

Evaluating Genome-Wide Association Study-Identified Breast Cancer Risk Variants in African-American Women

Jirong Long^{1*}, Ben Zhang¹, Lisa B. Signorello^{1,2}, Qiuyin Cai¹, Sandra Deming-Halverson¹, Martha J. Shrubsole¹, Maureen Sanderson³, Joe Dennis⁴, Kyriaki Michailiou⁴, Douglas F. Easton^{4,5}, Xiao-Ou Shu¹, William J. Blot^{1,2}, Wei Zheng¹

1 Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **2** International Epidemiology Institute, Rockville, Maryland, United States of America, **3** Meharry Medical College, Nashville, Tennessee, United States of America, **4** Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, **5** Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

Abstract

Genome-wide association studies (GWAS), conducted mostly in European or Asian descendants, have identified approximately 67 genetic susceptibility loci for breast cancer. Given the large differences in genetic architecture between the African-ancestry genome and genomes of Asians and Europeans, it is important to investigate these loci in African-ancestry populations. We evaluated index SNPs in all 67 breast cancer susceptibility loci identified to date in our study including up to 3,300 African-American women (1,231 cases and 2,069 controls), recruited in the Southern Community Cohort Study (SCCS) and the Nashville Breast Health Study (NBHS). Seven SNPs were statistically significant ($P \leq 0.05$) with the risk of overall breast cancer in the same direction as previously reported: rs10069690 (5p15/*TERT*), rs999737 (14q24/*RAD51L1*), rs13387042 (2q35/*TNP1*), rs1219648 (10q26/*FGFR2*), rs8170 (19p13/*BABAM1*), rs17817449 (16q12/*FTO*), and rs13329835 (16q23/*DYL2*). A marginally significant association ($P < 0.10$) was found for three additional SNPs: rs1045485 (2q33/*CASP8*), rs4849887 (2q14/*INHBB*), and rs4808801 (19p13/*ELL*). Three additional SNPs, including rs1011970 (9p21/*CDKN2A/2B*), rs941764 (14q32/*CCDC88C*), and rs17529111 (6q14/*FAM46A*), showed a significant association in analyses conducted by breast cancer subtype. The risk of breast cancer was elevated with an increasing number of risk variants, as measured by quintile of the genetic risk score, from 1.00 (reference), to 1.75 (1.30–2.37), 1.56 (1.15–2.11), 2.02 (1.50–2.74) and 2.63 (1.96–3.52), respectively, ($P = 7.8 \times 10^{-10}$). Results from this study highlight the need for large genetic studies in AAs to identify risk variants impacting this population.

Citation: Long J, Zhang B, Signorello LB, Cai Q, Deming-Halverson S, et al. (2013) Evaluating Genome-Wide Association Study-Identified Breast Cancer Risk Variants in African-American Women. PLoS ONE 8(4): e58350. doi:10.1371/journal.pone.0058350

Editor: Paolo Peterlongo, IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Italy

Received: September 17, 2012; **Accepted:** February 3, 2013; **Published:** April 8, 2013

Copyright: © 2013 Long et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Sample preparation was conducted at the Survey and Biospecimen Shared Resources, which is supported in part by Vanderbilt-Ingram Cancer Center (P30 CA68485). The SCCS dataset used for the present analyses was supported by U.S. National Institutes of Health grant R01CA92447. This project was also supported by US NIH grants R01CA100374 and U19 CA148065, as well as Ingram Professorship funds. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Jirong.Long@Vanderbilt.Edu

Introduction

Breast cancer is one of the most common malignancies diagnosed among women worldwide, including women of African descent. African American (AA) women experience a disproportionate burden of breast cancer. Age-adjusted mortality rate of this cancer is more than 40% higher in AAs than in European-ancestry populations. AA women tend to be diagnosed with breast cancer at a younger age and with more aggressive types of the disease, such as ER- (estrogen receptor negative) and ER-/PR-/HER2- (estrogen receptor negative, progesterone receptor negative, HER2 expression negative) breast cancer. Within AAs, women having a higher African ancestry level, estimated by ancestry informative markers (AIMs), have been shown to have an increased likelihood of ER-/PR- versus ER+/PR+ breast cancers [1]. However, such association was not observed in the Women's Contraceptive and Reproductive Experiences (CARE) study [2].

To date, four high-penetrance genes (*BRCA1*, *BRCA2*, *TP53*, and *P TEN*) and four moderate-penetrance genes (*CHEK2*, *ATM*,

BRIP1, and *PALB2*) have been discovered for breast cancer [3]. Candidate gene studies have largely failed to identify low-penetrance loci which can be robustly replicated in other studies [3]. Genome-wide association studies (GWAS) have emerged as the most widely used approach to identify genetic variants for complex diseases [4]. Since 2007, 67 common genetic susceptibility loci have been discovered, including 25 from several earlier GWAS [5–21] and 42 from a recent international Collaborative Oncological Gene-Environment Study (COGS) [22]. However, except the 5p15/*TERT* locus which was discovered among AA women [20], all other risk variants initially were identified in studies conducted in European or Asian descendants. Given the considerable differences in genetic architecture, including allele frequencies, linkage disequilibrium (LD) structure, and genetic diversity between the African-ancestry genome and genomes of Asian- and European-ancestry populations [23], it is important to investigate whether GWAS-identified variants are associated with breast cancer risk in African-ancestry populations. This investiga-

tion not only assesses the generalizability of initial GWAS findings, but also provides valuable data to guide fine-mapping efforts in the search for causal variants.

In this current study, we evaluated risk variants in all 67 breast cancer loci identified to date in an AA population of 1,231 cases and 2,069 controls.

Materials and Methods

This study uses resources from the Southern Community Cohort Study (SCCS) and the Nashville Breast Health Study (NBHS). The NBHS is a population-based case-control study [10]. Incident breast cancer cases were identified through the Tennessee State Cancer Registry and a network of major hospitals that provide medical care for patients with breast cancer. Controls were identified mostly via random-digit dialing of households in the same geographic area as cases and frequency-matched to cases on age (5-year group). All participants were phone-interviewed to obtain information related to personal and family medical history, and other lifestyle factors. A total of 437 cases and 252 controls from the NBHS who provided exfoliated buccal cell samples were included in the project. The SCCS is a prospective cohort study initiated in 2002 investigating racial disparities in the risk of cancer and other chronic diseases [24]. SCCS includes approximately 86,000 participants with two-thirds being AAs. In the SCCS, participants completed a comprehensive, in-person, baseline interview or completed a study questionnaire covering various aspects of health conditions, behavioral factors, personal and family medical history, and other lifestyle factors. In the SCCS, 679 breast cancer cases (217 incident and 462 prevalent cases) were included in the project. They were selected from those who were diagnosed with breast cancer and provided a blood or buccal cell sample. In the SCCS, controls ($n=680$) were selected randomly from those who were cancer-free and frequency-matched to cases in a 1:1 ratio on age at enrollment (± 1 year), recruitment method, and sample type (blood/buccal cell).

Ethics Statement

Written, informed consent was obtained from all participants prior to interview, and the study protocols have been approved by Institutional Review Boards at Vanderbilt University (for NBHS and SCCS) and Meharry Medical College (for SCCS).

Genotyping for samples described above was conducted using the protocol for the COGS Project as described elsewhere [22]. In brief, a custom Illumina Infinium BeadChip which contains 211,155 SNPs was genotyped. Individuals were excluded for any of the following reasons: 1) genotypically not female, 2) call rate $<95\%$, 3) low or high heterozygosity ($P < 10^{-6}$), 4) genotyping not concordant with previous data, 5) duplicates or “cryptic” duplicates, 6) first-degree relative, 7) ethnic outliers based on a subset of 37,000 uncorrelated markers which passed QC. SNPs were excluded for any of the following reasons: 1) call rate $<95\%$, 2) deviated from HWE in controls at $P < 10^{-7}$, 3) genotyping discrepancies in more than 2% of duplicate samples. Data cleaning was conducted within the whole COGS Project. After QC, a total of 199,961 SNPs for 1,116 cases and 932 controls were included in the dataset. We then performed principal component analyses (PCA) using a set of 4,613 uncorrelated SNPs (neighboring distance >500 kb, MAF >0.2 , $r^2 < 0.1$, and call rate $>99\%$). Five additional participants (three cases and two controls) were excluded due to $>6 \sigma$ away from the means of PCA1 and PCA2. In total, 1,113 cases and 930 controls from the SCCS and NBHS were successfully genotyped by the COGS SNP array.

In our previous project, nine GWAS-identified SNPs were genotyped using Taqman/Sequenom in 810 cases and 1,784 controls from the NBHS and SCCS [25]. Among them, 118 cases and 1,139 controls were not included in the study using COGS SNP array. Data for these nine SNPs, including rs13387042, rs10941679, rs889312, rs2046210, rs13281615, rs1219648, rs2981582, rs3817198, and rs3803662, were combined from the data genotyped through Taqman/Sequenom [25] and data newly obtained using COGS SNP array. In total, 1,231 cases and 2,069 controls from the SCCS (743 cases and 1,797 controls) and NBHS (488 cases and 272 controls) with genotype data available were included in the final analyses.

Statistical Analysis

Individual African ancestry level was estimated from 612 AIMs included in the COGS SNP array using the program frappe (<http://med.stanford.edu/tanglab/software/frappe.html>), which implements an Estimation-Maximization algorithm for simultaneously inferring each individual's ancestry proportion and allele frequencies in the ancestral populations [26]. Associations between individual SNP and breast cancer risk were assessed using odds ratios (ORs) and 95% confidence intervals (CIs) derived from logistic regression models and adjusted for age and study site. In the present study, data are available for ER, PR and Her2 among 564, 555, and 250 breast cancer women, respectively. Subgroup analyses were conducted within breast cancer subtypes including ER+, ER-, and ER-/PR-/HER2-. Principal component analyses (PCA) were conducted based on 4,613 uncorrelated SNPs using EIGENSTRAT [27]. The first ten principal components were included in the logistic regression model to test association for the SNPs in the present study.

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. Since data for a complete set of SNPs were only available for the 1,110 cases and 929 controls genotyped using COGS SNP array, GRS analysis was conducted among these subjects.

We constructed a GRS using SNPs that showed a statistically significant association with breast cancer risk in this study. In the present study, ten SNPs were associated with breast cancer ($P < 0.1$) with direction of association consistent with previous reports. SNP rs1219648 was not included in the COGS SNP array and only 66% women of the COGS subjects had data available for this SNP, genotyped in our previous project through Sequenom (23). SNP rs13387042 was genotyped in both by the COGS SNP array and Sequenom (23). This SNP, however, did not reach $P < 0.1$ in the samples analyzed using COGS SNP array and thus was not included in the GRS analyses. Thus, these two SNPs, rs1219648 and rs13387042, were excluded in GRS analyses. The remaining eight SNPs, rs10069690, rs999737, rs8170, rs17817449, rs13329835, rs1045485, rs4849887, and rs4808801 were included in the GRS analyses. All statistical analyses were conducted in SAS, version 9.3, with the use of two-tailed tests.

Results

The distributions of demographic characteristics and known breast cancer risk factors for cases and controls are shown in Table 1. Cases were more likely to have a family history of breast cancer, an earlier age at menarche, fewer live births, older age at first live birth and high body mass index.

Table 1. Characteristics of study participants.

Characteristic	Cases (N = 1,231)	Controls (N = 2,069)	P value
Sample sources			
NBHS ^a	488	272	
SCCS ^b	743	1,797	
Age (year, mean±sd)	54.83±9.74	55.14±9.64	0.37
Age at menarche (year, mean±sd)	12.60±1.87	12.83±2.04	0.002
Postmenopausal (%)	78.69	73.49	<0.001
Number of live births (median, range)	3 (1, 16)	3 (1, 18)	0.027
Age at first live birth (year, mean±sd) ^c	20.95±5.09	20.42±4.77	0.004
Breast cancer family history (%)	20.69	10.36	<0.001
BMI (kg/m ² , mean±sd)	31.95±7.34	31.49±8.07	0.108
BMI in postmenopausal women	32.16±7.21	31.45±7.97	0.03

^aThe Nashville Breast Health Study.

^bThe Southern Community Cohort Study.

^cAmong parous women.

doi:10.1371/journal.pone.0058350.t001

African Ancestry Level and Breast Cancer

We did not find any statistically significant association between African ancestry level and breast cancer risk. On average, cases had 83.22% African ancestry, and controls had 83.86%. No significant association was observed between African ancestry level with breast cancer subtype, either. The African ancestry proportion was 82.50%, 82.74%, and 82.25% for ER+/PR+, ER-/PR-, ER-/PR-/HER2- cases, respectively (Table S1). However, difference was observed between the two study cohorts with a higher African ancestry level in the SCCS than in the NBHS. On average, the African ancestry level is 85.23% and 80.14% in controls, and 84.50% and 81.22% in cases from the SCCS and NBHS, respectively.

Evaluation of SNPs in 25 Previously Reported Loci

Two SNPs have been discovered in the 10q21/*ZNF365* locus: rs10995190 in Europeans [21] and rs10822013 in East Asians [17]. These two SNPs are not in LD based on data from HapMap Africans ($r^2=0.001$), and both of them were included in the present study. Similarly, in the locus 16q12/*TOX3*, SNPs rs4784227 and rs3803662 were discovered in Asians [14] and Europeans [7], respectively. SNP rs4784227 was not included in the COGS SNP array. SNP rs17271951, which was in strong LD with rs4784227 based on HapMap Africans ($r^2=1.0$), was used as a substitute. SNPs rs17271951 and rs3803662 are not in LD in HapMap Africans ($r^2=0.03$), and both of them were included in the final analyses. In the 10q26/*FGFR2* locus, both rs1219648 and rs2981582 were discovered in Europeans but were in moderate LD in HapMap Africans ($r^2=0.25$). For the 6q25/*TAB2* locus, rs9485370 was used to replace the originally reported SNP rs9485372 [28] which was not included in the COGS SNP array ($r^2=1$ in HapMap Africans). For the other 21 loci, one index SNP (reported in previous GWAS) per locus was selected. Therefore, in total, 28 SNPs in previously reported 25 loci were investigated in the present study. Five of them were deviated from Hardy-Weinberg test with $P<0.05$, including rs1045485, rs10995190, rs13281615, rs889312, and rs999737.

Of the 28 SNPs, 19 SNPs had an OR of breast cancer risk in the same direction as initial reports. This is higher than expected under the null hypothesis ($P=0.04$, binomial sign test). Five SNPs were nominally statistically significant ($P\leq 0.05$) in

the same direction as previously reported (Table 2 and Table S2). SNP rs10069690 in the 5p15/*TERT* locus, previously discovered in African-ancestry population, showed a moderate-strong association in the present study with OR (95% CI) 1.19 (1.05–1.36) and P -value 0.007. Notably, the MAF is very low (0.047 in controls) for SNP rs999737 at 14q24/*RAD51L1*, however, a strong association was observed for this SNP, with OR (95% CI) 1.59 (1.15–2.19) and P -value 0.005. Allelic OR (95% CI) for the other 3 SNPs are 1.17 (1.04–1.33) for rs13387042 ($P=0.011$), 1.17 (1.04–1.33) for rs1219648 ($P=0.011$), and 1.25 (1.07–1.47) for rs8170 ($P=0.006$). The *CASP8* SNP (rs1045485) identified previously through a candidate gene approach has a low frequency (MAF=0.065) in AAs. A marginally significant association was found for this SNP with OR (95% CI) of 1.25 (0.96–1.62) and P -value 0.096. Significant association was observed for SNP rs4973768 at 3p24/*SLC447* in this study, however, in the opposite direction as previously reported. No association was observed with the other 21 SNPs. Among them, the minor allele in rs10771399 at 12p11/*PTHLH* is very rare in AAs with MAF=0.034 in controls.

Evaluation of SNPs in 42 Newly Identified Loci

Of the 42 SNPs, one for each of the 42 loci recently discovered in European-ancestry populations [22] were included in the present study. Among them, SNPs rs11780156, rs13329835, rs17356907, rs616488, and rs6762644 were deviated from HWE test with $P<0.05$. Among the 42 investigated SNPs, 27 had an OR of breast cancer risk in the same direction in Europeans (Table 2 and Table S2). However, only two of them showed a nominally statistically significant association ($P\leq 0.05$). The OR (95% CI) was 1.21 (1.06–1.37) for rs17817449 at 16q12/*FTO* ($P=0.004$) and 1.16 (1.01–1.32) for rs13329835 at 16q23/*DYL2* ($P=0.037$). A marginally significant association ($P<0.10$) in the same direction as in Europeans was found for two additional SNPs, with ORs (95% CI) of 1.13 (0.99–1.30) for rs4849887 at 2q14/*INHBBP* ($P=0.075$), and 1.13 (0.98–1.29) for rs4808801 at 19p13/*ELL* ($P=0.091$). The 22q13/*MKLI* SNP rs6001930 showed an association ($P=0.035$), however the direction was inconsistent with that observed in Europeans. No associations were identified for the other 37 SNPs, among which four SNPs have a MAF

Table 2. Associations of breast cancer risk with 10 SNPs located in reported breast-cancer susceptibility loci in African Americans with $P < 0.1$.

SNP	Chr./gene ^a	Allele ^b	RAF (cases/controls) ^c	N (cases/controls)	OR (95% CI) ^d		P_{trend}	Power (%) ^e
					Heterozygous	Homozygous		
rs1045485	2q33/CASP8	G/C	0.946/0.935	1,113/930	1.99 (0.57–6.93)	2.38 (0.70–8.04)	0.096	54
rs13387042	2q35/TNP1	A/G	0.762/0.739	1,230/2,059	1.14 (0.82–1.59)	1.35 (0.98–1.87)	0.011	86
rs10069690	5p15/TERT	T/C	0.629/0.592	1,112/930	1.39 (1.07–1.81)	1.50 (1.15–1.97)	0.007	86
rs1219648	10q26/FGFR2	G/A	0.456/0.416	826/1,769	1.12 (0.92–1.37)	1.39 (1.09–1.79)	0.011	83
rs999737	14q24/RAD51L1	C/T	0.966/0.953	1,113/930	2.08 (0.39–11.20)	3.24 (0.62–16.98)	0.005	97
rs8170	19p13/BABAM1	A/G	0.212/0.177	1,112/930	1.19 (0.98–1.44)	1.90 (1.15–3.12)	0.006	88
rs4849887	2q14/INHBB	C/T	0.725/0.697	1,113/930	1.15 (0.83–1.60)	1.30 (0.94–1.80)	0.075	52
rs17817449	16q12/FTO	T/G	0.636/0.597	1,113/930	1.02 (0.79–1.34)	1.37 (1.04–1.80)	0.004	92
rs13329835	16q23/DYL2	G/A	0.677/0.655	1,113/930	0.88 (0.65–1.20)	1.16 (0.86–1.58)	0.037	71
rs4808801	19p13/ELL	A/G	0.305/0.279	1,112/929	1.09 (0.90–1.31)	1.34 (0.96–1.85)	0.091	50

^aThe closest gene.

^bRisk/reference alleles based on NCBI Human Genome Build 36 forward strand.

^cRisk allele frequency of cases and controls.

^dAdjusted with age, study (NBHS and SCCS), and the first ten principal components.

^ePower to identify an association at an alpha level of 0.05 under additive model.

doi:10.1371/journal.pone.0058350.t002

< 0.05 , including rs11552449 at 1p13/*AP4BI*, rs1353747 at 5q11/*PDE4D*, rs11780156 at 8q24/*MYC*, and rs11571833 at 13q13/*BRC42*.

Evaluation of Associations by Breast Cancer Subtypes

Association results between SNPs with risk of breast cancer by subtypes, including ER+, ER-, and ER-/PR-/HER2-, are presented in Table 3. Only SNPs that showed significant ($P \leq 0.05$) or marginally significant ($P \leq 0.1$) association with any of these three subtypes of breast cancer are presented. Three SNPs that were not associated with overall breast cancer showed nominally statistical significance ($P \leq 0.05$) in analysis by subtypes in the same direction as previously reported. SNPs rs1011970 at 9p21/*CDKN2A/2B* and rs941764 at 14q32/*CCDC88C* were associated with ER+ breast cancer with ORs (95% CI) 1.27 (1.05–1.54) ($P = 0.014$) and 1.26 (1.02–1.56) ($P = 0.032$), respectively. SNP rs17529111 at 6q14/*FAM46A* was associated with ER-/PR-/HER2- tumor with OR (95% CI) of 1.97 (1.02–3.82) ($P = 0.043$). Differences in the strength of the association were also observed for three other SNPs across breast cancer subtypes. SNP rs17817449 at 16q12/*FTO* showed an association with ER+ tumor with OR (95% CI) of 1.32 (1.09–1.60) but not with ER- or ER-/PR-/HER2- tumors. Both rs10069690 at 5p15/*TERT* and rs999737 at 14q24/*RAD51L1* showed the strongest association for ER-/PR-/HER2- breast cancer though associations were also observed for ER+ and ER- cancers.

Genetic Risk Score (GRS) Analyses for Overall Breast Cancer

Significant associations were observed between GRS and risk of breast cancer (Table 4). ORs (95% CIs) for overall breast cancer risk across increasing quintiles of GRS were 1.00 (reference), 1.75 (1.30–2.37), 1.56 (1.15–2.11), 2.02 (1.50–2.74) and 2.63 (1.96–3.52) ($P = 7.8 \times 10^{-10}$). Such associations also were observed for all three subtypes of breast cancer.

Discussion

In the present study, we investigated associations of 70 index SNPs in 67 breast cancer susceptibility loci identified to date in up to 1,231 cases and 2,069 controls of AA women. We found that seven SNPs were significantly associated ($P < 0.05$) and three SNPs were marginally significantly associated ($P < 0.10$) with overall breast cancer risk in the same association direction as previously reported. Three additional SNPs showed a significant association ($P < 0.05$) when stratified by breast cancer subtype. GRS analyses showed significant associations with the risk of overall or subtype of breast cancer.

In the present study population, on average, approximate 83% of genetic ancestry is African origin, which is similar to the estimate in other studies [1,2,29,30]. Women in the SCCS have a higher African ancestry level than those in the NBHS. In the NBHS, most women were recruited in Tennessee, while SCCS women were recruited in 12 southern states including Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Tennessee, Mississippi, South Carolina, North Carolina, Virginia, and West Virginia. We did not find a significant difference in African ancestry level between breast cancer cases and controls or across breast cancer subtypes. These results are consistent with the recent finding in the CARE Study [2]. However, in another study, significant association was observed between genetic ancestry with ER+, PR+, or localized breast tumors [1]. In the present study, data for ER, PR and HER2 were available for only a portion of the subjects. Therefore, statistical power in regard to subtype of breast cancer analyses may be limited.

Among the ten SNPs that showed significant or marginally significant association with overall breast cancer risk in the present study, six have been investigated in previous studies of African-ancestry populations [20,25,31–36]. SNP rs10069690 (5p15/*TERT*) was originally discovered in a GWAS of AAs [20], so it is expected that this SNP should be replicated in the present study. For SNP rs1219648 at 10q26/*FGFR2*, association was observed in our previous study [25], the Carolina Breast Cancer Study (CBCS) [32], and the Women's Health Initiative (WHI) study [35], but not

Table 3. Association of breast cancer risk with selected GWAS-identified SNPs located in reported breast-cancer susceptibility loci in African Americans, by breast cancer subtypes.

SNP ^a	Chr./Gene ^b	ER+ (N = 369)		ER- (N = 195)		ER-/PR-/HER2- (N = 68)		<i>P</i> _{heterogeneity} ^d
		OR (95% CI) ^c	<i>P</i> _{trend} ^c	OR (95% CI) ^c	<i>P</i> _{trend} ^c	OR (95% CI) ^c	<i>P</i> _{trend} ^c	
rs13387042	2q35/ <i>TNP1</i>	1.20 (0.99–1.47)	0.064	1.10 (0.86–1.42)	0.441	1.00 (0.67–1.51)	0.985	0.591
rs10069690	5p15/ <i>TERT</i>	1.18 (0.98–1.42)	0.075	1.36 (1.07–1.72)	0.011	1.58 (1.07–2.32)	0.021	0.359
rs9485370	6q25/ <i>TAB2</i>	1.21 (0.97–1.52)	0.093	0.85 (0.65–1.10)	0.209	0.70 (0.46–1.05)	0.082	0.041
rs1011970	9p21/ <i>CDKN2A/2B</i>	1.27 (1.05–1.54)	0.014^e	0.96 (0.75–1.22)	0.733	1.08 (0.73–1.58)	0.700	0.074
rs999737	14q24/ <i>RAD51L1</i>	1.39 (0.88–2.18)	0.156	2.20 (1.11–4.36)	0.024	4.72 (1.10–20.28)	0.037	0.270
rs17271951	16q12/ <i>TOX3</i>	1.14 (0.79–1.66)	0.485	1.00 (0.62–1.62)	0.994	1.77 (0.95–3.31)	0.075	0.672
rs8170	19p13/ <i>BABAM1</i>	1.30 (1.04–1.64)	0.023	1.36 (1.03–1.81)	0.033	1.51 (0.95–2.40)	0.079	0.815
rs4849887	2q14/ <i>INHBB</i>	1.23 (1.00–1.50)	0.045	1.25 (0.96–1.61)	0.093	1.46 (0.95–2.24)	0.087	0.934
rs6762644	3p26/ <i>ITPR1</i>	1.18 (0.99–1.41)	0.067	1.09 (0.87–1.36)	0.440	0.87 (0.60–1.25)	0.436	0.593
rs9790517	4q24/ <i>TET2</i>	0.95 (0.64–1.40)	0.790	1.52 (0.98–2.35)	0.064	1.33 (0.65–2.70)	0.433	0.118
rs17529111	6q14/ <i>FAM46A</i>	0.97 (0.67–1.42)	0.893	1.32 (0.84–2.07)	0.229	1.97 (1.02–3.82)	0.043	0.312
rs9693444	8p21/ <i>RPL17P33</i>	0.96 (0.80–1.16)	0.675	1.21 (0.96–1.52)	0.099	1.32 (0.92–1.90)	0.131	0.122
rs11199914	10q26/ <i>FGFR2</i>	1.00 (0.83–1.19)	0.965	1.09 (0.87–1.37)	0.444	1.38 (0.96–1.98)	0.078	0.532
rs941764	14q32/ <i>CCDC88C</i>	1.26 (1.02–1.56)	0.032	0.92 (0.72–1.19)	0.536	0.89 (0.60–1.33)	0.570	0.063
rs17817449	16q12/ <i>FTO</i>	1.32 (1.09–1.60)	0.004	1.15 (0.91–1.46)	0.235	0.98 (0.67–1.44)	0.931	0.378
rs1436904	18q11/ <i>CHST9</i>	1.02 (0.83–1.27)	0.822	1.09 (0.83–1.44)	0.528	1.51 (0.92–2.47)	0.100	0.718
rs4808801	19p13/ <i>ELL</i>	0.95 (0.78–1.17)	0.644	1.24 (0.97–1.59)	0.083	1.41 (0.95–2.11)	0.091	0.103

^aOnly SNPs that showed significant ($P \leq 0.05$) or marginally significant ($P \leq 0.1$) association with any of these three subtypes of breast cancer are presented.

^bThe closest gene.

^cAdjusted with age, study (NBHS and SCCS), and the first ten principal components.

^d P for heterogeneity across ER positive and negative cases was calculated using a Cochran's Q test.

^eresults with $P < 0.05$ were bolded.

doi:10.1371/journal.pone.0058350.t003

in the Women's Insights and Shared Experiences study [36]. SNP rs13387042 was replicated in our previous study [25] and in a consortium study of 3,016 cases and 2,745 controls [33], but not in the WHI study [35], the CBCS [32], nor another pooled study [34]. SNP rs8170 at 19p13/*BABAM1* was only investigated in a pooled study from Africans and AAs and no association was observed [34]. Both *CASP8* SNP (rs1045485) and *RAD51L* SNP (rs999737) have a MAF $< 5\%$ in African-ancestry populations. Therefore, it is not surprising that these two SNPs were not replicated in all previous studies of African-ancestry populations [32–34]. The other four SNPs that showed associations with overall breast cancer in the present study, rs4849887, rs17817449,

rs13329835 and rs4808801, were recently discovered [22] and have not been evaluated in previous studies of African-ancestry populations.

We did not replicate associations for the other 60 reported SNPs with the risk of overall breast cancer. Inconsistent results were reported for some of them in previous studies of African-ancestry populations. For example, SNP rs3803662 at 16q12/*TOX3* was replicated in the WHI [35], but not in the Black Women's Health Study (BWHS) [37], the CBCS [32] and others [25,34]. In addition, a significant association has been identified for this SNP with association in the opposite direction as previously reported [9]. SNP rs2981582 (10q26/*FGFR2*) was significantly associated

Table 4. Association of genetic risk score (GRS) with breast cancer risk in African Americans.

GRS quintile ^a	Controls	All cases		ER+ cases		ER- cases		ER-/PR-/HER2- cases	
	N = 929	N = 1,110	OR (95% CI) ^b	N = 337	OR (95% CI) ^b	N = 188	OR (95% CI) ^b	N = 67	OR (95% CI) ^b
Q1	185	133	1.00 (reference)	38	1.00 (reference)	26	1.00 (reference)	10	1.00 (reference)
Q2	185	221	1.75 (1.30–2.37)	66	1.91 (1.20–3.02)	27	1.16 (0.64–2.09)	8	0.95 (0.36–2.54)
Q3	194	202	1.56 (1.15–2.11)	68	1.90 (1.20–2.99)	32	1.30 (0.74–2.30)	11	1.23 (0.50–3.06)
Q4	176	236	2.02 (1.50–2.74)	73	2.39 (1.51–3.77)	41	2.06 (1.19–2.57)	13	2.06 (0.84–5.05)
Q5	189	318	2.63 (1.96–3.52)	92	2.85 (1.83–4.45)	62	2.82 (1.68–4.75)	25	3.35 (1.48–7.54)
<i>P</i> _{for trend}			7.8×10^{-10}		5.2×10^{-6}		4.3×10^{-6}		4.5×10^{-4}

^aGRS was constructed based on 8 SNPs, including rs1045485, rs10069690, rs999737, rs8170, rs4849887, rs17817449, rs13329835 and rs4808801.

^bAdjusted for age, study, and the first ten principal components.

doi:10.1371/journal.pone.0058350.t004

with breast cancer risk in two studies [32,33], but not in the other studies [25,35,36,38]. The WHI study reported a significant association for rs10941679 at 5p12/*MRPS30* [35], and one study showed association for rs865686 at 9q31/*KLF4* [33], however, other studies did not replicate these two SNPs [25,34,39].

In general, our results that approximately 90% of index SNPs were not replicated in AAs are consistent with results from previous studies in African-ancestry populations [25,32–39]. Rebbeck et al [36] did not find any association for three investigated SNPs. Huo et al [34] evaluated 19 genetic loci and none of them were replicated. Five of the seven investigated SNPs in the CBCS [32], 21 of the 22 investigated SNPs in the WHI [35], and 14 of 19 SNPs in a consortium study [33] were not replicated.

Because of the large difference in genetic architecture between African-ancestry and European/Asian ancestry populations, failure to replicate most of the reported SNPs in AAs is not surprising. Most, if not all, index SNPs identified in GWAS are associated with breast cancer risk through their strong LD with causal variants. African-ancestry populations have shorter LD and more genetic variations than European/Asian ancestry populations and may have different SNPs in LD with the causal variant. This may be the major reason why index SNPs are not replicated in African descendants. For example, in the BWHs, originally reported index SNPs rs10941679 and rs3803662 were not replicated, but other SNPs in these regions, rs16901937 and rs3104746, were associated with breast cancer [37,39]. It has been reported that other markers were identified in AAs to better capture the association signal than the index SNPs originally discovered in the 2q35/*TNPI*, 5q11/*MAP3K1*, 10q26/*FGFR2*, and 19p13/*BABAM1* loci [33]. Second, allele frequencies for the index SNPs differ considerably across ethnic groups. Many index SNPs have lower MAF in AAs than in Europeans/Asians. Even if the effect size of the index SNP is the same across populations, larger sample size is required to detect association in AAs due to the lower MAF. Third, the vast majority of SNPs were originally discovered among European descendants, who have a much higher proportion of ER+ than ER- breast cancer. Because of this, most of the reported risk variants are, in general, more strongly associated with ER+ than ER- cancer [22]. African-ancestry women have a higher proportion of ER- breast cancer than European-ancestry women; this may be another reason for the non-replication in AAs.

To our knowledge, this is the first study in AAs that has evaluated index SNPs in all breast cancer susceptibility loci identified to date. However, the sample size in our study is

relatively small, especially when stratified by breast cancer subtype. Some of the null associations observed in this study could be due to inadequate statistical power. Meta-analysis by pooling together all existing data in the AA populations will increase the statistical power to evaluate the effects of these variants in AAs. The other limitation of this study is that we only investigated index SNP in each locus. Large-scale fine mapping studies are needed to identify genetic risk variants at these loci in African-ancestry populations. Such work will be very helpful to identify causal variants for breast cancer.

In summary, in this African-ancestry population study, we replicated approximately 10% of index SNPs in 67 breast-cancer susceptibility loci. Heterogeneity was observed across the breast cancer subtype. These results show the complexity in applying GWAS findings to African-ancestry populations. Large-scale studies in AAs are needed to discover genetic risk variants which impact this population.

Supporting Information

Table S1 African genetic ancestry proportion of study participants estimated by AIMs.
(DOCX)

Table S2 Association of breast cancer risk with 60 SNPs located in reported breast-cancer susceptibility loci in African Americans with P>0.1.
(DOCX)

Acknowledgments

The authors wish to thank all individuals who took part in these studies and all researchers, clinicians, technicians, and administrative staff who have enabled this work to be carried out. We thank Regina Courtney, Jie Wu, Mark Steinwandel (from International Epidemiology Institute), Jing He, and Mary Jo Daly for their help with sample preparation and genotyping, statistical analysis, and technical support for the project. The authors would like to acknowledge Dr. Doug Easton and his colleagues for designing the COGS SNP array that was used in the study. Sample preparation was conducted at the Survey and Biospecimen Shared Resources.

Author Contributions

Conceived and designed the experiments: JL WZ. Performed the experiments: QC. Analyzed the data: JL. Contributed reagents/materials/analysis tools: JL LDS QC SDH MJS MS XOS WJB. Wrote the paper: JL.

References

- Fejerman L, Haiman CA, Reich D, Tandon A, Deo RC, et al. (2009) An admixture scan in 1,484 African American women with breast cancer. *Cancer Epidemiol Biomarkers Prev* 18: 3110–3117. 1055–9965.
- Reding KW, Carlson CS, Khsai O, Chen CC, McDavid A, et al. (2012) Examination of ancestral informative markers and self-reported race with tumor characteristics of breast cancer among black and white women. *Breast Cancer Res Treat* 134: 801–809.
- Zhang B, Beeghly-Fadiel A, Long J, Zheng W (2011) Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol* 12: 477–488.
- Nathanson KL, Wooster R, Weber BL (2001) Breast cancer genetics: what we know and what we need. *Nat Med* 7: 552–556.
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447: 1087–1093.
- Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, et al. (2007) A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39: 870–874.
- Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, et al. (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 39: 865–869.
- Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, et al. (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A* 105: 4340–4345.
- Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, et al. (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 40: 703–706.
- Zheng W, Long J, Gao YT, Li C, Zheng Y, et al. (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 41: 324–328.
- Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, et al. (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* 41: 579–584. ng.353 [pii];10.1038/ng.353 [doi].
- Ahmed S, Thomas G, Ghousaini M, Healey CS, Humphreys MK, et al. (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 41: 585–590. ng.354 [pii];10.1038/ng.354 [doi].
- Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, et al. (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 42: 504–507. ng.586 [pii];10.1038/ng.586 [doi].
- Long J, Cai Q, Shu XO, Qu S, Li C, et al. (2010) Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet* 6: e1001002.

15. Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, et al. (2010) A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* 42: 19.
16. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, et al. (2011) Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst* 103: 425–435.
17. Cai Q, Long J, Lu W, Qu S, Wen W, et al. (2011) Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. *Hum Mol Genet* 20: 4991–4999. ddr405 [pii];10.1093/hmg/ddr405 [doi].
18. Long J, Cai Q, Sung H, Shi J, Zhang B, et al. (2012) Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet* 8: e1002532. 10.1371/journal.pgen.1002532 [doi];PGENETICS-D-11-01537 [pii].
19. Ghossaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, et al. (2012) Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet*. ng.1049 [pii];10.1038/ng.1049 [doi].
20. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, et al. (2011) A common variant at the TERT-CLPTMIL locus is associated with estrogen receptor-negative breast cancer. *Nat Genet* 43: 1210–1214. ng.985 [pii];10.1038/ng.985 [doi].
21. Lindstrom S, Vachon CM, Li J, Varghese J, Thompson D, et al. (2011) Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. *Nat Genet* 43: 185–187. ng.760 [pii];10.1038/ng.760 [doi].
22. Michailidou K, Hall K, Gonzalez-Niera A, Ghossaini M, Dennis J, et al. (2012) Large-scale genotyping identifies 42 new breast cancer susceptibility loci. *Nat Genet*, in press.
23. McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 17: R156–R165.
24. Signorello LB, Hargreaves MK, Steinwandl MD, Zheng W, Cai Q, et al. (2005) Southern community cohort study: establishing a cohort to investigate health disparities. *J Natl Med Assoc* 97: 972–979.
25. Zheng W, Cai Q, Signorello LB, Long J, Hargreaves MK, et al. (2009) Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev* 18: 2761–2764.
26. Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 28: 289–301. 10.1002/gepi.20064 [doi].
27. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909.
28. Long J, Cai QY, Sung H (2012) Genome-wide association study in East Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet* in press.
29. Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, et al. (2001) AID-AJPA1002>3.0.CO;2-2 [pii];10.1002/1096-8644(200101)114:1<18::AID-AJ-PA1002>3.0.CO;2-2 [doi].
30. Zakharia F, Basu A, Absher D, Assimes TL, Go AS, et al. (2009) Characterizing the admixed African ancestry of African Americans. *Genome Biol* 10: R141. gb-2009-10-12-r141 [pii];10.1186/gb-2009-10-12-r141 [doi].
31. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, et al. (2007) A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 39: 352–358. ng1981 [pii];10.1038/ng1981 [doi].
32. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, et al. (2010) FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. *Carcinogenesis* 31: 1417–1423. bgq128 [pii];10.1093/carcin/bgq128 [doi].
33. Chen F, Chen GK, Millikan RC, John EM, Ambrosio CB, et al. (2011) Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum Mol Genet* 20: 4491–4503. ddr367 [pii];10.1093/hmg/ddr367 [doi].
34. Huo D, Zheng Y, Ogundiran TO, Adebamowo C, Nathanson KL, et al. (2012) Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis* 33: 835–840. bgs093 [pii];10.1093/carcin/bgs093 [doi].
35. Hutter CM, Young AM, Ochs-Balcom HM, Carty CL, Wang T, et al. (2011) Replication of breast cancer GWAS susceptibility loci in the Women's Health Initiative African American SHARc Study. *Cancer Epidemiol Biomarkers Prev* 20: 1950–1959. 1055–9965. EPI-11-0524 [pii];10.1158/1055-9965.EPI-11-0524 [doi].
36. Rebbeck TR, Demichele A, Tran TV, Panossian S, Bunin GR, et al. (2009) Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis* 30: 269–274. bgn247 [pii];10.1093/carcin/bgn247 [doi].
37. Ruiz-Narvaez EA, Rosenberg L, Cozier YC, Cupples LA, Adams-Campbell LL, et al. (2010) Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. *Cancer Epidemiol Biomarkers Prev* 19: 1320–1327. 1055–9965.EPI-09–1250 [pii];10.1158/1055-9965.EPI-09–1250 [doi].
38. Udler MS, Meyer KB, Pooley KA, Karlins E, Struwing JP, et al. (2009) FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum Mol Genet* 18: 1692–1703. ddp078 [pii];10.1093/hmg/ddp078 [doi].
39. Ruiz-Narvaez EA, Rosenberg L, Rotimi CN, Cupples LA, Boggs DA, et al. (2010) Genetic variants on chromosome 5p12 are associated with risk of breast cancer in African American women: the Black Women's Health Study. *Breast Cancer Res Treat* 123: 525–530. 10.1007/s10549-010-0775-5 [doi].
40. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, et al. (2009) The genetic structure and history of Africans and African Americans. *Science* 324: 1035–1044.