A Multistage Genetic Association Study Identifies Breast Cancer Risk Loci at 10q25 and 16q24

Kathryn S. Higginbotham1, Joan P. Breyer1, Kate M. McReynolds1, Kevin M. Bradley5, Peggy A. Schuyler2, W. Dale Plummer6, Marcia E. Freudenthal2, Amy Trentham-Dietz7, Polly A. Newcomb8, Fritz F. Parl3, Melinda E. Sanders3, David L. Page3, Kathleen M. Egan6, William D. Dupont2, and Jeffrey R. Smith1,4

Introduction

Breast cancer is one of the most common malignancies among women in the United States, with an estimated 207,090 new cases and 39,840 deaths in 2010 (1). In recent years, large-scale association studies have identified multiple breast cancer susceptibility variants that have small effects on risk, but high population prevalence (2–9). Genetic risk models that include these variants could be clinically useful in the general population for multiple purposes, including risk-stratification to identify women who may benefit from more intensive breast cancer screening, or women who may be at increased risk of the development of breast cancer by taking hormone replacement therapy.

Several recent studies have explored risk models containing confirmed breast cancer associated variants (10–13). Predictive accuracy is expressed as the area under the receiver operating characteristic curve (AUC), which plots the sensitivity and specificity of a potential test. An AUC of 50% corresponds to completely random classification, although an AUC of 100% shows perfect classification of patient risk. A plausible maximum AUC for common diseases has been posited to be approx. 0.93 (14). Comparisons have been made by multiple studies between the Gail model (15), a nongenetic model that includes patient medical history and familial risk (AUC

10.1158/1055-9965.EPI-12-0386

©2012 American Association for Cancer Research.
0.557–0.607), genetic risk models including a set of common variants (AUC 0.574–0.587), and inclusive models which contain both nongenetic and genetic factors (AUC 0.589–0.632) (10–13). As more complete knowledge is obtained for risk-modifying genetic variants, clinically meaningful models may result. The top twelve low penetrance variants identified to date are estimated to account for only 8.3% of familial relative risk (11). The identification and incorporation of additional genetic loci confirmed to be associated with breast cancer could improve these risk models.

We describe a multistage association study investigating single nucleotide polymorphisms (SNPs) associated with breast cancer risk, seeking evidence of replication for SNPs of greatest nominal significance in stage I of the Cancer Genetic Markers of Susceptibility (CGEMS) genome wide association study (GWAS; ref. 8) in 3 additional independent breast cancer populations. The associations of greatest interest were those that were concordantly significant in the study.

Materials and Methods

Study subjects

The CGEMS GWAS, with 1,145 postmenopausal breast cancer cases and 1,142 postmenopausal controls from the Nurses’ Health Study, was an existing data set available for our analysis. All subjects of the study were Caucasian. Controls of that study were matched to cases by year of birth and by postmenopausal hormone use. This study is well described in the published literature. Age at diagnosis for cases was presented in 5-year intervals, with a median age of 66 (8, 16).

The Nashville Breast Cohort (NBC) is an ongoing retrospective cohort study of 17,017 women who underwent a breast biopsy revealing benign parenchyma or fibroadenoma at Vanderbilt, St. Thomas and Baptist Hospitals in Nashville, Tennessee since 1954. Approximately one third of these subjects had benign proliferative breast disease, which carries an increased risk for subsequent invasive breast cancer and is believed to be a nonobligate precursor lesion (17, 18). The germline DNA source for these subjects is the archival (formalin-fixed, paraffin-embedded [FFPE]) benign tissue biopsy. Additional details on the NBC have been published elsewhere (19). Entry biopsy FFPE blocks and complete follow up were available for 9,163 women in the cohort, of whom 646 developed incident breast cancer and were Caucasian. The mean age of cases at benign biopsy was 46 years (range 16–78 years) and at breast cancer diagnosis was 63 years (range 32–96 years); 80% were postmenopausal at breast cancer diagnosis. We conducted a nested case-control study as previously described (20, 21). Briefly, for each case we selected 2 controls from the risk set of women who had not been diagnosed with breast cancer in a similar period of observation. These controls were selected without replacement. Controls were individually matched to cases by age, race, and year of entry biopsy. DNA extraction was done by standard paraffin removal, proteinase K digestion, phenol/chloroform extraction, and ethanol precipitation. Successful DNA extractions from benign archival entry biopsy specimens were carried out for 622 of the 646 Caucasian cases (96.3%). A total of 599 of the 622 cases were matched to controls for whom DNA extractions were also successful. Successful DNA extraction was achieved for 1,161 of the 1,244 selected controls (93.3%). The study included a total of 562 trios (1 case matched to 2 controls), and 37 pairs (1 case matched to 1 control). The archival FFPE biopsy blocks (benign tissue) that were used for DNA extraction generally were quite old: 11.8% date to the 1950s, 25.6% to the 1960s, 34.3% to the 1970s, 22.0% to the 1980s, 6.2% to the 1990s, and 0.1% to the 2000s (2002 was the most recent).

Two additional independent Caucasian study populations were used to evaluate replication of significant associations. The Collaborative Breast Cancer Study (CBCS) is well described in the literature (21, 22). In brief, eligible breast cancer cases were female residents of Wisconsin, western Massachusetts, and New Hampshire, ages 20 to 74 years, with a recent primary diagnosis of invasive or in situ breast cancer reported to each state’s mandatory cancer registries from 1998 to 2001. Control women were randomly selected from population lists in each state and were frequency matched by 5-year categories to the age distribution of cases. DNA specimens prepared from buccal cells were available for genotyping from 1,552 cases and 1,185 controls of the CBCS study. The mean age at diagnosis of cases in the CBCS was 54 years (range 28–73 years) and 59% were postmenopausal at diagnosis (22).

The Vanderbilt biobank, BioVU, is composed of electronic medical records scrubbed of personal identifiers, linked to coded DNA samples prepared from whole blood from incident patients receiving care at Vanderbilt (107,991 DNA samples from 2009 to 2011; ref. 23). Robust replication of genotype–phenotype associations across multiple diseases has been showed in the biobank (24). We assembled a case-control study composed of Caucasian females: 1,172 invasive breast cancer cases and 1,172 controls. Data linkages were used to identify candidate subjects, followed by manual review of individual records to confirm inclusion and the exclusion criteria. Case inclusion criteria required the diagnosis of invasive breast cancer. Two thirds of these cases were diagnosed and/or treated at Vanderbilt (pathology confirmed). One third of the cases were diagnosed and treated at an outside hospital and received other care at Vanderbilt (pathology not reviewed). Control inclusion criteria required a negative screening mammogram conducted at Vanderbilt, and exclusion criteria included any prior abnormal screening mammogram (Bi-RADS other than 1 or 2), prior breast biopsy, prior surgery removing breast tissue, or prior diagnosis of in situ or invasive breast cancer. Controls were frequency matched on age at screen in 5-year intervals to age at diagnosis of cases. The mean age at diagnosis of cases in the BioVU was 54 years (range 20–95 years). Given wide dispersion in dates of subject accrual and...
Statistical analysis

Conditional logistic regression analysis of the individually matched subjects of the NBC was used to estimate breast cancer odds ratios (ORs), adjusted for age at entry biopsy and year of entry biopsy. Unconditional logistic regression was used to calculate these ORs for the frequency-matched subjects of the CGEMS, CBCS, and BioVU studies, adjusted for age. ORs, 95% confidence intervals, and P values were derived under a multiplicative model. Hardy–Weinberg equilibrium (HWE) analyses were carried out for cases and controls of each study population using Haploview. Three SNPs at one locus were not in HWE in any of the study populations; ORs for these SNPs were evaluated under a model that included 2 β parameters to concurrently assess the effect of the heterozygous and homozygous states. Associations between genotype and breast cancer were considered nominally significant if the associated 2-sided P value was less than 0.05. The associations of particular interest were those in which a significant CGEMS finding was replicated in the NBC as well as CBCS studies. These SNPs were further evaluated in the BioVU study, the last to be assembled. Combined analyses of all data sets were carried out for each variant by fixed-effects meta-analysis. Heterogeneity across studies was evaluated using a χ² statistic generated by the metan program (27).

BEAGLE v3.1.0 (28) was employed to impute missing genotypes and also to determine haplotypes of three regional SNPs at 5p12. Where a given genotype assay failed for a study subject, we imputed the missing data if it could be assigned with a probability of 1.0. Subjects with subsequent missing data for a given variant were excluded from analysis. Genotypes for the validated breast cancer susceptibility variant rs10941679 (26, 29), assayed in the NBC, CBCS, and BioVU studies but not directly assayed in the CGEMS study, were imputed for CGEMS subjects by BEAGLE with a mean probability of 96.6%. To accomplish this, BEAGLE used the known genotypes for NBC, CBCS, BioVU, and reference HapMap CEU trios, as well as HapMap genotypes of 500 SNPs to each flank as additional input.

Results

We employed single SNP tests to identify concordant associations with breast cancer risk in both the CGEMS and NBC breast cancer studies. We evaluated 1,162 SNPs of nominal P ≤ 0.003 from stage I of the CGEMS breast cancer GWAS. Twenty-six of these SNPs (2.2%) yielded concordant evidence of an association with breast cancer in the NBC study under a multiplicative model. Of the 26 SNPs, 8 were also significant in the independent CBCS population. These 8 SNPs included 4 previously published variants in linkage disequilibrium (LD; pairwise r² ≥ 0.84) in the second intron of FGFR2 (rs1219648, rs11200014, rs2420946, and rs2961579), 3 novel SNPs in LD (pairwise r² ≥ 0.9) near ENO4 (rs1626678, rs1681725, 10q25 and 16q24 Breast Cancer Risk Loci
and rs740363), and an additional novel SNP within WWOX (rs8046508). As expected, SNPs in strong pairwise LD at FGFR2 and at ENO4 yielded comparable estimates of effect size and significance, given in Table 1.

Table 2 presents association test results of these SNPs for each independent breast cancer study and for a combined analysis across all studies.

Novel SNPs rs1626678 (near ENO4) and rs8046508 (within WWOX) were further evaluated in the BioVU study. Table 2 presents association test results of these SNPs for each independent breast cancer study and for a combined analysis across all studies.
meta-analysis of the 4 combined studies. Supplementary Table S1 provides the number of case and control carriers of each risk allele for each study. Both SNPs are concordantly associated with breast cancer risk across CGEMS, NBC, and CBCS studies. In the BioVU study, odds ratios trended above one for both variants, although the effects were not individually significant. The meta-analysis supported a significant association of both SNPs with breast cancer risk. Neither of these associations has been previously reported. rs1626678 is located in a 156kb LD block at 10q25.3 containing the genes ENO4 (enolase family member 4) and KIAA1598 (shootin1). Breast cancer odds ratios for rs1626678 were 1.19 (NBC), 1.19 (CGEMS), 1.15 (CBCS), 1.03 (BioVU), and 1.13 in the meta-analysis ($P = 5.6 \times 10^{-5}$). rs8046508 is located in the eighth intron of WWOX (WW domain-containing oxidoreductase) in a 31kb LD block at 16q23.1. Breast cancer odds ratios for rs8046508 were 1.34 (NBC), 1.25 (CGEMS), 1.18 (CBCS), 1.09 (BioVU), and 1.20 in the meta-analysis ($P = 3.5 \times 10^{-5}$).

Three SNPs located between ANKS6 and GALNT12 (rs1543506, rs7861186, and rs10819308) were nominally significantly associated with breast cancer in the CGEMS study, and were each not in HWE. An excess of heterozygotes was observed among cases whereas a deficit of heterozygotes was observed among controls. The three SNPs are in LD ($r^2 > 0.8$). All three SNPs were also significantly out of HWE among subjects in the NBC and CBCS studies. Genotyping methods differed in the CGEMS and NBC studies, and three separate SNPs in the region yielded this result. For all 3 SNPs, the heterozygous state was associated with increased breast cancer risk in the CGEMS, NBC, and CBCS studies, but not in BioVU (Table 3). A test for interstudy heterogeneity of these odds ratios was significant.

We evaluated seven published breast cancer risk variants in the NBC, CBCS, and BioVU study populations (Table 4; refs. 5–9, 25, and 26), an early selection among a much larger set presently known. In view of the relatively small effect sizes previously observed, power in any single study to detect these associations is generally limited. Among the 7 examined variants, individual studies were estimated to have 80% or greater power to detect association only for FGFR2 rs1219648. We observed concordant association with breast cancer risk across all populations only for rs1219648, with ORs of 1.25 (NBC), 1.33 (CGEMS), 1.21 (CBCS), 1.15 (BioVU), and 1.24 in a meta-analysis ($P = 9.2 \times 10^{-12}$). Sequential replication as a means of discerning true from false positive observations in this study was conservative. FGFR2 rs1219648 was the only validated risk variant of greater significance than the novel variants of Table 2. The point estimates of the ORs for the 7 SNPs in the respective study populations generally agreed with published studies, with the exception of rs1562430 and rs999737 in the NBC study, and rs10941679 in the BioVU study. Heterogeneity of the various study populations could influence these results; for example, the mean age of diagnosis was relatively

<table>
<thead>
<tr>
<th>SNP Locus</th>
<th>Variant</th>
<th>Study</th>
<th>Cases</th>
<th>Controls</th>
<th>Hardy–Weinberg equilibrium</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1543506</td>
<td></td>
<td>CGEMS</td>
<td>0.47</td>
<td>0.46</td>
<td>0.017</td>
<td>0.049</td>
</tr>
<tr>
<td>rs1543506</td>
<td></td>
<td>CBCS</td>
<td>0.49</td>
<td>0.46</td>
<td>0.039</td>
<td>0.025</td>
</tr>
<tr>
<td>rs1543506</td>
<td></td>
<td>BioVU</td>
<td>0.41</td>
<td>0.43</td>
<td>0.022</td>
<td>0.045</td>
</tr>
</tbody>
</table>

The genes found in linkage disequilibrium with rs1543506 are GABBR2, ANKS6, and GALNT12. An odds ratio for all studies combined is not given because the test of odds ratio heterogeneity across studies was statistically significant. SNP nucleotide position is given for human genome build GRCh37/hg19.
greater for the NBC and CGEMS studies. Meta-analyses supported previously reported associations with breast cancer at all 7 loci (Table 4).

We evaluated three published risk SNPs on 5p12 between FGF10 and MRPS30 (rs4415084, ref. 26; rs920329, ref. 8; and rs10941679, ref. 26). SNP rs10941679 had a pairwise $r^2$ value of 0.513 with rs4415084, and of 0.507 with rs920329 among controls in the combined studies. Significant associations were observed at this locus in the CGEMS, NBC, and CBCS studies. Meta-analysis indicated significant associations for all 3 SNPs: rs4415084 (OR 1.11; $P = 9.3 \times 10^{-4}$), rs920329 (OR 1.11; $P = 1.0 \times 10^{-3}$), rs10941679 (OR 1.15; $P = 1.1 \times 10^{-5}$). Haplotype-based analyses by Stacey and colleagues had detected more significant association than single SNP analyses (26). Peak significance in a meta-analysis of our combined studies was observed for a risk haplotype encompassing rs920329 and rs10941679 (OR 1.15; $P = 6.1 \times 10^{-5}$).

We investigated the hypothesis that the 2 novel variants, rs1626678 and rs8046508, might alter risk of breast cancer associated with a history of proliferative disease (PD) in NBC subjects. The tests of interaction...
(i.e., departure from the multiplicative model) were not significant, but for rs8046508 the significant association with breast cancer risk was lost when corrected for PD, suggesting that the risk effect of the SNP and of PD were correlated. We also investigated the possibility of pairwise interactions between all SNPs from Tables 1 and 3. The tests of interaction were nominally significant between rs13387042 and rs1562430 ($P = 0.01$) and also between rs10941679 and rs1219648 ($P = 0.019$), although the tests were not significant after correction for multiple testing.

We investigated the ability of the evaluated known and novel SNPs to improve upon the Gail model (15). Gail model analysis yielded an AUC of 0.573 for the NBC and 0.588 for the CBCS. Covariate data were not available to us for the CGEMS or BioVU studies for comparison. Analysis of a genetic model based upon the 7 published SNPs in Table 3 yielded an AUC of 0.587 in the NBC and 0.565 in the CBCS. Further inclusion in the genetic model of the 2 novel SNPs yielded an AUC of 0.568 for the NBC and 0.569 for the CBCS. An inclusive model of both Gail and 7 published SNPs yielded an AUC of 0.601 for the NBC and 0.612 for the CBCS; additional inclusion of the 2 novel SNPs yielded an AUC of 0.601 for the NBC and 0.615 for the CBCS. An inclusive model of both Gail and 7 published SNPs yielded an AUC of 0.568 for the NBC and 0.565 in the CBCS. Further inclusion in the genetic model of the 2 novel SNPs appeared to improve the models, albeit insignificantly. The AUCs for these are given in Fig. 1 (Gail + all SNPs versus Gail alone: $P = 0.026$ [NBC] and $P = 0.003$ [CBCS]).

**Discussion**

This study sought to replicate nominally significant SNP associations within the CGEMS breast cancer study in sequentially investigated independent breast cancer study populations. Our results identified 2 novel loci with consistent evidence to support an association with risk of breast cancer.

The first novel SNP was rs1626678 located at 10q25 within an LD block of 156kb containing 2 genes, ENO4 and KIAA1598. Enolase is a metalloenzyme that catalyzes the final step of glycolysis, converting 2-phosphoglycerate to phosphoenopyruvate and generating ATP. Independent loci encode 5 enolase isoforms. Expression of ENO4 is described in the literature as being specific to spermatozoa (31). ENO2 is overexpressed in embryonic stem cells, and in breast cancer (32, 33). Other enolase isoforms, primarily ENO1, have been implicated in cancers with additional functions including cell surface expression as a plasminogen receptor to promote cell migration and metastasis, and alternative translation to produce a transcriptional repressor of MYC (34). Overexpression of enolase may also contribute to tumor development through the Warburg effect; cancer cells have a high glycolytic rate and use aerobic glycolysis to meet anabolic demand for precursors. KIAA1598 (also known as SHOOTIN1) is known to act in neuronal actin filament retrograde flow and to interact with L1 cell adhesion molecules (LICAM; ref. 35). Little is known of its role in breast tissue, but LICAM antibodies can block breast cancer cell adhesion and migration (36). KIAA1598 protein is differentially expressed in breast cancer tissues, with greater expression in estrogen receptor positive than negative tumors (37). Receptor status was not available in this study to allow us to investigate strength of association in these stratifications.

The second novel SNP was rs8046508 located at 16q23.1 within the eighth intron of WW domain-containing oxidoreductase (WWOX), a gene originally identified as a potential tumor suppressor in breast cancer and located at one of the most common human fragile sites, FRA16D (38). WWOX encodes a 46-kDa protein containing 2 N-terminal WW domains and a central short-chain dehydrogenase/reductase. Through its first WW domain, WWOX interacts with and modulates the function of multiple proteins implicated in cancer, such as p73 and ERBB4 intracellular domain (39). WWOX is inactivated in several types of cancer by varying methods, including genomic, epigenetic, and posttranslational modification, and studies have showed that restoration of WWOX to cell lines with low or no expression leads to inhibition of anchorage independent growth, enhanced apoptosis, and suppression of tumorigenicity in vivo (40–42). The WWOX
gene spans numerous LD blocks over 1.1 Mb. Somatic mutations of WWOX have been observed in breast cancer in exons 4 to 9 (43).

We identified 3 SNPs in a 186 kb LD block at 9q22.33 with an apparent excess of heterozygotes among breast cancer cases and deficit of heterozygotes among controls. Although commonly used to filter SNP data with assay artifact, deviation from expected genotype distribution in cases and controls (groups that are not randomly sampled) can also result from disease association (44–46). It was difficult to envision a source of assay error that might affect 3 independent SNPs of a region, assayed by more than one approach, and in several study populations. The tetra-allelic designation currently in dbSNP for 2 of these SNPs seems more likely to be because of an error in strand designation of a lone submission (among 37 for one, and 22 for the other), than to actual tetra-allelic polymorphism. If the region were subject to copy number variation, allele dosage and assay clusters could skew toward the heterozygote position. However, our data for these SNPs appeared unambiguous, with robust assays. The region contains GABBR2, ANK56, and GALNT12. The latter is altered in colon and breast tumors (47–49) and is linked to familial colon cancer (50).

Each of these novel loci are worthy of further investigation in additional independent breast cancer studies. Meta-analysis of the combined study populations supports the association of these novel loci, as well as previously described loci, with risk of breast cancer. The goal of this project has been to identify additional genetic loci that may provide mechanistic insights and have utility for improved breast cancer risk prediction. When we assessed the predictive ability of a model including validated GWAS SNPs it performed better than the Gail model alone, and inclusion of the 2 novel SNPs linked to breast cancer risk in our analyses modestly improved the model. The identification and inclusion of genetic risk variants, including additional ones that remain to be discovered, may improve current prediction models and assist in the identification of women at high risk for breast cancer who may benefit from informed decisions regarding prevention.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: K. S. Higginbotham, M. E. Sanders, D. L. Page, W. D. Dupont, J. R. Smith
Development of methodology: K. S. Higginbotham, J. P. Breyer, J. R. Smith
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): K. S. Higginbotham, J. P. Breyer, K. M. Egan, P. A. Newcom, K. M. Egans, D. L. Page, K. M. Egan, J. R. Smith
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K. M. Egan, P. A. Schuyler, M. E. Freudenthal, A. Trentham-Dietz, P. A. Newcom, K. M. Egan, J. R. Smith
Writing, review, and/or preparation of the manuscript: K. S. Higginbotham, J. P. Breyer, A. Trentham-Dietz, P. A. Newcom, P. F. Par, M. E. Sanders, K. M. Egan, W. D. Dupont, J. R. Smith
Study supervision: A. Trentham-Dietz, J. R. Smith

Acknowledgments
The authors thank the participating study subjects, investigators of the Cancer Genetic Markers of Susceptibility Study, and additional investigators of the Collaborative Breast Cancer Study, Drs. Montserrat Garcia-Closas and Linda T. Titus-Ernstoff.

Grant Support
This study was supported by DOD grant W81XWH-10-1-0784, NIH grants P50 CA098131, R01 CA050468, P30 CA68485, 1UL1 RR024975, and a MERIT grant from the US Department of Veterans Affairs. The CBCS was supported by NIH grants R01 CA105197, R01 CA47147, R01 CA72905, and R01 CA69664.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 1, 2012; revised June 6, 2012; accepted July 10, 2012; published OnlineFirst July 17, 2012.


A Multistage Genetic Association Study Identifies Breast Cancer Risk Loci at 10q25 and 16q24

Kathryn S. Higginbotham, Joan P. Breyer, Kate M. McReynolds, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-12-0386

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2012/07/18/1055-9965.EPI-12-0386.DC1

Cited articles
This article cites 48 articles, 17 of which you can access for free at:
http://cebp.aacrjournals.org/content/21/9/1565.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/21/9/1565.full#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.