Abstract

Triple-negative breast cancers are an aggressive subtype of breast cancer with poor survival, but there remains little known about the etiologic factors that promote its initiation and development. Commonly inherited breast cancer risk factors identified through genome-wide association studies display heterogeneity of effect among breast cancer subtypes as defined by the status of estrogen and progesterone receptors. In the Triple Negative Breast Cancer Consortium (TNBCC), 22 common breast cancer susceptibility variants were investigated in 2,980 Caucasian women with triple-negative breast cancer and 4,978 healthy controls. We identified six single-nucleotide polymorphisms, including rs2046210 (ESR1), rs8170 (19p13.1), and rs8100241 (19p13.1), significantly associated with the risk of triple-negative breast cancer.

Introduction

Triple-negative breast cancers are a biologically and clinically distinct subtype of breast cancer, defined as tumors that exhibit low or no expression of estrogen receptor (ER), progesterone receptor (PR), and HER2 (1). Women with triple-negative disease account for approximately 15% of all invasive breast cancers and are more likely to be younger, African American, have an earlier age at menarche, higher body mass index during premenopausal years, higher parity, and a lower lifetime duration of breastfeeding (2–4). In addition, triple-negative tumors are typically of higher histologic grade and are associated with more aggressive disease and poorer survival (1, 5, 6). These differences in tumor pathology, nongenetic risk factors, and survival among women with triple-negative disease suggest that the etiology of these tumors may differ from other breast cancer subtypes.
Genome-wide association studies (GWAS) have recently identified common, low-penetrance susceptibility variants that are associated with risk of breast cancer (7–16). Growing evidence suggests substantial heterogeneity by tumor subtype, defined by hormone receptor status, for associations with these single-nucleotide polymorphisms (SNP). In particular, variants in 5p12, FGFR2, 8q24, 1p11.2, 9q21.3, 10q21.2, and 11q13 are associated with the risk of developing ER-positive tumors (9–12, 14, 17, 18) but not ER-negative tumors, whereas variants in 2q35, TOX3, LSP1, MAP3K1, TFG99, and RAD51LI are associated with both ER-positive and ER-negative diseases (19). To date, no variants have been specifically associated with ER-negative or triple-negative disease. However, variants at TOX3, 2q35, and 2 distinct signals at 19p13.1 have been associated with breast cancer risk in BRCA1 mutation carriers, who predominantly develop tumors displaying an ER-negative and triple-negative phenotype (15, 20, 21). Thus, additional studies specifically investigating ER-negative and triple-negative disease are necessary to understand genetic susceptibility to these breast cancer subtypes.

Here, we report on the first Triple Negative Breast Cancer Consortium (TNBCC) study of genetic susceptibility to triple-negative breast cancer in which associations between 22 common breast cancer susceptibility loci and risk among 2,980 cases and 4,978 controls were evaluated. This comprehensive study included 21 common variants from all known susceptibility loci identified through currently published breast cancer GWAS (1p11.2, 2q35, 3p24/NEK10, 5p12/MAP3K3, MAP3K1, ER1, 8q24, 9q31.2, 10p15.1, TOX3, 11q13, RAD51LI, TOX3, 17q23/COX11, and 19p13.1) and a SNP from CASP8 identified in a candidate-gene study of CASP8 (22, 23). We show that SNPs from 4 of these loci are strongly associated with risk of triple-negative breast cancer.

Materials and Methods

Ethics statement

Study subjects were recruited on protocols approved by the Institutional Review Boards at each participating institution, and all subjects provided written informed consent.

Study populations

Samples from several triple-negative breast cancer case–control series, including 2,778 triple-negative breast cancer cases and 1,406 unaffected controls, were genotyped on the iPLEX platform. These subjects were ascertained by 22 studies in 9 different countries as follows: United States, Australia, Great Britain, Finland, Germany, Netherlands, Greece, Ireland, and Sweden. These included cases from the KBCCP and POSH cohort studies, cases and controls from the MCCS cohort study, and cases and controls from established population-based breast cancer case–control studies (BBCC, GENICA, MARIE, and SEARCH), hospital or clinic-based case–control studies (ABCS, BIGGS, LMBC, OBCS, SBCS, and RPCI), case-only studies with geographically matched controls (BBCC, KARBAC, SKDKFZS, and FCCC), and unselected cases identified in tumor collections (DFCI, ABCTB, and DEKORITOS). Data from an ongoing GWAS of triple-negative breast cancer, including cases and controls from several of the studies described earlier, and the triple-negative cases from the HEBCS GWAS along with population control data (n = 273) were also included (24). In addition, data from 4 publicly available control GWAS data sets [Wellcome Trust Case Control Consortium UK 1958 birth cohort (WTCCC), National Cancer Institute’s Cancer Genetic Markers of Susceptibility (CGEMS) project, Cooperative Health Research in...
the Region of Augsburg (KORA) study, and the Australian Twin Cohort study from the Queensland Institute of Medical Research (QIMR); n = 3,593 were used. Age distributions and years of diagnosis for individual study sites are provided in Supplementary Table S1, and these studies are described in more detail in Supplementary Material.

Pathology and tumor markers
A triple-negative breast cancer case was defined as an individual with an ER-negative, PR-negative, and HER2-negative [0 or 1 by immunohistochemical staining (IHC)] breast cancer diagnosed after age 18. Criteria used for defining ER, PR, and HER2 status varied by study. These are described in detail in Supplementary Table S2. IHC data for cytokeratin 5/6 and epidermal growth factor receptor for identification of basal tumors were not available.

Genotyping
The following 22 SNPs were genotyped on the iPLEX platform: rs11249433 (1p11.2), rs13387042 (2q35), rs4973768 (3p24), rs10941679 (5p12), rs889312 (MAP3K1), rs2046210 (ESR1), rs12662670 (ESR1, surrogate for rs9397435), rs13281615 (8q24), rs1011970 (9p21.3), rs865686 (9q31.2), rs2380205 (10p1.1), rs10509168 (10q21.2, surrogate for rs10995190), rs704010 (10q22.3), rs2981582 (FGFR2), rs3817198 (LSP1), rs614367 (11q13), rs999737 (RAD51L1), rs3803662 (TOX3), rs6504950 (17q23), rs8170 (19p13.1), and rs17468277 (tagSNP for rs6504950 (17q23), rs8170 (19p13.1), and rs816010 (10q22.3)). For 10q21.2, rs10509168 was re-genotyped as a surrogate for rs10995190 (14).

Genotype data for 22 SNPs were generated for 2,778 cases and 1,406 controls using a single multiplex on the iPLEX platform. SNPs and samples were excluded on the basis of the following criteria: an SNP call rate less than 95%, Hardy–Weinberg equilibrium (HWE) \( P < 0.01 \) among controls, and sample call rate of less than 95%. The final data set of 2,707 cases and 1,385 controls exhibited SNP call rates of more than 99%, HWE \( P > 0.01 \), and sample call rates of more than 95%.

In addition, genotype data from cases and controls included in a triple-negative GWAS were available to supplement the iPLEX genotypes. Cases from 10 study sites (ABCTB, BBC, DFCI, FCCC, GENICA, MARIE, MCBCS, MCCS, POSH, and SBCS) were genotyped using the Illumina 660-Quad SNP array. A subset of MARIE cases were genotyped using the Illumina CNV370 SNP array. HEBCS cases and controls were genotyped using the Illumina 550-Duo SNP array. GWAS data for public controls were generated using the following arrays: Illumina 660-Quad (QIMR), Illumina 550K (v1; CGEMS), Illumina 550K (KORA), and Illumina 1.2M (WTCCC). For HEBCS, population allele and genotype frequencies on 221 healthy population controls genotyped on Illumina HumanHap 370CNV in the NordicDB, a Nordic pool and portal for genome-wide control data, were obtained from the Finnish Genome Center (25). These GWAS data were independently evaluated by an iterative quality control process with the following exclusion criteria: minor allele frequency less than 0.01, call rate of less than 95%, HWE \( P < 1 \times 10^{-7} \) among controls, and sample call rate of more than 98%. When DNA was available (\( n = 1,402 \)), we re-genotyped samples from the triple-negative GWAS as part of the iPLEX study in an effort to obtain as much data as possible from a single platform. Therefore, following preferential selection of data from the iPLEX study, genotypes for an additional 273 cases and 3,393 controls were included from the GWAS data (Table 1). No GWAS genotype data were available for rs10941679 (5p12), rs2046210 (ESR1), and rs6504950 (17q23), and only partial data were available for 5 other SNPs because of the absence of these SNPs from some or all of the GWAS genotyping platforms (Table 1). As a further measure of genotype quality, genotype concordance was evaluated for the

<table>
<thead>
<tr>
<th>Table 1. Subjects by country and genotyping platform (iPLEX, GWAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country</strong></td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>United States</td>
</tr>
<tr>
<td>United Kingdom</td>
</tr>
<tr>
<td>Finland</td>
</tr>
<tr>
<td>Germany</td>
</tr>
<tr>
<td>Greece</td>
</tr>
<tr>
<td>Netherlands</td>
</tr>
</tbody>
</table>

*Study-specific distributions are shown in Supplementary Table S1.*
1,402 samples included in both the iPLEX and GWAS. Eighteen of 19 SNPs had concordance rates of more than 98% and rs8100241 showed concordance of 96.3%.

**Statistical methods**

Allele frequencies for each of the 22 SNPs included in these analyses were estimated using the iPLEX genotype data and the combined GWAS and iPLEX data for cases, controls, and all subjects (Supplementary Table S3). Associations for triple-negative breast cancer were estimated using unconditional logistic regression models. Homogeneity of ORs by country was tested using the \( F \) statistic \( (26) \), and the extent of heterogeneity was estimated by the \( I^2 \) statistic \( (27) \). All analyses were conducted using SAS version 9.2, R version 2.11.0, or Plink version 1.07.

**Results**

We evaluated 22 breast cancer susceptibility SNPs identified in breast cancer GWAS for associations with triple-negative disease using genotype data from an iPLEX study of the 22 SNPs supplemented with data from a triologically GWAS. The combined data resulted in a case–control study of 2,980 cases and 4,978 controls from 25 studies in 8 countries (Table 1). All 22 SNPs were in HWE among controls at \( P > 0.01 \). Only rs17468277, rs13387042, rs10941679, and rs614367 showed evidence of heterogeneity by country (rs17468277: \( P = 0.065 \), \( I^2 = 47.4 \% \); rs13387042: \( P = 0.037 \), \( I^2 = 53.1 \% \); rs10941679: \( P = 0.063 \), \( I^2 = 47.8 \% \); rs614367: \( 0.054 \), \( I^2 = 49.4 \% \)). Of the 22 SNPs from 20 loci, 7 were significantly associated with risk of triple-negative breast cancer (\( P < 0.05 \); Table 2). Six SNPs from 4 loci, rs2046210 (\( P = 4.38 \times 10^{-7} \)), rs12662670 (\( P = 1.13 \times 10^{-4} \)), rs999737 (\( P = 2.96 \times 10^{-4} \)), rs3803662 (\( P = 3.66 \times 10^{-3} \)), rs8170 (\( P = 2.25 \times 10^{-5} \)), and rs8100241 (\( P = 8.66 \times 10^{-6} \)), remained significant after correction for multiple testing (\( P < 2.27 \times 10^{-3} \)). Adjustments for age did not change the magnitude or significance of our results. In addition, we did not find evidence of significant interactions with age for any of the 22 SNPs.

rs2046210, located upstream of ESRI on chromosome 6q25.1, exhibited a strong association with triple-negative disease (OR = 1.29, 95% CI = 1.17–1.42; \( P = 4.38 \times 10^{-7} \); Fig. 1A), whereas rs12662670, located further upstream of ESRI, displayed a similar effect but slightly less significant association with triple-negative disease (OR = 1.33, 95% CI = 1.15–1.53; \( P = 1.13 \times 10^{-3} \); Fig. 1B). To assess the independence of these 2 ESRI SNPs, which are not correlated in HapMap subjects of European ancestry (\( r^2 = 0.09 \)), we included both SNPs in a multivariate model. rs2046210 was more strongly associated with triple-negative risk than rs12662670 (rs2046210: OR = 1.24, 95% CI = 1.12–1.38; \( P = 5.64 \times 10^{-5} \); rs12662670: OR = 1.20, 95% CI = 1.00–1.44; \( P = 0.053 \)) in this model, suggesting that rs2046210 may account in part for these 2 associations. In addition, 2 SNPs at 19p13.1, shown to have genome-wide significant associations with breast cancer in BRCA1 mutation carriers, were highly significantly associated with triple-negative breast cancer (rs8170: OR = 1.27, 95% CI = 1.17–1.38; \( P = 2.25 \times 10^{-5} \); rs8100241: OR = 0.84, 95% CI = 0.78–0.90; \( P = 8.66 \times 10^{-7} \); Fig. 1C and D). Multivariate modeling of these 2 SNPs, which are moderately correlated in HapMap subjects of European ancestry (\( r^2 = 0.74 \)), showed that rs8170 is more strongly associated with triple-negative breast cancer risk (rs8170: OR = 1.22, 95% CI = 1.10–1.34; \( P = 7.56 \times 10^{-5} \); rs8100241: OR = 0.90, 95% CI = 0.83–0.98; \( P = 0.014 \)), although both variants are retained in the model. In addition, rs3803662 (TOX3), which has been strongly associated with risk of ER-negative breast cancer (OR = 1.15, \( P = 2.1 \times 10^{-10} \); ref. 19), was associated with a 1.17-fold increase in risk of triple-negative disease (OR = 1.17, 95% CI = 1.09–1.26; \( P = 3.66 \times 10^{-5} \); Fig. 1E). Likewise, the rs999737 (RAD51L1) SNP was significantly associated with risk of triple-negative breast cancer (rs999737: OR = 0.86, 95% CI = 0.80–0.93; \( P = 2.96 \times 10^{-4} \); Fig. 1F). In contrast, rs17468277 (ALS2CR12/CASP8; \( P = 0.005 \)) was not significantly associated with triple-negative breast cancer risk after correction for multiple testing, suggesting that this result should be interpreted with caution. None of these 6 SNPs showed evidence of heterogeneity by country (Fig. 1). To further understand the influence of variants in the 6q25.1 and 19p13.1 loci on triple-negative risk, we looked for statistical interactions between the SNPs in these regions. Although there was no evidence for a statistical interaction between rs2046210 and rs1266270 (\( P = 0.820 \)) at 6q25.1, we found strong evidence of an interaction (\( P = 0.004 \)) between rs8170 and rs8100241 from 19p13.1 in a multiplicative model.

Next, we conducted a subset analysis using the iPLEX data alone (2,707 cases and 1,385 controls) for the 19 SNPs with both iPLEX and GWAS genotypes to assess the consistency of our results. Analysis of associations with triple-negative disease in the iPLEX-only data set showed that ORs for the 19 SNPs were consistent in both direction and magnitude of effect compared with the analysis using all available genotype data, although some variation in the significance of the associations was observed (Table 2). Four of the SNPs significantly associated with triple-negative breast cancer in the overall analysis retained statistical significance in the iPLEX-only analysis (rs12662670: \( P = 3.52 \times 10^{-4} \); rs3803662: \( P = 8.25 \times 10^{-3} \); rs8170: \( P = 7.30 \times 10^{-3} \); rs8100241: \( P = 1.81 \times 10^{-4} \)) after correction for multiple testing. Results were unchanged for rs2046210 from the ESRI locus, because the overall analysis was restricted to iPLEX data as a result of missing GWAS data for this variant. Finally, although the rs999737 (RAD51L1) SNP was only marginally associated with triple-negative breast cancer risk in the iPLEX-only analysis (rs999737: \( P = 0.053 \)), the estimate of effect for this SNP was consistent with the effect observed in the overall analysis.

Importantly, genotype data from a subset of these cases and controls have previously been used in association studies involving a number of these SNPs by the Breast Cancer
<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene/locus</th>
<th>Chromosome</th>
<th>Tested (minor) allele</th>
<th>Overall Cases Controls</th>
<th>OR (95% CI)</th>
<th>Published OR (95% CI)</th>
<th>iPlex Cases Controls</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11249433</td>
<td>1p11.2</td>
<td>G</td>
<td>2,976 4,968 0.27</td>
<td>0.96 (0.90–1.03)</td>
<td>2,707 1,385</td>
<td>0.54</td>
<td>0.97 (0.88–1.07)</td>
<td>1.16 (1.09–1.24)</td>
</tr>
<tr>
<td>rs17468277a</td>
<td>CASP8</td>
<td>T</td>
<td>2,979 4,977 0.005</td>
<td>0.87 (0.78–0.96)</td>
<td>2,707 1,385</td>
<td>0.16</td>
<td>0.90 (0.78–1.04)</td>
<td>0.88 (0.84–0.92)</td>
</tr>
<tr>
<td>rs13387042</td>
<td>2q35</td>
<td>G</td>
<td>2,977 4,976 0.26</td>
<td>0.96 (0.90–1.03)</td>
<td>2,705 1,384</td>
<td>0.92</td>
<td>0.99 (0.91–1.09)</td>
<td>1.20 (1.14–1.26)</td>
</tr>
<tr>
<td>rs4973768</td>
<td>SLC4A7/NEK10</td>
<td>T</td>
<td>2,960 4,974 0.24</td>
<td>1.04 (0.97–1.12)</td>
<td>2,688 1,382</td>
<td>0.21</td>
<td>1.06 (0.97–1.17)</td>
<td>1.11 (1.08–1.13)</td>
</tr>
<tr>
<td>rs10941679a</td>
<td>MRPS30/FGF10</td>
<td>G</td>
<td>2,705 1,385 0.43</td>
<td>1.04 (0.94–1.16)</td>
<td>2,705 1,385</td>
<td>0.43b</td>
<td>1.04 (0.94–1.16)</td>
<td>1.19 (1.11–1.28)</td>
</tr>
<tr>
<td>rs889312</td>
<td>MAP3K1</td>
<td>C</td>
<td>2,844 2,757 0.13</td>
<td>1.07 (0.98–1.17)</td>
<td>2,707 1,385</td>
<td>0.20</td>
<td>1.07 (0.97–1.19)</td>
<td>1.12 (1.08–1.16)</td>
</tr>
<tr>
<td>rs2046210</td>
<td>ESR1</td>
<td>A</td>
<td>2,707 1,385 4.38 × 10^-7</td>
<td>1.29 (1.17–1.42)</td>
<td>2,707 1,385</td>
<td>4.38 × 10^-7b</td>
<td>1.29 (1.17–1.42)</td>
<td>1.15 (1.03–1.28)</td>
</tr>
<tr>
<td>rs1266670</td>
<td>ESR1</td>
<td>G</td>
<td>2,707 2,759 1.13 × 10^-4</td>
<td>1.33 (1.15–1.53)</td>
<td>2,707 1,385</td>
<td>3.52 × 10^-4</td>
<td>1.37 (1.15–1.62)</td>
<td>1.18 (1.10–1.26)</td>
</tr>
<tr>
<td>rs13281615</td>
<td>8q44</td>
<td>G</td>
<td>2,841 3,413 0.79</td>
<td>0.99 (0.92–1.07)</td>
<td>2,707 1,385</td>
<td>0.70</td>
<td>0.98 (0.89–1.08)</td>
<td>1.08 (1.05–1.12)</td>
</tr>
<tr>
<td>rs1011970</td>
<td>CDKN2BAS:CDKN2A:CDKN2B</td>
<td>T</td>
<td>2,979 4,977 0.13</td>
<td>1.07 (0.98–1.17)</td>
<td>2,707 1,385</td>
<td>0.02</td>
<td>1.16 (1.02–1.31)</td>
<td>1.09 (1.04–1.14)</td>
</tr>
<tr>
<td>rs865686</td>
<td>LOC10100128567</td>
<td>G</td>
<td>2,979 4,971 0.65</td>
<td>1.02 (0.95–1.09)</td>
<td>2,707 1,385</td>
<td>0.96</td>
<td>1.00 (0.91–1.1)</td>
<td>0.89 (0.85–0.92)</td>
</tr>
<tr>
<td>rs2382025</td>
<td>ANKRD16</td>
<td>T</td>
<td>2,979 4,974 0.71</td>
<td>0.99 (0.92–1.06)</td>
<td>2,707 1,385</td>
<td>0.94</td>
<td>1.00 (0.91–1.1)</td>
<td>0.94 (0.89–0.89)</td>
</tr>
<tr>
<td>rs10509168</td>
<td>ZNF365</td>
<td>T</td>
<td>2,980 4,976 0.79</td>
<td>1.01 (0.94–1.08)</td>
<td>2,707 1,385</td>
<td>0.88</td>
<td>0.99 (0.90–1.09)</td>
<td>0.86 (0.82–0.91)</td>
</tr>
<tr>
<td>rs704010</td>
<td>ZMIZ1</td>
<td>T</td>
<td>2,964 4,963 0.80</td>
<td>0.99 (0.93–1.06)</td>
<td>2,892 1,370</td>
<td>0.99</td>
<td>1.00 (0.91–1.1)</td>
<td>1.07 (1.03–1.11)</td>
</tr>
<tr>
<td>rs2981582</td>
<td>FGFR2</td>
<td>A</td>
<td>2,707 2,756 0.24</td>
<td>0.95 (0.88–1.03)</td>
<td>2,707 1,385</td>
<td>0.64</td>
<td>0.98 (0.89–1.08)</td>
<td>1.26 (1.22–1.29)</td>
</tr>
<tr>
<td>rs3817198</td>
<td>LSP1</td>
<td>C</td>
<td>2,929 4,756 0.49</td>
<td>1.03 (0.95–1.10)</td>
<td>2,707 1,385</td>
<td>0.68</td>
<td>1.02 (0.92–1.13)</td>
<td>1.07 (1.04–1.11)</td>
</tr>
<tr>
<td>rs614367a</td>
<td>MYEOV:CCND1</td>
<td>T</td>
<td>2,926 4,749 0.17</td>
<td>1.07 (0.97–1.18)</td>
<td>2,707 1,385</td>
<td>0.12</td>
<td>1.12 (0.97–1.28)</td>
<td>1.15 (1.10–1.20)</td>
</tr>
<tr>
<td>rs999737</td>
<td>RAD51L1</td>
<td>T</td>
<td>2,978 4,977 2.96 × 10^-4</td>
<td>0.86 (0.80–0.93)</td>
<td>2,706 1,385</td>
<td>0.05</td>
<td>0.90 (0.80–1.00)</td>
<td>0.94 (0.88–0.99)</td>
</tr>
<tr>
<td>rs3803662</td>
<td>TOX3</td>
<td>A</td>
<td>2,980 4,973 3.66 × 10^-5</td>
<td>1.17 (1.08–1.26)</td>
<td>2,707 1,385</td>
<td>8.25 × 10^-4</td>
<td>1.20 (1.08–1.33)</td>
<td>1.19 (1.15–1.23)</td>
</tr>
<tr>
<td>rs6504950</td>
<td>COX1</td>
<td>A</td>
<td>2,707 1,385 0.54</td>
<td>0.97 (0.87–1.07)</td>
<td>2,707 1,385</td>
<td>0.54b</td>
<td>0.97 (0.87–1.07)</td>
<td>0.95 (0.92–0.97)</td>
</tr>
<tr>
<td>rs8170</td>
<td>C19orf62:ANKLE1</td>
<td>T</td>
<td>2,979 4,978 2.25 × 10^-8</td>
<td>1.27 (1.17–1.38)</td>
<td>2,707 1,385</td>
<td>7.30 × 10^-8</td>
<td>1.40 (1.24–1.58)</td>
<td>1.26 (1.17–1.35)</td>
</tr>
<tr>
<td>rs8100241</td>
<td>C19orf62:ANKLE1</td>
<td>A</td>
<td>2,980 4,320 8.66 × 10^-7</td>
<td>0.84 (0.78–0.90)</td>
<td>2,707 1,385</td>
<td>1.81 × 10^-6</td>
<td>0.79 (0.71–0.87)</td>
<td>0.84 (0.80–0.89)</td>
</tr>
</tbody>
</table>

*aThese SNPs showed evidence of country-based heterogeneity.

bNo additional samples included in overall analysis compared with iPlex-only.

cEstimated ORs in Europeans.
Association Consortium (BCAC). To avoid duplication and to assess the degree to which these BCAC samples influenced our results, we also conducted a subset analysis in which we excluded all cases and controls used in the BCAC studies ($n = 1,819$ cases and $n = 4,038$ controls; Supplementary Table S4). The effect estimates and significance of associations with triple-negative disease in either the iPLEX or combined analyses were not substantially modified following the removal of these cases and controls (Supplementary Table S5).

Discussion

Here, we report on the first study by the TNBC and the largest study to date of genetic susceptibility to triple-negative breast cancer, which is composed of 2,980 cases and 4,978 controls from 25 studies in 9 countries. We show that a subset of breast cancer susceptibility SNPs identified through GWAS is also associated with risk of triple-negative breast cancer. Specifically, we determined that 6 breast cancer susceptibility SNPs from 4 loci, rs2046210 (ESR1), rs12662670 (ESR1), rs999737 (RAD51L1), rs3803662 (TOX3), rs8170 (19p13.1), and rs800241 (19p13.1), are associated with risk of triple-negative breast cancer. Of these, rs8170 (19p13.1) achieved genome-wide significance ($P = 2.25 \times 10^{-8}$). Overall, these findings provide strong evidence of genetic susceptibility to triple-negative breast cancer.

We identified highly significant associations between SNPs at 6q25.1, including rs12662670 ($P = 1.13 \times 10^{-4}$) and rs2046210, which reached near genome-wide significance ($P = 4.38 \times 10^{-7}$), and risk of triple-negative breast cancer. These variants are located approximately 30 and 60 kb upstream of the first untranslated exon and 180 and 210 kb upstream of the first coding exon of ESR1, which encodes the ERα protein.

The rs2046210 SNP was originally reported in a breast cancer GWAS in Chinese women (13), where a stronger association was observed among ER-negative than among ER-positive breast cancer cases. Importantly, the magnitude of effect in this triple-negative study (OR = 1.29, 95% CI = 1.17–1.42) was identical to that reported for ER-negative breast cancer in the Chinese study (OR = 1.29, 95% CI = 1.21–1.37). In contrast, a study of women of European ancestry...
did not observe an association with breast cancer, although analyses were not stratified by ER status (28). When combined with our results, the suggestion is that this SNP may be specifically associated with triple- or ER-negative disease. The second variant in the ESRI locus rs12662670 was originally associated with breast cancer in the same study of women of European ancestry (OR = 1.12, 95% CI = 1.03–1.21) and was used as a surrogate for rs9397435, which is associated with breast cancer risk (OR = 1.15, 95% CI = 1.06–1.25) independently of rs2046210 (28). Here, rs12662670 showed a strong influence on triple-negative breast cancer risk (OR = 1.33, 95% CI = 1.15–1.53), again suggesting that variation in the ESRI locus is specifically associated with risk of ER-negative and/or triple-negative breast cancer. It remains to be determined whether a single locus represented by rs2046210 or 2 loci accounted for by rs2046210 and rs9397435 are associated with ER-negative and triple-negative breast cancer at chromosome 6q25.

Because triple-negative breast cancer is defined in part by the absence of expression of ERs, we can speculate that inherited variation may downregulate ESRI expression and promote formation of ERα-negative tumors. However, recent studies in mice have shown that the mammary stem cell compartment can be regulated by 17β-estradiol and progesterone through a paracrine-signaling mechanism from steroid receptor–positive luminal cells to steroid receptor–negative stem cells (29, 30). Thus, SNPs in the ESRI locus may promote expansion of receptor-negative precursors and subsequent development of triple-negative tumors. Interestingly, variation in the 5′ region of ESRI has been associated with an increased risk of breast cancer relapse in a British prospective cohort study (31), which was accounted for by including tumor grade and nodal status in multivariate models. Thus, the causal SNPs in this area may be associated with a more aggressive tumor phenotype.

The SNPs rs8170 (P = 2.25 × 10−5) and rs8100241 (P = 8.66 × 10−7) located at 19p13.1 were first identified both as modifiers of breast cancer risk in BRCA1 carriers (15) and as risk factors for ovarian cancer (32), as well as shown to be significantly associated with ER-negative breast cancer (15). In this study, we showed that rs8170 displayed a genome-wide significant association with triple-negative breast cancer, suggesting that we can now identify variation in the 19p13.1 locus as a risk factor for triple-negative disease. Interestingly, rs8170 attenuated the significance of rs8100241 when the SNPs were included in a multivariate regression model for breast cancer whereas both these SNPs retained significance in multivariate models evaluating effects on BRCA1-associated breast cancer and ER-negative breast cancer (15). In addition, our data suggest that these SNPs have a multiplicative effect on triple-negative breast cancer risk. Further studies are required to determine whether these SNPs represent independent signals in the 19p13.1 locus. Additional studies are also needed to identify the underlying causative genetic events in this locus and to determine whether the causative events for BRCA1, ER-negative, and triple-negative breast cancer as well as ovarian cancer are common.

These 19p13.1 variants are located in a cluster of genes including C19orf62, ANKLE1, and ABHD8. ABHD8 encodes the abhydrolase domain containing 8 protein, which is a gene of uncharacterized function, and is located about 13 kb downstream of both rs8170 and rs8100241. The SNP rs8170 is located within C19orf62, which encodes the MERIT40 protein, whereas rs8100241 is located within ANKLE1, a protein of unknown function that encodes ankyrin repeat and LEM domains. MERIT40 is the most plausible candidate in this region for breast cancer susceptibility because it is a component of the BRCA1-A complex and is required to ensure the integrity and localization of this complex during the repair of DNA double-strand breaks, specifically through the recruitment and retention of the BRCA1–BARD1 ubiquitin ligase and the BRCC36 deubiquitination enzyme (33–35). However, it remains to be determined whether the causal variants at 19p13.1 alter MERIT40 expression or function or influence other genes in the region such as ANKLE1 or ABHD8.

We also found that variants in RAD51L1 (rs999737: P = 2.96 × 10−4) and TOX3 (rs3803662: P = 3.66 × 10−5) were strongly associated with risk of triple-negative breast cancer. rs999737 (RAD51L1) was originally identified in a recent breast cancer GWAS of women of European ancestry (12). Detailed studies of breast tumors have suggested that rs999737 is associated with both ER-positive and ER-negative breast cancers, which is consistent with our findings. RAD51L1 is a member of the Rad51-like family and functions in the double-strand break repair and homologous recombination pathway (36). When coupled with the association of the 19p13.1/MERIT40 locus with triple-negative risk, the suggestion is that modification of DNA repair genes is an important mechanism involved in predisposition to triple-negative breast cancer. The SNP rs3803662, located telomeric to the gene TOX3, was also strongly associated with triple-negative breast cancer in our study (P = 3.66 × 10−5). This SNP was originally identified in 2 GWAS of breast cancer (7, 9) and has been associated with risk of developing both ER-positive and ER-negative tumors (9). The SNP is also associated with the risk of BRCA1-related breast cancers (15), which are primarily ER-negative or triple-negative. TOX3 encodes a protein containing an HMG-box that is speculated to be involved in the modification of DNA and chromatin structure (37).

Only a subset of the 22 susceptibility loci was associated with triple-negative disease in this study. This suggests that there may be heterogeneity in the predisposition loci associated with different breast tumor subtypes. However, it is important to consider whether limited statistical power may have influenced our results. Among the 16 SNPs that did not reach statistical significance in this study, the effect estimates for variants at 1p11.2, 2q35, 8q24, 9q31.2, 10p15.1, 10q21.2/ZNF365, 10q22.3/ZMIZ1, and FGFR2 either showed no evidence for association or were in the opposite direction compared with the original GWAS findings. Interestingly, 2q35 has been associated with both ER-negative (19) and BRCA1-related breast cancers (21) and was marginally significant in a smaller set of triple-negative breast cancer (19). However, we found no evidence for the association at 2q35 among
Our results are also consistent with a recent study reporting the most stringent criteria for defining triple-negative cases in a sensitivity analysis including only cases from studies with generally robust to the differences noted earlier. In addition, this analysis (Fig. 1) would suggest that our findings are with susceptibility loci that are specific to triple-negative or breast cancer may influence our ability to detect associations definition or the status of ER or PR across study sites (38, 39). Interpretation of positive immunostaining may also affect the staining in tumor cells. In addition to variability in thresholds for basis of IHC of tumors using thresholds of less than 1% of cells stained, less than 10% of cells stained, or an Allred score of 0 to minimize erroneous inclusion of HER2-positive cases. In general, cases were considered ER- or PR-negative on the basis of IHC of tumors using thresholds of less than 1% of cells stained, less than 10% of cells stained, or an Allred score of 0 to 2, which incorporates both intensity and percentage of staining in tumor cells. In addition to variability in thresholds for positivity, factors such as tissue fixation, antibody choice, and interpretation of positive immunostaining may also affect the definition or the status of ER or PR across study sites (38, 39). The resulting heterogeneity in the definition of triple-negative breast cancer may influence our ability to detect associations with susceptibility loci that are specific to triple-negative or ER-negative disease. However, we did successfully identify 6 genetic loci associated with triple-negative disease, and the lack of heterogeneity in effect estimates across study sites in this analysis (Fig. 1) would suggest that our findings are generally robust to the differences noted earlier. In addition, in a sensitivity analysis including only cases from studies with the most stringent criteria for defining triple-negative cases (<1% of cells stained positive for ER and PR, HER2 0 or 1+ on IHC), the effect estimates were very similar to those from the complete analysis for the 6 SNPs in \( ESRI \), 19p13.1, TOX3, and \( RAD51L1 \), with some attenuation of significance. Finally, it is important to note that the results of this study are specific to Caucasian women. Although greater proportions of African Americans and Latinas than do Caucasians develop triple-negative breast cancer, it is not known whether similar associations with the SNPs described here exist in these populations. Further studies are needed to address this question.

In conclusion, our study provides convincing evidence for genetic susceptibility to triple-negative breast cancer and suggests that susceptibility loci may differ by histologic breast tumor subtype, defined by the status of ER, PR, and HER2. These findings add to the evidence suggesting that these subtypes likely arise through distinct etiologic pathways. Additional studies, such as those from the BCAC, will be important for determining whether these SNPs are exclusively associated with ER-negative, triple-negative disease, or even basal breast cancer, a more refined subgroup of triple-negative tumors. Fine mapping and functional analyses of these susceptibility loci are needed to identify the causal variants and mechanisms underlying the associations with triple-negative breast cancer risk.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

Mammary Carcinoma Risk Factor Investigation (MARIE) MARIE thank Tracy Sanger and Elke Mutschelknauss for their valuable contributions and S. Behrens, R. Birr, W. Busch, U. Elber, B. Kasperzeit, N. Kneze, and K. Smit for their excellent technical assistance. Melbourne Collaborative Cohort Study (MCCS) The authors acknowledge the contribution of the MCCS investigators John L. Hopper, Dallas E. English, and Melissa C. Southey. Sheffield Breast Cancer Study (SBBC) The authors thank Helen Cramp, Dan Connolly, and Ian Brock for patient recruitment, database management, and DNA preparation, respectively. Prospective Study of Outcomes in Sporadic Versus Hereditary Breast Cancer (POSH) The authors thank the 126 participating investigators who recruited cases to the study and the NCRN for supporting recruitment to the study.

Leuven Multidisciplinary Breast Centre (LMBC) LMBC thanks Gillian Peutenier, Dominiek Smeets, and Sofie van Soest for technical assistance.

Mayo Clinic Breast Cancer Study (MCBCS) The authors thank Georgia Chenevix-Trench for her valuable contributions.

Helvetska Breast Cancer Study (HEBCS) HEBCS thanks R.N. Hansensdottir and Iria Ekkilä for their help with the patient data and samples and Drs. Päivi Heikkilä, Ari Ristimäki, Tuomas Heikkinen, Mira Heinonen, and Laura Hautala for their help with the tumor marker and pathology information and gratefully acknowledges the Finnish Cancer Registry for the cancer data. The population allele and genotype frequencies were obtained from the data source funded by the Nordic Center of Excellence in Disease Genetics based on samples regionally selected from Finland, Sweden, and Denmark.

Breast Cancer in Galway Genetic Study (BIGGS) The authors thank Drs. Gabrielle Colleran, Niall McNerney, Nicola Miller, and Prof. Michael Kerin, University Hospital Galway, for their help in collecting patient data and samples.

Amsterdam Breast Cancer Study (ABCS) The authors thank ABCS/BOSOM study collaborators, among others L.J. Van’t Veer, F.E. van Leeuwen, R. van Hien, S. Cornelissen, A. Broeks, and A.J. van den Broek, and the NKI-AVL Family Cancer Clinic, especially F.B. Hogervorst.

Australian Breast Cancer Tissue Bank (ABCTB) R.L. Balleine is a Cancer Institute New South Wales fellow.

Oulu Breast Cancer Study (OBCS) The authors thank Mervi Grip and Kari Mononen for their help with patient contacts and sample and data collection and Meeri Otsukka for assistance with sample and data handling.

Kuopio Breast Cancer Project (KRPC) KBCP is grateful to Mrs. Eija Myihainen and Mrs. Helena Krmilainen for their skillful assistance.

---

**References**

1. Kuopio Breast Cancer Project (KBCP)
2. Australian Breast Cancer Tissue Bank (ABCTB)
3. Mayo Clinic Breast Cancer Study (MCBCS)
4. Mariner Cancer Institute New South Wales fellow.
5. Oulu Breast Cancer Study (OBCS)
6. Kuopio Breast Cancer Project (KRPC)
7. Breast Cancer in Galway Genetic Study (BIGGS)
8. Mammary Carcinoma Risk Factor Investigation (MARIE)
9. Melbourne Collaborative Cohort Study (MCCS)
10. Sheffield Breast Cancer Study (SBBC)
11. Prospective Study of Outcomes in Sporadic Versus Hereditary Breast Cancer (POSH)
Grant Support

Mammary Carcinoma Risk Factor Investigation (MARIE)
The MARIE study was supported by the Deutsche Krebsfille e.V. grant 70-2892-Br 1, the Hamburg Cancer Society, the German Cancer Research Centre (DKFZ), and the DNA extraction and genotype work in part by the Federal Ministry of Education and Research (BMBF) Germany grant 01KH0402.

Gene Environment Interaction and Breast Cancer in Germany (GENICA)
The GENICA Network was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW975/5, 01KW976/8, 01KW977/7, 01KVW0114, the Robert Bosch Foundation of Medical Research, Stuttgart, Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University Tübingen, Germany (H. Brauch, Christina Justenhoven); Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany (U. Hamann); Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany (Y.-D. Ko, Christian Baisch); Institute of Pathology, Medical Faculty of the University of Bonn, Germany (Hans-Peter Fischer); and Institute for Prevention and Occupation Medicine of the German Social Accident Insurance (IPA), Bochum, Germany (Thomas Brüning, Beate Pesch, Volker Harth, and Sylvia Rabstein).

Melbourne Collaborative Cohort Study (MCCS)
The MCCS was supported by Australian NHMRC grants 209057, 251553, and 504711 and infrastructure provided by the Cancer Council Victoria.

Sheffield Breast Cancer Study (SBCS)
The SBCS was supported by the Breast Cancer Campaign (grant 2005Nov4 to A. Cox) and by Yorkshire Cancer Research core funding.

Dana Farber Cancer Institute (DFCI)
This work was supported in part by the DFCI Breast Cancer SPORE NIH P50 CA016056-32.

Central Hospital Research Fund, Academy of Finland (132473), the Finnish Cancer Society, and the Sigrid Juselius Foundation.

Fox Chase Cancer Center (FCCC)
A.K. Godwin was funded by SPORE P-50CA09368, U01CA09631, and the Eileen Stein Jacoby Fund.

Roswell Park Cancer Institute (RPCI)
Data and samples were obtained from the RPCI Data Bank and BioRepository (DBB), ref. 40, a Cancer Center Support Grant Shared Resource (F30 CA101605-32).

Städtisches Klinikum Karlsruhe und Deutsches Krebsforschungszentrum Breast Cancer Study (SKDKFZS)
The SKDKFZS study was financed by the Deutsches Krebsforschungszentrum.

Breast Cancer in Galway Genetic Study (BGGS)
E. Sawyer is funded by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Galway University Hospital (License Number: 1909).

Molecular Diagnostics Laboratory IRRP, National Centre for Scientific Research (DEMOKRITOS)
This work was supported by the Hellenic Cooperative Oncology Group research grant (HR B. BG/94) and the Greek General Secretariat for Research and Technology (GSRT) Program, Research Excellence II, funded at 75% by the European Union.

Bavarian Breast Cancer Cases and Controls (BBCC)
A. Fang was partly funded by the Dr. Mildred Scheel Stiftung of the Deutsche Krebsfille e.V.

British Breast Cancer Study (BICS)
The BBC NCRN study is funded by Cancer Research UK and Breastcancer.org and acknowledges NHS funding to the NIHR Biomedical Research Centre and the National Cancer Research Network (NCRN).

References

Common Breast Cancer Susceptibility Loci Are Associated with Triple-Negative Breast Cancer

Kristen N. Stevens, Celine M. Vachon, Adam M. Lee, et al.