HHS Public Access

Author manuscript

Breast Cancer. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Breast Cancer. 2015 September ; 22(5): 544–551. doi:10.1007/s12282-014-0518-2.

Evaluating 17 Breast Cancer Susceptibility Loci in the Nashville Breast Health Study

Mi-Ryung Han[1], **Sandra Deming-Halverson**[1], **Qiuyin Cai**[1], **Wanqing Wen**[1], **Martha J. Shrubsole**[1], **Xiao-Ou Shu**[1], **Wei Zheng**[1], and **Jirong Long**[1]

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA

Abstract

Background—Genome-wide association studies (GWAS) have discovered multiple genetic loci associated with breast cancer risk. Investigating these loci would be helpful to evaluate previous findings and identify causal variants for breast cancer. We evaluated index SNPs in17 of these loci in a study of 1,511 cases and 1,454 controls of European descent.

Methods—We investigated the overall association with breast cancer and among subtypes defined as ER+ (estrogen receptor positive), ER− (estrogen receptor negative) and TNBC (triple-negative breast cancer). Combined effects of SNPs on breast cancer risk were assessed via a genetic risk score (GRS). We evaluated the contribution of both genetic variants and traditional risk factors to a breast cancer risk assessment model.

Results—Five of the 17 SNPs were significantly associated ($P \leq 0.05$) with overall breast cancer in the same direction as previously reported: *rs13387042* (2q35/*TNP1*), *rs4973768* (3p24/*SLC4A7*), *rs2046210* (6q25/*ESR1*), *rs1219648* (10q26/*FGFR2*), and *rs4784227* (16q12/*TOX3*). When stratified by breast cancer subtype, all five SNPs were associated ($P<0.05$) with ER+ cancer, three with ER− cancer (*rs13387042*, *rs1219648*, and *rs4784227*), and one with TNBC (*rs1219648*). A GRS, based on those five significant SNPs, showed strong association with overall breast cancer with ORs (95% CI) of 1.48 (1.22–1.79), 1.85 (1.52–2.25) and 2.26 (1.82–2.80), respectively, for each quartile, ($P = 2.0\times10^{-15}$). Traditional risk factors, including previous benign breast disease, breast cancer family history and parity, were significantly associated with breast cancer risk in the present study. These factors, together with the GRS, were used to build a breast cancer risk assessment model with a *c* statistic of 0.6321 from receiver operating characteristic (ROC) analysis. The contribution of the GRS to the model was greater than prior benign breast disease, family history and parity with the c statistic change of 0.0374, 0.0324, 0.0103, 0.0012, respectively.

Conclusions—Our study demonstrates that five SNPs were associated with overall breast cancer, with stronger association for ER+ than ER− cancer as previously reported, and suggests

Corresponding author: Jirong Long, Ph.D., Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738, Phone: (615) 343-6741, Fax: (615) 936-8241, Jirong.Long@vanderbilt.edu.

Competing Interest

The authors have declared no competing financial interests.

that a risk assessment model incorporating the GRS from five loci is useful in identifying women at high risk of breast cancer.

Keywords

breast cancer; genetic susceptibility; GWAS loci; risk assessment model

Introduction

Breast cancer is one of the most common malignancies, ranking first in incidence and second in mortality for all cancers diagnosed among women in the United States. During 2012, approximately 229,060 women were diagnosed with breast cancer and 39,510 died from this disease in the United States (http://seer.cancer.gov/statfacts/html/breast.html). Breast cancer is a complex disease in which genetic factors play an important role [1,2]. To date, four high-penetrance genes (*BRCA1*, *BRCA2*, *TP53*, and *PTEN*) and four moderate-penetrance genes (*CHEK2*, *ATM*, *BRIP1*, and *PALB2*) have been discovered for breast cancer[3]. Approximately 70 genetic susceptibility loci have been discovered for breast cancer risk through genome wide association study (GWAS) [4–10].

Breast cancer is a heterogeneous disease. Estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) are the most commonly-used biomarkers for breast cancer subtyping. Women with ER negative (ER−) breast cancer, especially those with triple negative breast cancer (TNBC) characterized by tumors lacking expression of ER, PR or HER2, have more limited treatment options. With the availability of high-throughput genetic data, it is becoming increasingly apparent that genetic susceptibility to breast cancer varies by tumor subtype. Therefore, it is imperative to evaluate the associations between susceptibility loci and breast cancer by tumor subtypes.

Each of the genetic variants identified in GWAS has only a modest effect in increasing the risk of breast cancer (per allele OR<1.3). However, these variants combined may have a more substantial effect. Furthermore, the combination of such genetic factors with traditional risk factors may provide better risk prediction for breast cancer. In the present study, we investigated risk variants in 17 independent genetic loci previously discovered through GWAS, among participants in the Nashville Breast Health Study (NBHS), a large population-based case-control study. We assessed associations, for single variants and in combination via a genetic risk score, with breast cancer overall and by tumor subtypes. Lastly, we evaluated the contribution of both genetic variants and traditional risk factors to a breast cancer risk assessment model with the goal of better identifying women at higher risk for breast cancer.

Methods

Samples

The Nashville Breast Health Study (NBHS) is a case-control study of breast cancer conducted in the Nashville metropolitan area. Through a rapid case ascertainment system, we identified newly-diagnosed breast cancer cases through the Tennessee State Cancer

Registry and five major hospitals in the city that provide medical care for breast cancer patients. Eligible cases were women diagnosed between February 1, 2001 and December 31, 2008 with invasive breast cancer and ductal carcinoma in situ (DCIS), who were between the ages of 25 and 75, had no prior history of cancer other than non-melanoma skin cancer, had a residential telephone, spoke English, and were able to provide consent to the study. Information on ER, PR, and HER2 status of breast tumors was obtained from pathology records. Controls were identified primarily via random digit dialing (RDD) of households in the same geographic area as cases. Eligibility criteria for controls were the same as cases with the exception that controls did not have a prior cancer diagnosis other than simple skin cancer. Controls were frequency matched to cases on 5-year age group, race, and county of residence. Information on demographic factors, as well as known and suspected risk factors for breast cancer, was ascertained through a structured questionnaire administered via telephone interview. Included in the current project are 1,511 cases and 1,454 controls of European ancestry who participated in the study before August 2008 and have donated a buccal cell sample to the study. Buccal cell samples were collected via one of the two methods: Oragene saliva collection kits (DNA GenoteK, Ottawa, Canada) or mouthwash samples. Approval for this study was obtained from the institutional review boards of Vanderbilt University Medical Center and the other participating institutions. All participants provided informed consent prior to enrollment in this study.

Genotyping and sample quality control

Genotyping was conducted using the Illumina HumanExome-12v1_A Beadchip which contains 247k SNPs (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Only the GWAS identified variants that were reported in the NHGRI list by August 16, 2011 (http://www.genome.gov/gwastudies/), i.e., 17 GWAS loci for breast cancer, were included in the chip. Thus, 19 SNPs from the 17 GWAS loci were evaluated in the current study.

All samples were genotyped at the Genome Quebec Innovation Centre (Montreal, Quebec, Canada) following Illumina's protocol. On each 96-well plate, blind duplicate samples and two HapMap samples were included as quality control (QC). Genotype calling was carried out using Illumina's GenTrain version 2.0 clustering algorithm in GenomeStudio version 2011.1. Cluster boundaries were determined using study samples. The samples were excluded if: (i) consistency rates among duplicated samples were less than 99%, (ii) consistency rates between the HapMap samples with 1000 Genomes Project data were less than 99%, (iii) an heterozygosity rate outlier, (iv) an ethnic outlier, (v) samples indicated close genetic relationship, or (vi) samples suggested wrong gender. The SNPs were excluded if: (i) MAF=0, (ii) call rate < 98%, (iii) genotyping concordance rate < 98% in QC samples, (iv) HWE test $P<10^{-5}$, (v) redundant SNPs, or (vi) questionable SNPs discovered by the exomechip design group (http://genome.sph.umich.edu/wiki/Exome_Chip_Design#Cautious_Sites). After these exclusions were applied, 19 SNPs in 17 breast cancer GWAS loci were included in the final dataset. Strong LD ($r^2>0.8$) was observed between *rs2981582* and *rs1219648*, *rs3803662* and *rs4784227*; therefore, only one SNP in each of these two loci, *rs1219648* and *rs4784227*, respectively, were included in the final analyses.

Statistical analysis

Differences between cases and controls were compared using a Wilcoxon rank sum test (continuous variables) or χ^2 test (categorical variables). Association analysis between each SNP and breast cancer risk was assessed using odds ratios (ORs) and 95% confidence intervals (CIs) derived from logistic regression models. To evaluate the combined effect of SNPs on breast cancer risk, we created a weighted genetic risk score (GRS) for each study participant by multiplying the number of risk alleles (0/1/2) of each SNP by the weight (log scale of the pER−allele OR derived from the current study) for that SNP, and then summing them together. Persons with missing data for a SNP were assigned the average number of risk alleles at that SNP for cases and controls separately. We carried out principal components analysis (PCA) using EIGENSTRAT (http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm) based on the 3,200 ancestry informative markers (AIMs) on the exomechip (ftp://share.sph.umich.edu/exomeChip/IlluminaDesigns/AIM/). All analyses were adjusted for age and the top 5 principal components (PCs). Analyses were also stratified by tumor subtype.

Logistic regression was used to examine independent associations of breast cancer risk for non-genetic factors, including GRS, age at menarche, age at first live birth, parity, body mass index, breast cancer family history, and previous diagnosis of benign breast disease. Factors independently associated with breast cancer risk were used to construct a risk assessment model. The *c* (or concordance) statistic (ie, the area under the receiver operating characteristic curve) was used to assess the discriminatory accuracy of the model. A *c* statistic of 0.5 corresponds to random classification of women with breast cancer and women without breast cancer, whereas perfect classification provides a *c* statistic of 1.0. The added predictive value of each factor was evaluated by comparing models with and without that predictor with regard to their *c* statistic [11]. Nonparametric Mann–Whitney *U* test was used to test the difference in the *c* statistics between two models [11]. All statistical analyses were conducted in SAS, version 9.3, with the use of two-tailed tests, and R statistical language (http://www.r-project.org/).

Results

Descriptive characteristics for study participants are presented in Table 1. Five established breast cancer risk factors (age, menopausal status, parity, breast cancer family history, and previous benign breast disease diagnosis) differed significantly between breast cancer patients and controls ($P \leq 0.05$).

For overall breast cancer risk, SNPs at five loci, including 2q35/*TNP1*, 3p24/*SLC4A7*, 6q25/*ESR1*, 10q26/*FGFR2*, and 16q12/*TOX3* were associated ($P \leq 0.05$) in the same direction as previously reported (Table 2), with the *rs1219648* located in 10q26/*FGFR2* being the strongest predictor. Adjusted odds ratios (95% CI) were 1.56 (1.33–1.84) and 1.75 (1.41–2.17), respectively, for genotypes A/G and G/G versus A/A for the SNP rs1219648 (P for trend 1.4×10^{-8}). Allelic ORs (95% CI) for the other 4 SNPs were 1.22 (1.10–1.35) for *rs13387042* ($P = 2.0\times10^{-4}$), 1.14 (1.03–1.26) for *rs4973768* ($P = 0.02$), 1.14 (1.02–1.27) for *rs2046210* ($P = 0.02$), and 1.26 (1.12–1.42) for *rs4784227* ($P = 9.7\times10^{-5}$). No significant

associations were observed for the other 12 SNPs, however, ORs for 10 of the SNPs were in similar directions as previous reports.

The associations between SNPs and breast cancer subtypes are presented in Table 3. Among the five SNPs that were associated with overall breast cancer, all of them were significantly associated ($P<0.05$) with ER+ cancer, three with ER− cancer (*rs13387042, rs1219648,* and *rs4784227*), and one with TNBC (*rs1219648*). SNP *rs11249433* at 1p11/*FCGR1B* and *rs2380205* at 10p15/*ANKRD16* were associated with ER− breast cancer with ORs (95% CI) of 1.18 (0.98– 1.41) ($P = 0.07$) and 1.19 (0.99– 1.42) ($P = 0.06$), respectively. SNP *rs1045485* at 2q33/*CASP8* was associated with TNBC with an OR (95% CI) of 1.55 (1.02–2.34) ($P = 0.04$).

GRS analyses based on five significant SNPs showed significant associations with breast cancer, overall and by subtype (Table 4). The ORs (95% CIs) for overall breast cancer risk across increasing quartiles of GRS were 1.48 (1.22–1.79), 1.85 (1.52–2.25) and 2.26 (1.82–2.80), respectively, ($P = 2.0\times10^{-15}$) compared to the lowest quartile. Similar associations also were observed when stratified by breast cancer subtypes, menopausal status, age of onset, or breast cancer family history (Table 4).

Logistic regression analyses showed that GRS (based on its quartile distribution in the control subjects), parity, breast cancer family history, and previous diagnosis of benign breast disease were independently associated with breast cancer risk (Table 5). These factors were used to construct a risk assessment model. The contribution of these risk factors to the risk assessment model was presented in Table 6. The *c* statistic for the full model was 0.6321. The contribution of the GRS to the model, as measured by the reduction in adjusted *c* statistic from the full model, was 0.0374. Notably, the effect of GRS was stronger than breast cancer family history (reduction in adjusted *c* statistic = 0.0103), and prior benign breast disease (reduction in adjusted *c* statistic = 0.0324) with a statistically significant p-value ($P \leq 0.05$). We did not evaluate age at menarche, age at first live birth, and body mass index in the risk assessment model since these factors were not significantly associated with the risk of breast cancer in our study (data not shown).

Discussion

In the current study, we investigated associations between index SNPs in 17 breast cancer susceptibility loci and incident breast cancer in 1,511 cases and 1,454 controls from European-ancestry women. As expected, each SNP identified in GWAS was associated with a small to moderate increased risk of breast cancer. We found that five SNPs at 2q35/*TNP1*, 3p24/*SLC4A*7, 6q25/*ESR1*, 10q26/*FGFR2*, and 16q12/*TOX3* loci were significantly associated ($P \leq 0.05$) with overall breast cancer risk in the same direction as previously reported. GRS analyses, based on those five significant SNPs, showed significant associations with overall breast cancer and breast cancer stratified by subtype, menopausal status, age of onset or breast cancer family history. GRS and previous benign breast disease are the two top predictors in the risk assessment model.

Among the five loci that are significantly associated with breast cancer risk in our study population, all of them showed nominally significant associations with ER+ cancer, and three of them with ER− cancer (*rs13387042*, *rs1219648*, and *rs4784227*). The strength of association was stronger for ER+ than ER− cancer, which is consistent with previous studies [4,9,12,13]. In addition, for the other two SNPs, *rs11249433* at 1p11/*FCGR1B* and *rs2380205* at 10p15/*ANKRD16, a* suggestive association with ER−, but not ER+ cancer was observed. Recently, two SNPs (*rs2046210* at 6q25/*ESR1* and *rs8170* at 19p13/*BABAM1*) were found to be associated with ER− cancer [12]. In the present study, only a suggestive association was observed at SNP *rs2046210* and no association was found for *rs8170* with ER−cancer.

In the present study, two SNPs, including *rs1045485* (2q33/*CASP8)* and *rs1219648* (10q26/*FGFR2*) were associated with TNBC. Similarly, *rs1045485* showed a stronger association with ER− than ER+ in the Breast Cancer Association Consortium (BCAC) [14]. However, the association of *rs1219648* with TNBC was not replicated in the BCAC [14] nor the Triple Negative Breast Cancer Consortium (TNBCC) [15]. In the TNBCC [15], 6q25/*ESR1*, 16q12/*TOX3*, 14q24/*RAD51L1*, and 19p13/*BABAM1* were significantly associated with the risk of TNBC. However, only two of them, 19p13/*BABAM1* and 16q12/*TOX3*, were replicated in the BCAC [14,16]. In the present study, a similar, but non-significant association was observed with ORs (95% CI) of 1.19 (0.91–1.56) and 1.17 (0.86– 1.58) for 16q12/*TOX3* and 19p13/*BABAM1*, respectively. In the BCAC, significant associations with TNBC were observed at three loci, 5q11/*MAP3K1*, 11p15/*LSP1* and 2q35/*TNP1*. None of these associations were observed in the present study nor in the TNBCC.

In the present study, we were only able to replicate the association between five of the 17 loci in association with overall breast cancer risk. Small sample size is likely to be the major reason for the non-replication. Most of the GWAS-identified loci were associated with small effect size, which resulted in limited statistical power for replication in studies with a sample size like the present study. Notably the statistical power was limited for the breast cancer subtype analysis. Another limitation of the present study is that although approximately 67 GWAS loci have already been reported in the literature, only 17 of them were investigated in the present study. In the present study, the Exomechip was used as our genotyping platform. This chip was designed by a large consortium from multiple institutes based on sequencing data from over 12,000 subjects (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). All variants added to GWAS catalog by August 16, 2011 were included in the chip; however, variants identified after that time were not included. We will plan to investigate those loci identified after August 16, 2011 in future studies. Despite this, the GRS, the combination of five loci, was one of the strongest predictors in our risk assessment model. This suggests that in addition to the traditional risk factors, genetic variants are important in risk assessment. The full risk assessment model established in this study provided only moderate discriminatory accuracy with a *c* statistic of 0.6321, however, there is advantage to including genetic variants since they can be accurately measured at any time during a person's entire life. Including all 67 GWAS loci in the model should further improve the prediction accuracy and may serve as a useful tool for identifying high risk women for close monitoring and cancer screening.

Conclusion

In summary, we have evaluated index SNPs in 17 breast cancer susceptibility loci that were initially identified by GWAS studies for their association with breast cancer in the NBHS. Five of them were associated with overall breast cancer. Two SNPs were nominally significantly associated with TNBC. The GRS is one of the strongest predictors of breast cancer risk in our risk assessment model, suggesting the potential utility of using GRS in identifying women who are at high risk of breast cancer.

Acknowledgement

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agents. This work was supported in part by US National Institutes of Health grant R01CA100374 to WZ. The authors wish to thank the study participants and research staff for their contributions and commitment to this project, Drs. Alecia Fair and Alicia Beeghly-Fadiel for their contribution to the NBHS, Regina Courtney for DNA preparation, Jing He and Yanfeng Zhang for data processing, and Samantha Stansel for clerical support in manuscript preparation. Study recruitment, telephone interviews, sample preparation and genotyping assays were conducted at the Survey and Biospecimen Shared Resource and Vanderbilt Microarray Shared Resource, which are supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485).

References

1. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet. 2003; 33(Suppl): 238–244. [PubMed: 12610533]
2. Nathanson KN, Wooster R, Weber BL. Breast cancer genetics: What we know and what we need. Nat Med. 2001; 7:552–556. [PubMed: 11329055]
3. Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. The Lancet Oncology. 2011; 12:477–488. [PubMed: 21514219]
4. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet. 2009; 41:585–590. [PubMed: 19330027]
5. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, et al. Novel Breast Cancer Susceptibility Locus at 9q31.2: Results of a Genome-Wide Association Study. Journal of the National Cancer Institute. 2011; 103:425–435. [PubMed: 21263130]
6. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J, Friedman E, Narod S, Olshen AB, Gregersen P, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proceedings of the National Academy of Sciences. 2008; 105:4340–4345.
7. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007; 39:870–874. [PubMed: 17529973]
8. Long J, Shu XO, Cai Q, Gao YT, Zheng Y, Li G, Li C, Gu K, Wen W, Xiang YB, et al. Evaluation of Breast Cancer Susceptibility Loci in Chinese Women. Cancer Epidemiology Biomarkers & Prevention. 2010; 19:2357–2365.
9. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergthorsson JT, Gudmundsson J, Aben KK, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2008; 40:703–706. [PubMed: 18438407]
10. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009; 41:324–328. [PubMed: 19219042]

11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988; 44:837–845. [PubMed: 3203132]
12. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BrG, Axelsson CK, Arias JI, et al. Heterogeneity of Breast Cancer Associations with Five Susceptibility Loci by Clinical and Pathological Characteristics. PLoS Genet. 2008; 4:e1000054. [PubMed: 18437204]
13. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2007; 39:865–869. [PubMed: 17529974]
14. Broeks A, Schmidt MK, Sherman ME, Couch FJ, Hopper JL, Dite GS, Apicella C, Smith LD, Hammet F, Southey MC, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Human Molecular Genetics. 2011; 20:3289–3303. [PubMed: 21596841]
15. Stevens KN, Vachon CM, Lee AM, Slager S, Lesnick T, Olswold C, Fasching PA, Miron P, Eccles D, Carpenter JE, et al. Common Breast Cancer Susceptibility Loci Are Associated with Triple-Negative Breast Cancer. Cancer Research. 2011; 71:6240–6249. [PubMed: 21844186]
16. Stevens KN, Fredericksen Z, Vachon CM, Wang X, Margolin S, Lindblom A, Nevanlinna H, Greco D, AittomÃ¤ki K, Blomqvist C, et al. 19p13.1 Is a Triple-Negative-Specific Breast Cancer Susceptibility Locus. Cancer Research. 2012; 72:1795–1803. [PubMed: 22331459]

Table 1

Characteristics of study participants in the Nashville Breast Health Study (NBHS)

Characteristics	Cases (N=1,511)	Controls (N=1,454)	*P*-value [e]
Demographic factors			
Age (year, mean ± sd)	53.25 ± 9.08	52.50 ± 9.25	0.03
Education level, high school or higher (%)	99.07	99.52	0.15
Reproductive risk factors			
Age at menarche (year, mean ± sd)	12.53 ± 1.46	12.62 ± 1.57	0.19
Postmenopausal (%)	63.84	59.42	0.01
Age at menopause (year, mean ± sd) [a]	46.36 ± 6.87	46.20 ± 7.00	0.81
Number of live births (mean ± sd)	2.23 ± 1.16	2.27 ± 1.00	0.05
Age at first live birth (year, mean ± sd) [b]	24.28 ± 5.33	24.62 ± 5.21	0.06
Ever used hormone replacement therapy (%)	54.77	52.41	0.20
Other risk factors			
Breast cancer family history (%) [c]	19.99	14.37	<.0001
Prior BBD diagnosis (%) [d]	49.11	34.73	<.0001
BMI (kg/m^2, mean ± sd)	27.10 ± 5.98	27.00 ± 6.15	0.35
BMI in postmenopausal women (kg/m^2, mean ± sd)	27.49 ± 5.94	27.54 ± 5.97	0.96

[a] Among postmenopausal women.

[b] Among parous women.

[c] Among first-degree women.

[d] BBD = benign breast disease

[e] *P*-values (two-sided) were derived from Wilcoxon rank sum test (for continuous variables) and χ^2 tests (for categorical variables)

Table 2

Association of GWAS-Identified SNPs with breast cancer risk in NBHS

SNP	Chr./Gene[a]	Alleles[b]	RAF[c]		OR (95% CI)[d]			P_{trend}
			Cases	Controls	Heterozygous	Homozygous	Per-allele	
Rs11249433	1p11/*FCGR1B*	G/A	0.44	0.42	1.11(0.94– 1.30)	1.13(0.91– 1.39)	1.07(0.96– 1.19)	0.21
Rs1045485	2q33/*CASP8*	G/C	0.87	0.87	0.84(0.47– 1.52)	0.87(0.49– 1.55)	1.01(0.86– 1.17)	0.94
Rs13387042	2q35/*TNP1*	A/G	0.55	0.50	1.26(1.04– 1.51)	1.49(1.21– 1.84)	1.22(1.10– 1.35)	2.0×10^{-4}
Rs4973768	3p24/*SLC4A7*	T/C	0.50	0.47	1.16(0.98– 1.39)	1.30(1.06– 1.60)	1.14(1.03– 1.26)	0.02
Rs889312	5q11/*MAP3K1*	C/A	0.29	0.29	1.05(0.90– 1.22)	0.96(0.73– 1.27)	1.01(0.90– 1.13)	0.85
Rs2046210	6q25/*ESR1*	A/G	0.38	0.35	1.09(0.94– 1.28)	1.33(1.05– 1.69)	1.14(1.02– 1.27)	0.02
Rs13281615	8q24/*MYC*	G/A	0.42	0.40	1.08(0.92– 1.27)	1.14(0.92– 1.41)	1.07(0.97– 1.19)	0.19
Rs1011970	9p21/*CDKN2A/2B*	T/G	0.17	0.17	1.01(0.86– 1.19)	1.07(0.67– 1.70)	1.02(0.89– 1.17)	0.78
Rs2380205	10p15/*ANKRD16*	C/T	0.58	0.56	1.02(0.83– 1.24)	1.13(0.91– 1.39)	1.07(0.96– 1.18)	0.22
Rs10995190	10q21/*ZNF365*	G/A	0.86	0.86	1.06(0.63– 1.78)	1.03(0.62– 1.71)	0.98(0.85– 1.14)	0.84
Rs704010	10q22/*ZMIZ1*	T/C	0.40	0.39	1.11(0.94– 1.30)	1.02(0.82– 1.27)	1.03(0.93– 1.14)	0.57
Rs1219648	10q26/*FGFR2*	G/A	0.46	0.39	1.56(1.33– 1.84)	1.75(1.41– 2.17)	1.36(1.22– 1.51)	1.4×10^{-8}
Rs3817198	11p15/*LSP1*	C/T	0.35	0.33	1.06(0.91– 1.23)	1.22(0.96– 1.56)	1.09(0.98– 1.21)	0.12
Rs614367	11q13/*CCND1*	T/C	0.16	0.16	0.95(0.81– 1.12)	1.23(0.75– 2.00)	1.00(0.87– 1.15)	0.96
Rs999737	14q24/*RAD51L1*	C/T	0.77	0.76	1.22(0.87– 1.72)	1.26(0.90– 1.76)	1.07(0.95– 1.21)	0.25
Rs4784227	16q12/*TOX3*	T/C	0.29	0.24	1.21(1.04– 1.41)	1.72(1.28– 2.31)	1.26(1.12– 1.42)	9.7×10^{-5}
Rs8170	19p13/*BABAM1*	A/G	0.20	0.19	0.99(0.85– 1.16)	1.70(1.13– 2.55)	1.09(0.96– 1.25)	0.17

[a] The closest gene.

[b] Risk/referent alleles based on Human Genome hg 19.

[c] Risk allele frequency of cases and controls.

[d] Adjusted for age and top 5 principal components (PCs).

Table 3

Association of GWAS-Identified SNPs with breast cancer risk in NBHS, stratified by breast cancer subtypes

SNP	Chr./Gene[a]	ER+ (918 cases vs 1,454 controls)		ER− (299 cases vs 1,454 controls)		TNBC (148 cases vs 1,454 controls)	
		OR (95% CI)[b]	P_{trend}[b]	OR (95% CI)[b]	P_{trend}[b]	OR (95% CI)[b]	P_{trend}[b]
rs11249433	1p11/*FCGR1B*	1.08(0.96– 1.22)	0.20	1.18(0.98– 1.41)	0.07	1.02(0.80– 1.30)	0.86
rs1045485	2q33/*CASP8*	0.97(0.81– 1.15)	0.72	1.25(0.94– 1.65)	0.12	1.55(1.02– 2.34)	0.04
rs13387042	2q35/*TNP1*	1.25(1.10– 1.40)	3.0×10^{-4}	1.23(1.03– 1.47)	0.02	1.13(0.89– 1.45)	0.31
rs4973768	3p24/*SLC4A7*	1.14(1.03– 1.26)	1.0×10^{-2}	1.10(0.92– 1.31)	0.29	1.01(0.79– 1.28)	0.97
rs889312	5q11/*MAP3K1*	1.07(0.94– 1.21)	0.34	0.88(0.72– 1.07)	0.21	0.93(0.71– 1.22)	0.59
rs2046210	6q25/*ESR1*	1.13(1.00– 1.29)	5.0×10^{-2}	1.19(0.99– 1.44)	0.07	1.16(0.90– 1.50)	0.24
rs13281615	8q24/*MYC*	1.10(0.98– 1.24)	0.10	0.96(0.80– 1.15)	0.64	0.88(0.69– 1.12)	0.31
rs1011970	9p21/*CDKN2A/2B*	1.02(0.87– 1.19)	0.85	1.09(0.86– 1.38)	0.46	1.26(0.93– 1.71)	0.13
rs2380205	10p15/*ANKRD16*	1.07(0.96– 1.18)	0.22	1.19(0.99– 1.42)	0.06	1.20(0.94– 1.54)	0.14
rs10995190	10q21/*ZNF365*	0.97(0.83– 1.15)	0.77	1.07(0.83– 1.38)	0.60	1.14(0.80– 1.63)	0.46
rs704010	10q22/*ZMIZ1*	1.02(0.90– 1.15)	0.74	1.07(0.89– 1.28)	0.47	0.90(0.71– 1.15)	0.41
rs1219648	10q26/*FGFR2*	1.35(1.20– 1.53)	9.4×10^{-7}	1.31(1.09– 1.57)	4.0×10^{-3}	1.31(1.03– 1.67)	0.03
rs3817198	11p15/*LSP1*	1.09(0.96– 1.23)	0.17	1.04(0.86– 1.25)	0.71	0.93(0.72– 1.21)	0.59
rs614367	11q13/*CCND1*	1.00(0.85– 1.18)	0.97	1.00(0.79– 1.28)	0.97	0.95(0.68– 1.33)	0.77
rs999737	14q24/*RAD51L1*	1.09(0.94– 1.25)	0.25	1.01(0.82– 1.25)	0.90	0.92(0.70– 1.22)	0.57
rs4784227	16q12/*TOX3*	1.30(1.14– 1.49)	9.2×10^{-5}	1.32(1.08– 1.60)	4.0×10^{-3}	1.19(0.91– 1.56)	0.21
rs8170	19p13/*BABAM1*	1.05(0.91– 1.22)	0.50	1.17(0.93– 1.46)	0.18	1.17(0.86– 1.58)	0.32

[a] The closest gene.

[b] Adjusted for age and top 5 principal components (PCs).

Table 4

Association of GRS [a] with breast cancer, overall and by subtype, in NBHS

	GRS quartile[b]				
	Q1	**Q2**	**Q3**	**Q4**	$P_{\text{for trend}}$[c]
All cases					
N (Cases/Controls)	434/609	371/355	375/286	331/204	
OR (95% CI)	1.00 (reference)	1.48(1.22– 1.79)	1.85(1.52– 2.25)	2.26(1.82– 2.80)	2.0×10^{-15}
ER+ vs. Controls					
N (Cases/Controls)	253/609	228/355	230/286	207/204	
OR (95% CI)	1.00 (reference)	1.57(1.26– 1.97)	1.95(1.55– 2.45)	2.42(1.90– 3.09)	1.4×10^{-14}
ER− vs. Controls					
N (Cases/Controls)	84/609	75/355	77/286	63/204	
OR (95% CI)	1.00 (reference)	1.52(1.08– 2.13)	1.97(1.40– 2.77)	2.21(1.54– 3.19)	2.5×10^{-6}
TNBC vs. Controls					
N (Cases/Controls)	48/609	32/355	39/286	29/204	
OR (95% CI)	1.00 (reference)	1.12(0.70– 1.79)	1.71(1.10– 2.68)	1.80(1.10– 2.93)	0.01
Pre-menopausal					
N (Cases/Controls)	145/244	144/147	146/125	111/74	
OR (95% CI)	1.00 (reference)	1.63(1.19– 2.22)	1.95(1.42– 2.67)	2.46(1.71– 3.53)	1.8×10^{-8}
Post-menopausal					
N (Cases/Controls)	289/365	227/208	228/161	220/130	
OR (95% CI)	1.00 (reference)	1.38(1.08– 1.77)	1.79(1.38– 2.31)	2.12(1.62– 2.77)	2.3×10^{-8}
Early-Onset (≤45)					
N (Cases/Controls)	84/609	81/355	74/286	73/204	
OR (95% CI)	1.00 (reference)	1.67(1.12– 2.47)	1.99(1.33– 3.00)	3.02(1.97– 4.62)	1.4×10^{-6}
Late-Onset (>45)					
N (Cases/Controls)	350/609	290/355	301/286	258/204	
OR (95% CI)	1.00 (reference)	1.50(1.21– 1.85)	1.92(1.55– 2.39)	2.18(1.73– 2.76)	4.4×10^{-13}
Breast cancer family history					
N (Cases/Controls)	84/533	71/309	83/231	64/172	

	GRS quartile[b]				
	Q1	**Q2**	**Q3**	**Q4**	$P_{\text{for trend}}$[c]
OR (95% CI)	1.00 (reference)	1.48(1.05– 2.10)	2.33(1.65– 3.29)	2.28(1.58– 3.31)	3.9×10^{-8}
No breast cancer family history					
N (Cases/Controls)	308/418	352/364	316/300	233/163	
OR (95% CI)	1.00 (reference)	1.49(1.21– 1.83)	1.93(1.55– 2.41)	2.35(1.86– 2.97)	4.2×10^{-14}

[a] GRS = genetic risk score

[b] GRS (genetic risk score) was constructed based on 5 SNPs, including rs13387042, rs4973768, rs2046210, rs1219648, and rs4784227.

[c] Adjusted for age and top 5 principal components (PCs).

Breast Cancer. Author manuscript; available in PMC 2016 September 01.

Table 5

Associations between risk of breast cancer and GRS [a] and established risk factors in NBHS

Predictor (code)	Case	Control	OR (95% CI) [c]	OR (95% CI) [d]
Genetic risk score, quartile				
1 (0 [low])	434	609	1.00 (reference)	1.00 (reference)
2 (1)	371	355	1.47 (1.21– 1.78)	1.55 (1.25– 1.92)
3 (2)	375	286	1.85 (1.52– 2.25)	1.83 (1.47– 2.27)
4 (3 [high])	331	204	2.28 (1.84– 2.82)	2.42 (1.92– 3.07)
$P_{\text{for trend}}$			<.0001	<.0001
Parity				
1 (0)	280	231	1.00 (reference)	1.00 (reference)
2 (1)	605	595	0.84 (0.68– 1.03)	0.84 (0.68– 1.04)
≥ 3 (2)	375	402	0.73 (0.58– 0.92)	0.76 (0.61– 0.96)
$P_{\text{for trend}}$			7.4×10^{-3}	0.03
Breast cancer family history				
No (0)	1209	1245	1.00 (reference)	1.00 (reference)
Yes (1)	302	209	1.47 (1.21– 1.79)	1.57 (1.26– 1.95)
Prior BBD diagnosis [b]				
No (0)	769	949	1.00 (reference)	1.00 (reference)
Yes (1)	742	505	1.80 (1.55– 2.09)	1.77 (1.50– 2.09)

[a] GRS = genetic risk score

[b] BBD = benign breast disease

[c] Adjusted for age and education level.

[d] Adjusted for age, education level, and all variables listed in this table.

Table 6

Decrease in *c* statistic for each risk factor when it was removed from full model [†]

Predictor	Decrease in *c* statistic (*P*-value)
Prior benign breast disease	0.0324 (<.0001)
Genetic risk score	0.0374 (<.0001)
Breast cancer family history	0.0103 (0.0257)
Parity	0.0012 (0.3589)

[†]The full model includes all four risk factors in this table, and the *c* statistic for the full model was 0.6321. All *P*-values are two-sided and derived from the nonparametric Mann-Whitney *U* test [11].