Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer

Kristen S. Purrington1, Susan Slager1, Diana Eccles2, Drakoulis Yannoukakos1, Peter A. Fasching4,5, Penelope Miron6, Jane Carpenter7, Jenny Chang-Claude8, Nicholas G. Martin9, Grant W. Montgomery9, Vessella Kristensten8,9,11, Hoda Anton-Culver12, Paul Goodfellow13, William J. Tapper3, Sajjad Rafiq2, Susan M. Gerty2, Lorraine Durcan1, Irene Konstantopoulou3, Florentia Fostira3, Athanassios Vratisimos3, Paraskevi Apostolou3, Irene Konstanta4, Vassiliki Kotoula4, Sotiris Lakis5, Meletios A. Dimopoulous46, Dimosthenis Skarlos3, Dimitrios Pectasides13, George Fountzilas18, Peter A. Fasching4,5, Curtis Olswold1, Seth Slettedahl1, Jeanette E. Eckel-Passow1, Dimitrios Pectasides13, George Fountzilas18, Peter A. Fasching4,5, Lothar Haeberle5, Christine Clarke25, Ruediger Schulz-Wendtland22, Stefan P. Renner25, Drakoulis Yannoukakos3, Peter A. Fasching4,5, Celine M. Vachon1 and Fergus J. Couch1,29, *

SNP, single nucleotide polymorphism; TN, triple-negative; TNBCC, Triple

Abbreviations:

HMEC, human mammary epithelial cell; iCOGS, custom Illumina Infinium
eQTL, expression quantitative trait loci; GWAS, genome-wide association study; HER2, human epidermal growth factor receptor-2; HMEC, human mammary epithelial cell; iCOGS, custom Illumina Infinium array; LD, linkage disequilibrium; OR, odds ratio; PRS, polygenic risk score; SNP, single nucleotide polymorphism; TN, triple-negative; TNBC, Triple Negative Breast Cancer Consortium.

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Carcinogenesis vol.35 no.5 pp.1012–1019, 2014
Advance Access publication December 9, 2013

doi:10.1093/carcin/bgt404

Carcinogenesis
Triple-negative (TN) breast cancer is an aggressive subtype of breast cancer associated with a unique set of epidemiologic and genetic risk factors. We conducted a two-stage genome-wide association study of TN breast cancer (stage 1: 1529 TN cases, 3399 controls; stage 2: 2148 cases, 1309 controls) to identify loci that influence TN breast cancer risk. Variants in the 19p13.1 and PTHLH loci showed genome-wide significant associations (P < 5 × 10^{-8}) in stage 1 and 2 combined. Results also suggested a substantial enrichment of significantly associated variants among the single nucleotide polymorphisms (SNPs) analyzed in stage 2. Variants from 25 of 74 known breast cancer susceptibility loci were also associated with risk of TN breast cancer (P < 0.05). Associations with TN breast cancer were confirmed for 10 loci (LGR6, MDM4, CASP8, 2q35, 2p24.1, TERT-rs10069690, ESR1, TOX3, 19p13.1, RALY), and we identified associations with TN breast cancer for 15 additional breast cancer loci (P < 0.05: PEX14, 2q24.1, 2q31.1, ADAM29, EBF1, TCF7L2, 11q13.1, 11q24.3, 12p13.1, PTHLH, NTN4, 12q24, BRC2, RAD51L1-rs2588809, MKI1). Further, two SNPs independent of previously reported signals in ESR1 [rs12525163 odds ratio (OR) = 1.15, P = 4.9 × 10^{-12}] and 19p13.1 (rs1864112 OR = 0.84, P = 1.8 × 10^{-10}) were associated with TN breast cancer. A polygenic risk score (PRS) for TN breast cancer based on known breast cancer risk variants showed a 4-fold difference in risk between the highest and lowest PRS quintiles (OR = 4.03, 95% confidence interval 3.46–4.70, P = 4.8 × 10^{-49}). This translates to an absolute risk for TN breast cancer ranging from 0.8% to 3.4%, suggesting that genetic variation may be used for TN breast cancer risk prediction.

Introduction

Triple-negative (TN) breast cancer is a distinct histopathological subtype of breast cancer that accounts for approximately 15% of all invasive breast cancers (1,2). This disease subtype is defined by low or no expression of estrogen receptor (ER), progesterone receptor and human epidermal growth factor receptor-2 (HER2). In addition, TN tumors tend to be of higher histologic grade, more proliferative, and have medullary and metastatic features (1,3). Women with TN tumors are more likely to be BRCA1 mutation carriers, young or premenopausal, African American or Hispanic ethnicity, and experience higher rates of disease recurrence and progression, especially within the first 3 years following treatment, compared with other breast cancer subtypes (4). TN breast cancer is also associated with low socioeconomic status, an earlier age at menarche, higher body mass index during premenopausal years, higher parity and lower lifetime duration of breast feeding (1,5).

In addition to these epidemiologic factors, several common genetic variants have been established as risk factors for TN breast cancer (6). Among these, 19p13.1 (7), TERT-rs10069690 (8) and MDM4 (9) are specific to TN breast cancer, such that these loci are not associated with risk of ER-positive or ER-negative, HER2-positive breast cancer. Four other loci (RALY-rs12525163, LGR6-rs2588809, FTO-rs11075995) associated with ER-negative but not ER-positive breast cancer (9,10) may also influence TN breast cancer risk. More recently, a large study by the Breast Cancer Association Consortium identified 46 additional common breast cancer susceptibility loci (11–13). Although 26 of these loci were associated with ER-negative as well as ER-positive breast cancer, the influence of the loci on TN breast cancer and other histopathological subtypes of breast cancer has not yet been assessed.

Given the substantial heterogeneity in genetic risk profiles for different breast cancer subtypes that we and others have demonstrated (14–17), we hypothesized that additional genetic variants for TN breast cancer remain to be identified. These may include variants that could not be detected by previous breast cancer genome-wide association studies (GWAS) conducted predominantly with ER-positive breast cancer cases, and perhaps a subset of the 42 breast cancer hits recently identified by the Breast Cancer Association Consortium. In addition, recent evidence has shown that risk loci are often complex and may contain multiple independent risk-associated variants that influence different subtypes of breast cancer (11–13). Here, we present results from a comprehensive analysis of genetic variants and TN breast cancer within the Triple Negative Breast Cancer Consortium (TNBCC), including a two-stage GWAS of TN breast cancer, examining the contributions of known breast cancer risk loci to TN breast cancer in terms of overall associations, independent signals and expression quantitative trait loci (eQTL), and estimating the cumulative effect of all common genetic risk factors on TN breast cancer risk.

Materials and methods

Ethics statement

Study participants were recruited under protocols approved by the institutional review board at each institution and all subjects provided written informed consent.

Study participants: TNBCC

TNBCC subjects included in this analysis were recruited by 22 studies in 7 different countries (Supplementary Table S1, available at Carcinogenesis Online). In addition, data from four publicly available control GWAS data sets (Wellcome Trust Case Control Consortium UK 1958 Birth Cohort, National Cancer Institute’s Cancer Genetic Markers of Susceptibility project, Cooperative Health Research in the Region of Augsburg study and the Australian Twin Cohort study from the Queensland Institute of Medical Research; n = 3180) were utilized. These studies are described in more detail in Supplementary Material, available at Carcinogenesis Online, and have been described in detail elsewhere (8,10,14).

Pathology and tumor markers

A TN breast cancer case was defined as an individual with an ER-negative, progesterone-receptor-negative and HER2-negative (0 or 1 by immunohistochemical staining) breast cancer diagnosed after age 18 years. Criteria used for defining ER, progesterone receptor and HER2 status varied by study and have been described previously (8,10,14).

TN breast cancer GWAS

Stage 1 of the TNBCC GWAS has been described previously (8,10,14). Briefly, 1529 TN breast cancer cases and 3399 country-matched controls from 10 study sites were genotyped using the Illumina 660-Quad single nucleotide polymorphism (SNP) array, CNV370 SNP array and 550-Duo SNP array (10). GWAS data for public controls were generated using the Illumina 660-Quad (Queensland Institute of Medical Research), Illumina 550 (v1: National Cancer Institute’s Cancer Genetic Markers of Susceptibility), Illumina 550
(Cooperative Research in the Region of Augsburg) and Illumina 1.2M
(Wellcome Trust Case Control Consortium UK 1958 Birth Cohort). Genotype
data from the various GWAS were independently evaluated by an iterative
quality control process as described previously (10). Common SNP genotypes
were imputed to HapMap phase 2 (release 21). Quantile–quantile plots showed
no substantial evidence for cryptic population substructure or differential
genotype calling between cases and controls. We excluded all SNPs with a minor
allele frequency <0.05, imputation quality score <0.5 and effect size (β) with
absolute value <0.3.

TN breast cancer iCOGS (stage 2) genotyping

The design of the custom Illumina Infinium array (iCOGS) array (211 155
SNPs) and genotyping methods has been described previously (11). Briefly,
samples were genotyped as part of the COGS project using the iCOGS array
at two genotyping centers (Mayo Clinic and Genome Quebec). In this analy-
sis, 1263 cases and 1105 controls from the TNBCC were genotyped on the
iCOGS array at the Mayo Clinic, and 885 cases and 204 controls were geno-
typed at Genome Quebec. A total of 4628 from the 6087 TNBCC GWAS SNPs
provided for the iCOGS array yielded high-quality genotype data. A total of
147 762 SNPs from the iCOGS array overlapped with the TNBCC stage 1
GWAS data.

Complementary DNA-mediated annealing, selection, extension and ligation
expression data

Expression profiles were generated for a total of 702 TN tumors (Supplementary Table S2, available at Carcinogenesis Online) using the Illumina whole genome complementary DNA-mediated annealing, selection, extension and ligation assay (v4.0). Tumor samples were either whole 10 μm sections or 1 mm cores from formalin-fixed paraffin-embedded tumor blocks. Whole sections were macrodissected to select the tumor region on the slide, guided by a pathologist-read hematoxylin and eosin stained slide from the same block. RNA was extracted using the Roche High Pure RNA Isolation Kit (Indianapolis, IN). Samples were plated randomly by study on 96-well plates with two universal human reference samples and two duplicate tumor RNA samples. Complementary DNA-mediated annealing, selection, extension and ligation expression profiling was performed by the Mayo Clinic Medical Genome Facility Gene Expression Core (Rochester, MN).

<table>
<thead>
<tr>
<th>SNP</th>
<th>G/I</th>
<th>Chr</th>
<th>Position</th>
<th>Locus</th>
<th>Allele</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6781914</td>
<td>G</td>
<td>1</td>
<td>200453799</td>
<td>LGR6</td>
<td>A</td>
<td>0.90</td>
<td>0.84–0.97</td>
<td>3.31×10⁻⁶</td>
</tr>
<tr>
<td>rs4235710</td>
<td>G</td>
<td>1</td>
<td>202785465</td>
<td>MDM4</td>
<td>C</td>
<td>1.19</td>
<td>1.11–1.29</td>
<td>4.00×10⁻⁶</td>
</tr>
<tr>
<td>rs13387042</td>
<td>G</td>
<td>2</td>
<td>217614077</td>
<td>2q35</td>
<td>G</td>
<td>0.93</td>
<td>0.87–1.00</td>
<td>0.049</td>
</tr>
<tr>
<td>rs12710069</td>
<td>G</td>
<td>1</td>
<td>19184284</td>
<td>2p24.1</td>
<td>A</td>
<td>1.11</td>
<td>1.04–1.19</td>
<td>3.51×10⁻⁵</td>
</tr>
<tr>
<td>rs10069690</td>
<td>G</td>
<td>1</td>
<td>1332790</td>
<td>TERT</td>
<td>A</td>
<td>1.24</td>
<td>1.14–1.34</td>
<td>1.45×10⁻⁴</td>
</tr>
<tr>
<td>rs2736108*</td>
<td>G</td>
<td>1</td>
<td>1504088</td>
<td>TERT</td>
<td>T</td>
<td>0.77</td>
<td>0.69–0.87</td>
<td>8.33×10⁻⁴</td>
</tr>
<tr>
<td>rs3757318</td>
<td>G</td>
<td>6</td>
<td>151955806</td>
<td>ESR1</td>
<td>A</td>
<td>1.33</td>
<td>1.17–1.51</td>
<td>0.89b</td>
</tr>
<tr>
<td>rs2046210</td>
<td>G</td>
<td>6</td>
<td>151990059</td>
<td>ESR1</td>
<td>A</td>
<td>1.16</td>
<td>1.08–1.24</td>
<td>5.26×10⁻⁶</td>
</tr>
<tr>
<td>rs3803662</td>
<td>G</td>
<td>16</td>
<td>51143842</td>
<td>TOX3</td>
<td>A</td>
<td>1.09</td>
<td>1.01–1.17</td>
<td>0.022</td>
</tr>
<tr>
<td>rs8170</td>
<td>G</td>
<td>19</td>
<td>17250704</td>
<td>19p13.1</td>
<td>C</td>
<td>1.26</td>
<td>1.16–1.37</td>
<td>1.26×10⁻⁶</td>
</tr>
<tr>
<td>rs2363956</td>
<td>G</td>
<td>19</td>
<td>17255124</td>
<td>19p13.1</td>
<td>C</td>
<td>0.82</td>
<td>0.77–0.88</td>
<td>2.35×10⁻⁴</td>
</tr>
</tbody>
</table>

Newly identified TN associations

- rs616488 | G   | 1   | 10488802 | PEX14 | G     | 0.91 | 0.85–0.98  | 9.73×10⁻⁴ |
- rs489887 | G   | 2   | 12096152 | 2q14.2 | A    | 0.89 | 0.79–1.00  | 0.041   |
- rs2016394 | G   | 1   | 127681217 | 2q31.1 | A    | 1.10 | 1.03–1.18  | 6.90×10⁻³ |
- rs6828523 | G   | 4   | 167603001 | ADAM9 | A    | 0.84 | 0.75–0.93  | 1.33×10⁻³ |
- rs1432679 | G   | 5   | 158176661 | EBF1  | G    | 1.10 | 1.02–1.17  | 8.62×10⁻³ |
- rs7904519 | G   | 10  | 114763917 | TCF7L2 | G    | 1.12 | 1.05–1.20  | 9.95×10⁻⁴ |
- rs5903072 | G   | 11  | 65399642 | 11q13.1 | A     | 0.92 | 0.86–0.99  | 0.024   |
- rs11820646 | G   | 11  | 12868368 | 11q24.3 | A    | 0.92 | 0.86–0.98  | 0.016   |
- rs21422552 | G   | 12  | 14305198 | 12p13.1 | C    | 1.13 | 1.04–1.21  | 2.70×10⁻³ |
- rs10771199 | G   | 12  | 28046347 | PTHHL | G    | 0.72 | 0.64–0.80  | 1.55×10⁻³ |
- rs17356907 | G   | 12  | 94551890 | NTN4  | G    | 0.90 | 0.84–0.97  | 7.55×10⁻⁷ |
- rs1292011 | G   | 12  | 114320905 | 12q24 | G    | 1.08 | 1.01–1.16  | 0.035   |
- rs11571833 | G   | 13  | 31870626 | BRCA2 | T    | 1.44 | 1.05–1.96  | 0.023   |
- rs2588809 | G   | 14  | 67730181 | RAD51L | A    | 0.91 | 0.83–1.00  | 0.043   |
- rs6001930* | G   | 22  | 39206180 | MLK1  | C    | 1.21 | 1.02–1.45  | 0.025   |

1Genotyped in stage 2 only on the iCOGS platform (2148 cases, 1309 controls).
2ER-negative breast cancer risk results for rs2736108 from Bojesen et al. (12).
GWAS and known breast cancer risk loci in TN breast cancer

Bonferroni correction for 4785 tests, an additional five SNPs in MDM4, ESR1, PTHLH and 19p13.1 were significantly associated with risk of TN breast cancer (Supplementary Table S4, available at Carcinogenesis Online). Known associations between TN breast cancer and variants in the MDM4 and ESR1 loci (7,9,14) were also confirmed. The 10 SNPs with the lowest P-values not located in known breast cancer loci are shown in Supplementary Table S5, available at Carcinogenesis Online.

Known breast cancer susceptibility loci

Next, we evaluated whether any known breast cancer susceptibility SNPs that were genotyped or imputed in the combined TNBCC data were associated with risk of TN breast cancer (Supplementary Tables S3 and S6, available at Carcinogenesis Online). Genotype data were available for 74 of the 78 known breast cancer risk SNPs (Supplementary Table S3, available at Carcinogenesis Online). Of these, a total of 26 SNPs were associated with risk of TN breast cancer at P < 0.05 (Table I). These included 11 SNPs in the 2q35, LGR6, MDM4, TERT, ESR1, TOX3 and 19p13.1 loci that were previously associated with TN breast cancer. Of these, rs2588809 in the RADS1 locus replaced rs999737 from earlier studies as the SNP most significantly associated with TN breast cancer (Table I). A further 15 SNPs at the PEX14, 2q14.2, 2q31.1, ADAM29, EBF1, TCF7L2, 11q13.1, 11q24.3, 12p13.1, NTN4, PTHLH, 12q24, BRCA2 and MLK1 loci showed associations with TN breast cancer risk, which have not previously been described (Table I). In contrast, SNPs in CASP8, MAP3K1 and LSP1, which had been marginally associated with TN breast cancer in other studies (6), were not associated with TN disease in this combined analysis. Furthermore, the FTO locus that was recently associated with ER-negative disease (9) was not significantly associated with TN breast cancer in this study (rs11075995 OR = 1.08, 95% CI 1.00–1.17, P = 0.065).

Two of the TN breast cancer risk loci we identified contained additional SNPs with lower P-values for TN breast cancer than the reported SNP (ESR1 and PEX14; Supplementary Table S7a, available at Carcinogenesis Online). In 1000 genomes data from Caucasians (19), these new SNPs were in high linkage disequilibrium (LD) with the originally reported SNPs suggesting that the additional SNPs better capture the associations with TN breast cancer. Additionally, although the reported SNP in the CASP8 locus was not associated with TN breast cancer risk, another highly correlated SNP (rs37371171; R² = 0.93) was significantly associated with risk (P = 1.0 × 10⁻⁴) (Supplementary Table S7b, available at Carcinogenesis Online). Finally, a SNP in the RALY locus, for which the reported SNP was not genotyped in this study, was significantly associated with TN risk (rs6142050 P = 3.8 × 10⁻³, Supplementary Table S7c, available at Carcinogenesis Online). The RALY SNP was in high LD with the reported SNP in these regions.

To better understand the patterns of risk associated with genetic variation in these TN-associated loci, we looked for independent signals in each locus by adjusting each SNP in a 250 kb region for the SNP with the lowest P-value. We found evidence for additional independent associations in the 19p13.1 locus (Supplementary Figure S1, available at Carcinogenesis Online) and the ESR1 locus (Supplementary Figure S2, available at Carcinogenesis Online). In a multivariable model for 19p13.1, including rs8100241 and rs1864112, both SNPs remained strongly associated with risk of TN breast cancer (Table II). The newly identified rs1864112 is not in LD with
rs8100241 ($R^2 = 0.025$) or rs8170 ($R^2 = 0.093$). Using data from the ENCODE project (20), we found that rs1864112 is located in a region overlapping a DNase1 hypersensitivity site and promoter-associated histone mark (H3KMe1) site in primary human mammary epithelial cells (HMEC), indicating that this SNP may play a role in transcriptional regulation. In ESR1, both rs9397437 and rs12525163 were associated with TN risk, with the significance of the association for rs12525163 increasing in the multivariate model (Table II). This SNP is not in LD with either of the ESR1 SNPs previously associated with breast cancer risk (rs9397437, $R^2 = 0.005$; rs2046210, $R^2 = 0.021$) and does not overlap with any DNase hypersensitivity, H3K4Me1 or H3K4Me3 sites. These data provide evidence for two novel TN risk SNPs in 19p13.1 and ESR1.

eQTL for TN risk loci

To better understand the potential biological mechanisms that underlie the associations between SNPs in the 25 loci (Tables I and II, Supplementary Table S7b and c, available at Carcinogenesis Online) and risk of TN breast cancer, we conducted an eQTL analysis. Genome-wide messenger RNA expression data were available for 578 TN cases from corresponding clinically defined TN breast tumors, of which 62 were excluded because of ESR1 expression in the tumors (see Materials and methods), for a total of 516 TN cases included in the eQTL analysis (Supplementary Table S2, available at Carcinogenesis Online). We then examined each of the 30 SNPs present in the 25 TN loci of interest (Tables I and II, Supplementary Table S7b and c, available at Carcinogenesis Online) for associations with gene expression. We found evidence for 51 cis-associations with the 30 TN risk SNPs ($P < 0.05$) (Supplementary Table S8, available at Carcinogenesis Online), involving 46 genes in the 25 loci. Functional annotation of the eQTL SNPs by HaploReg (21) showed that eQTL SNPs were more likely located in normal mammary epithelial cell enhancer elements (HMEC: 9 observed versus 3.1 expected, $P = 3.6 \times 10^{-10}$) and DNase hypersensitivity sites (HMEC: 7 observed versus 1 expected, $P = 7.5 \times 10^{-5}$).

A recent study functionally annotated SNPs in high LD ($R^2 > 0.5$) with 71 known breast cancer risk SNPs (22) using histone modification ChIP-seq and DNase-seq data published as part of the ENCODE project (20), formaldehyde-assisted isolation of regulatory elements data and publically available eQTL data. Twenty-three of the 25 TN risk loci we describe here were included in this report (Supplementary Table S9, available at Carcinogenesis Online); among these, 8 (34.8% in TN versus 26.8% overall) had high-LD SNPs in transcription start site regions, 17 (73.9% in TN versus 77.5% overall) had high-LD SNPs in enhancers and 6 (26.1% in TN versus 22.5% overall) had high-LD SNPs in exons, suggesting a slight enhancement for TN risk SNPs in transcription start site regions. The vast majority of functional SNPs identified by Rhie et al. (22) were not genotyped or imputed in our data. The functional SNPs rs633800 and rs11227311 in the 11q13.1 locus were associated with CTSW expression, which we also observed with the correlated index SNP, rs9303072 (Supplementary Table S8, available at Carcinogenesis Online).

We next analyzed all other SNPs in the 25 TN risk loci for eQTLs (within 1 Mb flanking the top risk SNP) and identified 41 candidate cis-eQTLs in 14 TN risk loci, involving 35 unique SNPs and 26 unique genes, based on a 10% false discovery rate threshold (Supplementary Table S10, available at Carcinogenesis Online).

The 35 eQTL SNPs were enriched in HMEC enhancers (6 observed versus 1.9 expected, $P = 0.012$) and mammary ductal adenocarcinoma DNase hypersensitivity sites (T47D: 2 observed versus 0.4 expected, $P = 0.049$). Notably, the MDM4, TERT and 19p13.1 TN-specific risk loci contained cis-eQTLs (Supplementary Table S10, available at Carcinogenesis Online). Among these 35 eQTL SNPs, 8 were associated with CTSW expression and were in moderate LD ($0.084 \leq R^2 \leq 0.516$) with synonymous exonic mutations (Supplementary Table S11, available at Carcinogenesis Online), SNPs in transcription start site regions (Supplementary Table S12, available at Carcinogenesis Online) and SNPs in enhancers (Supplementary Table S13, available at Carcinogenesis Online) identified by Rhie et al. (22). No other eQTL SNPs we identified were correlated with putative functional SNPs.

Sensitivity analysis

We conducted a sensitivity analysis of all 30 TN risk SNPs identified in this study (Tables I and II, Supplementary Table S7b and c, available at Carcinogenesis Online) to evaluate the influence of potential misclassification with respect to ER status. We first examined the 30 SNPs in 578 TN cases with expression data and 4638 country-matched controls. The ORs for these SNPs were very similar to the ORs observed in the overall TN analysis (Supplementary Table S14, available at Carcinogenesis Online), although the reduction in sample size produced some variability. We then repeated the analysis after excluding 62 TN cases because of ESR1 expression in the tumors. All ORs were in the same direction and similar in magnitude for the majority of these SNPs, with the exception of 2q14.2 and ADAM29 moving slightly closer toward the null. Although the numbers are low, the results further strengthen the evidence that these 30 SNPs are associated with TN breast cancer risk.

Polygenic risk score

These results provide strong evidence that at least 24 of the 74 known breast cancer susceptibility SNPs are individually associated with risk of TN breast cancer (Table I). We implemented a PRS to approximate the combined effect of these SNPs on risk of TN disease. The PRS was calculated using all reported SNPs in known breast cancer loci for which genotype data were available ($n = 74$; Supplementary Table S3, available at Carcinogenesis Online), both to avoid bias from data-driven SNP selection and to account for SNPs that may be associated with TN risk that did not achieve significance in this study due to limited study size. Compared with the median quintile, an individual in the first or second quintile of the PRS was 0.51-fold or 0.76-fold less likely to have TN breast cancer, respectively (Table III). In contrast, an individual in the fourth or fifth quintile of the PRS was 1.29-fold or 2.05-fold more likely to have TN breast cancer compared with subjects in the median quintile. Further, our data show that there is more than 4-fold difference in risk comparing those in the highest versus lowest quintiles (Supplementary Table S15, available at Carcinogenesis Online). The receiver operating characteristic curves for predicting TN breast cancer using the 74 SNP PRS produced an AUC of 0.64 (95% CI 0.63–0.65; Supplementary Figure S3, available at Carcinogenesis Online). Applying the PRS to the population-based cumulative risk (up to age 90 years) of TN breast cancer among Caucasian women, defined as approximately 1.8% (see Materials and methods), yielded an estimated cumulative risk of TN breast cancer.

### Table III. PRS for TN breast cancer

<table>
<thead>
<tr>
<th>PRS quintile</th>
<th>74 SNPs</th>
<th>30 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintile definitions</td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>1</td>
<td>PRS ≤ 0.24</td>
<td>0.51</td>
</tr>
<tr>
<td>2</td>
<td>0.24 &lt; PRS ≤ 0.58</td>
<td>0.76</td>
</tr>
<tr>
<td>3</td>
<td>0.58 &lt; PRS ≤ 0.86</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>0.86 &lt; PRS ≤ 1.24</td>
<td>1.29</td>
</tr>
<tr>
<td>5</td>
<td>1.24 &lt; PRS</td>
<td>2.05</td>
</tr>
</tbody>
</table>
of 3.4% for women in the highest PRS quintile and 0.8% for women in the lowest PRS quintile (Figure 1).

To better understand how the additional TN risk SNPs reported in this study contribute to cumulative risk beyond the 74 overall breast cancer variants, the PRS was recalculated using all 30 TN risk SNPs identified in this study (Tables I and II, Supplementary Table S7b and c, available at Carcinogenesis Online). Estimates were slightly stronger for each PRS quintile compared with the 74 SNP PRS (Table III), and the discriminatory accuracy of the 30 SNP PRS was comparable with the 74 SNP PRS (Supplementary Figure S3, available at Carcinogenesis Online). This suggests that the identification of additional TN risk loci may improve the stratification of cumulative risk estimates for TN breast cancer (Supplementary Figure S4, available at Carcinogenesis Online). Given the complexity of known breast cancer risk loci such as CCND1 and TERT (12,13), further studies involving extensive fine mapping, haplotyping and functional characterization are needed for full understanding of the relationship between genetic variation in these loci and risk of TN breast cancer.

To gain some insight into whether the TN risk SNPs we identified have stronger effects for TN breast cancer compared with ER-negative breast cancer, we compared 25 of the SNPs in our combined analysis for which data were available from a recent ER-negative meta-analysis (9). As expected, stronger ORs were observed in our TN study compared with the ER-negative study for MDM4, TERT (rs10069690) and 19p13.1 (Table I), which have previously been shown to be TN-specific loci (7–9). In addition, stronger ORs were observed in our TN study for 2q14.2, ESR1, TCF7L2, 11q13.1, 12p13.1 and PTHLH in TN compared with the ER-negative study. Furthermore, four of the TN loci (2q31.1, ADAM29, 12q24 and RAD51L1-rs2588809) had no reported association with ER-negative breast cancer. Studies that directly compare ER-negative, non-TN to TN breast cancer are required to determine whether any of these loci are TN specific.

Discussion

In this report, we present results from the first 2-stage GWAS of TN breast cancer in Caucasian women. Variants in the PTHLH and 19p13.1 loci showed genome-wide significant associations (P < 5.0 × 10⁻⁸) with TN disease (Tables I and II). Ten SNPs with near-genome associations with TN breast cancer (Supplementary Table S5, available at Carcinogenesis Online) warrant follow-up in larger studies of TN breast cancer. In addition, 26 of 74 known overall breast cancer risk SNPs were associated with TN breast cancer (Table I, Supplementary Table S6, available at Carcinogenesis Online). Specifically, this study confirmed TN associations with SNPs in 10 loci (LGR6, MDM4, CASP8, 2q35, 2p24.1, TERT-rs10069690, ESR1, TOX3, 19p13.1 and RALY) and identified TN associations with 15 other loci. Furthermore, two novel signals that are independent of previously known risk-associated SNPs were identified in the ESR1 and 19p13.1 loci (Table II). Given the complexity of known breast cancer risk loci such as CCND1 and TERT (12,13), further studies involving extensive fine mapping, haplotyping and functional characterization are needed for full understanding of the relationship between genetic variation in these loci and risk of TN breast cancer.

Fig. 1. Cumulative incidence of TN breast cancer stratified by 74 SNP PRS. The effect of the 74 SNP PRS on cumulative risk of TN breast cancer among Caucasian women, stratified by PRS quintile, is shown. The population-based cumulative risk curve is shown as a solid black line, and the first through fifth quintile-specific cumulative risk estimates are shown as indicated by labels.
In addition, we have provided evidence for SNP-mediated regulation of gene expression in these TN risk loci through cis-eQTL analyses involving over 500 TN breast tumors. Many of the 27 TN risk SNPs (Supplementary Table S8, available at Carcinogenesis Online) and an additional 35 SNPs in the TN risk loci (Supplementary Table S10, available at Carcinogenesis Online) that were associated with gene expression were located in transcriptional enhancers and DNase hypersensitivity sites in normal mammary epithelial cell lines, suggesting direct effects on gene transcription. Several interesting candidate genes were identified as cis-eQTLs. PTHLH, which encodes parathyroid hormone-like hormone, influences mammary gland development through regulation of epithelial to mesenchymal cellular interactions, is involved in lactation and is expressed in 60% of breast cancers (23–25). IGFBP2 (insulin-like growth factor binding protein 2) in the 2q35 locus displays elevated expression in breast tumors and promotes the growth and survival of breast epithelial cells through regulation of the ER-α (26, 27). TRX3 in the 12q24 locus encodes T-box 3, a transcription factor involved in developmental regulation that is overexpressed in breast tumors (28) and can induce mammary stem-like cells and mammary gland hyperplasia in mice (29). Although the cis-eQTL results suggest mechanisms by which certain loci influence TN breast cancer risk, additional functional validation of these SNP–gene expression relationships in breast cancer cell lines is needed.

Beyond etiology, the identification of 30 TN risk SNPs provides an opportunity to better understand how genetic variation may inform TN breast cancer risk prediction. As we have shown through our PRS, where we observed a 4-fold difference in risk between the highest and lowest PRS quintiles of the TN breast cancer population, it may be possible to identify women who are substantially above or below population level risk of TN breast cancer. Our PRS had better discriminatory accuracy (AUC = 0.64) compared with that of the Gail model applied in the Women’s Health Initiative (overall AUC = 0.58, 95% CI 0.58–0.62; ER-negative AUC = 0.50, 95% CI 0.45–0.54) (30). It is also likely that the inclusion of additional TN breast cancer risk SNPs will further stratify these women with respect to cumulative incidence of TN breast cancer. It will also be important to combine these TN risk SNPs with known epidemiologic risk factors such as parity, age at menarche, body mass index during premenopausal years and duration of breast feeding (1.5) to understand the cumulative influence on TN breast cancer risk. An important limitation of this study was that the PRS was applied to the study population from which the TN breast cancer risk estimates were derived. Although our cross-validation approach mitigates potential bias arising from this approach, it will be important to develop a risk model with these SNPs and validate the model in an independent study population. Overall, the findings provide strong evidence that integration of SNPs into predictive models will have a substantial impact on our ability to identify women at elevated risk of TN breast cancer.

Supplementary material
Supplementary Materials, Tables 1–15 and Figures 1–4 can be found at http://carcin.oxfordjournals.org/

Funding
Breast Cancer Research Foundation, National Institutes of Health (R01 CA128978), National Cancer Institute specialized program of research excellence (SPORE) in Breast Cancer (P50 CA116201); Mayo Cancer Genetic Epidemiology Training Grant (R25 CA092049). MCBCS was supported by the David and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation; The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany (01KWK9775/5, 01KWK9768/5, 01KWK9770/0 and 01KWK0114), the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany; OSU studies were supported by the OSU Comprehensive Cancer Center and the Stefanie Spielman Fund; SBCS was supported by Yorkshire Cancer Research Programme (S050PA); SKKDKFZS is supported by the DKFZ. The Demokritos has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund; The University of Kansas Cancer Center’s (P30 CA168524) Biospecimen Repository Core Facility, A.K.G. was funded by National Institutes of Health (5U01CA113916 and R01CA14032), a grant from the Kansai Bioscience Authority Eminent Scholar Program and by the Chancellors Distinguished Chair in Biomedical Sciences Professorship; The Australian Twin Cohort Study was supported by National Health and Medical Research Council of Australia; The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the Academy of Finland and by the strategic funding of the University of Eastern Finland; The HEBCS study has been financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (132473), the Finnish Cancer Society and the Sigrid Juselius Foundation. 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.

Acknowledgements
We acknowledge the support of the Mayo Clinic Genotyping Core and Mayo Clinic Expression Core SKKDKFZS. We are grateful to all the patients for their participation. We thank the physicians, other hospital staff and research assistants who contributed to the patient recruitment, data collection and sample preparation. KBCP thank Eija Myöhänen and Helena Kemiläinen for technical assistance. HEBCS thank research nurses Hanna Jäntti and Irja Erkkilä for their help with the patient data and samples and Drs 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.

Conflict of Interest Statement: None declared.

References
GWAS and known breast cancer risk loci in TN breast cancer


Received August 5, 2013; revised November 4, 2013; accepted November 27, 2013