Cancer Epidemiology, Biomarkers & Prevention

Short Communication

A Genome-wide Association Study of Early-Onset Breast Cancer Identifies *PFKM* as a Novel Breast Cancer Gene and Supports a Common Genetic Spectrum for Breast Cancer at Any Age

Habibul Ahsan[1,2,3,4,5], Jerry Halpern[6], Muhammad G. Kibriya[1,2], Brandon L. Pierce[1,2,5], Lin Tong[1,2], Eric Gamazon[3], Valerie McGuire[6], Anna Felberg[6], Jianxin Shi[11], Farzana Jasmine[1,2], Shantanu Roy[1,2], Rachelle Brutus[1,2], Maria Argos[1,2], Stephanie Melkonian[1,2], Jenny Chang-Claude[19], Irene Andrulis[28], John L. Hopper[30], Esther M. John[6,7,8], Kathi Malone[13], Giske Ursin[33], Marilie D. Gammon[14], Duncan C. Thomas[9], Daniela Seminara[11], Graham Casey[9], Julia A. Knight[28], Melissa C. Southey[30,31], Graham G. Giles[30,32], Regina M. Santella[15], Eunjung Lee[9], David Conti[9], David Duggan[16], Steve Gallinger[29], Robert Haile[9], Mark Jenkins[32], Noralane M. Lindor[17], Polly Newcomb[13], Kyriaki Michailidou[34], Carmel Apicella[30], Daniel J. Park[31], Julian Peto[36], Olivia Fletcher[37], Isabel dos Santos Silva[36], Mark Lathrop[39,40], David J. Hunter[18], Stephen J. Chanock[12], Alfons Meindl[20], Rita K. Schmutzler[22], Bertram Müller-Myhsok[21], Magdalena Lochmann[20], Lars Beckmann[24], Rebecca Hein[19,23], Enes Makalic[30], Daniel F. Schmidt[30], Quang Minh Bui[30], Jennifer Stone[30], Dieter Flesch-Janys[25,26], Norbert Dahmen[27], Heli Nevanlinna[41], Kristiina Aittomäki[42], Carl Blomqvist[43], Per Hall[44], Kamila Czene[44], Astrid Irwanto[45], Jianjun Liu[45]; Nazneen Rahman[38] and Clare Turnbull[38] for the Familial Breast Cancer Study[38]; Alison M. Dunning[35], Paul Pharoah[34,35], Quinten Waisfisz[46], Hanne Meijers-Heijboer[46], Andre G. Uitterlinden[47], Fernando Rivadeneira[47], Dan Nicolae[3], Douglas F. Easton[34,35], Nancy J. Cox[3,4,5], and Alice S. Whittemore[6,10]

Abstract

Early-onset breast cancer (EOBC) causes substantial loss of life and productivity, creating a major burden among women worldwide. We analyzed 1,265,548 Hapmap3 single-nucleotide polymorphisms (SNP) among a discovery set of 3,523 EOBC incident cases and 2,702 population control women ages $\leq$ 51 years. The SNPs with smallest *P* values were examined in a replication set of 3,470 EOBC cases and 5,475 control women. We also tested EOBC association with 19,684 genes by annotating each gene with putative functional SNPs, and then combining their *P* values to obtain a gene-based *P* value. We examined the gene with smallest *P* value for replication in 1,145 breast cancer cases and 1,142 control women. The combined discovery and replication sets identified 72 new SNPs associated with EOBC ($P < 4 \times 10^{-8}$) located in six genomic regions previously reported to contain SNPs associated largely with later-onset breast cancer (LOBC). SNP rs2229882 and 10 other SNPs on chromosome 5q11.2 remained associated ($P < 6 \times 10^{-4}$) after adjustment for the strongest published SNPs in the region. Thirty-two of the 82 currently known LOBC SNPs were associated with EOBC ($P < 0.05$). Low power is likely responsible for the remaining 50 unassociated known LOBC SNPs. The gene-based analysis identified an association between breast cancer and the phosphofructokinase-muscle (*PFKM*) gene on chromosome 12q13.11 that met the genome-wide gene-based threshold of 2.5×10^{-6}. In conclusion, EOBC and LOBC seem to have similar genetic etiologies; the 5q11.2 region may contain multiple distinct breast cancer loci; and the *PFKM* gene region is worthy of further investigation. These findings should enhance our understanding of the etiology of breast cancer. *Cancer Epidemiol Biomarkers Prev; 23(4); 658–69. ©2014 AACR.*

Authors' Affiliations: [1]Center for Cancer Epidemiology and Prevention; Departments of [2]Health Studies, [3]Medicine, and [4]Human Genetics; [5]Comprehensive Cancer Center, University of Chicago, Chicago, Illinois; [6]Department of Health Research and Policy, Stanford University School of Medicine; [7]Stanford Cancer Institute, Stanford; [8]Cancer Prevention Institute of California, Fremont; [9]Department of Preventive Medicine, University of Southern California, Los Angeles; [10]Stanford Cancer Institute, Palo Alto, California; [11]Epidemiology and Genetics Research Program; [12]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; [13]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; [14]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; [15]Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York; [16]Integrated Cancer Genomics Division, Translational Genomics Research Institute, Phoenix; [17]Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona; [18]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts; [19]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg; [20]Clinic of Gynaecology and Obstetrics, Division for Gynaecological Tumor-Genetics, Technische Universität München; [21]Max Planck Institute of Psychiatry, Munich; [22]Department of Obstetrics and Gynaecology, Division of Molecular Gynaeco-Oncology, [23]PMV Research Group at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne; [24]Foundation for Quality and Efficiency in Health Care, Cologne; [25]Department of Cancer Epidemiology/Clinical Cancer Registry; [26]Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg; [27]Department of Psychiatry, University of

Introduction

Early-onset breast cancer (EOBC) leads to substantial loss of life and productivity, creating a major public health and economic burden in both developed and developing countries. Many patterns of breast cancer incidence, histopathologic characteristics, clinical behavior, and risk factors, including the increase in risk associated with a family history, differ between cases diagnosed during premenopausal and postmenopausal periods; a difference that has prompted speculation that there might be some genetic etiologies that are different for EOBC and later-onset breast cancer (LOBC; refs. 1–4). For example, a study of Utah families estimated that the risk of developing breast cancer for sisters of EOBC cases was 3.70 [95% confidence interval (CI), 2.5–5.2] times that for the general Caucasian population, nearly double the 1.83-fold relative risk (95% CI, 1.65–2.01) among sisters of cases of all ages (5). About 25% of the aggregation is explained by the high risks specific to carriers of deleterious mutations in the major susceptibility genes *BRCA1* and *BRCA2*, but even after excluding carrier families, risks are higher for relatives of EOBC cases than among relatives of LOBC cases (3, 6, 7). Recently, genome-wide association studies (GWAS) have reported many single-nucleotide polymorphisms (SNP) as associated with breast cancer risk (8). To date, however, no published GWASs have focused on EOBC. Here, we report findings from the first large-scale GWAS of EOBC involving a discovery set of 6,225 young Caucasian women from eight sites in the United States, Canada, Australia, and Germany and two replication sets of Caucasian women from Australia, the United States, the United Kingdom, and other European countries.

Materials and Methods

We used a case–control design to investigate EOBC risk among Caucasian women in relation to 1,265,546 SNPs included in the HapMap3 project (http://hapmap.ncbi.nlm.nih.gov/downloads/phasing/2009-02_phaseIII/HapMap3_r2/). Specifically, we used Illumina SNP arrays to genotype 3,523 EOBC cases and 2,702 control women and to impute their genotypes for the HaMap3 SNPs (hereafter called the discovery set). We then conducted two SNP-based analyses and a gene-based analysis. The results of these analyses were examined in two sets of independent data (called replication sets). We begin with a description of subject recruitment, genotyping, and quality control for the discovery set. We then describe the SNP-based analysis and replication, followed by the gene-based analysis and replication.

Discovery set

Subject recruitment. Population-based subjects were recruited from the eight sites described in Supplementary Table S1, some of which oversampled cases with a personal or family history, suggesting a heritable basis for their disease (9–14). Eligible cases were non-Hispanic White (NHW) women diagnosed with invasive breast cancer when 51 years or younger and not known to carry pathogenic mutations in *BRCA1* or *BRCA2*. Eligible controls were NHW women ages 20 to 51 years without a history of breast cancer, who were identified largely by random-digit dialing. Supplementary Table S2 shows the number of eligible subjects from each of the eight contributing sites after quality control.

Genotyping and quality control. DNA samples for subjects from all but one of the sites were genotyped at the University of Chicago on Illumina 610-Quad and Cyto12 v2 BeadChips (Illumina Inc.), using the protocol described in the Supplementary Methods. Two hundred and twenty-seven population control subjects from the Colon Cancer Family Registry (CCFR) were genotyped at TGEN (Translational Genomics Research Institute; http://www.tgen.org) using the Illumina Human1M and HumanOmni1-Quad BeadChips. In addition, 27 blinded and 22 unblinded quality control replicates from the study sample were genotyped on the Human1M. Replicates showed concordance of called genotypes >99.94% (for samples with call rates >90%). Standard laboratory quality control procedures were applied and have been described previously (15). Quality control was implemented using a combination of PLINK (16) and custom programs written in C, R, Perl, and the Unix bash shell. Data quality control procedures are described in more detail in Supplementary Methods and summarized in Supplementary Table S3. This table shows that 555,254 of the 1,298,078 SNPs remained after quality control, and

Mainz, Mainz, Germany; [28]Samuel Lunenfeld Research Institute; [29]Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Toronto, Ontario, Canada; [30]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, Melbourne School of Population Health; [31]Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne; [32]Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Victoria, Australia; [33]Norway Cancer Registry, Norway; Departments of [34]Public Health and Primary Care and [35]Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge; [36]Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine; [37]Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London; [38]Section of Cancer Genetics, Institute of Cancer Research, Sutton, United Kingdom; [39]Centre National de Genotypage, Evry; [40]Fondation Jean Dausset–CEPH, Paris, France; Departments of [41]Obstetrics and Gynecology, [42]Clinical Genetics, and [43]Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; [44]Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [45]Human Genetics Division, Genome Institute of Singapore, Singapore, Singapore; [46]Department of Clinical Genetics, VU University Medical Center, Division of Oncogenetics, Amsterdam; and [47]Department of Internal Medicine and Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands

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

Corresponding Author: Habibul Ahsan, University of Chicago Medical Center, 5841 South Maryland Avenue, Chicago, IL 60615. Phone: 773-834-9956; Fax: 773-834-0139; E-mail: habib@uchicago.edu

doi: 10.1158/1055-9965.EPI-13-0340

Table 1. Newly identified SNPs associated with EOBC at combined significance level $P < 4 \times 10^{-8}$

Region	Published SNP with smallest *P* value	Newly identified SNPs	Combined data					
			RAF[a]	OR[b]	95% CI[b]	Unadjusted *P* value[c]	Adjusted *P* value[c]	R^{2} [d]
3p24	rs4973768	rs653465	0.54	1.18	1.12–1.23	4.73E−12	1.85E−02	0.75
		rs487930	0.54	1.18	1.12–1.23	5.04E−12	1.98E−02	0.74
		rs552647	0.54	1.18	1.12–1.23	5.28E−12	2.03E−02	0.76
		rs2100006	0.55	1.17	1.12–1.23	1.13E−11	2.92E−02	0.72
		rs2034190	0.55	1.17	1.12–1.23	1.23E−11	3.24E−02	0.72
		rs12487340	0.55	1.17	1.12–1.23	3.88E−11	2.39E−02	0.60
		rs7653795	0.55	1.17	1.11–1.22	8.96E−11	4.90E−02	0.62
		rs2370946	0.57	1.16	1.11–1.22	1.14E−10	7.26E−02	0.65
		rs1445111	0.57	1.16	1.11–1.22	1.32E−10	7.66E−02	0.65
		rs11129270	0.57	1.16	1.11–1.22	1.42E−10	8.15E−02	0.65
		rs10049490	0.57	1.16	1.11–1.22	2.93E−10	1.14E−01	0.63
		rs1472254	0.54	1.15	1.1–1.21	5.76E−10	7.44E−01	0.80
		rs7634878	0.51	1.15	1.1–1.21	1.13E−9	4.17E−01	0.64
5q11	rs889312	**rs2229882**[e]	0.06	1.45	1.32–1.6	1.02E−14	**3.63E−07**	0.10
		rs16886181	0.18	1.26	1.18–1.34	9.01E−14	**5.72E−04**	0.53
		rs961847	0.30	1.21	1.15–1.27	1.04E−13	5.33E−01	0.96
		rs252913	0.37	1.20	1.14–1.26	4.66E−13	9.34E−02	0.42
		rs832540	0.36	1.19	1.14–1.25	8.38E−13	1.25E−01	0.42
		rs702691	0.36	1.19	1.14–1.25	8.67E−13	2.19E−01	0.50
		rs252905	0.36	1.19	1.14–1.25	9.22E−13	2.21E−01	0.50
		rs832585	0.36	1.19	1.14–1.25	1.04E−12	2.21E−01	0.50
		rs832566	0.36	1.19	1.14–1.25	1.06E−12	2.19E−01	0.50
		rs252906	0.36	1.19	1.14–1.25	1.07E−12	2.33E−01	0.50
		rs11960484	0.36	1.19	1.14–1.25	1.08E−12	2.45E−01	0.50
		rs832552	0.36	1.19	1.14–1.25	1.11E−12	2.45E−01	0.50
		rs252925	0.36	1.19	1.14–1.25	1.16E−12	1.41E−01	0.46
		rs6890270	0.20	1.24	1.17–1.31	1.47E−12	3.40E−03	0.18
		rs832577	0.36	1.19	1.13–1.25	1.82E−12	2.74E−01	0.50
		rs16886448	0.07	1.37	1.25–1.49	2.24E−12	**5.21E−06**	0.05
		rs16886397	0.07	1.36	1.25–1.49	3.84E−12	**8.05E−06**	0.05
		rs3822625	0.07	1.36	1.24–1.48	5.01E−12	**7.79E−06**	0.05
		rs16886364	0.07	1.36	1.25–1.48	5.28E−12	**1.03E−05**	0.05
		rs16886113	0.08	1.35	1.23–1.47	3.79E−11	**1.71E−05**	0.10
		rs1017226	0.08	1.33	1.22–1.45	5.73E−11	**2.57E−05**	0.05
		rs7726354	0.06	1.37	1.24–1.5	6.52E−11	**4.70E−05**	0.08
		rs1445996	0.36	1.17	1.12–1.23	1.6E−10	5.43E−01	0.36
		rs832529	0.36	1.17	1.12–1.23	1.64E−10	5.54E−01	0.36
		rs252890	0.36	1.17	1.12–1.23	1.81E−10	5.75E−01	0.35
		rs12655019	0.10	1.27	1.18–1.37	3.06E−10	**6.53E−04**	0.07
		rs331498	0.37	1.17	1.11–1.23	6.35E−10	7.11E−01	0.35
		rs331497	0.37	1.16	1.11–1.22	1.23E−9	8.10E−01	0.35
		rs2662024	0.37	1.16	1.11–1.22	1.65E−9	8.06E−01	0.35
		rs16886034	0.08	1.36	1.23–1.51	1.8E−9	**7.00E−05**	0.12
		rs10940511	0.43	1.15	1.1–1.21	4.85E−9	8.62E−01	0.43
		rs4700008	0.43	1.15	1.1–1.2	8.65E−9	9.08E−01	0.47
		rs6862199	0.44	1.14	1.09–1.2	1.43E−8	9.55E−01	0.45
		rs10940518	0.34	1.15	1.1–1.21	3.55E−8	7.54E−01	0.30
		rs10039338	0.34	1.15	1.1–1.21	3.96E−8	8.42E−01	0.30

(Continued on the following page)

Table 1. Newly identified SNPs associated with EOBC at combined significance level $P < 4 \times 10^{-8}$ (Cont'd)

Region	Published SNP with smallest *P* value	Newly identified SNPs	Combined data					
			RAF[a]	OR[b]	95% CI[b]	Unadjusted *P* value[c]	Adjusted *P* value[c]	R^{2} [d]
8q24	rs1562430	rs2392780	0.60	1.15	1.10–1.20	1.48E−8	8.45E−01	1.00
		rs673745	0.42	1.14	1.09–1.20	2.02E−8	2.10E−02	0.42
		rs7002826	0.60	1.14	1.09–1.20	2.06E−8	8.55E−01	0.94
		rs418269	0.41	1.14	1.09–1.20	2.13E−8	2.09E−02	0.43
		rs7007568	0.60	1.14	1.09–1.20	2.23E−8	9.26E−01	0.94
		rs7815100	0.59	1.14	1.09–1.20	2.56E−8	7.78E−01	0.95
		rs7826557	0.60	1.14	1.09–1.19	3.42E−8	9.49E−01	0.94
		rs10098985	0.60	1.14	1.09–1.19	3.72E−8	8.94E−01	0.95
10q26	rs2981579	rs2912774	0.44	1.29	1.23–1.35	2.72E−27	8.73E−01	0.86
		rs3750817	0.63	1.23	1.17–1.3	3.39E−16	5.02E−01	0.43
		rs17102287	0.20	1.20	1.13–1.28	3.05E−9	7.11E−01	0.26
11q13	rs614367	**rs537626**	0.18	1.29	1.21–1.37	1.8E−15	**3.74E−04**	0.41
		rs680618	0.24	1.20	1.14–1.27	1.04E−10	8.01E−02	0.40
		rs567488	0.18	1.21	1.14–1.29	3.95E−10	7.29E−03	0.26
		rs493786	0.34	1.16	1.1–1.21	1.23E−8	1.17E−02	0.19
		rs559664	0.35	1.15	1.1–1.21	1.43E−8	9.06E−03	0.19
		rs510754	0.35	1.15	1.1–1.21	1.46E−8	9.36E−03	0.19
16q12	rs3803662	rs4784223	0.29	1.27	1.21–1.34	6.2E−21	7.41E−01	0.95
		rs4784220	0.39	1.21	1.14–1.27	1.52E−12	6.34E−01	0.37
		rs8046979	0.48	1.16	1.11–1.21	2.68E−10	7.53E−01	0.41
		rs9933638	0.48	1.16	1.10–1.21	6.85E−10	7.35E−01	0.41
		rs2193094	0.48	1.15	1.10–1.21	9.09E−10	7.18E−01	0.42
		rs1420533	0.48	1.15	1.10–1.21	9.23E−10	7.17E−01	0.42
		rs9931232	0.48	1.15	1.10–1.22	1.21E−9	6.75E−01	0.42

[a]RAF, risk allele frequency in discovery set controls.

[b]OR, per allele OR from combined discovery and validation data, unadjusted for published SNPs.

[c]*P* value from models unadjusted and adjusted for published SNP with the smallest *P* value.

[d]Squared correlation coefficient with published SNP having the smallest *P* value.

[e]SNPs in bold represent those that are associated with EOBC with $P < 0.001$ after adjusting for published SNPs.

that most SNPs were deleted because they appeared only on the 1M and 1M Omni chips that were used to type only 227 controls.

Analysis. We first identified principal components representing axes of ancestral variation to adjust for population stratification (17) and imputed untyped SNPs using the HapMap3 data. We then conducted two SNP-based analyses and a gene-based analysis. The first SNP-based analysis consisted of SNP-specific logistic regressions for each of the 1,265,548 typed or imputed HapMap3 SNPs using BEAGLE (18). We checked for population stratification using graphical plots of test statistics and the λ measure of overdispersion (19). We used an additive regression model in which the logit of EOBC risk was linearly related to the number of SNP minor alleles, and noted the SNPs with nominal *P* values less than 4×10^{-8}. These SNPs and their minor allele frequencies (MAFs) for cases and controls as well as the discovery set *P* values are shown in Supplementary Table S4. The second SNP-based analysis was conducted to examine association between EOBC and each of the 82 breast cancer–associated SNPs currently reported and validated in the literature. (We were unable to impute one SNP that was not polymorphic in the HapMap3 data.) Here, we used SNP-specific logistic regressions for each of the 82 SNPs in which the logit of EOBC risk was linearly related to the number of SNP risk alleles, as reported in the literature.

The gene-based analysis was conducted in two steps. First, we attempted to annotate each known human gene with one or more of the SNPs in the discovery set that could affect its expression and/or function. Then, we combined the EOBC discovery set *P* values of these expression-related SNPs into summary gene-based *P* values. For step 1, we used expression-quantitative trait locus (eQTL) mapping of SNPs to genes, as implemented in the online database SCAN (20, 21) and used the eQTL significance levels to quantify the likelihood that a SNP [or one in strong linkage disequilibrium (LD) with it]

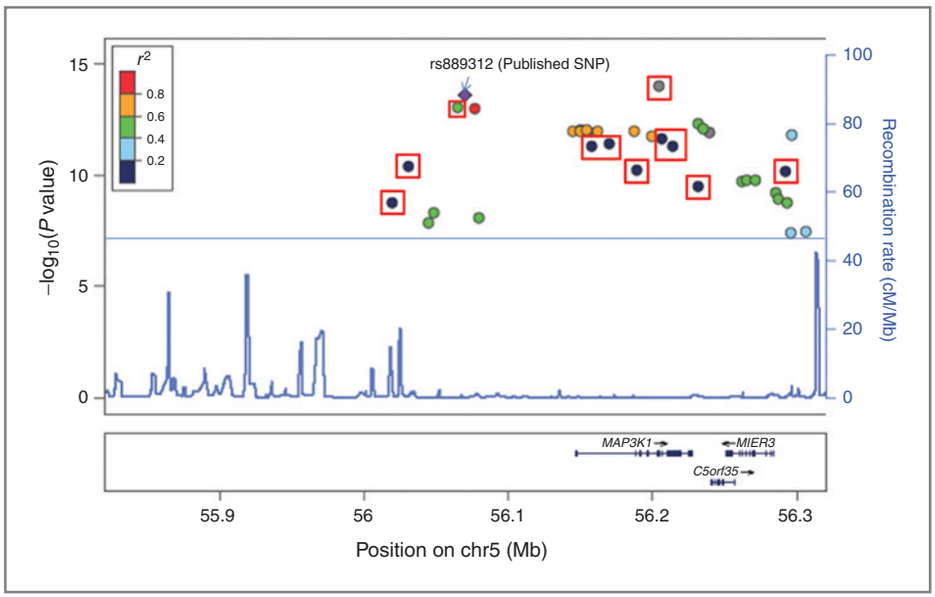

Figure 1. Manhattan plot of significance levels from combined discovery and replication data for SNPs in the 5q11.2 region. *Y*-axis, minus log *P* value for association with EOBC; *X*-axis, chromosomal position; and SNP color reflects its correlation with SNP rs889312 (SNP with smallest *P* value in discovery set, marked by arrow). SNPs in red boxes are associated with *P* value < 0.001 from regression analyses adjusting for rs889312. Horizontal bar, genome-wide significance threshold $P = 4 \times 10^{-8}$; blue curve, recombination rate.

regulates gene transcript levels (22). That is, we assigned an SNP to a gene if the SNP encoded a missense, nonsense, or frameshift (MNF) variant in the gene, or if it met our criteria for an eQTL SNP for the gene. Although not all the SNPs annotated to a gene are likely to be functional, they are clearly enriched for those with functional consequence. We were able to annotate 19,684 genes with one or more putative functional SNPs, 11,040 of which were annotated with at least one e-QTL SNP. In step 2, we calculated a gene-based *P* value for each of the 19,684 genes by combining the EOBC-association *P* values for all its putative functional SNPs using methods described elsewhere (23).

The Supplementary Methods contains additional details about both SNP-based and gene-based analyses.

Replication sets

Replication of SNP-based results. Primary genotype data were obtained from three early-onset breast cancer GWAS in populations of European ancestry (3, 24–28) as described in Supplementary Tables S5 and S6. For each typed or imputed SNP using ProbABEL (29), we combined the SNP-specific regression coefficients obtained for the discovery and replication sets using the commonly-deemed inverse-weighted summary statistic proposed by Cochran (30).

Replication of gene-based results. To replicate the gene-based association analyses, we used available GWAS data from the CGEMS breast cancer study of 1,145 Caucasian case women and 1,142 Caucasian control women ages 55 to 74 years. Details of subject selection, genotyping methods, and quality control analyses for CGEMS breast cancer project have been published (31, 32). The identical gene-based analytic method, described above and in Supplementary Methods, was applied to the CGEMS data obtained from dbGaP. The gene-based *P* values from both discovery and replication datasets were combined for the gene with smallest gene-based *P* value in discovery data using Fisher's method for meta-analysis (33).

Further details of both the SNP-based and gene-based replication sets can be found in Supplementary Method.

Results

SNP-based analysis

Analysis of combined discovery and replication sets identified 96 SNPs from six chromosomal regions as associated with EOBC risk with $P < 4 \times 10^{-8}$ (the threshold for genome-wide significance at level 0.05 with 1.2 million independent tests). These results were not driven by data from a single site. The six regions lie on chromosomes 3p24.1, 5q11.2, 8q24, 10q26.13, 11q13.2, and 16q12.1. Previous GWASs have associated SNPs in these regions with (largely later-onset) breast cancer; however, they have reported only 24 of these 96 SNPs (Table 1; refs. 28, 31, 32, 34–60). To investigate how many of the remaining 72 unpublished SNPs are independently associated with EOBC, we evaluated each of them using a regression model that also contained the published SNP in the region having the smallest *P* value in the combined discovery and replication data (called the index SNP). These regressions identified 12 of the 72 SNPs as independently associated with EOBC at significance level $P < 0.001$ (listed in bold in Table 1). Eleven of these 12 SNPs are in the 5q11.2 region and almost all are within or near the *MAP3K1* gene; eight are downstream of the published index SNP (Fig. 1). The strongest of these SNPs is rs2229882 with unadjusted *P* value 1.02×10^{-14} and squared correlation $r^2 = 0.10$ with the published SNP rs889312 (Fig. 1 and Table 1).

To further explore the 5q11.2 association, we examined 2,889 SNPs (278 typed and 2,611 imputed using 1 KG data)

Table 2. EOBC discovery set risk allele frequencies (RAFs) and per-allele odds-ratios (ORs) for known breast cancer hits (GWAS Catalogue) in subjects of European ancestry ($P \leq 5 \times 10^{-8}$)

					Published data		EOBC discovery set			
S. No.	Region	SNP rs number	bp position	Gene	RAF[a]	OR[b]	RAF[c]	OR[d]	*P* value	Power[e]
1	1p36	rs616488	10488802	*PEX14*	0.67	1.06	0.68	1.06	1.61E−01	0.32
2	1p13	rs11552449	114249912	*SYT6*	0.17	1.08	0.17	1.00	9.66E−01	0.37
3	1p11	rs11249433	120982136	*FCGR1B*	0.40	1.12	**0.42**[f]	**1.14**	**2.20E−03**	0.88
4	1q32	rs6678914	202187176	*LGR6*	0.59	1.08	0.59	1.06	1.50E−01	0.55
5	1q32	rs4245739	204518842	*MDM4*	0.26	1.13	0.27	0.97	4.70E−01	0.87
6	2p24	rs12710696	19320803	*OSR1*	0.36	1.11	0.38	1.04	3.40E−01	0.81
7	2p14	rs4849887	120961592	*LOC84931*	0.90	1.10	0.90	1.02	6.89E−01	0.33
8	2q31	rs2016394	172681217	*DLX2*	0.52	1.05	0.55	0.98	6.71E−01	0.27
9	2q31	rs1550623	173921140	*CDCA7*	0.84	1.08	0.85	1.04	3.99E−01	0.32
10	2q33	rs1045485	202149589	*CASP8*	0.85	1.03	0.92	0.9	1.54E−01	0.06
11	2q35	rs13387042	217614077	*TNP1*	0.50	1.16	**0.54**	**1.12**	**2.78E−03**	0.98
12	2q35	rs16857609	218004753	*TNS1*	0.26	1.09	0.25	1.09	5.68E−02	0.56
13	3p26	rs6762644	4717276	*ITPR1*	0.40	1.07	0.37	1.01	8.50E−01	0.45
14	3p24	rs4973768	27391017	*SLC4A7*	0.47	1.12	**0.50**	**1.13**	**7.05E−04**	0.88
15	3p24	rs12493607	30657943	*TGFBR2*	0.35	1.05	0.34	1.06	1.60E−01	0.25
16	4q24	rs9790517	106304227	*TET2*	0.23	1.07	0.22	1.08	1.23E−01	0.35
17	4q34	rs6828523	176083001	*ADAM29*	0.87	1.11	0.89	1.04	4.89E−01	0.41
18	5p15	rs10069690	1279790	*TERT*	0.26	1.05	0.27	0.94	3.85E−01	0.23
19	5p15	rs7734992	1280128	*TERT*	0.43	1.05	0.42	1.04	6.23E−01	0.27
20	5p15	rs3215401	1296255	*TERT*	0.70	1.06	NA[g]	NA	NA	NA
21	5p12	rs4415084	44662515	*MRPS30*	0.40	1.16	0.43	1.04	2.87E−01	0.98
22	5p12	rs10941679	44742255	*MRPS30*	0.25	1.15	0.23	1.05	4.15E−01	0.92
23	5q11	rs889312	56067641	*MAP3K1*	0.28	1.14	**0.29**	**1.29**	**1.16E−08**	0.92
24	5q11	rs10472076	58219818	*RAB3C*	0.38	1.06	0.35	1.03	4.09E−01	0.34
25	5q11	rs1353747	58373238	*PDE4D*	0.90	1.10	0.91	1.04	5.19E−01	0.31
26	5q33	rs1432679	158176661	*EBF1*	0.43	1.07	**0.43**	**1.11**	**7.21E−03**	0.46
27	6p25	rs11242675	1263878	*FOXQ1*	0.61	1.04	0.65	1.04	3.66E−01	0.18
28	6p23	rs204247	13830502	*RANBP9*	0.43	1.06	0.44	1.07	7.21E−02	0.36
29	6q14	rs17530068	82249828	*FAM46A*	0.22	1.07	0.24	0.99	7.97E−01	0.37
30	6q25	rs3757318	151955806	*ESR1*	0.07	1.19	**0.08**	**1.16**	**2.90E−02**	0.80
31	6q25	rs2046210	151990059	*C6orf97-ESR1*	0.34	1.10	**0.36**	**1.18**	**2.55E−05**	0.73
32	7q25	rs720475	143705862	*ARHGEF5*	0.75	1.06	**0.75**	**1.12**	**1.14E−02**	0.28
33	8p12	rs9693444	29565535	*DUSP4*	0.32	1.07	0.33	1.03	4.23E−01	0.43
34	8q21	rs6472903	76392856	*HNF4G*	0.82	1.11	0.83	1.05	3.23E−01	0.55
35	8q21	rs2943559	76580492	*HNF4G*	0.07	1.15	0.09	1.14	5.12E−02	0.65
36	8q24	rs13281615	128424800	*POU5F1B*	0.41	1.10	**0.43**	**1.14**	**6.12E−04**	0.75
37	8q24	rs1562430	128457034	—	0.59	1.17	**0.59**	**1.14**	**4.28E−04**	0.99
38	8q24	rs11780156	129263823	*MIR1208*	0.16	1.10	0.19	1.07	1.48E−01	0.57
39	9p21	rs1011970	22052134	*CDKN2A/B*	0.17	1.08	0.18	1.08	1.05E−01	0.39
40	9q31	rs10759243	109345936	*KLF4*	0.39	1.07	0.29	1.08	8.02E−02	0.41
41	9q31	rs865686	109928299	*KLF4*	0.62	1.12	**0.64**	**1.09**	**1.93E−02**	0.84
42	10p15	rs2380205	5886734	*ANKRD16*	0.56	1.02	0.57	1.04	2.80E−01	0.08
43	10p12	rs7072776	22072948	*MLLT10*	0.29	1.09	0.30	1.07	1.16E−01	0.60
44	10p12	rs11814448	22355849	*DNAJC1*	0.020	1.31	0.01	1.23	2.69E−01	0.40
45	10q21	rs10995190	63948688	*ZNF365*	0.85	1.18	**0.85**	**1.13**	**2.24E−02**	0.87
46	10q22	rs704010	80511154	*ZMIZ1*	0.39	1.10	**0.40**	**1.10**	**1.29E−02**	0.74
47	10q25	rs7904519	114763917	*TCF7L2*	0.46	1.06	**0.39**	**1.10**	**4.26E−03**	0.35
48	10q26	rs11199914	123083891	*FGFR2*	0.68	1.05	0.68	1.08	7.63E−02	0.24

(Continued on the following page)

Table 2. EOBC discovery set risk allele frequencies (RAFs) and per-allele odds-ratios (ORs) for known breast cancer hits (GWAS Catalogue) in subjects of European ancestry ($P \leq 5 \times 10^{-8}$) (Cont'd)

					Published data		EOBC discovery set			
S. No.	Region	SNP rs number	bp position	Gene	RAF[a]	OR[b]	RAF[c]	OR[d]	*P* value	Power[e]
49	10q26	rs2981579	123327325	*FGFR2*	0.40	1.31	**0.45**	**1.24**	**1.94E−08**	1.00
50	10q26	rs1219648	123336180	*FGFR2*	0.42	1.29	**0.43**	**1.22**	**1.42E−07**	1.00
51	10q26	rs2981582	123342307	*FGFR2*	0.38	1.23	**0.43**	**1.21**	**3.13E−07**	1.00
52	11p15	rs3817198	1865582	*LSP1*	0.31	1.07	**0.33**	**1.15**	**5.14E−04**	0.43
53	11q13	rs3903072	65339642	*SNX32*	0.53	1.06	0.54	1.04	3.00E−01	0.36
54	11q13	rs614367	69037945	*CCND1*	0.15	1.26	**0.16**	**1.34**	**1.14E−08**	1.00
55	11q13	rs554219	69331642	*CCND1*	0.12	1.33	**0.14**	**1.35**	**1.33E−07**	1.00
56	11q13	rs494406	69344241	*CCND1*	0.26	1.07	**0.26**	**1.11**	**2.23E−02**	0.39
57	11q13	rs75915166	69379161	*FGF3*	0.06	1.38	**0.08**	**1.28**	**2.39E−03**	1.00
58	11q24	rs11820646	128966381	*BARX2*	0.59	1.09	**0.61**	**1.11**	**7.58E−03**	0.63
59	12p13	rs12422552	14305198	*ATF7IP*	0.26	1.08	0.24	1.07	1.47E−01	0.46
60	12p11	rs10771399	28046347	*PTHLH*	0.88	1.19	**0.95**	**1.30**	**4.58E−03**	0.49
61	12q22	rs17356907	94551890	*NTN4*	0.70	1.11	0.71	0.99	8.94E−01	0.73
62	12q24	rs1292011	114320905	*TBX3*	0.58	1.09	0.60	1.10	1.33E−02	0.64
63	13q13	rs11571833	31870626	*BRCA2/N4BP2L1/2*	0.008	1.33	0.01	1.03	8.94E−01	0.44
64	14q13	rs2236007	36202520	*PAX9*	0.79	1.10	0.79	1.02	6.28E−01	0.55
65	14q24	rs2588809	67730181	*RAD51B*	0.16	1.08	0.18	1.00	9.59E−01	0.39
66	14q24	rs999737	68104435	*RAD51B*	0.77	1.12	**0.78**	**1.13**	**6.66E−03**	0.71
67	14q32	rs941764	90910822	*CCDC88C*	0.34	1.06	0.34	1.02	6.35E−01	0.34
68	16q12	rs8051542	51091668	*TOX3*	0.44	1.09	**0.45**	**1.11**	**1.06E−02**	0.66
69	16q12	rs12443621	51105538	*TOX3*	0.46	1.11	**0.51**	**1.16**	**1.47E−04**	0.82
70	16q12	rs4783780	51128937	*TOX3*	0.49	1.16	**0.50**	**1.16**	**1.06E−04**	0.98
71	16q12	rs3803662	51143842	*TOX3*	0.26	1.26	**0.31**	**1.25**	**3.26E−08**	1.00
72	16q12	rs3112612	52635164	*TOX3*	0.43	1.15	**0.44**	**1.10**	**7.66E−03**	0.97
73	16q12	rs17817449	52370868	*FTO*	0.60	1.06	0.61	1.02	6.54E−01	0.35
74	16q12	rs11075995	53855291	*FTO*	0.24	1.10	0.22	1.02	6.91E−01	0.61
75	16q23	rs13329835	79208306	*CDYL2*	0.22	1.11	**0.24**	**1.13**	**7.98E−03**	0.72
76	17q22	rs6504950	50411470	*STXBP4*	0.72	1.08	0.73	1.03	4.81E−01	0.46
77	18q11	rs527616	22591422	*AQP4*	0.62	1.08	**0.66**	**1.13**	**3.40E−03**	0.51
78	18q11	rs1436904	22824665	*CHST9*	0.60	1.05	0.61	1.05	2.23E−01	0.26
79	19q13	rs4808801	18432141	*ELL*	0.65	1.06	0.67	1.04	2.61E−01	0.32
80	19q13	rs3760982	48978353	*KCNN4*	0.46	1.06	0.47	1.03	4.13E−01	0.36
81	21q21	rs2823093	15442703	*NRIP1*	0.73	1.09	0.73	1.07	1.01E−01	0.55
82	22q12	rs132390	27951477	*EMID1*	0.036	1.24	0.02	0.90	4.95E−01	0.46
83	22q13	rs6001930	39206180	*MKL1*/SGSM3	0.11	1.15	0.11	1.10	1.39E−01	0.72

[a]Mean minor allele frequency over all European controls in previous GWAs and iCOGs studies.

[b]Mean per-allele OR over all European participants in previous GWAs and iCOGs studies.

[c]Minor allele frequency in controls.

[d]Per allele frequency for the minor allele relative to the major allele.

[e]Probability of obtaining $P < 0.05$ with discovery data.

[f]SNPs in bold represent *P* value < 0.05.

[g]Not included in either the HapMap2 or 1000 Genome imputation sets.

within a 2 Mb region centered at rs889312, the strongest published SNP in the region. We found rs7709971 to have the smallest *P* value (1.01×10^{-9}). Adjusting for this SNP in bivariate regressions did not produce strong new associations for any of the other SNPs in the region (results not shown).

Association with known breast cancer SNPs

Table 2 shows 83 SNPs reported in the GWAS catalog http://www.genome.gov/26525384#1 as associated with breast cancer at $P < 4 \times 10^{-8}$ in studies of predominantly LOBC (28, 31, 32, 34–60). We used the discovery set to examine association between EOBC and the 82 SNPs

that we could impute using HapMap3 and/or 1 KG data. Table 2 shows that 32 SNPs were associated at $P < 0.05$ (listed in bold in the table). We also computed the probability that a test of size 0.05 using 3,523 cases and 2,702 controls would detect association with each of the 82 SNPs, given its published effect size as shown in the table. We found that the mean power to detect the 50 missed SNPs was 44%, appreciably lower than the mean power of 77% for the 32 we detected. Thus our failure to confirm the remaining 50 SNPs seems due to insufficient power to detect their small effect sizes. These results suggest that the genetic etiology of EOBC is not different than that of LOBC.

Gene-based analysis

Analysis of the discovery set identified the phosphofructokinase muscle-type (*PFKM*) gene on chromosome 12q13.11 region as associated with EOBC with *P* value of 9×10^{-7}, which meets the genome-wide threshold of $P < 2.5 \times 10^{-6}$ for the 19,684 statistical tests performed. This region is distinct from the regions 12q22 and 12q24 containing SNPs known to be associated with breast cancer (Table 2). When we repeated the same analysis using the predominantly LOBC breast cancer replication data from the CGEMS study, the *PFKM* gene also was associated with breast cancer ($P = 3 \times 10^{-2}$). Combined analysis of the two data sets yielded an overall gene-based Fisher's meta *P* value of 5×10^{-7} for the *PFKM* gene. No other genes met the genome-wide significance threshold.

The association between *PFKM* and breast cancer risk was based on its annotation with the 35 putative functional SNPs shown in Table 3. This set consists largely of trans e-QTL SNPs rather than MNF SNPs in the coding region of the gene. Nevertheless, we also found evidence implicating SNPs in the 1 M region centered at the *PFKM* gene. We found that 27 of the 966 SNPs in this region that were included in the EOBC GWAS discovery set were associated with EOBC at $P < 0.01$. These SNPs are listed in Fig. 2. Also shown in the figure are the genes in this region (Fig. 2A), a Manhattan plot of the 966 *P* values (Fig. 2B), and the D' measure of linkage disequilibrium between pairs of SNPs (Fig. 2C). To evaluate the statistical significance of this finding, we permuted subjects' case–control statuses 1,000-times, and in each permutation, we evaluated how many of the 966 SNPs were associated with EOBC at $P < 0.01$. We found that none of the 1,000 permutations yielded 27 or more such SNPs, giving a significance level of $P < 0.001$. Most of the 27 EOBC-associated SNPs were located within other nearby genes, suggesting that EOBC risk could be due to some complex gene expression pattern in this gene-rich region (see A and B of Fig. 2). Fig. 2C of the figure shows the correlations among the SNPs in the region.

Table 3. Significance levels from discovery and replication sets for association of EOBC with 35 putatively functional *PFKM* SNPs

	GWAS *P* value		
SNP	Discovery data	Replication data	Combined
rs9895850	1.59E−04	1.35E−03	9.52E−07
rs6892066	3.85E−04	1.21E−01	1.44E−04
rs16959569	5.31E−04	9.70E−04	2.17E−06
rs4462967	1.03E−03	2.80E−02	9.26E−05
rs12190699	6.92E−02	4.23E−02	6.02E−03
rs4242252	2.00E−01	7.11E−02	2.36E−02
rs12442176	2.20E−01	5.38E−01	1.34E−01
rs16881917	3.23E−01	8.08E−01	2.43E−01
rs7096642	3.28E−01	8.83E−01	2.63E−01
rs10091208	3.31E−01	8.38E−01	2.54E−01
rs7006101	3.72E−01	5.95E−01	2.15E−01
rs2377800	4.14E−01	5.11E−01	2.08E−01
rs999450	4.85E−01	6.35E−01	2.74E−01
rs7597958	4.85E−01	4.82E−01	2.24E−01
rs2228500	4.89E−01	4.99E−01	2.32E−01
rs1468195	5.40E−01	8.59E−01	3.66E−01
rs16955826	5.51E−01	6.99E−01	3.22E−01
rs8095381	5.57E−01	4.51E−01	2.37E−01
Rs11114379	5.70E−01	4.99E−01	2.59E−01
rs6999405	5.81E−01	NA[a]	NA[a]
rs1245012	5.86E−01	5.04E−01	2.66E−01
rs7199193	6.01E−01	1.85E−01	1.27E−01
rs11777718	6.76E−01	6.30E−01	3.45E−01
rs12470945	7.00E−01	8.83E−01	4.43E−01
rs1376386	7.45E−01	6.22E−01	3.66E−01
rs12476834	7.50E−01	9.22E−01	4.76E−01
rs9952079	8.14E−01	3.86E−01	2.78E−01
rs17775523	8.14E−01	6.57E−01	4.03E−01
rs17505688	8.38E−01	1.12E−01	1.12E−01
rs1920398	8.50E−01	5.16E−01	3.52E−01
rs8057807	9.08E−01	4.32E−01	3.26E−01
rs7031588	9.36E−01	1.48E−01	1.51E−01
rs12306431	9.43E−01	4.54E−01	3.46E−01
rs17505369	9.61E−01	1.53E−01	1.58E−01
rs6984368	9.98E−01	8.69E−01	5.48E−01

[a]Missing in replication dataset.

Discussion

This study identified and replicated EOBC associations with 72 previously unpublished SNPs in six regions known to harbor variants affecting breast cancer risk. Twelve of the 72 SNPs remained associated with EOBC after adjusting for the SNP with smallest published *P* value in the same region. Eleven of these 12 SNPs lie on chromosome 5q11.2 near the *MAP3K1* gene. Their lack of strong correlation with the strongest published SNP rs889312 suggests the presence of multiple causal variants in this region. Future sequence-based studies, coupled with functional experiments, can exploit these associations to identify the causal variant(s) in the region.

Order	SNPs	P	$-Log_{10}(P)$
1	rs11168142	0.005	2.3
2	rs12830810	0.003	2.5
3	rs7303895	0.001	3.0
4	rs7304743	0.001	3.0
5	rs734706	0.008	2.1
6	rs17122311	0.002	2.7
7	rs875023	0.004	2.4
8	rs2544029	0.006	2.2
9	rs2544030	0.005	2.3
10	rs2070615	0.007	2.2
11	rs1054442	0.003	2.5
12	rs7975791	0.009	2.0
13	rs10875914	0.010	2.0
14	rs11168830	0.008	2.1
15	rs12580349	0.007	2.2
16	rs2241726	0.007	2.2
17	rs7969091	0.001	3.0
18	rs12821008	0.001	3.1
19	rs6580698	0.001	3.0
20	rs6580699	0.001	3.0
21	rs7296288	0.001	3.2
22	rs11168857	0.008	2.1
23	rs3741621	0.008	2.1
24	rs10783302	0.010	2.0
25	rs2279596	0.009	2.0
26	rs10747561	0.009	2.0
27	rs1274726	0.010	2.0

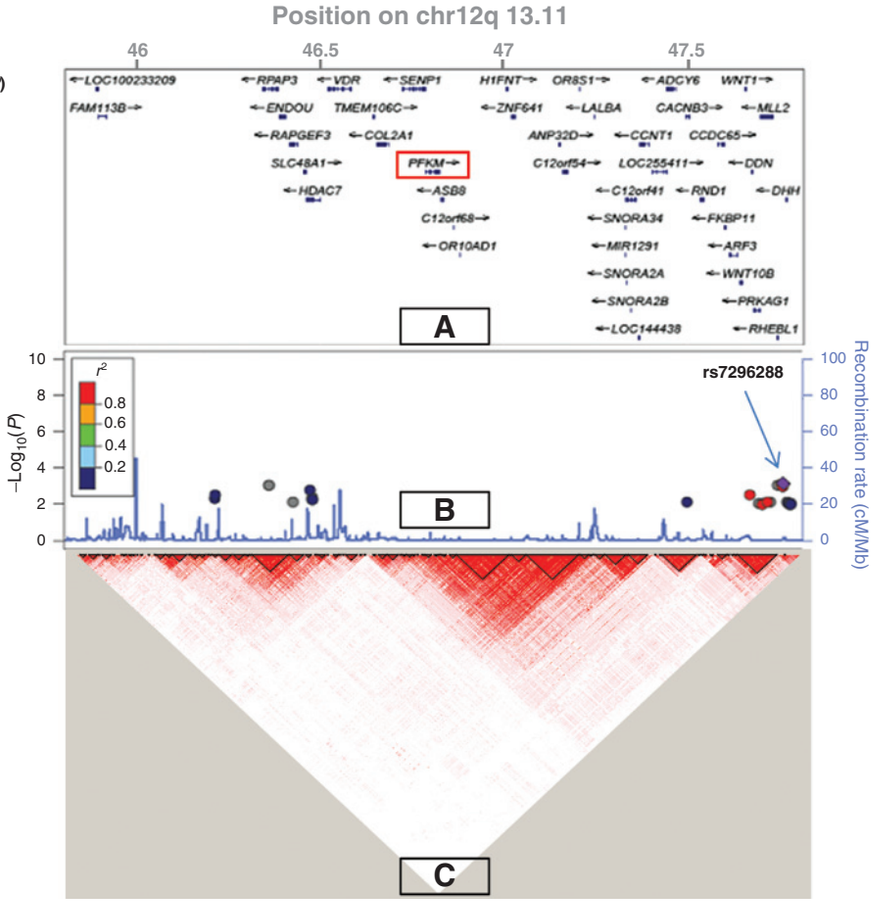

Figure 2. A, chromosomal positions of genes on chr12q13.11 in the 2 Mb region surrounding the *PFKM* gene. B, Manhattan plot of 27 SNPs in the region associated with EOBC with combined discovery and replication *P* values of 0.01 or less. These 27 SNPs and their *P* values are listed on the left in their order of appearance from left to right. SNP colors reflect magnitudes of their squared correlation coefficients with SNP rs7296288 (marked by arrow), which had the smallest discovery set *P* value. C, linkage disequilibrium measures *D′* for all 966 HapMap3 imputed SNPs in the chr12q13.11 region. Dark red squares represent *D′* values near 1 and white squares represent *D′* values near zero.

We examined association between EOBC and 82 of the 83 common SNPs currently known to be associated with largely LOBC. We found evidence for association with only 32 (39%) of these SNPs. However, comparison of detected and missed SNPs with respect to effect size and power suggests that this low confirmation rate reflects the inadequate power to detect the missed SNPs rather than systematic etiologic differences between EOBC and LOBC. These findings suggest that the genetic factors responsible for breast cancer affect risk at all ages.

The gene-based GWAS analyses identified the *PFKM* gene region 12q13.11 to be associated with breast cancer risk, independently of the 12q22 and 12q24 regions previously associated with breast cancer. *PFKM*, one of the three PFK isoenzymes, is the key regulator of cellular glycolysis catalyzing the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate. Disabling *PFKM* mutations leads to glycogen storage diseases (especially type VII—Tarui's disease) as well as cardiac and hematologic disorders (61–63). The association of *PFKM* expression with breast cancer risk is plausible for several reasons. First, this gene is expressed in breast cancer cell lines (64). Second, variants in the gene have been related to posttranslational modifications, which have been shown to alter the metabolism and promote the growth of cancer cells (65). Third, an association between breast cancer risk and this gene is consistent with observations that tumor cells can consume large amounts of glucose due to aberrant glucose metabolism, especially through a glycolytic pathway that produces lactate (65). Finally, tumor-suppressor protein p53 has been shown to suppress *PFKM* expression in model system (66). Because the biology of the *PFKM* gene and its modulators and inhibitors are well characterized (67, 68), identification of *PFKM* gene region as a breast cancer susceptibility locus has potential

translational implications for breast cancer prevention and treatment.

The present study has several strengths, including its large sample size, its focus on EOBC, its homogenous Caucasian study population, and its novel gene-based analysis involving the functional characteristics of gene-related SNPs. Study limitations include use of somewhat different types of study populations between the discovery (population-based) and replication (both population- and clinic-based) phases, and our inability to replicate the gene-based analysis in an EOBC replication set due to lack of access to necessary relevant data from replication cohorts.

In conclusion, the study identified EOBC risks to be associated with 72 new SNPs in six chromosomal regions that were previously associated with LOBC risks. Eleven of the 72 SNPs, all on chromosome 5q11.2, were associated with EOBC independently of previously reported SNPs. These EOBC-associated SNPs may help in the search for causal variants in the 5q11.2 region. In addition, we found little evidence to support genetic heterogeneity between EOBC and LOBC. Finally, the gene-based analysis identified a region containing the key glycolysis regulation gene *PFKM* that is worthy of further investigation as a susceptibility locus for breast cancer in Caucasian women of all ages. Future studies need to determine whether the current findings apply to non-Caucasian women.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The content of this article does not necessarily reflect the views or policies of the NCI or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or the BCFR.

Authors' Contributions

Conception and design: H. Ahsan, M.G. Kibriya, D.C. Thomas, D. Seminara, G.G. Giles, M. Lathrop, R.K. Schmutzler, N. Dahmen, C. Blomqvist, N. Rahman, A.M. Dunning, N.J. Cox, A.S. Whittemore

Development of methodology: J. Halpern, B.L. Pierce, D.C. Thomas, M. Lochmann, N. Rahman, N.J. Cox, A.S. Whittemore

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H. Ahsan, M.G. Kibriya, F. Jasmine, J. Chang-Claude, I. Andrulis, J.L. Hopper, E.M. John, K. Malone, G. Ursin, M.D. Gammon, D. Seminara, G. Casey, J.A. Knight, M.C. Southey, G.G. Giles, R.M. Santella, E. Lee, D. Duggan, R. Haile, M. Jenkins, N.M. Lindor, P. Newcomb, C. Apicella, J. Peto, O. Fletcher, I. dos Santos Silva, M. Lathrop, D.J. Hunter, S.J. Chanock, A. Meindl, R.K. Schmutzler, M. Lochmann, J. Stone, D. Flesch-Janys, N. Dahmen, H. Nevanlinna, K. Aittomäki, P. Hall, K. Czene, J. Liu, C. Turnbull, A.M. Dunning, P. Pharoah, Q. Waisfisz, F. Rivadeneira, D.F. Easton

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H. Ahsan, J. Halpern, M.G. Kibriya, B.L. Pierce, L. Tong, E. Gamazon, J. Shi, F. Jasmine, S. Melkonian, J. Chang-Claude, D.C. Thomas, M.C. Southey, D. Conti, S. Gallinger, K. Michailidou, M. Lathrop, D.J. Hunter, B. Müller-Myhsok, M. Lochmann, L. Beckmann, E. Makalic, D.F. Schmidt, Q.M. Bui, P. Hall, A. Irwanto, Q. Waisfisz, D. Nicolae, D.F. Easton, N.J. Cox, A.S. Whittemore

Writing, review, and/or revision of the manuscript: H. Ahsan, J. Halpern, M.G. Kibriya, B.L. Pierce, E. Gamazon, V. McGuire, M. Argos, J. Chang-Claude, I. Andrulis, J.L. Hopper, E.M. John, K. Malone, G. Ursin, M.D. Gammon, D.C. Thomas, D. Seminara, J.A. Knight, M.C. Southey, G.G. Giles, E. Lee, D. Conti, R. Haile, M. Jenkins, J. Peto, I. dos Santos Silva, R. Hein, H. Nevanlinna, C. Blomqvist, P. Hall, K. Czene, A.M. Dunning, P. Pharoah, H. Meijers-Heijboer, A.G. Uitterlinden, F. Rivadeneira, D.F. Easton, N.J. Cox, A.S. Whittemore

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H. Ahsan, J. Halpern, V. McGuire, A. Felberg, M. Argos, S. Melkonian, E.M. John, D. Seminara, D.J Park, A. Meindl, R.K. Schmutzler, J. Stone, H. Meijers-Heijboer, F. Rivadeneira, A.S. Whittemore

Study supervision: H. Ahsan, M.G. Kibriya, M. Argos, J.L. Hopper, D.J Park, A.S. Whittemore

Other: Running laboratory assay and data generation, S. Roy; laboratory work, R. Brutus

Acknowledgments

The authors are thankful for the use of data from the Database of Genotypes and Phenotypes (dbGaP), and thank all investigators who contributed the phenotype data and DNA samples to the CGEMS project, and the NCI, the primary funder of the CGEMS genome-wide association study.

Grant Support

This study was supported by NIH Grants U01CA122171, RC1CA145506, R01CA094069, and U19CA148065. The Breast Cancer Family Registry (BCFR) is supported by the NCI, NIH under RFA-CA-06-503 and through cooperative agreements with members of the BCFR and principal investigators, including Cancer Care Ontario (U01 CA69467), Cancer Prevention Institute of California (CPIC; U01 CA69417), and University of Melbourne (U01 CA69638). Samples from the CPIC were processed and distributed by the Coriell Institute for Medical Research. The Colon Cancer Family Registry (CCFR) is supported by the NCI, NIH under RFA-CA-95-011.

Received April 2, 2013; revised January 29, 2014; accepted January 29, 2014; published OnlineFirst February 3, 2014.

References

1. Narod SA. Early-onset breast cancer: what do we know about the risk factors?: a countercurrents series. Curr Oncol 2011;18:204–5.
2. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet 2008;371:569–78.
3. Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MR, et al. Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. J Natl Cancer Inst 2003;95:448–57.
4. Whittemore AS, Gong G, Itnyre J. Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: results from three U.S. population-based case-control studies of ovarian cancer. Am J Hum Genet 1997;60:496–504.
5. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J Natl Cancer Inst 1994;86:1600–8.
6. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001;358:1389–99.
7. Lee JS, John EM, McGuire V, Felberg A, Ostrow KL, DiCioccio RA, et al. Breast and ovarian cancer in relatives of cancer patients, with and without BRCA mutations. Cancer Epidemiol Biomarkers Prev 2006;15:359–63.
8. Hindorff LA, MacArthur J, Morales J, Junkins HA, Hall PN, Klemm AK, et al. A catalog of published genome-wide association studies. [accessed Sept 2012]. Available from: www.genome.gov/gwastudies.
9. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the

genetic epidemiology of breast cancer. Breast Cancer Res 2004;6: R375–89.
10. Chang-Claude J, Eby N, Kiechle M, Bastert G, Becher H. Breastfeeding and breast cancer risk by age 50 among women in Germany. Cancer Causes Control 2000;11:687–95.
11. Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, et al. The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. Breast Cancer Res Treat 2002;74:235–54.
12. Friedrichsen DM, Malone KE, Doody DR, Daling JR, Ostrander EA. Frequency of CHEK2 mutations in a population-based case-control study of breast cancer in young women. Breast Cancer Res 2004;6: R629–35.
13. Lee E, Ma H, McKean-Cowdin R, Van Den Berg D, Bernstein L, Henderson BE, et al. Effect of reproductive factors and oral contraceptives on breast cancer risk in BRCA1/2 mutation carriers and noncarriers: results from a population-based study. Cancer Epidemiol Biomarkers Prev 2008;17:3170–8.
14. Newcomb PA, Baron J, Cotterchio M, Gallinger S, Grove J, Haile R, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomarkers Prev 2007;16:2331–43.
15. Figueiredo JC, Lewinger JP, Song C, Campbell PT, Conti DV, Edlund CK , et al. Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. Cancer Epidemiol Biomarkers Prev 2011;20: 758–66.
16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 2007;81:559–75.
17. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
18. Browning BL, Browning SR. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. Genet Epidemiol 2007;31:365–75.
19. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997–1004.
20. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010;6:e1000888.
21. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, et al. SCAN: SNP and copy number annotation. Bioinformatics 2010;26:259–62.
22. Duan S, Huang RS, Zhang W, Bleibel WK, Roe CA, Clark TA, et al. Genetic architecture of transcript-level variation in humans. Am J Hum Genet 2008;82:1101–13.
23. De la Cruz O, Wen X, Ke B, Song M, Nicolae DL. Gene, region and pathway level analyses in whole-genome studies. Genet Epidemiol 2010;34:222–31.
24. Osborne RH, Hopper JL, Kirk JA, Chenevix-Trench G, Thorne HJ, Sambrook JF. kConFab: a research resource of Australasian breast cancer families. Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer. Med J Aust 2000;172:463–4.
25. Fletcher O, Johnson N, Palles C, dos Santos Silva I, McCormack V, Whittaker J, et al. Inconsistent association between the STK15 F31I genetic polymorphism and breast cancer risk. J Natl Cancer Inst 2006;98:1014–8.
26. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol 2006;35:34–41.
27. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.
28. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 2010;42:504–7.
29. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 2010;11:134.
30. Cochran BG. The combination of estimates from different experiments. Biometrics 1954;10:101–29.
31. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4.
32. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41:579–84.
33. Fisher RA. Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd; 1925.
34. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys M, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 2009;41:585–90.
35. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93.
36. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Vilae A, Garber J, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A 2008; 105:4340–5.
37. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ, et al. FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. Carcinogenesis 2010;31:1417–23.
38. Boyarskikh UA, Zarubina NA, Biltueva JA, Sinkina TV, Voronina EN, Lazarev AF, et al. Association of FGFR2 gene polymorphisms with the risk of breast cancer in population of West Siberia. Eur J Hum Genet 2009;17:1688–91.
39. Jia C, Cai Y, Ma Y, Fu D. Quantitative assessment of the effect of FGFR2 gene polymorphism on the risk of breast cancer. Breast Cancer Res Treat 2010;124:521–8.
40. Raskin L, Pinchev M, Arad C, Lejbkowicz F, Tamir A, Rennert HS, et al. FGFR2 is a breast cancer susceptibility gene in Jewish and Arab Israeli populations. Cancer Epidemiol Biomarkers Prev 2008;17:1060–5.
41. Li J, Humphreys K, Darabi H, Rosin G, Hannelius U, Heikkinen T, et al. A genome-wide association scan on estrogen receptor-negative breast cancer. Breast Cancer Res 2010;12:R93.
42. Long J, Cai Q, Shu XO, Qu S, Li C, Zheng Y, et al. Evaluation of breast cancer susceptibility loci in Chinese women. Cancer Epidemiol Biomarkers Prev 2010;19:2357–65.
43. Reeves GK, Travis RC, Green J, Bull D, Tipper S, Baker K, et al. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. JAMA 2010;304: 426–34.
44. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J Natl Cancer Inst 2011;103:425–35.
45. Udler MS, Ahmed S, Healey CS, Meyer K, Struewing J, Maranian M, et al. Fine scale mapping of the breast cancer 16q12 locus. Hum Mol Genet 2010;19:2507–15.
46. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility loci at 6q25.1. Nat Genet 2008;41:324–8.
47. Liang J, Chen P, Hu Z, Zhou X, Chen L, Li M, et al. Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. Carcinogenesis 2008;29:2341–6.
48. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson A, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39:865–9.
49. Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. PLoS Genet 2010;6:e1001183.
50. Long J, Cai Q, Shu XO, Qu S, Li C, Zheng Y, et al. Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. PLoS Genet 2010;6: e1001002.

51. Udler MS, Meyer KB, Pooley KA, Karlins KA, Struewing J, Zhang J, et al. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. Hum Mol Genet 2009;18:1692–703.
52. Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Hum Mol Genet 2011;20:4991–9.
53. Li J, Humphreys K, Heikkinen T, Aittomaki K, Blomqvist C, Pharoah PDP, et al. A combined analysis of genome-wide association studies in breast cancer. Breast Cancer Res Treat 2011;126:717–27.
54. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson S, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2008;40:703–6.
55. Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, et al. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. PLoS Genet 2010;6: e1001029.
56. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Hum Mol Genet 2012;21:5373–84.
57. Chen F, Chen GK, Stram DO, Millikan RC, Ambrosone CB, John EM, et al. A genome-wide association study of breast cancer in women of African ancestry. Hum Genet 2013;132:39–48.
58. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013;45: 392–8.
59. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet 2013;45:371–84.
60. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45:353–61.
61. Tarui S, Okuno G, Ikura Y, Tanaka T, Suda M, Nishikawa M. Phosphofructokinase deficiency in skeletal muscle. A new type of Glycogenosis. Biochem Biophys Res Com 1965;19:517–23.
62. Vasconcelos O, Sivakumar K, Dalakas MC, Quezado M, Nagle J, Leon-Monzon M, et al. Nonsense mutation in the phosphofructokinase muscle subunit gene associated with retention of intron 10 in one of the isolated transcripts in Ashkenazi Jewish patients with Tarui disease. Proc Natl Acad Sci U S A 1995;92:10322–6.
63. Garcia M, Pujol A, Ruzo A, Riu E, Ruberte J, Arbos A, et al. Phosphofructo-1-kinase deficiency leads to a severe cardiac and hematological disorder in addition to skeletal muscle glycogenosis. PLoS Genet 2009;5:e1000615.
64. Zancan P, Sola-Penna M, Furtado CM, Da Silva D. Differential expression of phosphofructokinase-1 isoforms correlates with the glycolytic efficiency of breast cancer cells. Mol Genet Metab 2010;100:372–8.
65. Smerc A, Sodja E, Legisa M. Posttranslational modification of 6-phosphofructo-1-kinase as an important feature of cancer metabolism. PLoS ONE 2011;6:e19645.
66. Danilova N, Kumagai A, Lin J. p53 upregulation is a frequent response to deficiency of cell-essential genes. PLoS ONE 2010;5: e15938.
67. Deng H, Yu F, Chen J, Zhao Y, Xiang J, Lin A. Phosphorylation of Bad at Thr-201 by JNK1 promotes glycolysis through activation of phosphofructokinase-1. J Biol Chem 2008;283:20754–60.
68. Usenik A, Legisa M. Evolution of allosteric citrate binding sites on 6-phosphofructo-1-kinase. PLoS ONE 2010;5:e15447.