allele with AF = 0.001.

As an example of a study utilizing this design, Flannick et al. [48] interrogated exonic variants in 115 type 2 diabetes genes by sequencing 758 Scandinavian cases and controls selected from phenotypic extremes, and found no evidence for association. They then genotyped 71 rare variants that either had nominal significance or were predicted to affect protein structure in 13,884 individuals, and still found no evidence for association. They subsequently followed up a single variant in *SLC30A8* in an additional 33,000 individuals, and found it was nominally significant. These results are quite sobering, given the enormous sample size needed to discover only a single rare variant, as well as the ad-hoc criteria employed for variant selection.

To overcome these barriers, the second NGS study design is a staged study in which unrelated cases selected for a presumed high "genetic load" for disease are sequenced in stage one, and only variants with evidence for association are genotyped in stage two in a much larger number of cases and controls. This approach assumes that the genetic complexity in patients with high "genetic load" is considerably reduced as compared with unselected patients because of the high-penetrance mutations. One study using this approach was performed by Cybulski et al. [49], who sequenced the exomes of women with familial breast cancer from two populations harboring founder mutations, Quebec-based French-Canadians and Poles. A total of 195 patients were selected based on family history or early age of breast cancer diagnosis and no mutation in *BRCA*, *CHEK2*, *NBN*, or *PALB2*. Multiple rare truncating variants were found in *RECQL*, a previously identified cancer-related gene, in both populations. Fourteen *RECQL* exons were then sequenced in 950 *BRCA*-negative Polish and French-Canadian familial breast cancer patients. Two previously unknown germline truncating mutations were discovered in four patients; one only in Polish individuals, and the other only in French-Canadian individuals. The Polish mutation was then genotyped in 13,136 unselected Polish cases and 4702 controls, and the French-Canadian mutation in 538 French-Canadian cases with familial or early-onset breast cancer and 7136 controls. In the Polish set the risk AF was 0.23 % in cases and 0.04 % in controls ($P = 0.008$), while in the French-Canadian set the frequencies were 0.69 % and 0.014 %, respectively ($P = 3 \times 10^{-6}$). Thus, by performing a discovery stage using cases who, based on clinical characteristics and/or family history, were likely to have high-penetrance mutations, the number of hypotheses tested in a subsequent validation stage was limited, thereby minimizing the penalty for multiple testing.

The purpose of selecting cases based on their presumed genetic load is to reduce the genetic complexity of the analysis. This concept can be expanded to other clinical or genetic features, which forms the basis of the third NGS study design. For example, in cancer the presence of a tumor genome provides the opportunity to use the mutational landscape of a patient's somatic genome as supportive evidence to guide discovery of novel candidate germline risk variants, as Kanchi et al. [50] demonstrated in ovarian cancer.

Other studies have also shown relationships between the tumor genome and specific germline predispositions. Liu et al. [51], for example, demonstrated in patients with nonsmall-cell lung cancer that known functional germline polymorphisms in *EGFR* predict both higher somatic mutational burden and also specific somatic exonic microdeletions within *EGFR*. Additionally, Rausch et al. [52] found that chromothripsis (the occurrence of massive somatic chromosomal rearrangements within localized regions of the genome) in some, but not all, cancers, was associated with the presence of high-penetrance germline mutations in *TP53* that have been associated with LFS. Similarly, in breast cancer the recognition that some subtypes are enriched for germline deleterious mutations in predisposition genes can help prioritize individuals and families for NGS investigations either to identify known or to discover novel high-penetrance risk variants even in the absence of family history [53].

Finally, building upon the idea of leveraging genetically loaded individuals to improve power, the fourth NGS strategy is to sequence multigenerational families enriched for disease. Sequencing multiple affected family members as opposed to only probands simplifies considerably the analysis by limiting the number of candidate disease-causing mutations to those shared by affected family members and obligate carriers. Spurrell performed whole exome sequencing on 144 affected individuals from 54 breast cancer families with no germline mutations in known breast cancer genes to identify genes with truncating mutations shared by at least two affected family members. The study found germline mutations in *ATR*, *CHEK1*, and *GEN1* in three separate families. Another 2544 sporadic cases and 7652 controls were sequenced to identify additional rare variants in these three genes. An excess of truncating mutations in all three genes was found in cases, although the total number of cases with deleterious mutations was only 11

[54] (dissertation; not peer reviewed). A similar design was used by Kiiski et al. [55], who performed whole exome sequencing on 11 Finnish families enriched for breast cancer and identified 22 rare deleterious variants in 21 DNA repair genes. Of these, one variant in *FANCM* was significantly more common in a set of more than 3500 breast and ovarian cancer cases as compared with 2000 controls.

These examples illustrate the value of family context for genetic studies. If a gene harbors a mutation segregating in near-Mendelian fashion in one family, then there may be other highly penetrant mutations in the same gene in other families. Furthermore, if mutations identified in a family are also found in the general population, then an extension of this family-based study design is to investigate their contribution to sporadic disease risk. As noted for both the linkage and candidate gene examples discussed earlier, family history can modify the contribution to risk of even highly penetrant variants. This may suggest the existence of genetic modifiers in families that potentiate the effect of risk alleles in these families, or attenuate their affect in unaffected mutation carriers. Thus, these studies may provide the best opportunity to convert insights from rare variants into discoveries of clinical and biological significance.

An important caveat to NGS studies is that they are not agnostic; all approaches assume that variants must be filtered based on functional or population characteristics. This limits analysis only to variants with high a-priori likelihood of being functionally important, reducing the number investigated and the burden of multiple testing. Without filtering, all NGS studies would be woefully underpowered. Filters include: minor AF; functional consequence (nonsynonymous, missense, nonsense, splice-site, frame-shift indel); and functional annotation, such as predicted importance for protein function (SIFT [56], Polyphen-2 [57]), conservation across species (GERP [58], PhyloP [58, 59], and SiPhy [60]), or overall predicted importance abstracted from multiple sources (MutationTaster [61], CADD [62]). In some studies, further filtering restricts attention to genes or pathways previously implicated in disease.

Another important consideration is that not all familial disease aggregates are due to high-penetrance mutations. This is especially true for common diseases such as breast cancer, in which it is not unusual to observe familial clusters simply by chance. Additionally, nongenetic factors such as shared environment also contribute to familial risk, independent of genetics.

Risk variants in context: the role of environment

Breast cancer, like all complex diseases, results from genetic factors, environmental factors, and interactions between the two. While it is clear that the environment contributes significantly to risk, it is unclear how to incorporate environment into genetic models. Because exposures are essentially impossible to quantify for any individual, genetic association studies largely disregard them. The inevitable consequence is that cases arising via multiple distinct mechanisms are lumped and analyzed together, resulting in an overall attenuation of genetic signals. In rare instances when an etiologic exposure is known and measured, restricting case and control selection to those with the exposure is an effective strategy both to account for this exposure as a risk factor and to simplify the underlying genetic vulnerability. While this study design yields smaller studies, the fact that cases and controls are more homogeneous improves power to detect exposure-specific risk variants [63].

In practice, matching exposures is challenging and can be confounded by changes over time in the contribution of an exposure to disease risk. Nonetheless, there are opportunities to design studies controlling for exposure. Adolescents with Hodgkin lymphoma (HL) treated with ionizing radiation (IR) are at high risk for IR-induced second primary tumors, particularly breast cancer [64, 65]. Best et al. [66] hypothesized that because both cases (HL survivors who did develop a second cancer) and controls (HL survivors who did not develop a second cancer) were exposed to IR, the major factors distinguishing cases from controls were genetic. They performed discovery GWAS using just 96 cases and 82 controls, and succeeded in identifying and replicating a variant that they then demonstrated functionally regulates IR-mediated *PRDM1* induction. Importantly, although this variant was highly penetrant in the context of IR, it did not contribute to risk in the absence of IR. These results underscore the importance of environmental context in genetic studies. Success in designing studies incorporating exposures, however, is predicated upon a substantive epidemiological understanding of the role of these environmental factors in disease susceptibility.

Conclusion

The search for germline breast cancer predisposing variants has been highly successful. Linkage analysis and fine-mapping led to the discovery of *BRCA1/2* and other high-penetrance and medium-penetrance genes, which are mutated in about a quarter of all familial breast cancer cases. Eighty-four independent loci have been identified by GWAS, which account for a small but meaningful proportion of sporadic breast cancer risk. New technologies such as NGS have catalyzed a new era of discovery by making possible studies unimaginable even a few years ago.

With new technologies, however, come new challenges, perhaps foremost of which is the urgent need to rethink study design. Complex diseases such as breast

cancer lie at the interface of human genetics and epidemiology. Here, we propose that one powerful framework for future investigation may be studying the genetics of breast cancer risk in context; in particular, the context of either a homogeneous exposure or familiality. An important implication of epidemiology-guided study design is that the contribution of variants to disease is not monolithic; variants that are highly penetrant in one context may not be associated with disease in another.

Finally, genetics is predicated upon collaborative research. The sine qua non is shared access to large numbers of samples with well-annotated clinical and exposure data. However, genetics is only the starting point. Moreover, breakthroughs in breast cancer prevention and treatment can only come through functional follow-up of genetic discoveries. This is an exciting time for genetics, but only through concerted multidisciplinary efforts will the clinical promise of genetics will be realized.

Abbreviations

AA: African American; AF: Allele frequency; AT: Ataxia-telangiectasia; EA: European American; ER: Estrogen receptor; FA: Fanconi anemia; GWAS: Genome-wide association studies; HDGC: Hereditary diffuse gastric cancer; HL: Hodgkin lymphoma; ILCB: Invasive lobular carcinoma of the breast; IR: Ionizing radiation; LFS: Li–Fraumeni syndrome; MAF: Minor allele frequency; NGS: Next-generation sequencing; OR: odds ratio; PR: Progesterone receptor; PRS: Polygenic risk score; SNP: Single nucleotide polymorphism; UK: United Kingdom

Author contributions

ADS and KO wrote the manuscript with substantial contributions from MMS. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Author details

[1]Department of Pediatrics, The University of Chicago, Chicago, IL 60637, USA. [2]Department of Biology, Hamilton College, Clinton, NY 13323, USA. [3]Section of Hematology/Oncology, Department of Pediatrics, The University of Chicago, KCBD 5140, 900 East 57th Street, Chicago, IL 60637, USA.

Published online: 03 October 2016

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. Intl J Canc J Intl Canc. 2014;136(5):E359–86.
2. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. Intl J Canc J Intl Canc. 1997;71(5):800–9.
3. Singletary SE. Rating the risk factors for breast cancer. Ann Surg. 2003;237:474–82.
4. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 1994;266(5182):66–71.
5. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. Identification of the breast cancer susceptibility gene BRCA2. Nature. 1995;378(6559):789–92.
6. Thompson D, Easton D. The genetic epidemiology of breast cancer genes. J Mammary Gland Biol Neoplasia. 2004;9(3):221–36.
7. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet. 2015;47(4):373–80.
8. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990;250(4988):1684–9.
9. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science. 1994;265(5181):2088–90.
10. Smith SA, DiCioccio RA, Struewing JP, Easton DF, Gallion HH, Albertsen H, Mazoyer S, Johansson B, Steichen-Gersdorf E, Stratton M, et al. Localisation of the breast-ovarian cancer susceptibility gene (BRCA1) on 17q12-21 to an interval of < or = 1 cM. Genes Chromosomes Cancer. 1994;10(1):71–6.
11. Albertsen HM, Smith SA, Mazoyer S, Fujimoto E, Stevens J, Williams B, Rodriguez P, Cropp CS, Slijepcevic P, Carlson M, et al. A physical map and candidate genes in the BRCA1 region on chromosome 17q12-21. Nat Genet. 1994;7(4):472–9.
12. Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. Oncogene. 2006;25(43):5898–905.
13. Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg Å, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet. 2003;72(5):1117–30.
14. Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Am J Hum Genet. 1995;56(1):265–71.
15. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet. 1998;62(3):676–89.
16. Neuhausen S, Gilewski T, Norton L, Tran T, McGuire P, Swensen J, Hampel H, Borgen P, Brown K, Skolnick M, et al. Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish women affected by breast cancer. Nat Genet. 1996;13(1):126–8.
17. Malkin D, Li FP, Strong LC, Fraumeni Jr JF, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250(4985):1233–8.
18. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, van Veghel-Plandsoen M, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55–9.
19. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. 2007;39(2):165–7.
20. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med. 2014;371(6):497–506.
21. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet. 2006;38(11):1239–41.
22. Easton DF, Lesueur F, Decker B, Michailidou K, Li J, Allen J, Luccarini C, Pooley KA, Shah M, Bolla MK, et al. No evidence that protein truncating variants in BRIP1 are associated with breast cancer risk: implications for gene panel testing. J Med Genet. 2016;53(5):298–309.
23. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 2006;38(8):873–5.
24. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, Lesueur F, Byrnes GB, Chuang SC, Forey N, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. Am J Hum Genet. 2009;85(4):427–46.
25. Bernstein JL, Haile RW, Stovall M, Boice Jr JD, Shore RE, Langholz B, Thomas DC, Bernstein L, Lynch CF, Olsen JH, 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):475–83.
26. Berx G, Cleton-Jansen AM, Nollet F, de Leeuw WJ, van de Vijver M, Cornelisse C, van Roy F. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. EMBO J. 1995;14(24):6107–15.

27. Dossus L, Benusiglio PR. Lobular breast cancer: incidence and genetic and non-genetic risk factors. Breast Cancer Res. 2015;17:37.
28. Pharoah PD, Guilford P, Caldas C. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology. 2001;121(6):1348–53.
29. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet. 2007;40(1):17–22.
30. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol. 2011;35(5):310–7.
31. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2007;39(7):865–9.
32. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007;447(7148):1087–93.
33. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007;39(7):870–4.
34. GWAS catalog. 2016. https://www.ebi.ac.uk/gwas/search?query=breast cancer.
35. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45(4):353–61.
36. Feng Y, Stram DO, Rhie SK, Millikan RC, Ambrosone CB, John EM, Bernstein L, Zheng W, Olshan AF, Hu JJ, et al. A comprehensive examination of breast cancer risk loci in African American women. Hum Mol Genet. 2014;23(20): 5518–26.
37. Stevens KN, Vachon CM, Couch FJ. Genetic susceptibility to triple-negative breast cancer. Cancer Res. 2013;73(7):2025–30.
38. Broeks A, Schmidt MK, Sherman ME, Couch FJ, Hopper JL, Dite GS, Apicella C, Smith LD, Hammet F, Southey MC, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Hum Mol Genet. 2011;20(16):3289–303.
39. Figueroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC, Apicella C, Hammet F, Schmidt MK, Broeks A, et al. Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. Hum Mol Genet. 2011;20(23):4693–706.
40. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, Michailidou K, Stram DO, Beckmann L, Rhie SK, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Hum Mol Genet. 2012;21(24):5373–84.
41. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013;45(4):392–8.
42. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Canc Inst. 2015;107(5).
43. Vachon CM, Pankratz VS, Scott CG, Haeberle L, Ziv E, Jensen MR, Brandt KR, Whaley DH, Olson JE, Heusinger K, et al. The contributions of breast density and common genetic variation to breast cancer risk. J Natl Canc Inst. 2015;107(5).
44. Li J, Holm J, Bergh J, Eriksson M, Darabi H, Lindstrom LS, Tornberg S, Hall P, Czene K. Breast cancer genetic risk profile is differentially associated with interval and screen-detected breast cancers. Ann Oncol. 2016;27(6):1181.
45. Dite GS, MacInnis RJ, Bickerstaffe A, Dowty JG, Allman R, Apicella C, Milne RL, Tsimiklis H, Phillips KA, Giles GG, et al. Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged under 50 years: Australian Breast Cancer Family Registry. Cancer Epidemiol Biomarkers Prev. 2015;25(2):359–65.
46. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES. Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A. 2014;111(4):E455–64.
47. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95(1):5–23.
48. Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burtt NP, Mahajan A, Fuchsberger C, Atzmon G, Benediktsson R, et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nat Genet. 2014; 46(4):357–63.
49. Cybulski C, Carrot-Zhang J, Kluzniak W, Rivera B, Kashyap A, Wokolorczyk D, Giroux S, Nadaf J, Hamel N, Zhang S, et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet. 2015;47(6):643–6.
50. Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, Zhang Q, Koboldt DC, Xie M, Kandoth C, et al. Integrated analysis of germline and somatic variants in ovarian cancer. Nat Commun. 2014;5:3156.
51. Liu W, He L, Ramirez J, Krishnaswamy S, Kanteti R, Wang YC, Salgia R, Ratain MJ. Functional EGFR germline polymorphisms may confer risk for EGFR somatic mutations in non-small cell lung cancer, with a predominant effect on exon 19 microdeletions. Cancer Res. 2011;71(7):2423–7.
52. Rausch T, Jones DT, Zapatka M, Stutz AM, Zichner T, Weischenfeldt J, Jager N, Remke M, Shih D, Northcott PA, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell. 2012;148(1-2):59–71.
53. Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, Miron P, Olson JE, Godwin AK, Pankratz VS, Olswold C, et al. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. J Clin Oncol. 2014;33(4):304–11.
54. Spurrell C: Identifying New Genes for Inherited Breast Cancer by Exome Sequencing. Seattle: University of Washington; 2013.
55. Kiiski JI, Pelttari LM, Khan S, Freysteinsdottir ES, Reynisdottir I, Hart SN, Shimelis H, Vilske S, Kallioniemi A, Schleutker J, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci U S A. 2014;111(42):15172–7.
56. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11(5):863–74.
57. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; Chapter 7:Unit7 20.
58. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005;15(7):901–13.
59. Siepel A, Pollard KS, Haussler D. Proceedings of the 10th International Conference on Research in Computational Molecular Biology (RECOMB 2006) New methods for detecting lineage-specific selection. 2006. p. 190–205.
60. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics. 2009;25(12):i54–62.
61. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7(8):575–6.
62. 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(3):310–5.
63. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PLoS One. 2013;8(10):e76295.
64. Travis LB, Hill DA, Dores GM, Gospodarowicz M, van Leeuwen FE, Holowaty E, Glimelius B, Andersson M, Wiklund T, Lynch CF, et al. Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. JAMA. 2003;290(4):465–75.
65. Moskowitz CS, Chou JF, Wolden SL, Bernstein JL, Malhotra J, Novetsky Friedman D, Mubdi NZ, Leisenring WM, Stovall M, Hammond S, et al. Breast cancer after chest radiation therapy for childhood cancer. J Clin Oncol. 2014;32(21):2217–23.
66. Best T, Li D, Skol AD, Kirchhoff T, Jackson SA, Yasui Y, Bhatia S, Strong LC, Domchek SM, Nathanson KL, et al. Variants at 6q21 implicate PRDM1 in the etiology of therapy-induced second malignancies after Hodgkin's lymphoma. Nat Med. 2011;17(8):941–3.