Human Molecular Genetics, 2016, Vol. 25, No. 21 4835–4846

doi: 10.1093/hmg/ddw305
Advance Access Publication Date: 4 September 2016
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Genome-wide association studies in women of African ancestry identified 3q26.21 as a novel susceptibility locus for oestrogen receptor negative breast cancer

Dezheng Huo[1], Ye Feng[2], Stephen Haddad[3], Yonglan Zheng[4], Song Yao[5], Yoo-Jeong Han[4], Temidayo O. Ogundiran[6], Clement Adebamowo[7], Oladosu Ojengbede[8], Adeyinka G. Falusi[9], Wei Zheng[10], William Blot[10], Qiuyin Cai[10], Lisa Signorello[11], Esther M. John[12,13], Leslie Bernstein[14], Jennifer J. Hu[15], Regina G. Ziegler[16], Sarah Nyante[17], Elisa V. Bandera[18], Sue A. Ingles[2], Michael F. Press[19], Sandra L. Deming[10], Jorge L. Rodriguez-Gil[15], Katherine L. Nathanson[20], Susan M. Domchek[20], Timothy R. Rebbeck[21], Edward A. Ruiz-Narváez[3], Lara E. Sucheston-Campbell[22], Jeannette T. Bensen[17], Michael S. Simon[23], Anselm Hennis[24,25], Barbara Nemesure[25], M. Cristina Leske[25], Stefan Ambs[26], Lin S. Chen[1], Frank Qian[4], Eric R. Gamazon[27,28], Kathryn L. Lunetta[29], Nancy J. Cox[27], Stephen J. Chanock[16], Laurence N. Kolonel[30], Andrew F. Olshan[17], Christine B. Ambrosone[5], Olufunmilayo I. Olopade[4,*], Julie R. Palmer[3,*] and Christopher A. Haiman[2,*]

[1]Department of Public Health Sciences, University of Chicago, Chicago, IL, USA, [2]Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA, [3]Slone Epidemiology Center, Boston University, Boston, MA, USA, [4]Center for Clinical Cancer Genetics and Global Health, Department of Medicine, University of Chicago, Chicago, IL, USA, [5]Roswell Park Cancer Institute, Buffalo, NY, USA, [6]Department of Surgery, College of Medicine, University of Ibadan, Ibadan, Nigeria, [7]Department of Epidemiology & Preventive Medicine, University of Maryland, Baltimore, MD, USA, [8]Center for Population and Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Nigeria, [9]Institute for Medical Research and Training, College of Medicine, University of Ibadan, Ibadan, Nigeria, [10]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN, USA, [11]Cancer Prevention Fellowship Program, National Cancer Institute, Bethesda, MD, USA, [12]Cancer Prevention Institute of California, Fremont, CA, USA, [13]Department of Health Research & Policy and Stanford Cancer Institute, Stanford University School of Medicine Stanford, CA, USA, [14]Division of Cancer Etiology, Department of Population Sciences, Beckman

Received: April 1, 2016. **Revised:** July 21, 2016. **Accepted:** August 26, 2016

Research Institute of City of Hope, Duarte, CA, USA, [15]Sylvester Comprehensive Cancer Center and Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, USA, [16]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MA, USA, [17]Department of Epidemiology, Gillings School of Global Public Health, and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA, [18]The Cancer Institute of New Jersey, New Brunswick, NJ, USA, [19]Department of Pathology, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA, [20]Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA, [21]Dana Farber Cancer Institute & Harvard T.H. Chan School of Public Health, Boston, MA, USA, [22]Colleges of Pharmacy and Veterinary Medicine, Ohio State University, Columbus, OH, USA, [23]Karmanos Cancer Institute, Department of Oncology, Wayne State University, Detroit, MI, USA, [24]Chronic Disease Research Centre, Tropical Medicine Research Institute, University of the West Indies, Bridgetown, Barbados, [25]Department of Preventive Medicine, State University of New York at Stony Brook, Stony Brook, NY, USA, [26]Laboratory of Human Carcinogenesis, National Cancer Institute, Bethesda, MD, USA, [27]Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA, [28]Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, [29]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA and [30]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA

*To whom correspondence should be addressed at: Olufunmilayo I. Olopade, MBBS, Center for Clinical Genetics & Innovation in Global Health, University of Chicago, Chicago, IL, USA. Email: folopade@medicine.bsd.uchicago.edu, Julie R. Palmer, ScD, Slone Epidemiology Center, Boston University, Boston MA, USA. Email: jpalmer@bu.edu, Christopher A. Haiman, PhD, Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA. Email: Christopher.Haiman@med.usc.edu

Abstract

Multiple breast cancer loci have been identified in previous genome-wide association studies, but they were mainly conducted in populations of European ancestry. Women of African ancestry are more likely to have young-onset and oestrogen receptor (ER) negative breast cancer for reasons that are unknown and understudied. To identify genetic risk factors for breast cancer in women of African descent, we conducted a meta-analysis of two genome-wide association studies of breast cancer; one study consists of 1,657 cases and 2,029 controls genotyped with Illumina's HumanOmni2.5 BeadChip and the other study included 3,016 cases and 2,745 controls genotyped using Illumina Human1M-Duo BeadChip. The top 18,376 single nucleotide polymorphisms (SNP) from the meta-analysis were replicated in the third study that consists of 1,984 African Americans cases and 2,939 controls. We found that SNP rs13074711, 26.5 Kb upstream of *TNFSF10* at 3q26.21, was significantly associated with risk of oestrogen receptor (ER)-negative breast cancer (odds ratio [OR]=1.29, 95% CI: 1.18-1.40; $P = 1.8 \times 10^{-8}$). Functional annotations suggest that the *TNFSF10* gene may be involved in breast cancer aetiology, but further functional experiments are needed. In addition, we confirmed SNP rs10069690 was the best indicator for ER-negative breast cancer at 5p15.33 (OR = 1.30; $P = 2.4 \times 10^{-10}$) and identified rs12998806 as the best indicator for ER-positive breast cancer at 2q35 (OR = 1.34; $P = 2.2 \times 10^{-8}$) for women of African ancestry. These findings demonstrated additional susceptibility alleles for breast cancer can be revealed in diverse populations and have important public health implications in building race/ethnicity-specific risk prediction model for breast cancer.

Introduction

As the most common cancer among women both in the United States and in the world (1), breast cancer is a heterogeneous disease with several molecular subtypes defined by biomarkers such as oestrogen receptor (ER). Oestrogen receptor-negative breast cancer is more likely to be early-onset and is associated with worse clinical outcomes. African Americans are more likely to have ER-negative breast cancer, for reasons that remain unknown and understudied (2). Indigenous Africans have an even higher proportion of ER-negative breast cancer than African Americans (3,4), suggesting that previously unrecognized genetic factors play an important role in this racial difference. Multiple genome-wide association studies (GWAS) of breast cancer have been conducted but mainly in populations of European ancestry, which have revealed more than 90 common susceptibility loci (5–21). However, only a few of these loci appear to be specific for ER-negative disease (19,21). Because of its aggressive nature, there is an unmet clinical need for predictive genetic markers of susceptibility to ER-negative breast cancer, especially among women of African ancestry.

Results

The study consists of a discovery stage and a validation stage. In the discovery stage, we conducted a meta-analysis of two GWAS (Table 1). One study (GWAS of Breast Cancer in the African Diaspora, ROOT consortium) consists of 1,657 cases and 2,029 controls, while the other study (African American Breast Cancer Consortium, AABC) consists of 3,016 cases and 2,745 controls. Genotyping was conducted using the Illumina HumanOmni2.5-8v1 and Illumina Human1M-Duo BeadChip in ROOT and AABC studies, respectively. To combine genotyping

Table 1. Characteristics of study participants in the GWAS of breast cancer for women of African ancestry

Study	Study Name	No. of subjects		Mean (SD) of age in years		Oestrogen receptor (n, %)	
		Case	Control	Case	Control	Positive	Negative
ROOT consortium							
NBCS	Nigerian Breast Cancer Study	711	624	48 (12)	45 (12)	99 (70)	42 (30)
BNCS	Barbados National Cancer Study	92	229	57 (15)	55 (13)		
RVGBC	Racial Variability in Genotypic Determinants of Breast Cancer Risk Study	145	257	46 (11)	41 (12)	25 (48)	27 (52)
CCPS	Chicago Cancer Prone Study	394	387	47 (12)	45 (11)	140 (45)	171 (55)
BBCS	Baltimore Breast Cancer Study	95	102	54 (14)	52 (13)	44 (49)	45 (51)
SCCS	Southern Community Cohort Study	220	430	57 (9)	57 (9)	66 (36)	118 (64)
subtotal		**1657**	**2029**	**49 (12)**	**48 (13)**	**374 (48)**	**403 (52)**
AABC consortium							
MEC	Multiethnic Cohort Study	694	990	66 (9)	67 (9)	408 (70)	176 (30)
CARE	Women's Contraceptive and Reproductive Experiences Study, Log Angeles component	357	215	49 (8)	48 (8)	183 (58)	130 (42)
WCHS	Women's Circle of Health Study	261	239	50 (10)	50 (9)	131 (62)	80 (38)
SFBCS	San Francisco Bay Area Breast Cancer Study	165	220	56 (12)	55 (12)	84 (63)	50 (37)
NC-BCFR	Northern California site of the Breast Cancer Family Registry	424	50	50 (9)	49 (9)	219 (64)	121 (36)
CBCS	Carolina Breast Cancer Study	635	589	51 (12)	52 (11)	272 (46)	317 (54)
PLCO	Prostate, Lungs, Colorectal and Ovarian Cancer Screening Trial Cohort	56	116	68 (7)	68 (6)	14 (70)	6 (30)
NBHS	Nashville Breast Health Study	304	182	54 (11)	52 (10)	143 (69)	65 (31)
WFBC	Wake Forest University Breast Cancer Study	120	144	55 (12)	55 (11)	66 (61)	43 (39)
subtotal		**3016**	**2745**	**55 (12)**	**58 (13)**	**1520 (61)**	**988 (39)**
AMBER consortium							
BWHS	Black Women's Health Study	752	2249	54 (10)	53 (11)	412(65)	218(35)
CBCS	Carolina Breast Cancer Study	727	61	51 (11)	53 (9)	444(64)	252(36)
WCHS	Women's Circle of Health Study	505	629	54 (11)	51 (10)	237(70)	103(30)
subtotal		**1984**	**2939**	**53 (11)**	**53 (10)**	**1093 (66)**	**573 (34)**
Total		**6657**	**7713**	**53 (12)**	**53 (13)**	**2987 (60)**	**1964 (40)**

results from the two studies, we imputed genotypes for single nucleotide polymorphisms in the 1000 Genomes Project using IMPUTE2 (22). Quantile-quantile plots of P-values showed little inflation of test statistics in ROOT, AABC, or in the meta-analysis (all genomic control inflation factor λ's < 1.04; Supplementary Material, Fig. S1), indicating that there was no residual cryptic population substructure in the analysis. Results of the meta-analysis are summarized in the Manhattan plots (Fig. 1). The top 20,000 SNPs from the meta-analysis of overall breast cancer or ER-specific disease were considered for stage 2 replication genotyping in an independent sample of 1,984 invasive cases and 2,939 controls assembled by the African American Breast Cancer Epidemiology and Risk (AMBER) consortium (Table 1). Of the ~20,000 SNPs, 18,376 SNPs passed the design stage for the Illumina custom chip and were successfully genotyped.

In the combined analysis of GWAS discovery and validation stages, associations of SNPs in three regions attained genome-wide significance (Table 2), with one of these regions (3q26.21) not identified in previous GWAS. All three SNPs were significantly associated with overall risk of breast cancer, but the strength of the association varied by breast cancer subtype. SNP rs12998806 at 2q35 was associated with ER-positive breast cancer (G-allele, OR = 1.34; $P = 2.2 \times 10^{-8}$), but not with ER-negative disease (OR = 0.99). SNP rs13074711 at 3q26.21 was significantly associated with ER-negative breast cancer (T-allele, OR = 1.29; $P = 1.8 \times 10^{-8}$), but only weakly with ER-positive breast cancer (OR = 1.10, $P = 0.0094$). Similarly, rs10069690 at 5p15.33 was associated with ER-negative breast cancer (T-allele, OR = 1.30; $P = 2.4 \times 10^{-10}$), but only weakly associated with ER-positive disease (OR = 1.08, $P = 0.03$). SNP rs12998806 was imputed in ROOT and AABC (imputation information score $r^2 > 0.99$) and rs13074711 was imputed in ROOT ($r^2 = 0.987$). We genotyped these two SNPs in 96 samples from ROOT using sequencing technique, and found that the genotyping results were 100% and 99% in concordance with imputation for rs12998806 and rs13074711, respectively, suggesting that the results based on imputed SNPs were reliable.

The rs13074711 SNP is located 26.5 Kb upstream of the Tumour Necrosis Factor Superfamily, Member 10 (*TNFSF10*) gene and the association signal for ER-negative breast cancer maps to a linkage disequilibrium (LD) block of ~100 Kb between 172.2 Mb and 172.3 Mb on chromosome 3, which covers the *TNFSF10* gene (Fig. 2A, Supplementary Material, Fig S2). The *TNFSF10* gene, also known as *TRAIL* or *APO2L*, encodes a cytokine that belongs to the tumour necrosis factor ligand family. TNFSF10 cytokine preferentially induces apoptosis in tumour cells or transformed cells, but is not toxic to normal cells (23). The *TNFSF10* gene is expressed in a wide range of tissues, including blood lymphocytes, breast, prostate, lungs, minor salivary grand and spleen (24). Interestingly, triple-negative breast cancer cell lines have been shown to be very sensitive to TNFSF10-induced apoptosis, while ER-positive cell lines were resistant to TNFSF10-induced apoptosis (25). Because of its important role in tumour apoptosis, several TNFSF10 receptor agonists or recombinant forms of TNFSF10 are in clinical

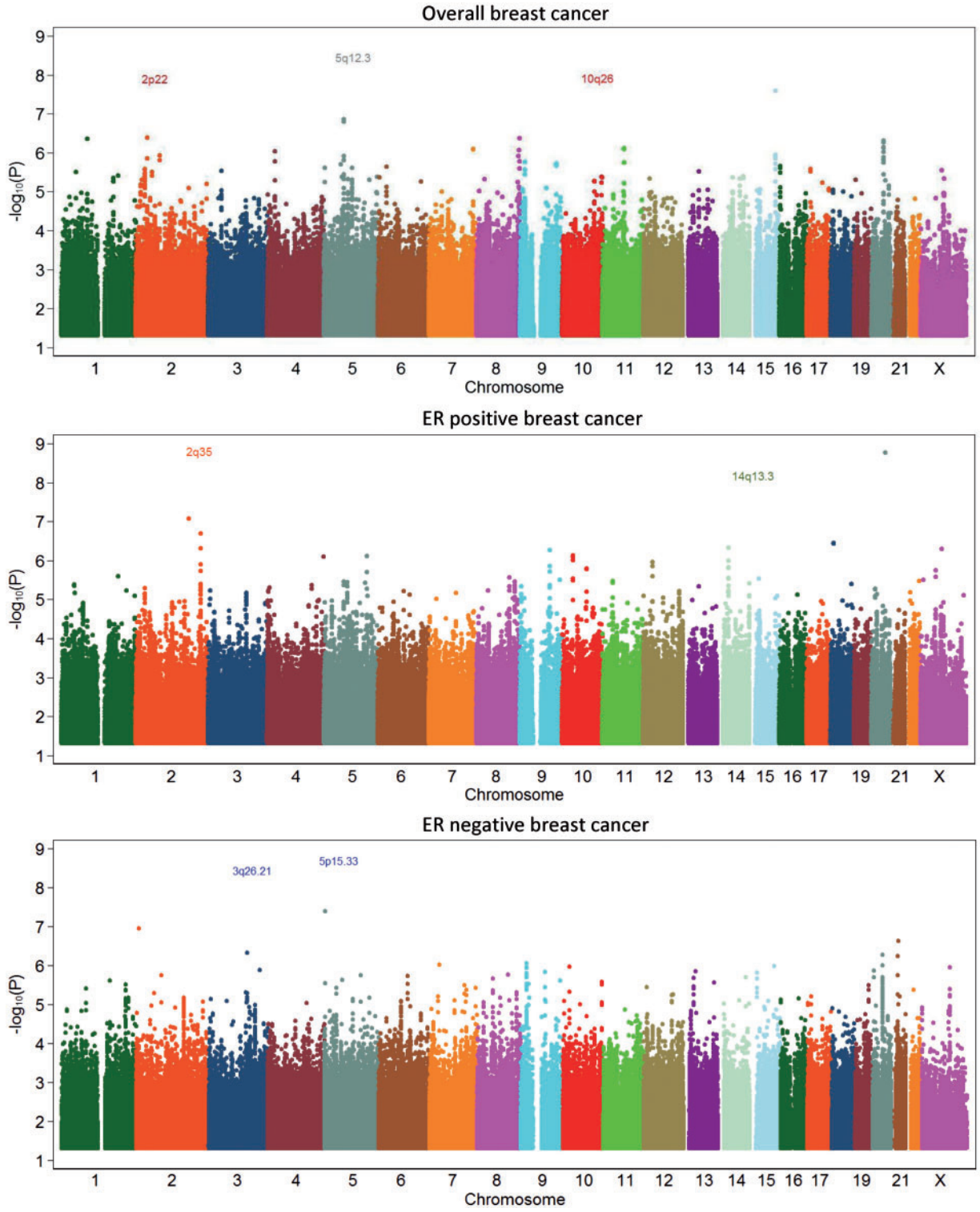

Figure 1. Manhattan plots of –log10 of P-values in the meta-analysis of ROOT and AABC for overall, ER+ and ER- breast cancer risk.

development as molecularly targeted therapy for cancers (26). Thus, *TNFSF10* gene is a plausible causal gene in the region.

There appeared to be a second association signal in the 3q26.21/*TNFSF10* region: rs9833271 was associated with ER-negative breast cancer (A-allele, OR = 1.39, $P = 3.7 \times 10^{-5}$) and it was in weak LD with the top signal SNP, rs13074711 ($r^2 = 0.06$). After adjusting rs13074711, the association between rs9833271 and ER-negative breast cancer remained statistically significant (p = 0.002) though the adjusted odds ratio was smaller (Table 3). Interestingly, the risk allele of rs9833271 exists only in African populations. Furthermore, a missense variant in exon 1 of *TNFSF10* (rs6763816, c.97G > A, Val33Ile) has a frequency of 0.04 in African populations, but is monomorphic in other populations; the minor allele of this missense variant was associated

Table 2. Association between three loci and breast cancer risk in African ancestry, by type of breast cancer

Cytoband, gene	SNP	Position*	Alleles (reference/test)	Study	Imputation score	RAF	Overall breast cancer		ER-positive breast cancer		ER-negative breast cancer	
							OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
2q35	rs12998806	217893763	A/G	ROOT	0.995	0.87	1.09 (0.95-1.26)	0.22	1.21 (0.95-1.54)	0.12	0.88 (0.70-1.12)	0.30
				AABC	0.992	0.86	1.20 (1.08-1.34)	0.0011	1.45 (1.26-1.67)	3.0E-07	1.00 (0.86-1.16)	0.99
				AMBER	Typed	0.86	1.14 (0.99-1.31)	0.076	1.24 (1.03-1.49)	0.022	1.07 (0.85-1.34)	0.56
				Combined		0.86	1.15 (1.07-1.24)	0.00017	**1.34 (1.21-1.48)**	**2.2E-08**	0.99 (0.88-1.10)	0.84
				P for heterogeneity		0.57			0.27		0.51	
3q26.21 *TNFSF10*	rs13074711	172267803	C/T	ROOT	0.987	0.67	1.03 (0.93-1.14)	0.53	0.86 (0.73-1.02)	0.09	1.11 (0.93-1.33)	0.24
				AABC	Typed	0.70	1.23 (1.13-1.34)	1.9E-06	1.18 (1.06-1.31)	0.0017	1.37 (1.21-1.55)	4.1E-07
				AMBER	Typed	0.69	1.15 (1.03-1.28)	0.015	1.15 (1.00-1.32)	0.049	1.30 (1.09-1.56)	0.0038
				Combined		0.69	1.14 (1.08-1.21)	2.1E-06	1.10 (1.02-1.19)	0.0094	**1.29 (1.18-1.40)**	**1.8E-08**
				P for heterogeneity		0.04			0.007		0.16	
5p15.33 *TERT*	rs10069690	1279790	C/T	ROOT	Typed	0.63	1.15 (1.04-1.26)	0.0049	1.11 (0.94-1.31)	0.21	1.27 (1.08-1.50)	0.0048
				AABC	Typed	0.59	1.13 (1.04-1.22)	0.0022	1.08 (0.98-1.19)	0.11	1.31 (1.17-1.47)	2.3E-06
				AMBER	Typed	0.60	1.11 (1.00-1.23)	0.049	1.06 (0.93-1.20)	0.37	1.31 (1.11-1.54)	0.0016
				Combined		0.60	1.13 (1.07-1.19)	4.8E-06	1.08 (1.01-1.16)	0.03	**1.30 (1.20-1.41)**	**2.4E-10**
				P for heterogeneity		0.88			0.91		0.94	

*NCBI Build 37 position.

Abbreviation: ER, oestrogen receptor; OR, allelic odds ratio; CI, confidence interval; RAF, risk (test) allele frequency.

with reduced risk of ER-negative breast cancer (OR = 0.75; $P = 0.0012$). Other SNPs that were in moderate LD with rs13074711 were no longer statistically significant after adjusting for rs13074711 (Table 3).

The expression quantitative trait locus (eQTL) analysis showed that several significant SNPs in the 3q26.21/*TNFSF10* region were associated with the expression of *TNFSF10* in peripheral blood (27) or brain (28), but not in breast tumour (Table 4, Supplementary Material, Fig. S3). We also examined the possibility that the breast cancer-associated SNPs influence *TNFSF10* expression by serving as enhancer variants. The top associated SNP, rs13074711, and its LD-linked SNPs were located in regions of high histone H3 lysine 4 monomethylation (H3K4Me1) and histone H3 lysine 27 acetylation (H3K27Ac) features that are characteristic of enhancer elements (Fig. 2B). The second associated SNP, rs9833271, was also located in an enhancer element. Notably, strong enhancer features of these SNPs were found in human mammary epithelial cells (HMEC) and breast myoepithelial cells (MYO) (Table 4). The three SNPs close to the transcriptional start site of *TNFSF10* (rs6763816, rs16845798, and rs17601879) also resided in transcription factor (GATA3) binding sites in a ductal breast cancer cell line (T47D). The minor allele of rs13074711 could alter multiple motifs. Taken together, these lines of evidences suggest that the association between variants in the 3q26.21 region and breast cancer risk may be mediated by regulating *TNFSF10* gene expression through enhancer activities.

The association signal around rs12998806 was confined to a region of ~87 Kb between 217.855 Mb and 217.942 Mb on chromosome 2q35 (Supplementary Material, Fig S4). Variant rs13387042 in this region has been associated with ER-positive breast cancer in previous GWAS conducted mainly in populations of European ancestry (9,12). rs13387042 was associated with breast cancer risk in some studies among African Americans (29,30), but not in other studies among women of African ancestry (9,31,32) or Asians (33). In the current study, the A-allele of rs13387042 was significantly associated with ER-positive breast cancer (OR = 1.16; $P = 0.002$), but the association was weaker than with rs12998806. The two SNPs are in the same LD block, separated by 12 Kb and correlated in both YRI ($r^2 = 0.63$) and CEU ($r^2 = 0.31$) populations in the 1000 Genomes Project. No gene was found in the LD block and the nearest known genes include *TNP1* (169 Kb), *IGFBP5* (334 Kb) and *IGFBP2* (365 Kb). One study proposed that rs4442975 (G-allele) regulation of *IGFBP5* gene expression is the underlying mechanism for 2q35 region's effect on ER-positive breast cancer risk (33), but rs4442975 was not associated with breast cancer risk in Asians (33) and only weakly associated with breast cancer in our study (OR = 1.12, $P = 0.009$). For the 2q35 locus, the LD pattern in African populations is weaker than in European populations, which may help to pinpoint the causal variants (Supplementary Material, Fig. S5).

The rs10069690 SNP is located in intron 4 of the *TERT* gene and the association signal was confined to a small region on chromosome 5p15 (Supplementary Material, Fig. S6). The LD pattern surrounding rs10069690 was weak in both African and European populations, and few SNPs were correlated with rs10069690 at $r^2 > 0.5$ (Supplementary Material, Fig. S7). The rs10069690 SNP was previously identified in a GWAS meta-analysis of ER-negative breast cancer which included data from the AABC consortium (21). The strength of association between rs10069690 and ER-negative breast cancer in ROOT and AMBER is very similar to that observed in AABC, suggesting that the T-allele of rs10069690 is the best indicator of ER-negative breast cancer at 5p15/*TERT* for women of African ancestry.

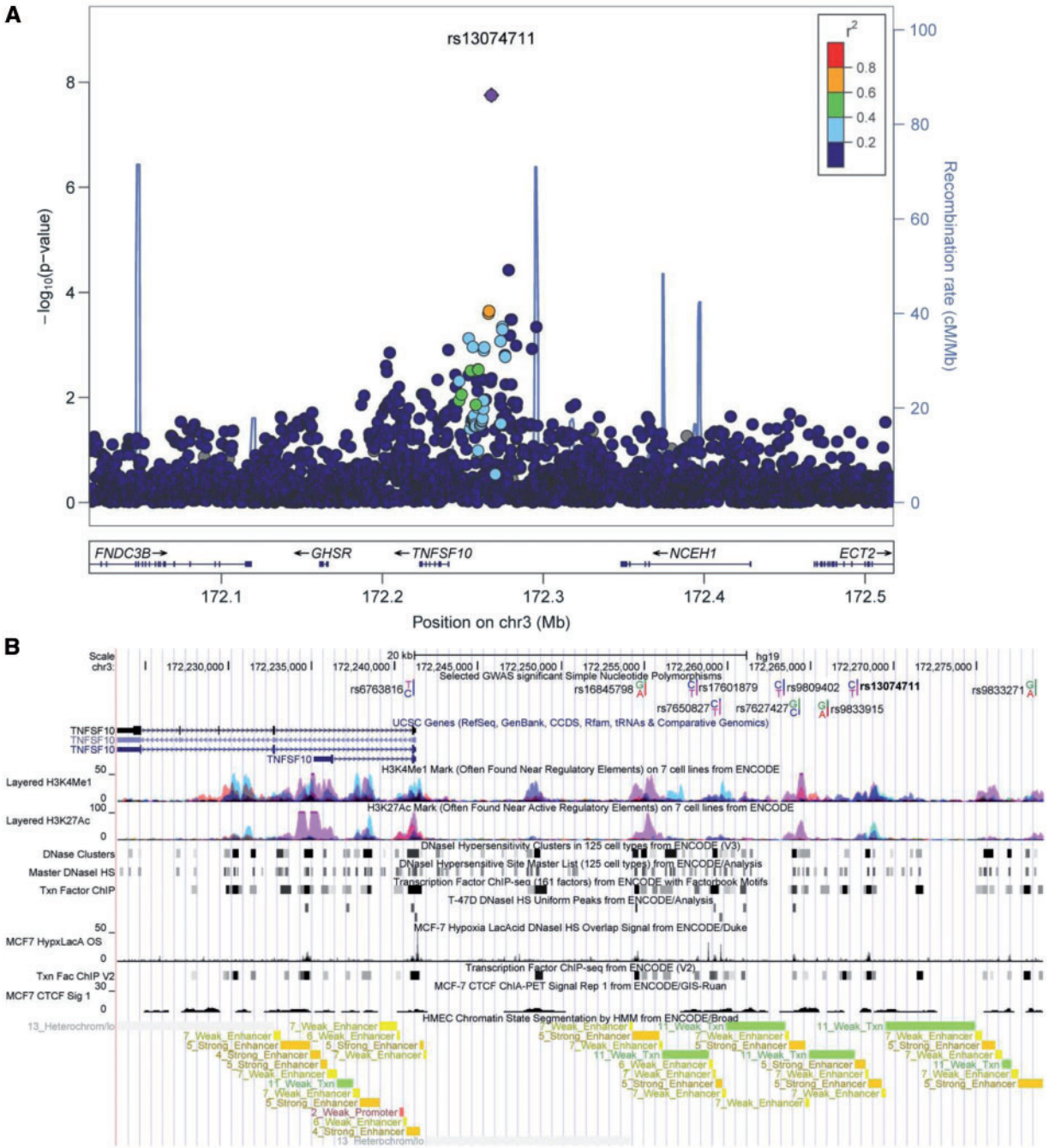

Figure 2. The 3q26.21/*TNFSF10* locus tagged by rs13074711. **(A)** The regional Manhattan plot shows the –log10 P values for testing SNPs in 250kb region with ER- breast cancer risk in the meta-analysis of the ROOT, AABC and AMBER studies. The colours depict the strength of the correlation (r^2) between SNP rs13074711 and the SNPs tested in the region. The correlation is estimated using the African panel of the 1000 Genomes Project data. The blue line indicates the recombination rates in centimorgans per megabase. **(B)** Analysis of regulation enhancer with data from ENCODE through UCSC Genome Browser, including histone modification marks for H3K4Me1 and H3K27Ac of seven cell types, transcription factor binding sites and DNase hypersensitivity sites of human mammary epithelial cells (HMEC), breast cancer cells (MCF7, T-47D). Chromosomal coordinates are in NCBI build 37.

Discussion

In the largest breast cancer GWAS conducted in women of African ancestry with more than 14,000 cases and controls, we have identified 3q26.31 locus as a novel risk region for ER-negative breast cancer. The second signal SNP and a missense variant in 3q26.31 locus only exists in African ancestry populations, which may explain why this locus had not been detected in previous GWAS of other populations. Functional annotation of the associated variants suggests that the nearby *TNFSF10* gene may be the causal gene underlying this association. We also found and confirmed breast cancer risk variants that are specific for African Americans at two known susceptibility regions. These findings showed that genome-wide scans of common and rare genetic variants carried out in diverse

Table 3. Selected variants in 3q26/*TNFSF10* region: their association with oestrogen-receptor negative breast cancer, and their linkage disequilibrium with rs13074711 (the top SNP in the region)

Variant	Position*	Allele†	r^2 in YRI	r^2 in CEU	Study	RAF	OR (95% CI)	P value	AOR (95% CI)‡	P value‡
rs6763816	Val33Ile Exon 1	C/T	0.08	NA	ROOT	0.06	0.92 (0.66-1.28)	0.62	0.98 (0.68-1.39)	0.90
					AABC	0.05	0.73 (0.56-0.94)	0.014	0.85 (0.65-1.11)	0.22
					AMBER	0.06	0.60 (0.41-0.87)	0.0079	0.65 (0.44-0.98)	0.038
					Combined	0.06	**0.75 (0.62-0.89)**	**0.0012**	**0.83 (0.69-1.01)**	**0.058**
rs16845798	13767	G/A	0.44	0.75	ROOT	0.73	1.03 (0.85-1.24)	0.79	0.88 (0.67-1.16)	0.37
					AABC	0.74	1.25 (1.10-1.42)	6.7E-04	0.95 (0.78-1.15)	0.57
					Combined	0.74	1.17 (1.06-1.30)	0.0031	0.92 (0.79-1.08)	0.33
rs17601879	16833	C/T	0.25	0.32	ROOT	0.81	1.00 (0.81-1.22)	0.98	0.87 (0.67-1.14)	0.32
					AABC	0.81	1.24 (1.08-1.43)	0.0025	0.97 (0.80-1.16)	0.71
					Combined	0.81	1.16 (1.03-1.30)	0.014	0.93 (0.80-1.09)	0.38
rs7650827	18258	C/T	0.36	0.32	ROOT	0.78	1.04 (0.85-1.27)	0.69	0.92 (0.70-1.22)	0.58
					AABC	0.80	1.27 (1.10-1.46)	7.9E-04	0.98 (0.81-1.19)	0.86
					Combined	0.79	1.19 (1.06-1.34)	0.0029	0.96 (0.82-1.13)	0.65
rs9809402	22074	C/T	0.25	0.09	ROOT	0.30	1.14 (0.96-1.35)	0.15	1.10 (0.92-1.33)	0.30
					AABC	0.32	1.19 (1.06-1.34)	0.0031	1.05 (0.92-1.20)	0.43
					Combined	0.31	1.17 (1.07-1.29)	0.0011	1.07 (0.96-1.19)	0.22
rs7627427	23005	G/C	0.25	0.09	ROOT	0.29	1.04 (0.87-1.25)	0.66	1.00 (0.83-1.21)	0.99
					AABC	0.29	1.19 (1.05-1.35)	0.0053	1.07 (0.94-1.22)	0.29
					Combined	0.29	1.14 (1.03-1.26)	0.011	1.05 (0.94-1.17)	0.38
rs9833915	24723	G/A	0.75	1.0	ROOT	0.60	1.06 (0.90-1.26)	0.47	0.93 (0.69-1.26)	0.64
					AABC	0.61	1.26 (1.12-1.42)	8.1E-05	0.95 (0.77-1.15)	0.58
					Combined	0.61	1.20 (1.09-1.32)	2.5E-04	0.94 (0.80-1.11)	0.47
rs13074711	26505	C/T	1.0	1.0	ROOT	0.67	1.11 (0.93-1.33)	0.24	NA	NA
					AABC	0.69	1.37 (1.21-1.55)	4.1E-07		
					AMBER	0.69	1.30 (1.09-1.56)	0.0038		
					Combined	0.69	**1.29 (1.18-1.40)**	**1.8E-08**		
rs9833271	37244	G/A	0.06	NA	ROOT	0.07	1.63 (1.19-2.24)	0.0024	1.59 (1.15-2.18)	0.0046
					AABC	0.06	1.32 (1.06-1.64)	0.012	1.18 (0.94-1.47)	0.15
					AMBER	0.06	1.32 (0.95-1.82)	0.095	1.23 (0.89-1.71)	0.21
					Combined	0.06	**1.39 (1.19-1.63)**	**3.7E-05**	**1.28 (1.09-1.50)**	**0.002**

*base pair upstream of the transcription starting site of *TNFSF10*.

†reference allele/test allele.

‡conditional analysis adjusted for SNP rs13074711.

Abbreviation: NA, not applicable; OR, allelic odds ratio; CI, confidence interval; RAF, risk (test) allele frequency; YRI, Yoruba in Ibadan, Nigeria; CEU, Utah residents with Northern and Western European ancestry.

populations can reveal additional susceptibility alleles for breast cancer, and population-specific risk variants are important for building appropriate risk prediction models.

Interestingly, the associations with the three breast cancer risk susceptibility loci depend on ER status, which reinforces the need for even larger discovery efforts for etiologically distinct breast cancer subtypes. The risk allele frequency for rs10069690 in 5p15/TERT locus is greater in African populations (0.62) than in European ancestry populations (0.27), and the risk allele of rs9833271 exists only in African populations. These may contribute to the higher proportion of ER-negative breast cancer in women of African ancestry. By contrast, the risk allele frequency for rs13074711 is smaller in African populations (0.61) than in European ancestry populations (0.89), and the risk allele frequency for rs12998806 is greater in African populations (0.78) than in women of European ancestry (0.66), suggesting that these two loci may not account for racial differences in subtype distribution but more definitive studies are warranted.

The aetiology of ER-negative breast cancer is largely unknown, and our study findings that the association for all three top loci varies by ER status provide important insight into distinct biological pathways for ER-positive and ER-negative breast cancers, which will benefit both breast cancer treatment and prevention. In particular, theses data can be used to improve risk prediction model for aggressive forms of breast cancer in diverse populations. We provide several lines of evidence that the association at 3q26.31 region may be mediated by regulation of *TNFSF10* gene expression, and *TNFSF10* is important for tumour apoptosis. Further functional experiments will be needed to confirm whether the *TNFSF10* gene is a susceptibility gene for ER-negative breast cancer and to fully understand how genetic variants affect risk of breast cancer in diverse populations and why the associations vary by breast cancer subtype.

Materials and Methods

Study populations

All the studies included in the GWAS have been approved by their corresponding Institutional Review Boards. Informed consent has been obtained from the participants. The discovery stage (stage 1) of the study is a meta-analysis of genome-wide association studies in two breast cancer consortia among women of African ancestry. The ROOT (GWAS in Breast Cancer in the African Diaspora) consortium consists of six epidemiologic studies and the AABC (African American Breast Cancer)

Table 4. Correlation of selected variants with expression of *TNFSF10* (from eQTL analysis (27,28)) and regulatory element annotation of selected variants in 3q26/*TNFSF10* region using HaploReg (37)

Variant	Position*	Alleles	Minor allele frequency				eQTL	Promoter histone marks†	Enhancer histone marks	DNAse hypersensitivity	Proteins bound	Motifs changed by the variant
		Ref/Alt	AFR	AMR	ASN	EUR						
rs6763816	Exon 1, Val33Ile	C/T	0.04	0	0	0	Brain	24 tissues (HMEC.35, MYO, HMEC)	6 tissues (HMEC.35, MYO)	26 tissues (HMEC.35, HEMC)	7 bound proteins (ERALPHA_A and GATA3 in T-47D)	Irf, Pax-5, XBP-1
rs16845798	13767	A/G	0.34	0.14	0.12	0.12	Brain	3 tissues	7 tissues (HMEC.35, MYO, HMEC)	9 tissues (HEME.35, HMEC)	GATA1,GATA3 (T-47D)	Pou2f2, Pou5f1
rs17601879	16833	T/C	0.28	0.1	0.12	0.08	Blood, brain		Blood, GI and breast (HEMC.35, MYO)	Cervix (HELAS3)	FOXA1 (T-47D) GATA3 (T-47D)	SEF-1
rs7650827	18258	T/C	0.27	0.1	0.12	0.08	Blood, brain		10 tissues (HMEC.35, MYO, HMEC)	Skin, Cervix		HNF4, Mef2, Pax-2
rs9809402	22074	C/T	0.24	0.34	0.12	0.44	Blood, brain		(HMEC, MYO)			6 altered motifs
rs7627427	23005	G/C	0.23	0.33	0.12	0.44	Blood, brain	4 tissues	12 tissues (HMEC.35, MYO, HMEC)	6 tissues (HMEC.35, HMEC)		
rs9833915	24723	A/G	0.45	0.13	0.12	0.11	Brain		3 tissues (HMEC.35, MYO)			8 altered motifs
rs13074711	26505	T/C	0.39	0.12	0.12	0.11	Brain		7 tissues (HMEC.35, MYO, HMEC)			29 altered motifs
rs9833271	37244	G/A	0.05	0	0	0			8 tissues (HMEC.35, MYO, HMEC)	4 tissues (HMEC)		DMRT2

*base pair upstream of the transcription starting site of *TNFSF10*.

†Normal mammary or breast cancer cell lines are indicated in parenthesis.

HMEC.35: breast variant human mammary epithelial cells; MYO: breast myoepithelial primary cells; HMEC: mammary epithelial primary cells; T-47D: mammary ductal carcinoma cells.

consortium consists of nine epidemiological studies (34). The replication stage (stage 2) of the study is based on the AMBER (African American Breast Cancer Epidemiology and Risk) consortium, which consists of four epidemiologic studies (35). Selected characteristics of study participants for each study are summarized in Table 1. The study design for each study is described in the Supplementary Materials.

Genome-wide genotyping and quality control

Genotyping in the ROOT consortium was conducted using the Illumina HumanOmni2.5-8v1 array. Genotyping was attempted for a total of 3,909 study samples, of which 3,859 were successful. The 3,859 study samples were derived from 3,774 subjects, with 85 subjects with duplicate samples. After genotyping, quality control analysis was carried out by checking chromosomal anomalies, participant relativeness, population structure, missing call rates, batch effects, duplicate sample discordance, Mendelian errors, Hardy-Weinberg equilibrium, and duplicate SNP probes. First, duplicate samples from 85 pairs were excluded. Chromosomal anomalies >5 Mb and anomalies on chromosomes where the sum of the anomaly lengths >10 Mb were identified. No subjects were excluded for this reason but all genotypes were set to missing for regions with chromosome anomalies. The relatedness between each pair of participants was evaluated by the identical by descent (IBD) analysis. According to the IBD analysis, samples from 59 subjects were excluded because they were possibly relatives of other subjects. Additionally, 20 subjects (18 African Americans and 2 African Barbadian) were excluded because principal components analysis showed that they were clustered with the CEU/TSI HapMaps (i.e. high percentage of European ancestry). We further filtered out 7 samples with a missing call rate > 2% and 2 samples with possible tumour contamination. After these exclusions, a total of 3,686 subjects (1,657 cases and 2,029 controls) were included for the final analysis.

A total of 2,379,855 SNP probes were attempted in Illumina HumanOmni2.5-8v1 array. In the SNP level quality control analysis, 32,192 SNPs with technical errors, 119,924 SNPs with minor allele frequency = 0 and 5,631 duplicate SNPs were excluded. We further filtered out 99,511 SNPs with missing call rate > = 2%, 563 SNPs with >1 discordant calls, 1,078 SNPs with >1 Mendelian errors and 4,591 SNPs with Hardy-Weinberg equilibrium test $P < 10^{-4}$. After these exclusions, 2,116,365 SNP remained.

Genotyping in the AABC consortium was conducted using the IlluminaHuman1M-Duo BeadChip. Of the 5,984 samples available to the AABC consortium (3,153 cases and 2,831 controls), we attempted genotyping of 5,932, removing samples ($n = 52$) with DNA concentrations <20 ng/ul. Following genotyping, we removed samples based on the following exclusion criteria: 1) unknown replicates (≥98.9% genetically identical) that we were able to confirm (only one of each duplicate was removed, $n = 15$); 2) unknown replicates that we were not able to confirm through discussions with study investigators (pair or triplicate removed, $n = 14$); 3) samples with call rates <95% after a second attempt ($n = 100$); 4) samples with ≤ 5% African ancestry ($n = 36$); and, 5) samples with <15% mean heterozygosity of SNPs in the X chromosome and/or similar mean allele intensities of SNPs on the X and Y chromosomes ($n = 6$) (these are likely to be males).

In the analysis, we removed SNPs with <95% call rate ($n = 21,732$) or minor allele frequencies (MAFs) <1% ($n = 80,193$). To assess genotyping reproducibility, we included 138 replicate samples; the average concordance rate was 99.95% (>99.93% for all pairs). We also eliminated SNPs with genotyping concordance rates <98% based on the replicates ($n = 11,701$). The final analysis dataset included 1,043,036 SNPs genotyped on 3,016 cases and 2,745 controls, with an average SNP call rate of 99.7% and average sample call rate of 99.8%. Hardy-Weinberg equilibrium (HWE) was not used as a criterion for removing SNPs at this stage.

Imputation and technical validation by sequencing

Genotype imputation for the ROOT consortium was conducted by the University of Washington Genetics Coordinating Center (GCC) using the *IMPUTE2* software (23). Following filters for genotype imputation, including no ambiguity in strand alignment and ability to be mapped to NCDB build 37, 2,018,833 SNPs were selected as the basis for imputation. Using the 1000 Genomes Project phase I integrated variant set as the reference panel (October 2011 release), 20,109,249 SNPs were imputed. Of these, 16,147,413 have passed the imputation quality filter (imputation score >0.3) and were included along with the genotyped SNPs in the final association analysis. Genotype imputation in AABC was conducted using *IMPUTE2* software (23) to a cosmopolitan panel of all 1000 Genome Project subjects (March 2012 release).

To provide technical validation, we sequenced 96 samples that have been included in the final analysis of the ROOT GWAS. We conducted whole genome sequencing on the Illumina platform, with average depth of 30x. The concordance rate between whole genome sequencing and array-based genotyping was 99.5%.

Genotyping in the replication stage

Genotyping of the AMBER cases and controls was performed on the Illumina Human Exome Beadchip v1.1 with additional custom content at the Center for Inherited Disease Research (CIDR). The standard content of this array includes more than 240,000 coding variants, as well as ancestry informative markers (AIMs). Twenty thousand variants identified in the AABC/ROOT meta-analysis, including 2 or more variants per associated region for the top ~5000 loci, were included as part of the custom content.

Of 405,555 SNPs attempted for genotyping, 381,212 were released by CIDR and 299,873 of these remained after removing SNPs that were monomorphic ($n = 70,761$), were positional duplicates, were on the Y chromosome, had Hardy-Weinberg Equilibrium (HWE) $P < 10^{-4}$, had call rate < 0.98, had > 1 Mendelian errors in trios from HapMap (http://hapmap.ncbi.nlm.nih.gov/), or had > 2 discordant calls in duplicate samples. IMPUTE2 was used to impute additional variants using the 1000 Genomes Phase I reference panel (5/21/2011 1000G data, December 2013 haplotype release in IMPUTE2 site). For the meta-analysis, 18,376 SNPs passed quality control and included in the study.

Statistical analysis

In the ROOT GWAS, we analysed genotyped SNPs and imputed SNPs with imputation score >0.3 using SNPTEST software to account for uncertainty in imputation (36). We examined the association of each SNP and breast cancer risk using unconditional logistic regression, adjusting for age, study site and the first four eigenvectors from principal component analysis. The first four eigenvectors were used to control for population stratification as only the first 4 eigenvectors were associated with case status.

Odds ratios (OR) and 95% confidence intervals (CI) were calculated from the multivariable logistic regressions. All tests of statistical significance were two-sided. Using similar methods, we conducted an analysis to compare ER-positive breast cancers with controls and compare ER-negative breast cancers with controls, in order to identify subtype-specific SNPs.

In the AABC GWAS, we analysed genotyped SNPs and imputed SNPs with imputation score >0.3 and minor allele frequency >0.01. Unconditional logistic regression was used to examine the association between each SNP and overall breast cancer risk, ER-positive and ER-negative breast cancer risk separately. Age, the first 10 eigenvectors from principal component analysis, and study site were adjusted for in the logistic regression. To be conservative, all the first 10 eigenvectors were used to control population stratification. The SFBCS and NC-BCFR studies were conducted in the same San Francisco Bay Area population and were combined as one study site in the analyses.

In the replication study, principal components were computed using the smartpca program in the EIGENSOFT package. No eigenvectors from the principal component analysis were strongly associated with case status after controlling for the matching factors, DNA source and study. ORs and 95% CIs for each SNP were estimated in the analyses adjusted for study, age, geographic region, DNA source and eigenvectors 5, 6 and 8, which were associated with cases status at $P < 0.10$. Regression coefficients (log odds ratios) and their standard errors, from the two GWAS and the replication study were combined using a fixed effect meta-analysis method.

Functional annotation

For each top variant and region identified in this study, we used HaploReg (37) and USCS Genome Browser to explore functional annotations of noncoding variants. Chromatin states (promoters and enhancers), conserved regions, variant effect on regulatory motifs, and protein binding sites were assessed from available data from the ENCODE (38) and the Roadmap Epigenomics Consortium (39). Data from breast tumour cell lines (MCF-7, T-47D) and normal mammary epithelial cells (HMEC, MYO, vMHEC) were emphasized.

Data on expression quantitative trait locus (eQTL) analysis in blood were obtained from Westra *et al.* (27), in which eQTLs were identified in a meta-analysis in 5,311 peripheral untransformed blood samples and replication analysis in another 2,775 samples (27). Data on eQTL analysis in brain were obtained from ~400 autopsied subjects (28). Analysis on eQTL in breast tumours was conducted using 46 ER+ breast tumours and 45 ER- breast tumours from The Cancer Genome Atlas (TCGA). Only samples from African Americans were included. Copy number change and methylation status were controlled for in the eQTL analysis in tumours with MatrixEQTL package in Bioconductor.

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

Funding

The ROOT Consortium was supported by National Cancer Institute grants R01-CA142996, P50-CA125183, R01-CA89085 and U01-CA161032. DH and OIO were also supported by American Cancer Society (MRSG-13-063-01-TBG and CRP-10-119-01-CCE). AABC was supported by a Department of Defense Breast Cancer Research Program Era of Hope Scholar Award to CAH [W81XWH-08-1-0383] and the Norris Foundation. Each of the participating AABC studies was supported by the following grants: MEC (National Institutes of Health grants R01-CA63464 and R37-CA54281); CARE (National Institute for Child Health and Development grant NO1-HD-3-3175, K05 CA136967); WCHS (U.S. Army Medical Research and Material Command (USAMRMC) grant DAMD-17-01-0-0334, the National Institutes of Health grant R01-CA100598, and the Breast Cancer Research Foundation; SFBCS (National Institutes of Health grant R01-CA077305 and United States Army Medical Research Program grant DAMD17-96-6071); NC-BCFR (National Institutes of Health grant U01-CA069417); CBCS (National Institutes of Health Specialized Program of Research Excellence in Breast Cancer, grant number P50-CA58223 and Center for Environmental Health and Susceptibility National Institute of Environmental Health Sciences, National Institutes of Health, grant number P30-ES10126); PLCO (Intramural Research Program, National Cancer Institute, National Institutes of Health); NBHS (National Institutes of Health grant R01-CA100374). The Breast Cancer Family Registry (BCFR) was supported by the National Cancer Institute, National Institutes of Health under RFA-CA-06-503 and through cooperative agreements with members of the Breast Cancer Family Registry and Principal Investigators. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or the BCFR. AMBER was supported by the National Cancer Institute grants P01-CA151135, R01-CA098663, R01-CA058420, UM1-CA164974, R01-CA100598, UM1-CA164973, R01-CA54281, P50-CA58223, and the University Cancer Research Fund of North Carolina. Pathology data were obtained from numerous state cancer registries (Arizona, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, New Jersey, New York, North Carolina, Oklahoma, Pennsylvania, South Carolina, Tennessee, Texas, Virginia). The results reported do not necessarily represent their views or the views of the National Institutes of Health.

References

1. Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J. and Jemal, A. (2015) Global cancer statistics, 2012. *CA. Cancer J. Clin.*, **65**, 87–108.
2. Carey, L.A., Perou, C.M., Livasy, C.A., Dressler, L.G., Cowan, D., Conway, K., Karaca, G., Troester, M.A., Tse, C.K., Edmiston, S., *et al.* (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*, **295**, 2492–2502.
3. Huo, D., Ikpatt, F., Khramtsov, A., Dangou, J.M., Nanda, R., Dignam, J., Zhang, B., Grushko, T., Zhang, C., Oluwasola, O., *et al.* (2009) Population differences in breast cancer: survey in indigenous african women reveals over-representation of triple-negative breast cancer. *J. Clin. Oncol.*, **27**, 4515–4521.
4. Eng, A., McCormack, V. and dos-Santos-Silva, I. (2014) Receptor-defined subtypes of breast cancer in indigenous populations in Africa: a systematic review and meta-analysis. *PLoS Med.*, **11**, e1001720.

5. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K., *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, **45**, 353–361. 361e351-352.
6. Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M.J., Maranian, M.J., Bolla, M.K., Wang, Q., Shah, M., *et al.* (2015) Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.*, **47**, 373–380.
7. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087–1093.
8. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., *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.
9. Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., *et al.* (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, **39**, 865–869.
10. Stacey, S.N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S.A., Jonsson, G.F., Jakobsdottir, M., Bergthorsson, J.T., Gudmundsson, J., Aben, K.K., *et al.* (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, **40**, 703–706.
11. Ahmed, S., Thomas, G., Ghoussaini, M., Healey, C.S., Humphreys, M.K., Platte, R., Morrison, J., Maranian, M., Pooley, K.A., Luben, R., *et al.* (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat. Genet.*, **41**, 585–590.
12. Zheng, W., Long, J., Gao, Y.T., Li, C., Zheng, Y., Xiang, Y.B., Wen, W., Levy, S., Deming, S.L., Haines, J.L., *et al.* (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.*, **41**, 324–328.
13. Thomas, G., Jacobs, K.B., Kraft, P., Yeager, M., Wacholder, S., Cox, D.G., Hankinson, S.E., Hutchinson, A., Wang, Z., Yu, K., *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.
14. Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C.S., *et al.* (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.*, **42**, 504–507.
15. Antoniou, A.C., Wang, X., Fredericksen, Z.S., McGuffog, L., Tarrell, R., Sinilnikova, O.M., Healey, S., Morrison, J., Kartsonaki, C., Lesnick, T., *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**, 885–892.
16. Fletcher, O., Johnson, N., Orr, N., Hosking, F.J., Gibson, L.J., Walker, K., Zelenika, D., Gut, I., Heath, S., Palles, C., *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. Ghoussaini, M., Fletcher, O., Michailidou, K., Turnbull, C., Schmidt, M.K., Dicks, E., Dennis, J., Wang, Q., Humphreys, M.K., Luccarini, C., *et al.* (2012) Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat. Genet.*, **44**, 312–318.
18. Siddiq, A., Couch, F.J., Chen, G.K., Lindstrom, S., Eccles, D., Millikan, R.C., Michailidou, K., Stram, D.O., Beckmann, L., Rhie, S.K., *et al.* (2012) A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum. Mol. Genet.*, **21**, 5373–5384.
19. Garcia-Closas, M., Couch, F.J., Lindstrom, S., Michailidou, K., Schmidt, M.K., Brook, M.N., Orr, N., Rhie, S.K., Riboli, E., Feigelson, H.S., *et al.* (2013) Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.*, **45**, 392–398. 398e391-392.
20. Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E., *et al.* (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.*, **45**, 371–384. 384e371-372.
21. Haiman, C.A., Chen, G.K., Vachon, C.M., Canzian, F., Dunning, A., Millikan, R.C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C.B., *et al.* (2011) A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.*, **43**, 1210–1214.
22. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, **5**, e1000529.
23. Wiley, S.R., Schooley, K., Smolak, P.J., Din, W.S., Huang, C.P., Nicholl, J.K., Sutherland, G.R., Smith, T.D., Rauch, C., Smith, C.A., *et al.* (1995) Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity*, **3**, 673–682.
24. Consortium, G.T. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.
25. Rahman, M., Davis, S.R., Pumphrey, J.G., Bao, J., Nau, M.M., Meltzer, P.S. and Lipkowitz, S. (2009) TRAIL induces apoptosis in triple-negative breast cancer cells with a mesenchymal phenotype. *Breast Cancer Res. Treat.*, **113**, 217–230.
26. Lemke, J., von Karstedt, S., Zinngrebe, J. and Walczak, H. (2014) Getting TRAIL back on track for cancer therapy. *Cell Death Differ.*, **21**, 1350–1364.
27. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238–1243.
28. Zou, F., Chai, H.S., Younkin, C.S., Allen, M., Crook, J., Pankratz, V.S., Carrasquillo, M.M., Rowley, C.N., Nair, A.A., Middha, S., *et al.* (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genetics*, **8**, e1002707.
29. Chen, F., Chen, G.K., Millikan, R.C., John, E.M., Ambrosone, C.B., Bernstein, L., Zheng, W., Hu, J.J., Ziegler, R.G., Deming, S.L., *et al.* (2011) Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum. Mol. Genet.*, **20**, 4491–4503.
30. Zheng, W., Cai, Q., Signorello, L.B., Long, J., Hargreaves, M.K., Deming, S.L., Li, G., Li, C., Cui, Y. and Blot, W.J. (2009) Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer. Epidemiol. Biomarkers Prev.*, **18**, 2761–2764.
31. Huo, D., Zheng, Y., Ogundiran, T.O., Adebamowo, C., Nathanson, K.L., Domchek, S.M., Rebbeck, T.R., Simon, M.S.,

John, E.M., Hennis, A., *et al.* (2012) Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis*, **33**, 835–840.
32. Palmer, J.R., Ruiz-Narvaez, E.A., Rotimi, C.N., Cupples, L.A., Cozier, Y.C., Adams-Campbell, L.L. and Rosenberg, L. (2013) Genetic susceptibility loci for subtypes of breast cancer in an African American population. *Cancer. Epidemiol. Biomarkers Prev.*, **22**, 127–134.
33. Ghoussaini, M., Edwards, S.L., Michailidou, K., Nord, S., Cowper-Sal Lari, R., Desai, K., Kar, S., Hillman, K.M., Kaufmann, S., Glubb, D.M., *et al.* (2014) Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat. Commun.*, **4**, 4999.
34. Chen, F., Chen, G.K., Stram, D.O., Millikan, R.C., Ambrosone, C.B., John, E.M., Bernstein, L., Zheng, W., Palmer, J.R., Hu, J.J., *et al.* (2013) A genome-wide association study of breast cancer in women of African ancestry. *Hum. Genet.*, **132**, 39–48.
35. Palmer, J.R., Ambrosone, C.B. and Olshan, A.F. (2014) A collaborative study of the etiology of breast cancer subtypes in African American women: the AMBER consortium. *Cancer Causes Control*, **25**, 309–319.
36. Marchini, J. and Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nat. Rev.*, **11**, 499–511.
37. Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–D934.
38. Encode Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.
39. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.