

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 Plummer², Marcia E. Freudenthal², Amy Trentham-Dietz⁷, Polly A. Newcomb⁸, Fritz F. Parf⁹, Melinda E. Sanders³, David L. Page³, Kathleen M. Egan⁶, William D. Dupont², and Jeffrey R. Smith^{1,4}

Abstract

Background: Heritable risk for breast cancer includes an increasing number of common, low effect risk variants. We conducted a multistage genetic association study in a series of independent epidemiologic breast cancer study populations to identify novel breast cancer risk variants.

Methods: We tested 1,162 SNPs of greatest nominal significance from stage I of the Cancer Genetic Markers of Susceptibility breast cancer study (CGEMS; 1,145 cases, 1,142 controls) for evidence of replicated association with breast cancer in the Nashville Breast Cohort (NBC; 599 cases, 1,161 controls), the Collaborative Breast Cancer Study (CBCS; 1,552 cases, 1,185 controls), and BioVU Breast Cancer Study (BioVU; 1,172 cases, 1,172 controls).

Results: Among these SNPs, a series of validated breast cancer risk variants yielded expected associations in the study populations. In addition, we observed two previously unreported loci that were significantly associated with breast cancer risk in the CGEMS, NBC, and CBCS study populations and had a consistent, although not statistically significant, risk effect in the BioVU study population. These were rs1626678 at 10q25.3 near *ENO4* and *KIAA1598* (meta-analysis age-adjusted OR = 1.13 [1.07–1.20], $P = 5.6 \times 10^{-5}$), and rs8046508 at 16q23.1 in the eighth intron of *WVVOX* (meta-analysis age-adjusted OR = 1.20 [1.10–1.31], $P = 3.5 \times 10^{-5}$).

Conclusions: Our data supports the association of two novel loci, at 10q25.3 and 16q23.1, with risk of breast cancer.

Impact: The expanding compendium of known breast cancer genetic risk variants holds increasing power for clinical risk prediction models of breast cancer, improving upon the Gail model. *Cancer Epidemiol Biomarkers Prev*; 21(9); 1565–73. ©2012 AACR.

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

Authors' Affiliations: Departments of ¹Medicine, ²Biostatistics, ³Pathology, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine; ⁴Medical Research Service, Veteran's Affairs Tennessee Valley Healthcare System, Nashville, Tennessee; ⁵Foundation for Applied Molecular Evolution, Gainesville; ⁶Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida; ⁷Department of Population Health Sciences, Paul P. Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, Madison, Wisconsin; and ⁸Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Authors: William D. Dupont, Department of Biostatistics, Vanderbilt University School of Medicine, S-2323 Medical Center North, 1161 21st Avenue South, Nashville, TN 37232. Phone: 615-343-4100; Fax: 615-343-4924; E-mail: william.dupont@vanderbilt.edu; and Jeffrey R. Smith, Department of Medicine, Vanderbilt University School of Medicine, 529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232. Phone: 615-936-2171; Fax: 615-936-2661; E-mail: jeffrey.smith@vanderbilt.edu

doi: 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 additional study populations.

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 breast 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 states' 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 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

geography of the study populations, minimal or no subject overlap was expected between the 4 breast cancer studies. Subjects of these studies were not selected for a family history of breast cancer; carrier status of potential Mendelian mutations was unknown.

SNP selection and genotyping

One thousand one hundred and sixty-two variants from stage I of the CGEMS breast cancer GWAS with nominally significant associations (P value of less than 0.003) were selected for replication in the NBC. CGEMS GWAS data and analyses were available by permission through dbGaP authorized access. SNPs that were concordantly significant in CGEMS and NBC were further genotyped in the CBCS and BioVU study populations. We also included 7 well-established breast cancer risk loci identified in published GWAS (5–9, 25, 26). Genotyping in NBC, CBCS, and BioVU was carried out using commercial Illumina GoldenGate and Applied Biosystems TaqMan assays.

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 χ^2 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.

Variants that concordantly replicated as breast cancer risk factors across multiple studies were further analyzed for evidence of statistical interaction with a history of benign proliferative breast disease. Benign breast disease histology was uniquely available for the subjects of the NBC, all of whom had different types of biopsy-proven benign breast disease. We derived the combined effects on breast cancer risk of SNP variant and proliferative disease (PD) in the patient's entry biopsy. These models contained a parameter for the variant under consideration, a parameter for PD, an interaction parameter for the joint effects of the variant with PD, age at entry biopsy, and year of entry biopsy. Models with pairwise interactions between all variants in Tables 1–3 were also run. These models contained the product of the number of alleles of each variant as a covariate in the model.

We used the Gail model (15) to estimate breast cancer ORs for the NBC study based on age, age at menarche, age at first birth, number of first degree relatives with breast cancer, number of previous breast biopsies, and biopsy histopathology. In replicating the Gail model for the CBCS data, we came as close as possible to the model described in Gail and colleagues given the available data; our model was identical to Gail's with 2 exceptions. First, the CBCS data coded a history of benign breast surgery as a dichotomous variable whereas Gail and colleagues coded this history as 0, 1, or 2, with 2 denoting more than one biopsy. Similarly, in the CBCS a first degree family history of breast cancer was coded as a dichotomous variable whereas Gail et al. coded this history as 0, 1, or more than one first-degree relatives with this history. We evaluated the Gail model only, the Gail model with the 7 examined GWAS SNPs (Table 3), and the Gail model with both GWAS SNPs and novel SNPs identified in this analysis (Table 1). We also evaluated models containing only the genetic factors. The linear predictors from our logistic regression models were used to calculate the AUCs. Tests of the difference between these areas were done using the method of Delong and colleagues (30).

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 $P \leq 0.003$ 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 $r^2 \geq 0.84$) in the second intron of *FGFR2* (rs1219648, rs11200014, rs2420946, and rs2981579), 3 novel SNPs in LD (pairwise $r^2 \geq 0.9$) near *ENO4* (rs1626678, rs1681723,

Table 1. SNPs in pairwise LD at *FGFR2*, and in pairwise LD at *ENO4*, are consistently associated with breast cancer

SNP	Locus	Variant	Study	OR (95% CI)	P
rs1219648	10q26.13	G	NBC	1.25 (1.08–1.44)	0.0025
	123,346,190 bp		CGEMS	1.33 (1.19–1.50)	1.49×10^{-6}
	<i>FGFR2</i>		CBCS	1.23 (1.10–1.38)	3.57×10^{-4}
rs11200014	10q26.13	T	NBC	1.25 (1.08–1.44)	0.0024
	123,334,930 bp		CGEMS	1.31 (1.17–1.48)	4.41×10^{-6}
	<i>FGFR2</i>		CBCS	1.20 (1.07–1.34)	0.0018
rs2420946	10q26.13	T	NBC	1.24 (1.07–1.43)	0.0042
	123,351,324 bp		CGEMS	1.34 (1.19–1.51)	1.18×10^{-6}
	<i>FGFR2</i>		CBCS	1.19 (1.06–1.33)	0.0024
rs2981579	10q26.13	T	NBC	1.25 (1.08–1.44)	0.0024
	123,337,335 bp		CGEMS	1.31 (1.17–1.48)	4.41×10^{-6}
	<i>FGFR2</i>		CBCS	1.20 (1.07–1.34)	0.0018
rs1626678	10q25.3	G	NBC	1.19 (1.03–1.37)	0.016
	118,576,146 bp		CGEMS	1.19 (1.06–1.33)	0.0036
	<i>ENO4</i>		CBCS	1.15 (1.03–1.29)	0.013
rs1681723	10q25.3	A	NBC	1.17 (1.02–1.34)	0.030
	118,591,615 bp		CGEMS	1.17 (1.04–1.31)	0.0090
	<i>ENO4</i>		CBCS	1.12 (1.00–1.25)	0.049
rs740363	10q25.3	C	NBC	1.18 (1.02–1.36)	0.022
	118,575,606 bp		CGEMS	1.17 (1.05–1.32)	0.0065
	<i>ENO4</i>		CBCS	1.14 (1.02–1.28)	0.021

The four SNPs in the second intron of *FGFR2* have pairwise LD values ≥ 0.84 , whereas those near *ENO4* have pairwise LD values ≥ 0.90 . *FGFR2* and *ENO4* are 4.7 MB apart, not in linkage disequilibrium with each other. SNP nucleotide positions are given for human genome build GRCh37/hg19.

and rs740363), and an additional novel SNP within *WVVOX* (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.

Novel SNPs rs1626678 (near *ENO4*) and rs8046508 (within *WVVOX*) 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

Table 2. Novel loci associated with breast cancer risk

SNP	Locus	Variant	^a Frequency		Study	OR (95% CI)	P	
			Ca	Co				
^b rs1626678	10q25.3 118,576,146 bp <i>ENO4</i>	G			NBC	1.19 (1.03–1.37)	0.016	
					CGEMS	1.19 (1.06–1.33)	0.0036	
					CBCS	1.15 (1.03–1.29)	0.013	
					BioVU	1.03 (0.92–1.15)	0.65	
					Combined	1.13 (1.07–1.20)	5.6×10^{-5}	
^c rs8046508	16q23.1 79,078,175 bp <i>WVVOX</i>	T			NBC	1.34 (1.09–1.64)	0.0049	
					CGEMS	1.25 (1.06–1.47)	0.0081	
					CBCS	1.18 (1.00–1.39)	0.048	
					BioVU	1.09 (0.92–1.30)	0.33	
					Combined	1.20 (1.10–1.31)	3.5×10^{-5}	

^aRisk variant frequency in combined cases (Ca) or controls (Co). Genes in linkage disequilibrium with SNP.

^bSNP between *ENO4* and *KIAA1598*.

^cSNP in intron eight of *WVVOX*. SNP nucleotide positions are given for human genome build GRCh37/hg19.

Table 3. rs1543506 is not in Hardy-Weinberg equilibrium and may affect breast cancer risk

SNP	Locus	Variant	Study	Cases			Controls			Association			
				Obs Het	Pred Het	P	Obs Het	Pred Het	P	ORHet (95% CI)	P	ORHom (95% CI)	P
rs1543506	9q22.33	T	NBC	0.47	0.45	0.22	0.40	0.45	2.0×10^{-4}	1.25 (1.01-1.54)	0.039	0.80 (0.57-1.11)	0.18
	101,567,178 bp		CGEMS	0.49	0.46	0.017	0.42	0.45	0.025	1.28 (1.08-1.53)	0.0056	0.88 (0.67-1.15)	0.35
	GALNT12		CBCS	0.46	0.44	0.09	0.41	0.44	0.049	1.26 (1.06-1.49)	0.032	0.92 (0.71-1.20)	0.55
			BioVU	0.41	0.43	0.032	0.45	0.45	0.85	0.78 (0.65-0.93)	0.0063	0.91 (0.69-1.20)	0.50

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.

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 *WFOX* (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 (pairwise $r^2 > 0.8$). 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 4. Previously identified loci associated with breast cancer risk

SNP	Locus	Variant	^a Frequency		Study	OR (95% CI)	P	
			Ca	Co				
rs13387042	2q35 217,905,832 bp	A	0.55	0.52	NBC	1.01 (0.87–1.17)	0.89	
					CGEMS	1.19 (1.06–1.34)	0.0030	
					CBCS	1.18 (1.05–1.32)	0.0055	
					BioVU	1.04 (0.93–1.17)	0.50	
					Combined	1.11 (1.05–1.18)	5.8×10^{-4}	
rs4973768 S	3p24.1 27,416,013 bp <i>SLC4A7</i>	T	0.50	0.48	NBC	1.09 (0.95–1.25)	0.22	
					CGEMS	1.10 (0.98–1.23)	0.11	
					CBCS	1.09 (0.97–1.22)	0.13	
					BioVU	1.08 (0.96–1.22)	0.17	
					Combined	1.09 (1.03–1.16)	0.0050	
rs16886165	5q11.2 56,023,083 bp <i>MAP3K1</i>	G	0.17	0.16	NBC	1.18 (0.98–1.42)	0.083	
					CGEMS	1.22 (1.04–1.42)	0.013	
					CBCS	1.04 (0.89–1.21)	0.64	
					BioVU	1.10 (0.94–1.28)	0.23	
					Combined	1.13 (1.04–1.22)	0.0037	
^b rs10941679	5p12 44,706,498 bp <i>MRPS30</i>	G	0.28	0.25	NBC	1.16 (1.00–1.36)	0.057	
					CGEMS	1.26 (1.10–1.45)	7.1×10^{-4}	
					CBCS	1.23 (1.08–1.40)	0.0020	
					BioVU	0.96 (0.84–1.09)	0.52	
					Combined	1.15 (1.07–1.23)	1.1×10^{-4}	
rs1562430	8q24.21 128,387,852 bp	A	0.60	0.58	NBC	0.93 (0.80–1.07)	0.29	
					CGEMS	1.17 (1.04–1.31)	0.010	
					CBCS	1.06 (0.95–1.19)	0.32	
					BioVU	1.13 (1.00–1.27)	0.051	
					Combined	1.08 (1.02–1.15)	0.010	
rs1219648	10q26.13 123,346,190 bp <i>FGFR2</i>	G	0.44	0.39	NBC	1.25 (1.09–1.45)	0.0022	
					CGEMS	1.33 (1.19–1.50)	1.5×10^{-6}	
					CBCS	1.21 (1.07–1.36)	0.0016	
					BioVU	1.15 (1.02–1.29)	0.022	
					Combined	1.24 (1.16–1.31)	9.2×10^{-12}	
rs999737	14q24.1 69,034,682 bp <i>RAD51B</i>	C	0.78	0.76	NBC	0.95 (0.80–1.13)	0.56	
					CGEMS	1.20 (1.05–1.38)	0.0089	
					CBCS	1.09 (0.95–1.25)	0.20	
					BioVU	1.05 (0.91–1.20)	0.53	
					Combined	1.08 (1.00–1.16)	0.038	

^aRisk variant frequency in combined cases (Ca) or controls (Co).

^brs10941679 was not genotyped in the CGEMS study, but was imputed with a probability of 96.6%. SNP nucleotide positions are 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; $P = 1.1 \times 10^{-4}$). 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 0.557 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 0.594 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. The addition of the 2 novel SNPs appeared to improve the models, albeit insignificantly. The AUCs for these is given in Fig. 1 (Gail + all SNPs versus Gail alone: $P = 0.026$ [NBC] and $P = 0.003$ [CBCS]).

Discussion

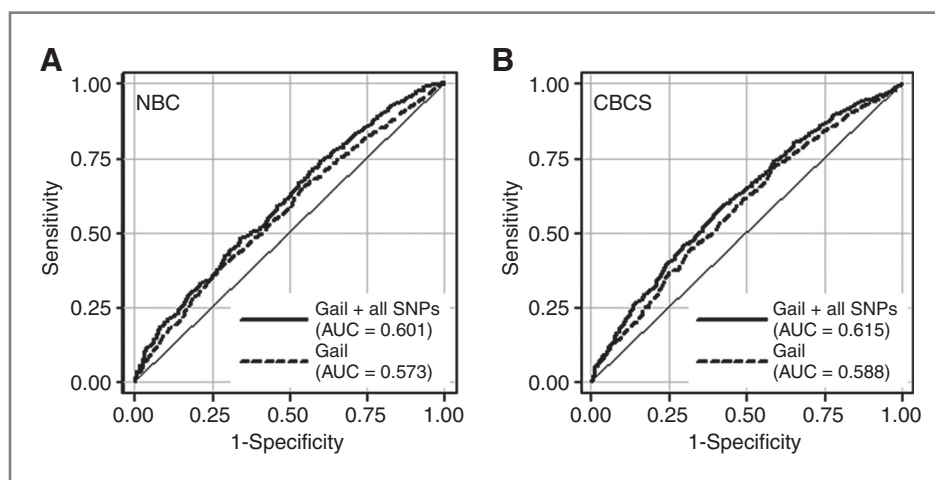
This study sought to replicate nominally significant SNP associations within the CGEMS breast cancer study in sequentially investigated independent breast cancer study populations. Our results identified 2 novel loci with consistent evidence to support an association with risk of breast cancer.

The first novel SNP was rs1626678 located at 10q25.3 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 (L1CAM; ref. 35). Little is known of its role in breast tissue, but L1CAM 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*

Figure 1. Receiver operating characteristic curves and the area under these curves (AUCs) for the Gail model (dashed line) and the model with the Gail covariates and all 9 study SNPs of Tables 1 and 3 (solid black line). A, NBC study: Gail + all SNPs versus Gail alone; $P = 0.026$. B, CBCS study: Gail + all SNPs versus Gail alone; $P = 0.003$.



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*, *ANKS6*, 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. McReynolds, P. A. Schuyler, M. E. Freudenthal, A. Trentham-Dietz, M. E. Sanders, D. L. Page, K. M. Egan, J. R. Smith

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K. M. McReynolds, P. A. Schuyler, M. E. Freudenthal, A. Trentham-Dietz, P. A. Newcomb, K. M. Egan, J. R. Smith

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): K. S. Higginbotham, K. M. Bradley, W. Dale Plummer, Jr, P. A. Newcomb, K. M. Egan, W. D. Dupont, J. R. Smith

Writing, review, and/or revision of the manuscript: K. S. Higginbotham, J. P. Breyer, A. Trentham-Dietz, P. A. Newcomb, F. F. Parl, 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 CA068485, 1UL1RR024975, and a MERIT grant from the US Department of Veterans Affairs. The CBCS was supported by NIH grants R01 CA105197, R01 CA47147, R01 CA47305, 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.

References

- Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin* 2010;60:277–300.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009;106:9362–7.
- Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, et al. A common variant at the *TERT-CLPTM1L* locus is associated with estrogen receptor-negative breast cancer. *Nat Genet* 2011;43:1210–4.
- Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst* 2011;103:425–35.
- Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 2010;42:504–7.
- Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). *Nat Genet* 2009;41:579–84.
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087–93.
- Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007;39:870–4.
- Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2007;39:865–9.
- Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst* 2008;100:1037–41.
- Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. *Mol Oncol* 2010;4:174–91.

12. Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst* 2010; 102:1618–27.
13. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR, et al. Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 2010;362:986–93.
14. Jostins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet*. 2011;20:R182–8.
15. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–86.
16. Azzato EM, Pharoah PD, Harrington P, Easton DF, Greenberg D, Caporaso NE, et al. A genome-wide association study of prognosis in breast cancer. *Cancer Epidemiol Biomarkers Prev* 2010;19:1140–3.
17. Hartmann LC, Sellers TA, Frost MH, Lingle WL, Degnim AC, Ghosh K, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005;353:229–37.
18. London SJ, Connolly JL, Schnitt SJ, Colditz GA. A prospective study of benign breast disease and the risk of breast cancer. *JAMA* 1992; 267:941–4.
19. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med* 1985;312:146–51.
20. Dupont WD, Breyer JP, Bradley KM, Schuyler PA, Plummer WD, Sanders ME, et al. Protein phosphatase 2A subunit gene haplotypes and proliferative breast disease modify breast cancer risk. *Cancer* 2010;116:8–19.
21. Higginbotham KS, Breyer JP, Bradley KM, Schuyler PA, Plummer WD Jr, Freudenthal ME, et al. A multistage association study identifies a breast cancer genetic locus at NCOA7. *Cancer Res* 2011;71:3881–8.
22. Zhang Y, Newcomb PA, Egan KM, Titus-Ernstoff L, Chanock S, Welch R, et al. Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2006;15: 353–8.
23. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008;84:362–9.
24. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* 2010;86:560–72.
25. Ahmed S, Thomas G, Ghousaini M, Healey CS, Humphreys MK, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 2009;41:585–90.
26. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2008;40:703–6.
27. Harris RJ, Bradburn MJ, Deeks JJ, Harbord RM, Altman DG, Sterne JAC. meta: fixed- and random-effects meta-analysis. *Stata J* 2008; 8:3–28.
28. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009;84:210–23.
29. Garcia-Closas M, Chanock S. Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin Cancer Res* 2008;14:8000–9.
30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
31. Edwards YH, Grootegoed JA. A sperm-specific enolase. *J Reprod Fertil* 1983;68:305–10.
32. Assou S, Le Carrouer T, Tondeur S, Strom S, Gabelle A, Marty S, et al. A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* 2007; 25:961–73.
33. Durany N, Joseph J, Jimenez OM, Climent F, Fernandez PL, Rivera F, et al. Phosphoglycerate mutase, 2,3-bisphosphoglycerate phosphatase, creatine kinase and enolase activity and isoenzymes in breast carcinoma. *Br J Cancer* 2000;82:20–7.
34. Capello M, Ferri-Borgogno S, Cappello P, Novelli F. alpha-Enolase: a promising therapeutic and diagnostic tumor target. *FEBS J* 2011; 278:1064–74.
35. Shimada T, Toriyama M, Uemura K, Kamiguchi H, Sugiura T, Watanabe N, et al. Shootin1 interacts with actin retrograde flow and L1-CAM to promote axon outgrowth. *J Cell Biol* 2008;181:817–29.
36. Li Y, Galileo DS. Soluble L1CAM promotes breast cancer cell adhesion and migration *in vitro*, but not invasion. *Cancer Cell Int* 2010;10:34.
37. Rezaul K, Thumar JK, Lundgren DH, Eng JK, Claffey KP, Wilson L, et al. Differential protein expression profiles in estrogen receptor-positive and -negative breast cancer tissues using label-free quantitative proteomics. *Genes Cancer* 2010;1:251–71.
38. Bednarek AK, Laffin KJ, Daniel RL, Liao Q, Hawkins KA, Aldaz CM. WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3–24.1, a region frequently affected in breast cancer. *Cancer Res* 2000;60:2140–5.
39. Del Mare S, Salah Z, Aqeilan RI. WWOX: its genomics, partners, and functions. *J Cell Biochem* 2009;108:737–45.
40. Bednarek AK, Keck-Waggoner CL, Daniel RL, Laffin KJ, Bergsagel PL, Kiguchi K, et al. WWOX, the FRA16D gene, behaves as a suppressor of tumor growth. *Cancer Res* 2001;61:8068–73.
41. Fabbri M, Iliopoulos D, Trapasso F, Aqeilan RI, Cimmino A, Zanasi N, et al. WWOX gene restoration prevents lung cancer growth *in vitro* and *in vivo*. *Proc Natl Acad Sci U S A* 2005;102:15611–6.
42. Iliopoulos D, Guler G, Han SY, Johnston D, Druck T, McCorkell KA, et al. Fragile genes as biomarkers: epigenetic control of WWOX and FHIT in lung, breast and bladder cancer. *Oncogene* 2005;24: 1625–33.
43. Ekizoglu S, Muslumanoglu M, Dalay N, Buyru N. Genetic alterations of the WWOX gene in breast cancer. *Med Oncol* 2011 Oct 8. [Epub ahead of print].
44. Li M, Li C. Assessing departure from Hardy-Weinberg equilibrium in the presence of disease association. *Genet Epidemiol* 2008;32: 589–99.
45. Nielsen DM, Ehm MG, Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet* 1998;63:1531–40.
46. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005;76:967–86.
47. Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, et al. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc Natl Acad Sci U S A* 2009;106:12921–5.
48. Guo JM, Zhang Y, Cheng L, Iwasaki H, Wang H, Kubota T, et al. Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAc-T12. *FEBS Lett* 2002;524:211–8.
49. Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep* 2006;7:599–604.
50. Gray-McGuire C, Guda K, Adrianto I, Lin CP, Natale L, Potter JD, et al. Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. *Cancer Res* 2010;70:5409–18.