Abstract

Breast cancer is a heterogeneous disease, and risk factors could be differentially associated with the development of distinct tumor subtypes that manifest different biological behavior and progression. In support of this view, there is growing evidence that known breast cancer risk factors vary by hormone receptor status and perhaps other pathologic characteristics of disease. Recent work from large consortial studies has led to the discovery of novel breast cancer susceptibility loci in genic (CASP8, FGFR2, TNRC9, MAP3K1, LSP1) and nongenic regions (8q24, 2q35, 5p12) of the genome, and to the finding of substantial heterogeneity by tumor characteristics. In particular, susceptibility loci in FGFR2, TNRC9, 8q24, 2q35, and 5p12 have stronger associations for estrogen receptor – positive (ER+) disease than estrogen receptor – negative (ER-) disease. These findings suggest that common genetic variants can influence the pathologic subtype of breast cancer, and provide further support for the hypothesis that ER+ and ER- disease result from different etiologic pathways. Current studies had limited power to detect susceptibility loci for less common tumor subtypes, such as ER- disease including triple-negative and basal-like tumors. Ongoing work targeting uncommon subtypes is likely to identify additional tumor-specific susceptibility loci in the near future. Characterization of etiologic heterogeneity of breast cancer may lead to improvements in the understanding of the biological mechanisms for breast cancer, and ultimately result in improvements in prevention, early detection, and treatment.

Expression profiling studies indicate that expression of hormone receptors is associated with consistent "molecular portraits" in tumor tissues that are established early in tumorigenesis and represent critical determinants of tumor biology (7 – 9). Accordingly, risk factors could be linked not only to the development of tumors but also to their biology and progression. Therefore, determining whether risk factor associations for breast cancer differ by morphologic and molecular characteristics of the tumors represents a critical etiologic question.

Recent discoveries have led to the identification of at least eight novel common breast cancer susceptibility loci derived primarily from genome-wide association studies, and it is expected that further genome-wide association studies will confirm additional loci in the near future. The first wave of genome-wide association studies have been conducted primarily in Caucasian populations with a predominance of ER+ disease, and it is expected that further genome-wide association studies will confirm additional loci in the near future. The first wave of genome-wide association studies have been conducted primarily in Caucasian populations with a predominance of ER+ disease, and are not well powered to identify susceptibility factors that might be more common in other ethnicities or that are specific to less common tumor subtypes. In this article we discuss the evidence for the heterogeneity of genetic associations with breast cancer risk by tumor subtypes, in particular those defined by the expression of estrogen receptor in the tumors.

Genetic Susceptibility to Breast Cancer

Although environmental factors, primarily hormonal and reproductive factors such as pregnancy history, age at...
menarche, and age at menopause, are major contributors to breast cancer risk (10), there is also strong evidence for a genetic component. First-degree relatives of breast cancer patients have approximately a 2-fold increase in the risk of developing breast cancer, and most of the excess risk is likely to be caused by genetic factors rather than shared environment (11, 12). Different approaches have led to the identification of susceptibility loci that contribute to explain the excess familial risk. The loci identified range from those rare in the population associated with very high risk (relative risk of carriers versus noncarriers of 8 to 10) to common susceptibility loci associated with small increases in risk (relative risk <2; Fig. 1).

**High-penetrance mutations.** Family linkage studies have identified rare mutations in genes associated with very high risks (high-penetrance mutations) that are responsible for inherited breast cancer syndromes. These include mutations in BRCA1 and BRCA2 causing breast and ovarian cancer syndromes, PTEN in Cowden syndrome, TP53 in Li-Fraumeni syndrome, and STK11/LKB1 in Peutz-Jegher syndrome (13, 14). In spite of the high risks, because these variants are rare in the population, they account for a relatively small percentage of the familial risk, estimated at 20% to 25% (14, 15). Tumors in BRCA1 carriers have different morphologic and immunohistochemical characteristics from those tumors occurring in BRCA2 mutation carriers or in noncarrier tumors (16–21). Specifically, breast tumors in BRCA1 mutation carriers tend to be triple negative (i.e., ER-, progesterone receptor negative, and human epidermal growth factor receptor-2 negative), high grade, and with lymphatic node invasion. These tumors are primarily of infiltrating ductal histology, and they also show a higher percentage of medullary tumors than in noncarriers. The high percentage of ER- tumors is not explained by differences in the age at onset or other tumor characteristics (17, 22), suggesting that BRCA1 tumors might originate from ER- cells. More recently, profiling studies have shown that most triple-negative tumors among BRCA1 carriers show expression patterns corresponding to the “basal-like” gene cluster (7, 21, 23, 24). These studies also noted similarities in the expression and immunohistochemical profiles of BRCA1 and basal-like nonhereditary tumors, suggesting similar origins of these tumors (7, 21, 23, 24). These observations underscore the etiologic heterogeneity of breast cancer subtypes by pathologic characteristics.

**Moderate-penetrance variants.** A combination of family-based and population-based approaches has identified relatively uncommon variants associated with modest increases in risk (moderate-penetrance variants; Fig. 1). These include the 1100delC protein-truncating variant in CHEK2 (25–27) and rare variants in ATM causing ataxia-telangiectasia (28), in the BRIP1 gene encoding a BRCA1-interacting protein (29), and the PALB2 gene encoding a BRCA2-interacting protein (30). Because of the modest increases in risk and relatively low frequency of this class of genetic variants, their contribution to familial risk is estimated to be <3% (31). Current studies have been too small to be able to evaluate differences in the association between these susceptibility loci and tumor subtypes, and additional studies are needed to address this question.

**Low-penetrance variants.** Current modeling suggests that the majority of the unexplained fraction of familial risk is likely to be explained by a polygenic model implying a combination of many individual variants with weak associations with risk (32–34). Most of these variants are likely to be common (minor allele frequency >0.05) genetic susceptibility loci associated with small increases in risk (low-penetrance variants), which are best studied in population-based studies. This class of genetic variants will be the focus of the remainder of this review.

![Fig. 1. Breast cancer susceptibility loci according to the approximate magnitude of their associated relative risk (per risk allele) and frequency of the risk allele. This figure shows that low-penetrance variants in susceptibility loci discovered to date fall into the bottom right corner (risk allele frequencies >20% and relative risk per risk allele <1.3), and that common risk alleles associated with higher relative risk (top right corners) are unlikely to exist. Although additional high-penetrance mutations in susceptibility loci are unlikely, moderate- and low-penetrance variants are likely to be discovered in the near future as the genetic coverage of whole genome scans improves for uncommon variants and the size of studies increases.](image-url)
Discovery of Common, Low-Penetrance Variants in Susceptibility Loci

The approaches to studying common genetic susceptibility factors have evolved very quickly over the last several years, owing to the completion of the sequencing of the human genome (35) and the mapping of haplotypes of a large subset of the most common genetic variation, namely, the single nucleotide polymorphism (SNP; refs. 36, 37). Rapid advances in annotating genomes in populations coupled with efficient developments in genotyping technologies together and substantial reductions in genotyping costs now enable the determination of hundreds of thousands of SNPs simultaneously. Genetic variants are determined for each individual from a source of genomic DNA (usually lymphocytes or buccal cells) using different genotyping technologies such as TaqMan assays for single SNPs or multiplexed genotyping platforms to determine several SNPs simultaneously. The technical advances have enabled investigators to move beyond evaluating a few candidate variants in key genes, to conduct more comprehensive as well as exploratory evaluation of common genetic variation in candidate pathways to cancer, and do genome-wide association studies.

Candidate Gene Approach

Over the last 15 years, many studies have used candidate gene approaches to identify common genetic susceptibility loci for breast cancer and other diseases but with somewhat disappointing results. Because of the central role of hormones and reproductive history in the development of breast cancer, genes involved in hormone biosynthesis, metabolism, and bioavailability have been prime candidates for study. DNA double-strand break repair and related cell cycle checkpoints are also important candidate pathways because of the links between these processes and high- to moderate-penetrance breast cancer genes. Other etiologic pathways evaluated include those related to known or suspected risk factors for breast cancer such as carcinogen metabolism, alcohol metabolism, and obesity; and genes involved in key carcinogenic processes such as inflammation, immunity, DNA repair, apoptosis, cell signaling, methylation, regulation of telomere length, as well as genes with frequent somatic alterations in breast tumors.

Most findings reported as “positive” from individual studies evaluating candidate genes end up failing to replicate across studies and thus are considered to be false-positive findings (38-44). The lack of replication, assuming the absence of biases, can be explained by two main factors, namely low statistical power of individual studies to detect weak associations with risk, and low prior probability of a disease association for a given variant. The latter is true even for genes considered to be strong candidates because of limited knowledge of carcinogenic processes and the functional implications of genetic variants. Therefore, large efforts aimed at replicating findings from individual studies using stringent criteria (e.g., very low P values) are required to identify variants conclusively associated with risk (45). An example of such effort is the Breast and Prostate Cancer Cohort Consortium (BPC3)1 that is pooling data from 10 prospective cohort studies to evaluate common variation in candidate genes in the steroid metabolism and insulin-like growth factor pathways in relation to the risk of breast and prostate cancers (46).

The Breast Cancer Association Consortium (BCAC)2 attempted to replicate previously reported associations for 15 candidate gene variants that had been evaluated in at least three independent studies with at least 10,000 subjects in total (47, 48). The best finding that was successfully replicated is for a coding variant (D302N) in the caspase-8 (CASP8) gene (48). The N allele was found in 13% of women of Caucasian origin, with a relative risk of 1.0.

Table 1. Established breast cancer susceptibility loci

<table>
<thead>
<tr>
<th>rs number</th>
<th>Gene</th>
<th>Chromosome</th>
<th>MAF</th>
<th>Per allele OR</th>
<th>Heterozygote, OR (95% CI)</th>
<th>Rare homzygote, OR (95% CI)</th>
<th>Trend test, P*</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1045485</td>
<td>CASP8</td>
<td>2q</td>
<td>0.13</td>
<td>0.89 (0.85-0.94)</td>
<td>0.74 (0.62-0.87)</td>
<td>1.1 x 10^-7</td>
<td>Cox et al. (48)</td>
<td></td>
</tr>
<tr>
<td>rs2981582</td>
<td>FGFR2</td>
<td>10q</td>
<td>0.38</td>
<td>1.23 (1.18-1.28)</td>
<td>1.63 (1.53-1.72)</td>
<td>2.0 x 10^-6</td>
<td>Easton et al. (59)</td>
<td></td>
</tr>
<tr>
<td>rs1219648</td>
<td>FGFR2</td>
<td>10q</td>
<td>0.39</td>
<td>1.32 (1.07-1.42)</td>
<td>1.64 (1.42-1.90)</td>
<td>1.1 x 10^-10</td>
<td>Hunter et al (60)</td>
<td></td>
</tr>
<tr>
<td>rs10941679</td>
<td>LSP1</td>
<td>5p12</td>
<td>0.24</td>
<td>1.19 (1.13-1.26)</td>
<td>Not reported</td>
<td>2.9 x 10^-11</td>
<td>Stacey et al. (62)</td>
<td></td>
</tr>
<tr>
<td>rs3803662</td>
<td>TNRC9</td>
<td>16q</td>
<td>0.25</td>
<td>1.23 (1.18-1.29)</td>
<td>1.39 (1.26-1.45)</td>
<td>10^-6</td>
<td>Easton et al. (59)</td>
<td></td>
</tr>
<tr>
<td>rs13387042</td>
<td>MAP3K1</td>
<td>2q34</td>
<td>0.50</td>
<td>1.11 (1.03-1.20)</td>
<td>1.44 (1.30-1.58)</td>
<td>1.3 x 10^-13</td>
<td>Stacey et al. (61)</td>
<td></td>
</tr>
<tr>
<td>rs13281615</td>
<td>FGFR2</td>
<td>8q24</td>
<td>0.40</td>
<td>1.06 (1.01-1.11)</td>
<td>1.18 (1.10-1.25)</td>
<td>5 x 10^-12</td>
<td>Easton et al. (59)</td>
<td></td>
</tr>
<tr>
<td>rs889312</td>
<td>TNRC9</td>
<td>8q24</td>
<td>0.40</td>
<td>1.06 (1.01-1.11)</td>
<td>1.18 (1.10-1.25)</td>
<td>5 x 10^-12</td>
<td>Easton et al. (59)</td>
<td></td>
</tr>
<tr>
<td>rs3817198</td>
<td>LSP1</td>
<td>5p12</td>
<td>0.30</td>
<td>1.06 (1.02-1.11)</td>
<td>1.17 (1.08-1.25)</td>
<td>3 x 10^-9</td>
<td>Easton et al. (59)</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Numbers are from published data cited in references shown in the study column.
Abbreviations: MAF, minor allele frequency among controls; OR, odds ratio; 95% CI, 95% confidence interval.

*P value for trend test under a log additive model.
1 The two reported SNPs have an r² of 1.0.
2 Per-allele OR calculated under a log additive model.

1http://epi.grants.cancer.gov/BPC3/
2http://www.srl.cam.ac.uk/consortia/bcac/index.html
and was associated with an estimated 12% reduction in breast cancer risk (the relative risk measured by the per-allele odds ratio was 0.88; 95% confidence interval, 0.84-0.92; \( P \) for trend = \( 10^{-7} \); see Table 1 for more details). This corresponds to an odds ratio of 1.14 (= 1/0.88) for carriers of the common D allele compared with the N allele, which is a substantially weaker association with risk compared with previously identified high- and moderate-penetrance loci such as \( BRCA1, \) \( BRCA2, \) \( CHEK2, \) and others (Fig. 1). This finding was based on a pooled analysis of 16,432 cases and 17,106 controls from 14 studies participating in BCAC. This work illustrated the value of large consortia in the replication of findings from initial studies, particularly when the strength of the association is weak. Caspase-8 is a cysteine protease that plays an important role in the initiation of apoptosis or programmed cell death in response to DNA damage (49). The functional implications of the D302N variant are unknown, and therefore it is possible that other variants in linkage disequilibrium are the causative variants. The D302N variant is very rare in Asian populations, but a 6-bp deletion polymorphism (-652 6N del) in the promoter of \( CASP8 \) has been associated with the risk of multiple cancers, including breast cancer in a Chinese population (50). This variant abolished an Sp1 transcription-binding site and is associated with decreased mRNA expression in lymphocytes, and with lower caspase-8 activity after activation-induced cell death in T-lymphocytes (50). Subsequent studies have not been able to confirm the association with breast cancer risk in Caucasian populations (51-54), nor in populations of Asian and African origin in the United States (51). A relatively small study in Italy suggested that this variant might be associated with age at diagnosis in familial breast cancer cases (53). Ongoing studies are further evaluating genetic variants in \( CASP8 \) and related
genes, and these efforts might result in the identification of additional variants associated with breast cancer risk. BCAC also reported a weak association with a coding variant (L10P) in the transforming growth factor-b (TGBF1) gene coding of a cytokine involved in the regulation of normal mammary gland development and function (48). This association was present only in progesterone receptor–negative tumors and remains to be confirmed.

Studies of candidate genes in breast cancer have evaluated a few hundred “favorite” genes, which is a small percentage of the more than 25,000 genes across the genome. In some respects, genome-wide association studies seek to identify regions or genes that should then be considered as candidate genes in subsequent studies. Previously, most candidate gene studies did not do comprehensive evaluations of genetic variation in these genes. Therefore, it is possible that additional associations can be found using this approach in large collaborative efforts. The candidate gene approach is likely to be particularly important for the discovery of relatively uncommon variants in genes with high probability of being related to disease as these are unlikely to be captured by current approaches in genome-wide association studies described further on.

### Agnostic Approach: Genome-wide Association Studies

Although genotyping costs have decreased dramatically in the last few years, it is still too expensive to genotype all known SNPs (over 10 million SNPs have been identified by the Human Genome Project with a minor allele frequency >1% in at least one population) or sequence the entire human genome. The genome-wide association studies strategy selects subsets of genetic markers that serve as surrogates for untested markers, taking advantage of the linkage disequilibrium, defined as the correlation among genetic variants located close together (refs. 55–57; Fig. 2). The current generation of SNP panels used to conduct genome-wide scans includes up to 1 million SNPs selected using data from the International HapMap Project. These SNP panels should be able to capture the majority of common genetic variation in human populations, although the extent of coverage can vary by population genetics history (58). Doing whole genome scans in large numbers of individuals can be very costly; therefore, most genome-wide association studies use multistage designs to reduce genotyping costs (56). In these designs, a relatively small proportion of samples from cases and controls are scanned in a discovery study (Fig. 3). In subsequent stages, only those markers showing the most significant associations with disease risk are genotyped in additional samples from cases and controls. Because of the large number of markers being tested, very stringent statistical criteria (e.g., $P < 10^{-7}$) are needed to confirm findings in the final stages, thus minimizing the number of false-positive findings.

Over the last year, we have witnessed an explosion of new discoveries of susceptibility loci for a wide range of diseases derived from genome-wide association studies. Three large-scale genome-wide association studies in breast cancer (59–62) have discovered seven novel genetic susceptibility loci (FGFR2, TNRC9, MAP3K1, LSP1, 8q24, 2q35, and 5p12). Each of these variants shows independent associations with risk of breast cancer, although statistical gene-gene interactions resulting in larger joint effects than expected by their individual relative risks could exist. None of these variants were in coding regions of genes, four variants were in nongenic regions (8q24, 2q35, and 5p12), and the magnitude of the association with breast cancer risk was small (Table 1).

The strongest association was for a SNP in the second intron of the fibroblast growth factor receptor 2 (FGFR2; rs number, rs2981582; Table 1). The high-risk allele was present in 38% of Caucasian populations, and women with two copies of this allele (rare homozygote genotype) had a relative risk (measured by the odds ratio) of 1.63 compared with women with two copies of the low-risk allele (common homozygote genotype; Table 1). Although this association was found by an agnostic approach, the FGFR signaling pathway has been shown to be important in mammary tumorigenesis by mouse models (63) and FGFR2, in particular, encodes a transmembrane tyrosine kinase that has been involved in mammary gland development and breast carcinogenesis (64, 65). FGFR2 is overexpressed or amplified in up to 10% of breast tumors (65–68), and its expression is associated with ER+ tumors (69), suggesting a hormone-dependent action of this gene.

Fine mapping studies of the FGFR2 region around the initial SNP discovered in genome-wide association studies identified a haplotype of eight strongly linked SNPs that could be the as-yet unidentified causative SNPs (70). Gene expression studies have shown increasing FGFR2 expression levels associated with the rare homozygote genotype, and functional studies identified the OCT1/RUNX2 binding site as the main determinant of the increased expression levels (70). This work illustrates how initial discoveries of genetic markers associated with risk in genome-wide association studies can lead to additional work aimed at studying the biological mechanisms underlying the observed associations.

The rs13281615 variant lies in a nongenic region of chromosome 8q24, more than 350 kb from the proto-oncogene c-MYC [V-myc myelocytomatosis viral oncogene homologue (avian)]. This variant was found in 40% of Caucasians, and those with the homozygous variant genotype had a relative risk of 1.18 compared with the homozygous common genotype (Table 1). This chromosomal region of 8q24 is of great interest because multiple independent variants in this nongenic region have been associated with the risk of prostate (71–79), colorectal (73, 77, 80–82), and ovarian (77) cancer. Genetic markers in this region are located in five distinct haplotype blocks, one associated only with breast cancer risk, three blocks associated with prostate cancer risk, and one block associated with prostate, colorectal, and ovarian cancer risk (73, 77). The underlying mechanism that explains the striking associations is not well understood. As mentioned above, MYC is the closest gene in this region, and functional studies suggest that it might be implicated in prostate cancer by down-
regulating the prostate tumor suppressor KLF6 gene (83). Ongoing and future association and functional studies in this region are likely to lead to a better understanding of mechanisms of carcinogenesis.

Little is known about the functions of the other loci identified in genome-wide association studies. The mechanisms of association with the other two SNPs in nongenic regions 2q35 and 5p12 are unknown. The closest gene to rs10941679 in 5p12 is MRPS30, also known as programmed cell death protein (9), or PDCD9, which is implicated in apoptosis (84) and is associated with ER+ tumors (85) and with tumor characteristics associated with good prognosis (86). In addition, the fibroblast growth factor 10 (FGF10), a breast cancer oncogene and FGFR2 ligand, is also in the vicinity of this region. The identification of SNPs in nongenic regions associated with the risk of breast cancer and other common disorders through genome-wide association studies highlights the role of intragenic regions in the regulation of transcription, and suggests that polymorphisms in noncoding RNAs might also play a role in these diseases (87).

The rs889312 variant is in a linkage disequilibrium block that contains the mitogen-activated protein kinase 3 K1 (MAP3K1) and two hypothetical genes (MGC33648 and mesoderm induction early response 1, family member 3 or MIER3). MAP3K1 forms part of the MAPK cell signaling pathway implicated in cellular response to mitogens, but MAP3K1 has not been previously linked to breast carcinogenesis. The rs383662 variant is in an area of linkage disequilibrium that contains the trinucleotide repeat containing 9 [TNRC9; also known as TOX high mobility group box family member 3 (TOX3)] and a hypothetical gene LOC643714. The function of TNRC9 is unknown although it contains a putative high mobility group, suggesting that it might function as a transcription factor. In addition, TNRC9 expression has been associated with the presence of bone metastasis derived from breast tumors (88). Rs38017198 lies in intron 10 of the lymphocyte-specific protein 1 (LSP1) gene (also known as WP43) that encodes an F-actin-bundling cytoskeletal protein that is expressed in hematopoietic and endothelial cells. LSP1 has been implicated in malignant lymphoma and Hodgkin’s disease (89), and other variants in this gene have been associated with the risk of developing non-Hodgkin’s lymphoma (90).

Because genome-wide association studies encompass dense data sets, it is critical for other bona fide investigators to access the data through a registered system to explore new hypotheses and also use the resources to confirm possible associations, as

---

**Fig. 3.** Multistage design for genome-wide association studies. In a multistage design, a large number of SNPs selected to capture most common genetic variations across the genome (genome-wide scan chip) are tested in a relatively small number of cases and controls in a “discovery study.” The SNPs showing the most significant associations with disease risk in the discovery study (e.g., P value from an association test <0.05) are retested in subsequent replication studies including large independent sets of cases and controls. In the example shown in the figure, SNPs with P values <0.001 in a first replication study are retested in a second replication study. SNPs showing strong evidence of an association with disease risk based on data from the three phases (e.g., P value from an association test <10^-7) are selected as markers for chromosomal regions likely to contain disease causing variants. Very large studies and stringent statistical criteria are necessary to have sufficient power to detect associations while minimizing the probability of false-positive findings. The selected markers in genome-wide association studies are further evaluated in fine mapping studies to identify causal variants, and in functional studies to understand the biological mechanism of the observed associations with disease. Red and blue individuals represent cases of breast cancer and controls, respectively, being tested at different stages of the design. Green and red dots in the inverted cone represent SNPs being tested at each stage. The red dots are markers for disease susceptibility alleles.

---

www.aacajournals.org

Breast Cancer Loci by Estrogen Receptor Status


8005


Downloaded from clincancerres.aacrjournals.org on April 15, 2017. © 2008 American Association for Cancer Research.
has been done by the Cancer Genetic Markers of Susceptibility initiative of the U.S. National Cancer Institute (60).5

**Heterogeneity of Genetic Associations by Tumor Estrogen Receptor Status**

Recent work from large consortial studies has shown that the associations between novel breast cancer susceptibility loci and breast cancer described above could vary by clinically important tumor characteristics. Most notably, associations with most of the susceptibility loci identified to date are evidently stronger for ER+ disease than ER- disease, although some of these differences are small and not statistically significant (Table 2). Susceptibility loci significantly modified by estrogen receptor status include FGFR2, TNRC9, 8q24, 2q35, and 5p12 (61, 62, 91). The strongest evidence is for a variant in FGFR2 that was primarily associated with ER+ disease in two separate reports (62, 91). This finding is consistent with the involvement of FGFR2 in estrogen-related breast carcinogenesis (68, 92–94), and with higher levels of FGFR2 expression in ER+ than ER- cell lines and tumors (69, 93). The variant in FGFR2 was also associated with low tumor grade independent of estrogen receptor status (91). Interestingly, the variant in the 5p12 region, which is close to the FGFR2 ligand FGF10, also shows strong evidence of an association primarily with ER+ tumors (62).

The variant in TNRC9 was found to be associated more strongly with ER+ disease by Stacey at al. based on analyses of 2,128 ER+ and 589 ER- tumors (61). A much larger study including 12,974 ER+ and 3,765 ER- cases from BCAC found an association with both tumor subtypes, although the association was slightly weaker for ER- tumors (91). Based on these data, the TNRC9 variant shows the strongest association with ER+ tumors among all loci identified to date. Finally, the variant in the 8q24 region also showed a significantly stronger association for ER+ than ER- disease in the BCAC analyses (91).

The above susceptibility loci also showed associations with progesterone receptor status, but these did not seem to be independent of estrogen receptor. It is also possible that the observed associations between susceptibility loci and estrogen receptor status reflect stronger underlying associations with other correlated tumor markers, or with particular marker profiles such as molecular subtypes previously defined by expression profiling (7, 9). For instance, triple-negative tumors (estrogen receptor negative, progesterone receptor negative, human epidermal growth factor receptor negative), which include a large percentage of basal-like tumors, are a clinically distinct subgroup of breast cancer and could have different etiology (21). Limited epidemiologic data support differences in non-genetic risk factors for triple-negative tumors compared with “luminal A” tumors (estrogen receptor positive, progesterone receptor positive, human epidermal growth factor receptor negative; refs. 21, 96, 97). In addition, initial analyses using data from studies participating in BCAC support the notion that additional heterogeneity on genetic risk factors can be detected within hormone receptor–positive or hormone receptor–negative tumors according to the expression of human epidermal growth factor receptor-2 and basal markers (epidermal growth factor receptor and basal cytokeratins; ref. 98).

The Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA),6 has recently evaluated whether variants in FGFR2 (rs2981582), TNRC9 (rs3803662), and MAP3K1 (rs889312) are associated with the risk of breast cancer in over 10,000 BRCA1 and BRCA2 mutation carriers from 23 studies (99). The minor alleles in SNPs in FGFR2 and MAP3K1 were associated with increases in risk among BRCA2 mutation carriers similar to those previously reported in noncarriers; these SNPs, however, were not associated with an increased risk in BRCA1 carriers. On the other hand, the SNP in TNRC9 was associated with the risk of both BRCA1- and BRCA2-

---

5 http://cgems.cancer.gov
6 http://www.srl.cam.ac.uk/consortia/cimba/index.html

### Table 2. Established breast cancer susceptibility loci by estrogen receptor status

<table>
<thead>
<tr>
<th>Gene/region</th>
<th>rs number</th>
<th>MAF</th>
<th>ER+ tumors (95% CI)</th>
<th>ER- tumors (95% CI)</th>
<th>Heterogeneity, P</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGFR2</td>
<td>rs2981582</td>
<td>0.38</td>
<td>1.31 (1.27-1.36)</td>
<td>1.08 (1.03-1.14)</td>
<td>10^-13</td>
<td>Garcia-Closas et al. (91)</td>
</tr>
<tr>
<td>5p12</td>
<td>rs10941679</td>
<td>0.24</td>
<td>1.27 (1.19-1.35)</td>
<td>1.05 (0.92-1.18)</td>
<td>0.004</td>
<td>Stacey et al. (62)</td>
</tr>
<tr>
<td>TNRC9</td>
<td>rs3803662</td>
<td>0.27</td>
<td>1.23 (1.19-1.27)</td>
<td>1.14 (1.09-1.24)</td>
<td>0.015</td>
<td>Garcia-Closas et al. (91)</td>
</tr>
<tr>
<td>2q35</td>
<td>rs13387042</td>
<td>0.54</td>
<td>1.22 (1.14-1.31)</td>
<td>1.07 (0.94-1.22)</td>
<td>0.0098</td>
<td>Stacey et al. (61)</td>
</tr>
<tr>
<td>8q24</td>
<td>rs13281615</td>
<td>0.41</td>
<td>1.13 (1.10-1.17)</td>
<td>1.03 (0.96-1.08)</td>
<td>0.001</td>
<td>Garcia-Closas et al. (91)</td>
</tr>
<tr>
<td>MAP3K1</td>
<td>rs889312</td>
<td>0.28</td>
<td>1.12 (1.09-1.16)</td>
<td>1.07 (1.01-1.13)</td>
<td>0.11</td>
<td>Garcia-Closas et al. (91)</td>
</tr>
<tr>
<td>LSP1</td>
<td>rs3817198</td>
<td>0.30</td>
<td>1.07 (1.04-1.11)</td>
<td>1.04 (0.99-1.10)</td>
<td>0.31</td>
<td>Garcia-Closas et al. (91)</td>
</tr>
<tr>
<td>CASP8</td>
<td>rs1045485</td>
<td>0.13</td>
<td>0.89 (0.82-0.96)*</td>
<td>0.95 (0.84-1.07)*</td>
<td>0.24</td>
<td>Cox et al. (48)</td>
</tr>
</tbody>
</table>

**NOTE:** Numbers are from published data cited in references shown in the study column. *Per-allele OR not shown in original paper. ORs are for heterozygous versus homozygous common genotype.
related tumors. Over 90% of BRCA1-related tumors are estrogen receptor negative, in contrast with BRCA2 tumors that tend to be estrogen receptor positive, similarly to tumors in noncarriers (100). Thus, these findings further support the observed heterogeneity of these genetic associations by estrogen receptor status of tumors in noncarriers reported by the BCAC (91). Unfortunately, the CIMBA report did not have detailed information on the estrogen receptor status of the tumors and thus could not evaluate if estrogen receptor status could explain the observed differences by carrier status.

An important limitation of studies of etiologic heterogeneity is the very large sample size required to study less common subtypes and the use of nonstandardized tissue collection/processing protocols and immunohistochemical assays done at different study centers. Tissue microarray blocks consist of recipient paraffin blocks that contain small tissue cores removed from targets of many individual donor paraffin blocks, thus offering an attractive method for obtaining standardized, rapid, and cost-effective immunohistochemical characterization of many tumors (101). This technique is facilitating the evaluation of etiologic heterogeneity in large epidemiologic studies. In addition, improvements in automated image analysis technologies (102) and web-based systems allowing access to pathologic images from multiple centers might facilitate the daunting task pathologists face when scoring thousands of tissue cores, while providing more standardized and quantitative measures.

Current genome-wide association studies have been conducted in breast cancer cases unselected by estrogen receptor status. Because ER+ tumors are more common than ER- tumors, these studies had better statistical power to detect SNPs preferentially associated with the more common subtype rather than less common groups, such as ER tumors or specific ER subtypes (e.g., triple-negative or basal-like disease). Ongoing genome-wide association studies including large numbers of ER+ tumors will be better suited to identify estrogen receptor-negative-specific susceptibility loci. Understanding the etiology of basal-like tumors is of particular importance because these tumors tend to occur early in life, are harder to detect by screening mammography, and are associated with poor prognosis (103).

The observed differences in genetic associations by tumor subtypes support the notion that ER+ and ER- tumors result from different etiologic pathways (some of which might be shared), rather than different stages of tumor evolution within a common carcinogenic pathway (104). Although the magnitude of the observed differences is small, and by themselves these findings are unlikely to have any immediate clinical implications, the observed differences provide clues to the biological mechanisms that underpin tumor heterogeneity.

Concluding Remarks

In the last year, at least eight novel susceptibility loci have been discovered primarily by genome-wide association studies. Theses recent discoveries are shedding light on important mechanisms in breast carcinogenesis, and provide support for the presence of etiologic heterogeneity across breast cancer subtypes. Under a polygenic model of disease susceptibility, seven of the eight low-penetration genes were estimated to explain about 5% of the genetic risk for breast cancer (31), with an additional small contribution expected from the eighth SNP in 5p12 not included in that report. Given that high-penetration genes explain 20% to 25% and moderate-penetration genes explain <3% of the genetic risk, a large number of additional variants are likely to exist (31). Each of the unknown susceptibility loci is expected to be weakly associated with risk and thus very large studies are required. Studies evaluating different ethnic groups and tumor subtypes should increase the power to find variants that are more common in these population subgroups. Finally, the study of strong candidate genes and pathways such as those already implicated in breast cancer risk, as well as the study of gene-gene and gene-environment interactions might enhance the ability to identify and characterize additional variants. However, evaluation of interactions as a means of genetic discovery can dramatically increase the chances of false-positive findings and thus will require even larger replication studies to be confirmed.

Because of the weak associations with risk, low-penetration loci are unlikely to have utility for individualized prevention strategies (31). The value of these studies is to identify novel loci in the genome that could shed light on new mechanisms of disease, and thus become targets for therapy or prevention. Although the currently identified loci provide low discriminatory accuracy to distinguish between low- and high-risk groups in the population (105), many additional loci are yet to be discovered, and the combination of loci might prove useful in population screening strategies (31).7

Current studies have focused on SNPs as markers of genetic susceptibility, and other less common forms of genetic variation such as copy number variants might also contribute to susceptibility to breast and other cancers. In addition, integrated biomarkers that reflect multiple genetic, epigenetic, and environmental challenges such as DNA repair capacity, telomere length, or methylation patterns (106) are also promising markers to identify susceptible groups in the population. Characterization of common genetic variation in breast cancer cases can also lead to the discovery of variants that influence response to therapy and survival that could prove useful in the clinical management of patients (107).

In conclusion, the discovery of disease susceptibility loci and characterization of underlying etiologic heterogeneity may lead to improvements in the understanding of the biological mechanisms for breast cancer, and ultimately result in improvements in prevention, early detection and treatment.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Dr. Mark Sherman (Division of Cancer Epidemiology and Genetics, National Cancer Institute) for his contribution to the discussions on etiologic heterogeneity of breast cancer.

7 http://www.patrocles.org
References


Breast Cancer Loci by Estrogen Receptor Status

Clinical Cancer Research

Genetic Susceptibility Loci for Breast Cancer by Estrogen Receptor Status

Montserrat Garcia-Closas and Stephen Chanock


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/14/24/8000

Cited articles
This article cites 97 articles, 24 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/14/24/8000.full.html#ref-list-1

Citing articles
This article has been cited by 34 HighWire-hosted articles. Access the articles at:
/content/14/24/8000.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.