DOI: 10.1093/jnci/djq057
Advance Access publication on March 22, 2010.

ARTICLE

Association Between a Germline *OCA2* Polymorphism at Chromosome 15q13.1 and Estrogen Receptor–Negative Breast Cancer Survival

Elizabeth M. Azzato, Jonathan Tyrer, Peter A. Fasching, Matthias W. Beckmann, Arif B. Ekici, Rüdiger Schulz-Wendtland, Stig E. Bojesen, Børge G. Nordestgaard, Henrik Flyger, Roger L. Milne, José Ignacio Arias, Primitiva Menéndez, Javier Benítez, Jenny Chang-Claude, Rebecca Hein, Shan Wang-Gohrke, Heli Nevanlinna, Tuomas Heikkinen, Kristiina Aittomäki, Carl Blomqvist, Sara Margolin, Arto Mannermaa, Veli-Matti Kosma, Vesa Kataja, The Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer, Jonathan Beesley, Xiaoqing Chen, Georgia Chenevix-Trench, Fergus J. Couch, Janet E. Olson, Zachary S. Fredericksen, Xianshu Wang, Graham G. Giles, Gianluca Severi, Laura Baglietto, Melissa C. Southey, Peter Devilee, Rob A. E. M. Tollenaar, Caroline Seynaeve, Montserrat García-Closas, Jolanta Lissowska, Mark E. Sherman, Kelly L. Bolton, Per Hall, Kamila Czene, Angela Cox, Ian W. Brock, Graeme C. Elliott, Malcolm W. R. Reed, David Greenberg, Hoda Anton-Culver, Argyrios Ziogas, Manjeet Humphreys, Douglas F. Easton, Neil E. Caporaso, Paul D. P. Pharoah

Manuscript received July 29, 2009; revised February 1, 2010; accepted February 15, 2010.

Correspondence to: Elizabeth M. Azzato, MPH, 6120 Executive Plaza, Rm 7106, Rockville, MD 20852 (e-mail: azzatoe2@mail.nih.gov).

Background Traditional prognostic factors for survival and treatment response of patients with breast cancer do not fully account for observed survival variation. We used available genotype data from a previously conducted two-stage, breast cancer susceptibility genome-wide association study (ie, Studies of Epidemiology and Risk factors in Cancer Heredity [SEARCH]) to investigate associations between variation in germline DNA and overall survival.

Methods We evaluated possible associations between overall survival after a breast cancer diagnosis and 10 621 germline single-nucleotide polymorphisms (SNPs) from up to 3761 patients with invasive breast cancer (including 647 deaths and 26 978 person-years at risk) that were genotyped previously in the SEARCH study with high-density oligonucleotide microarrays (ie, hypothesis-generating set). Associations with all-cause mortality were assessed for each SNP by use of Cox regression analysis, generating a per rare allele hazard ratio (HR). To validate putative associations, we used patient genotype information that had been obtained with 5′ nuclease assay or mass spectrometry and overall survival information for up to 14 096 patients with invasive breast cancer (including 2303 deaths and 70 019 person-years at risk) from 15 international case–control studies (ie, validation set). Fixed-effects meta-analysis was used to generate an overall effect estimate in the validation dataset and in combined SEARCH and validation datasets. All statistical tests were two-sided.

Results In the hypothesis-generating dataset, SNP rs4778137 (C>G) of the *OCA2* gene at 15q13.1 was statistically significantly associated with overall survival among patients with estrogen receptor–negative tumors, with the rare G allele being associated with increased overall survival (HR of death per rare allele carried = 0.56, 95% confidence interval [CI] = 0.41 to 0.75, $P = 9.2 \times 10^{-5}$). This association was also observed in the validation dataset (HR of death per rare allele carried = 0.88, 95% CI = 0.78 to 0.99, $P = .03$) and in the combined dataset (HR of death per rare allele carried = 0.82, 95% CI = 0.73 to 0.92, $P = 5 \times 10^{-4}$).

Conclusion The rare G allele of the *OCA2* polymorphism, rs4778137, may be associated with improved overall survival among patients with estrogen receptor–negative breast cancer.

J Natl Cancer Inst 2010;102:650–662

Breast cancer is the most common malignancy in women (1). Overall, prognosis is generally good, with a 5-year breast cancer–specific survival rate of more than 80% in England and Wales (http://www.statistics.gov.uk); however, breast cancer survival can vary considerably. This variation can be partially explained by established prognostic and predictive indicators, which include clinical stage at diagnosis (that is based on tumor size, lymph node status, and presence of metastasis) and related tumor characteristics, such as histopathologic grade and hormone receptor status (2). We (3) and others (4) have hypothesized that germline genetic variation might provide additional prognostic information by contributing to both tumor and host heterogeneity.

Various evidence support the role of inherited factors in breast cancer prognosis. For example, in mice, mammary tumor progression differs according to strain (5), and results from murine models have implicated germline polymorphisms as potential markers of metastasis risk and prognosis (6). A Swedish population-based cohort study (7) reported evidence of heritability of breast cancer–specific mortality. Previous research by our group (8–13) and others (14–16) has identified additional common germline genetic polymorphisms that are associated with breast cancer–specific and/or overall survival. Furthermore, a functional, homozygous common missense polymorphism of *NQO1*2* (rs1800566) that disables NQO1 protein activity has been associated with breast cancer prognosis and also with response to anthracycline therapy (17). However, these studies have focused on candidate polymorphisms and genes, which are often chosen initially as candidate breast cancer susceptibility genes. This approach is limited by our incomplete knowledge of breast tumor biology and often ignores genetic variants that are not implicated in breast cancer susceptibility.

Identification of novel common germline genetic markers of breast cancer prognosis has the potential to help to elucidate mechanisms of tumor progression and metastasis and the role of the genome in tumor characteristics (such as tumor grade and hormone receptor status) and to improve understanding of tumor–host interactions and the host immune response. This research also has the potential to identify markers for risk of metastasis and relapse, determinants of treatment response, and indicators for the targeting of unique therapies (3).

We recently performed a two-stage genome-wide association study of breast cancer susceptibility among 4398 case patients with breast cancer and 4316 control subjects. In the first stage, we used a panel of 266 722 single-nucleotide polymorphisms (SNPs) selected to tag the majority of known common SNPs in the human genome. In the second stage, we selected 12 711 SNPs on the basis of the statistical significance of the difference in genotype frequency between case patients and control subjects. This second stage was then followed by a third stage that evaluated the 30 most statistically significant SNPs in 21 860 case patients and 22 578 control subjects from 22 studies (18). In that study, we identified SNPs at five loci (10q26.13, 16q12.1, 5q11.2, 11p15.5, and 8q24.3) that were associated with risk of breast cancer.

In this study, we investigated whether common germline genetic variants (SNPs with a minor allele frequency of >5%) were associated with overall survival after a diagnosis of breast cancer. We included 10 621 of the 12 711 SNPs that were genotyped for patients in stage II of the genome-wide association study (18), hypothesizing that these SNPs, some of which may be associated with susceptibility, may also be associated with tumor behavior and prognosis. We used available follow-up and genotype information for 3761 patients with invasive breast cancer who were participating in the Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) breast cancer study, a population-based cohort study (19) as the hypothesis-generating dataset. The most statistically significant associations were then analyzed in a validation dataset including 14 096 patients with breast cancer from 15 international case–control studies (10,17,20–41) that participated in the Breast Cancer Association Consortium (BCAC).

CONTEXT AND CAVEATS

Prior knowledge

Traditional prognostic factors for survival and treatment response of patients with breast cancer do not fully account for observed variations in survival.

Study design

Genotype data from a previously conducted, two-stage, breast cancer susceptibility genome-wide association study were used as the hypothesis-generating dataset to identify additional genes that were associated with overall survival. Data from 15 international case–control studies of breast cancer patients were used as the validation dataset.

Contribution

A single-nucleotide polymorphism of the *OCA2* gene was found in the hypothesis-generating dataset to be statistically significantly associated with overall survival among patients with estrogen receptor–negative tumors, with the rare G allele being associated with increased overall survival. This association was also observed in the validation dataset and in the combined hypothesis-generating and validation dataset.

Implications

The rare G allele of the *OCA2* gene may be associated with improved survival among patients with estrogen receptor–negative breast cancer.

Limitations

The observed association among patients with estrogen receptor–negative tumors did not reach nominal genome-wide statistical significance. A false-positive association caused by confounding cannot be ruled out. Treatment data were not available for most patients in this study. The power to detect associations was modest.

From the Editors

Patients, Materials, and Methods

Patients With Breast Cancer

Hypothesis-Generating Dataset. The hypothesis-generating set was a convenience sample of breast cancer patients from the SEARCH study who had both follow-up data and genotype data from the second stage of the breast cancer susceptibility genome-wide association study. The SEARCH breast cancer study (19) is an ongoing population-based study of women who were diagnosed with breast cancer in the region of England included in the Eastern Cancer Registration and Information Centre (formerly the East Anglian Cancer Registry). The study started on July 1, 1996. Eligible participants included women who were diagnosed with invasive breast cancer and who were either 1) younger than 70 years at the beginning of the study (ie, patients considered incident cases of breast cancer) or 2) aged 55 years or younger on January 1, 1991, and alive at the beginning of this study (ie, patients considered prevalent cases of breast cancer). Because of boundary changes, some cases of breast cancer that were diagnosed before 1995 were identified in the North Thames Cancer Registry. The Eastern

Cancer Registration and Information Centre and the North Thames Cancer Registry have active follow-up at years 3 and 5 after diagnosis and then at 5-year intervals thereafter. All participants in the study provided informed consent, and the study was approved by the Eastern Multicentre Research Ethics Committee.

Follow-up and vital status information was obtained by searching hospital information systems for recent visits. If a patient did not have a recent visit, the patient's general practitioner was contacted to obtain the vital status. Flagging of death certificates through the Office of National Statistics also provided the registries with notification of deaths (ie, all-cause mortality). All patients who were still alive at the end of the study were censored on November 30, 2006. Breast cancer–specific mortality was defined as a death for which breast cancer was listed as the cause of death on the death certificate. All-cause mortality was chosen as the endpoint of interest because most validation studies (see below) had information only on all-cause mortality. Breast cancer–specific mortality in the SEARCH hypothesis-generating dataset was used to confirm that effect sizes of breast cancer mortality were consistent for SNPs chosen for further study in the validation dataset. TNM (42) stage (I, II, III, or IV), which is based on tumor size, number of positive lymph nodes, and the presence of distant metastasis (42), and histopathologic grade (well differentiated, moderately differentiated, or poorly differentiated) were obtained through Eastern Cancer Registration and Information Centre. Estrogen receptor (ER) status was determined on paraffin-embedded breast tumor sections by immunohistochemistry with monoclonal antibodies against the ER (Novocastra clone 6F11; Leica Microsystems Ltd, Milton Keynes, UK). The Allred system (43) was used for scoring, with scores of greater than 2 being considered positive. We included 3761 patients with invasive breast cancer from the SEARCH study who had both follow-up information and genotyping data from the second stage of the breast cancer susceptibility genome-wide association study (18) (Table 1). Genotype was determined by use of genomic DNA isolated from blood. Patients in the SEARCH dataset provided 26978 person-years at risk; there were 647 deaths from any cause, of which 518 were breast cancer–specific deaths, within 15 years after diagnosis. Mean age at diagnosis was 52 years (standard deviation [SD] = 8.7 years). Stage data were available for 3683 (98%) of the 3761 SEARCH patients, histopathologic grade was available for 3033 (81%), and ER status was available for 2480 (66%). More than 99% of the patients in the SEARCH dataset were of European ancestry. Information on surgery, chemotherapy, endocrine therapy, and radiation therapy was available for 3702 patients (97%). Of those with available information, 3378 (91%) underwent surgery, 2674 (72%) received radiotherapy, 772 (21%) received combined chemoendocrine therapy, 1811 (50%) received endocrine therapy only, and 430 (12%) received chemotherapy only.

Table 1. Characteristics of patients with available follow-up information in the Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) study who were included in stage II of the breast cancer susceptibility genome-wide association study*

Variable	Value
Total No. of patients	3761
Total time at risk, person-years	26978.54
Median follow-up† (range), y	8.45 (0.47–15)
Median time at risk (range), y	7.45 (0.04–11.48)
Median time from diagnosis to study entry (range), y	1.15 (0–11.74)
No. of deaths	647
Annual mortality rate (95% CI)	0.024 (0.022 to 0.026)
5-y overall survival rate (95% CI)	0.88 (0.87 to 0.89)
Mean age at diagnosis (range), y	52 (23–69)
Age at diagnosis, No. (%)	
<35 y	122 (3.24)
35–49 y	1331 (35.39)
50–64 y	2030 (53.98)
65–74 y	278 (7.39)
Histopathologic grade, No. (%)	
Well differentiated	729 (19.38)
Moderately differentiated	1433 (38.10)
Poorly differentiated	871 (23.16)
Unknown	728 (19.36)
Morphological type, No. (%)	
Ductal	2777 (73.84)
Lobular	564 (15.00)
Other	386 (10.26)
Unknown	34 (0.90)
Clinical stage, No. (%)	
I	1850 (49.19)
II	1677 (44.59)
III or IV	156 (4.15)
Missing	78 (2.07)
ER status, No. (%)	
Positive	1982 (52.70)
Negative	498 (13.24)
Missing	1281 (34.06)

* CI = confidence interval; ER = estrogen receptor.

† Follow-up censored at 15 years.

Validation Dataset. For the validation set, we identified 14096 patients with invasive breast cancer through 15 case–control studies (10,17,20–41) from Europe, North America, and Australia, whose research groups participated in the BCAC (Supplementary Table 1, available online). All BCAC studies with available patient follow-up information were used in this analysis. Recruitment for these studies began as early as 1990; some studies are still recruiting patients (Supplementary Table 1, available online). All 15 studies had information on disease status, vital status (all-cause mortality), length of follow-up, time from diagnosis to study entry (if applicable), and age at diagnosis and/or enrollment. Patients in the BCAC validation dataset provided 70019 person-years at risk; there were 2303 deaths from any cause within 15 years after diagnosis. More than 99% of patients in the validation set were of European origin for all studies, except for those in the University of California Irvine Breast Cancer Study (UCIBCS), in which 87% were of European descent, 4% were of Asian descent, and 9% were of other descent. Where available, studies provided information on tumor grade, TNM stage, and ER status. The mean age at diagnosis was 58 years (SD = 12.2 years). Stage data were available for 10046 (71%) of the 14096 patients, histopathologic grade was available for 11767 (84%), and ER status was available for 11786 (84%). Characteristics of patients by study are in Table 2.

Patients in the BCAC validation dataset tended to be older (mean age = 58 years, SD = 12.2 years) than those in the SEARCH

Table 2. Tumor and clinical characteristics for the 15 studies in the validation set*

Study	ER status, No. (%)			Stage, No. (%)				Grade†, No. (%)			
	Negative	Positive	Missing	I	II	III or IV	Missing	1	2	3	Missing
BBCC	260 (21.70)	704 (58.76)	234 (19.53)	517 (43.16)	328 (27.38)	83 (6.93)	270 (22.54)	107 (8.93)	671 (56.01)	301 (25.13)	119 (9.93)
CGPS	313 (16.40)	1443 (75.59)	153 (8.01)	993 (52.02)	673 (35.25)	68 (3.56)	175 (9.17)	426 (22.32)	818 (42.85)	310 (16.24)	355 (18.60)
CNIO-BCS	32 (15.24)	93 (44.29)	85 (40.48)	109 (51.90)	66 (31.43)	2 (0.95)	33 (15.71)	47 (22.38)	61 (29.05)	43 (20.48)	59 (28.10)
GESBC	157 (28.65)	257 (46.90)	134 (24.45)	223 (40.69)	232 (42.34)	62 (11.31)	31 (5.66)	42 (7.66)	238 (43.43)	188 (34.31)	80 (14.60)
HEBCS	402 (18.21)	1722 (78.02)	83 (3.76)	1326 (60.08)	700 (31.72)	158 (7.16)	23 (1.04)	559 (25.33)	937 (42.46)	605 (27.41)	106 (4.08)
KARBAC	76 (16.00)	370 (77.89)	29 (6.11)	265 (55.79)	158 (33.26)	46 (9.68)	6 (1.26)	115 (24.12)	190 (40.00)	104 (21.89)	66 (13.89)
KBCP	100 (22.27)	320 (71.27)	29 (6.46)	185 (41.20)	235 (52.34)	15 (3.34)	14 (3.12)	109 (24.28)	193 (42.98)	114 (25.39)	33 (7.35)
kConFab/AOCS	50 (19.23)	114 (43.85)	96 (36.92)	N/A	N/A	N/A	299 (100)	49 (18.85)	87 (33.46)	69 (26.54)	55 (21.15)
MCBCS	140 (14.61)	714 (74.53)	104 (10.86)	259 (27.04)	119 (12.42)	41 (4.28)	539 (56.26)	211 (22.03)	300 (31.32)	179 (18.68)	268 (27.97)
MCCS	149 (22.41)	449 (67.52)	67 (10.08)	333 (50.08)	119 (17.89)	9 (1.35)	204 (30.68)	140 (21.05)	273 (41.05)	193 (29.02)	59 (8.87)
ORIGO	88 (18.14)	269 (55.46)	128 (26.39)	191 (39.38)	232 (47.84)	39 (8.04)	23 (4.74)	81 (16.70)	176 (36.29)	120 (24.74)	108 (22.27)
PBCS	524 (29.62)	1011 (57.15)	234 (13.23)	N/A	N/A	N/A	780 (100)	328 (18.54)	777 (43.92)	386 (21.82)	278 (15.72)
SASBAC	155 (12.63)	700 (57.05)	372 (30.32)	703 (57.29)	501 (40.83)	19 (1.55)	4 (0.33)	127 (10.35)	390 (31.78)	322 (26.24)	388 (31.62)
SBCS	107 (16.04)	296 (44.38)	264 (39.58)	N/A	N/A	N/A	682 (100)	125 (18.74)	287 (43.03)	207 (31.03)	48 (7.20)
UCIBCS	165 (15.43)	606 (56.69)	298 (27.88)	730 (68.29)	261 (24.42)	46 (4.30)	32 (2.99)	124 (11.60)	353 (33.02)	285 (26.66)	307 (28.72)
TOTAL	2718 (19.28)	9068 (64.33)	2310 (16.39)	5834 (41.39)	3624 (25.71)	588 (4.17)	4050 (28.73)	2590 (18.37)	5751 (40.80)	3426 (24.30)	2329 (16.52)

* BBCC = Bavarian Breast Cancer Cases and Controls; CGPS = Copenhagen Breast Cancer Study and Copenhagen General Population Study; CNIO-BCS = Spanish National Cancer Centre Breast Cancer Study; ER = estrogen receptor; GESBC = Genetic Epidemiology Study of Breast Cancer by Age 50; HEBCS = Helsinki Breast Cancer Study; KARBAC = Karolinska Breast Cancer Study; KBCP = Kuopio Breast Cancer Project; kConFab/AOCS = The Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer/Australian Ovarian Cancer Study; MCBCS = Mayo Clinic Breast Cancer Study; MCCS = Melbourne Collaborative Cohort Study; N/A = not available; ORIGO = Leiden University Medical Centre Breast Cancer Study; PBCS = Polish Breast Cancer Study; SASBAC = Singapore and Swedish Breast Cancer Study; SBCS = Sheffield Breast Cancer Study; UCIBCS = University of California Irvine Breast Cancer Study.

† Grade designations of 1, 2, and 3 represent well differentiated, moderately differentiated, and poorly differentiated, respectively.

hypothesis-generating set (mean age = 52 years, SD = 8.7 years). Among those with available information on grade and ER status, patient distributions were similar in the BCAC validation dataset and in the SEARCH hypothesis-generating dataset for histopathologic grade (well differentiated = 22% vs 24%, respectively; moderately differentiated = 49% vs 47%, respectively; and poorly differentiated = 29% vs 29%, respectively) and ER status (ER positive = 77% vs 80%, respectively, and ER negative = 23% vs 20%, respectively). Among those with known clinical stage information, a higher proportion of patients in the BCAC validation set than in the SEARCH hypothesis-generating set were TNM stage I (58% vs 50%, respectively) and TNM stages III and IV (6% vs 4%, respectively) and a lower proportion of those in the BCAC validation set were TNM stage II (36% vs 45%, respectively).

Methods for Genotyping

Genotyping procedures for the breast cancer susceptibility genome-wide association study have been previously described (18). Briefly, in stage I, 390 patients with invasive breast cancer who had a family history of breast cancer were identified through UK clinical genetics centers and a national study of bilateral breast cancer (18) and 364 control subjects from the EPIC-Norfolk study, a population-based cohort study of diet and cancer that was based in Norfolk, East Anglia, UK (44), were genotyped for a genome-wide panel of 266 722 SNPs by use of high-density oligonucleotide photolithographic arrays at Perlegen Sciences (Mountain View, CA) using genomic DNA obtained from blood samples. In stage II, 3990 case patients from the SEARCH study (19) and 3916 control subjects from the EPIC-Norfolk study (44) were genotyped for a set of 12 711 SNPs that were selected on the basis of the statistical significance of the difference in genotype frequency between case patients and control subjects ($P < .052$) and the number of SNPs that could fit on the chip. SEARCH samples were genotyped by use of 2.5 μg of genomic DNA isolated from blood and a custom-designed oligonucleotide array (information is available from the authors on request). Because follow-up information is not available for stage I of the susceptibility study, only patient and genotyping information from stage II were included in this analysis. We excluded SNPs with a call rate of less than 95% or those with a genotype frequency that deviated from Hardy–Weinberg equilibrium at a P value of less than 10^{-5} for control subjects, leaving 10 621 SNPs for this analysis (Supplementary Table 2, available online).

For the BCAC validation studies, either a 5′-nuclease assay (Taqman; Applied Biosystems, Inc, Foster City, CA) or matrix-assisted laser desorption–ionization time-of-flight mass spectrometry (iPLEX; Sequenom, Inc, San Diego, CA) was used for genotyping the two SNPs, rs6626269 and rs4778137 (Supplementary Table 1, available online), using genomic DNA isolated from blood samples. Both Taqman and iPLEX technologies require an initial polymerase chain reaction step. In Taqman assays, a perfectly matching allele-specific probe, labeled with a fluorophore, is displaced and cleaved by *Taq* polymerase and genotype is determined by fluorescence detection. Genotype calls for iPLEX are based on determining the mass of primer extension products, which are designed to differ substantially for each allele, by matrix-assisted laser desorption–ionization time-of-flight mass spectrometry.

All genotyping centers for studies in the validation dataset genotyped a HAPMAPPT01 Coriell plate (Coriell Institute for Medical Research, Camden, NJ) that included 90 samples from 30 trios from the International HapMap Project, Centre d'Etude du Polymorphisme Humain (CEPH) population (Utah residents with ancestry from northern and western Europe) with 100% concordance. Independent validation of SEARCH genotyping (oligonucleotide array) was not carried out. All genotyping for the SEARCH and validation studies was blinded to patient outcome (all-cause mortality). We removed any SNP–study combination that had a call rate of 95% or less (after removing any plate with a call rate of <90% and any samples that could not be genotyped for 20% of the SNPs in the BCAC collaborative studies). We also eliminated all data for a given SNP–study combination in which the reproducibility for duplicate samples was less than 97% or in which there was a marked deviation from the Hardy–Weinberg equilibrium (ie, $P < 10^{-5}$).

Statistical Methods

Associations with all-cause mortality were assessed for each SNP by use of Cox regression analysis that modeled the time from diagnosis to death. To account for variable time from diagnosis to recruitment in some studies, analyses allowing for left truncated data (ie, prevalent cases of breast cancer) were conducted in which patients were considered at risk only after the date of study entry. This method generates an unbiased estimate of the association, provided that the proportional hazards assumption has not been violated (45). The proportional hazards assumption was evaluated by visual inspection of log–log plots and was tested analytically by use of Schoenfeld residuals. Follow-up was censored at the earlier of the date last known to be alive or 15 years after diagnosis because most individuals (>97%) had a follow-up of at most 15 years.

A per-allele hazard ratio (HR) of death was estimated for each SNP (in the log-additive codominant model), which was based on the number of rare alleles carried; statistical significance was assessed with a trend test with 1 *df*. SNPs were chosen for validation on the basis of two criteria: 1) a *P* value of less than 5×10^{-8} (the genome-wide level of statistical significance) and 2) comparing the distribution of observed trend test χ^2 values with expected values in a normal χ^2 distribution (quantile–quantile [Q–Q] plot). SNPs with higher χ^2 values observed than expected by the distribution were selected for validation study. Associations between each SNP selected for validation study and breast cancer–specific mortality were assessed in the SEARCH hypothesis-generating dataset by use of Cox regression analysis, modeling the time from diagnosis to death from breast cancer. These additional analyses were performed to confirm that effect sizes associated with breast cancer mortality were consistent with associations with all-cause mortality to reduce the possibility of spurious associations with competing mortality. Hazard ratios were estimated from log-additive models that were unstratified or stratified by major prognostic factors (including stage, histopathologic grade, and ER status) to assess heterogeneity. A test for heterogeneity was performed by generation of a *z* statistic that was calculated as the difference in stratified hazard ratios for each prognostic factor divided by the square root of the sum of their squared standard errors. A test for the statistical interaction between an SNP and a prognostic factor (effect beyond additive) was performed by inclusion of a SNP–prognostic factor cross-product term in the Cox model and assessed by use of a likelihood ratio test with 1 *df*.

In multivariable models, ER status was modeled as a dichotomous variable (positive or negative) and age at diagnosis was modeled as a categorical variable (<35, 35–49, 50–64, 65–74, and >75 years). Grade and stage were modeled as ordinal variables. Because stage, grade, and ER status have been shown to have substantial deviation from the proportional hazards assumption in the SEARCH breast cancer study (45), adjusted analyses were conducted by including age as a covariate and stage, histopathologic grade, and ER status as strata.

For the BCAC validation set, Cox proportional hazards models, modeling time from diagnosis until death from any cause, were used to generate study-specific hazard ratios of death for selected SNPs. For the Polish Breast Cancer Study only, the Cox model was stratified by study site (Warsaw or Lodz). Fixed-effects meta-analysis was used to generate an overall estimate of effect for the validation dataset and for the combined dataset containing the SEARCH hypothesis-generating set and the BCAC validation set. Heterogeneity (ie, the consistency of HR estimates across studies) was assessed by use of the I^2 statistic (46), which describes the percentage of total variation across studies that is due to heterogeneity rather than chance. The I^2 value lies between 0% and 100%, with a value of 0% indicating no observed heterogeneity and with larger values indicating increasing heterogeneity. For adjusted analyses, some validation studies did not have sufficient numbers of patients with covariate information to generate study-specific hazard ratios. Therefore, for adjusted analyses, data were pooled for all validation studies and a Cox proportional hazards model that was stratified by study was used to generate an overall estimate of effect in the validation set and in the combined dataset (ie, the hypothesis-generating dataset plus the validation dataset). Cumulative overall survival by genotype was determined by a Kaplan–Meier analysis. Predicted cumulative survival curves that were generated by genotype and adjusted for study were estimated by adjusting to the baseline hazard function of the SEARCH study.

Possible bias from population stratification was evaluated for rs4778137 by plotting study-specific minor allele frequencies, weighted by the total person-years at risk provided by each study, against study-specific annual mortality rates and assessed by a test of pairwise correlation. Because study-specific allele frequencies must track with annual mortality rates for an association to show substantial bias from population stratification, a lack of correlation between these factors indicates that population stratification is not a major threat to the validity of conclusions (47). All statistical tests were two-sided. The α level for the validation dataset was set at .05. All analyses were performed in Intercooled Stata, version 9.2 (STATACorp, College Station, TX).

Results

SEARCH Hypothesis-Generating Dataset

The 3761 patients with breast cancer from the SEARCH study who had available follow-up information and who participated in the second stage of the breast cancer susceptibility genome-wide

association study provided 26978 person-years at risk (median time at risk = 7.5 years, range = 0.04–11.48 years, and median time from diagnosis to enrollment = 1.15 years, range = 0–11.74 years). Of the 3761 patients, 647 had died within 15 years of their diagnosis; the annual overall mortality rate was 2.4% (95% confidence interval [CI] = 2.2% to 2.6%), the 5-year overall survival rate was 88% (95% CI = 87.0% to 89.4%), and the 10-year overall survival rate was 80% (95% CI = 78.3% to 81.3%) (Table 1).

We tested the 10621 SNPs that were selected as candidates for breast cancer susceptibility in the analysis of the first stage of the genome-wide association study for association with all-cause mortality. Although no SNPs reached nominal genome-wide statistical significance (ie, $P < 5 \times 10^{-8}$), two SNPs (rs6626269, $P_{trend} = 2.2 \times 10^{-6}$, and rs4778137, $P_{trend} = 1.9 \times 10^{-5}$) were strongly associated with overall survival after a diagnosis of breast cancer (with lower *P* values than expected by chance, Figure 1, and survival curves by genotype in Supplementary Figure 1, available online). The SNP with the strongest association with all-cause mortality, rs6626269 (A>G), is located on the X chromosome approximately 300 kilobase pairs upstream from the fragile X mental retardation 1 gene (*FMR1*) at Xq27.3; the rare (G) allele of SNP rs6626269 was associated with an increased risk of death compared with the common (A) allele (HR of death per rare allele carried = 1.35, 95% CI = 1.19 to 1.52, $P_{trend} = 2.2 \times 10^{-6}$). The other SNP with a strong association with overall survival, rs4778137 (C>G), was located on chromosome 15 in the oculocutaneous albinism II gene (*OCA2*) at 15q13.1. The rare (G) allele of SNP rs4778137 was associated with a decreased risk of death compared with the common (C) allele (HR of death per rare allele carried = 0.76, 95% CI = 0.67 to 0.86, $P_{trend} = 1.9 \times 10^{-5}$). When we repeated the analysis by using breast cancer–specific mortality as the endpoint, the strength of the association between rs6626269 and breast cancer–specific survival was slightly reduced (HR of death per rare allele carried = 1.29, 95% CI = 1.13 to 1.49, $P = 2.8 \times 10^{-4}$) but the association between rs4778137 and breast cancer–specific survival was slightly larger (HR of death per rare allele carried = 0.74, 95% CI = 0.64 to 0.85, $P = 2.4 \times 10^{-5}$), both compared with the associations with overall survival. Neither SNP showed evidence of violating the proportional hazards assumption in either analysis (ie, all $P > .05$).

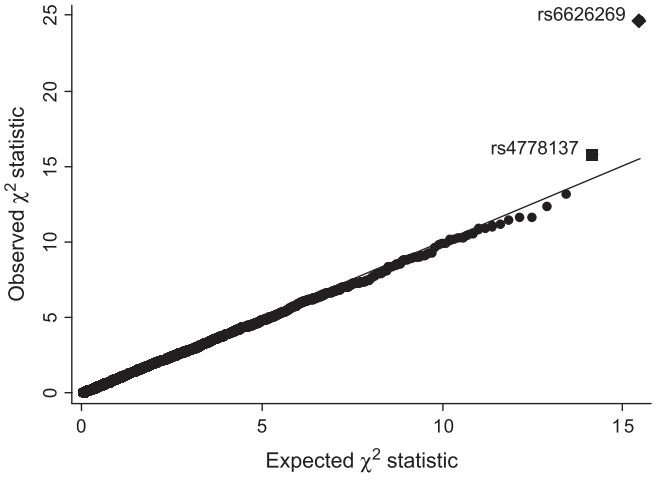

Figure 1. Quantile–quantile plot for the test statistics for the 10621 single-nucleotide polymorphisms (SNPs) evaluated in the Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) dataset. The test statistic was the χ^2 statistic from two-sided χ^2 trend tests with 1 *df*. **Solid circles** represent the expected and observed test statistics for each SNP. Under the null hypothesis of no association at any locus, the points would be expected to follow the **straight line**. SNPs rs6626269 and rs4778137 are represented by a **solid diamond** and a **solid square**, respectively.

We also evaluated associations for these two SNPs after stratifying the analyses for each SNP by stage, histopathologic grade, and ER status. The strongest association between rs4778137 and overall survival was observed among patients with ER-negative tumors (HR of death per rare allele carried = 0.56, 95% CI = 0.41 to 0.75, $P = 9.2 \times 10^{-5}$) (Table 3). For SNP rs4778137, we found evidence for heterogeneity by ER status ($P_{interaction} = .0024$). We found essentially no evidence for interaction between SNP rs4778137 and grade or stage or between SNP rs6626269 and stage, histopathologic grade, or ER status.

BCAC Validation Dataset

The two SNPs, rs4778137 and rs6626269, have been genotyped in 15 international BCAC studies that had follow-up information available. Patient characteristics are presented in Table 2. These 15 studies contained data for a total of 14096 patients diagnosed with invasive breast cancer who provided 70019 person-years at risk (median time at risk = 4.6 years, range = 0.03–15 years, and median time from diagnosis to enrollment = 0.11 years, range = 0–14.97 years). The total number of deaths within 15 years of diagnosis was 2303; the annual mortality rate for the combined studies was 3.3% (95% CI = 3.2% to 3.4%), the 5-year overall survival rate was 85% (95% CI = 84.7% to 86.1%), and the 10-year overall survival rate was 73% (95% CI = 71.4% to 73.6%) (Supplementary Table 3, available online). Genotyping information was not available for SNP rs6626269 from three studies, two of which had a failed assay design (Genetic Epidemiology Study of Breast Cancer by Age 50 [GESBC] and University of California Irvine Breast Cancer Study [UCIBCS]) and one of which failed the quality control requirements (Bavarian Breast Cancer Cases and Controls [BBCC]), but genotyping information was available for SNP rs4778137 from all 15 studies. Information as to whether patients were alive or dead within 15 years of their diagnosis and genotype information stratified by major prognostic factors for the combined hypothesis-generating and validation datasets are shown in Supplementary Table 4 (available online).

Associations between SNP rs6626269 or rs4778137 and overall survival for the hypothesis-generating, validation, and combined datasets are shown in Figure 2, Table 4, Supplementary Table 5 (available online), and Supplementary Figure 2 (available online). We investigated the relationships between the two SNPs and overall survival. Given that we found evidence for heterogeneity by ER status for rs4771837 in the SEARCH dataset, we stratified the validation dataset by ER status for rs4778137. In the validation set, the association of the rare (G) allele of rs4778137 with better overall survival was strongest among patients with ER-negative breast cancer (HR of death per rare allele carried = 0.88, 95% CI = 0.78 to 0.99, $P = .030$), as in the SEARCH hypothesis-generating set. When we repeated the analysis in a combined dataset containing data from the hypothesis-generating and validation sets, we again

Table 3. Associations between single-nucleotide polymorphisms, rs6626269 and rs4778137, and overall survival, stratified by breast cancer prognostic factors: Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) study*

	rs6626269			rs4778137		
Prognostic factor	**HR (95% CI)**	***P* for interaction†**	***P* for heterogeneity‡**	**HR (95% CI)**	***P* for interaction†**	***P* for heterogeneity‡**
Stage						
I	1.33 (1.04 to 1.69)	.79	.95	0.90 (0.71 to 1.15)	.19	.17
II	1.36 (1.16 to 1.60)			0.68 (0.57 to 0.80)		
III or IV	1.42 (0.98 to 2.06)			0.73 (0.51 to 1.05)		
Grade						
Well differentiated	1.52 (1.00 to 2.31)	.35	.57	0.81 (0.53 to 1.24)	.67	.91
Moderately differentiated	1.44 (1.16 to 1.79)			0.74 (0.59 to 0.93)		
Poorly differentiated	1.25 (1.01 to 1.55)			0.73 (0.59 to 0.90)		
ER status						
Positive	1.44 (1.18 to 1.75)	.83	.53	0.91 (0.75 to 1.10)	.0024	.0056
Negative	1.29 (1.00 to 1.68)			0.56 (0.41 to 0.75)		

* Hazard ratios (HRs) were calculated within strata of each prognostic factor and were not adjusted for other prognostic factors. CI = confidence interval; ER = estrogen receptor.

† Statistical significance based on a two-sided likelihood ratio test with 1 *df*.

‡ Statistical significance based on a two-sided test for heterogeneity with 1 *df*.

found the strongest association among patients with ER-negative breast cancer (HR of death per rare allele carried = 0.82, 95% CI = 0.73 to 0.92, $P = 5.0 \times 10^{-4}$). No statistically significant heterogeneity was observed across the studies in the validation studies ($I^2 = 0.0\%$, $P = .70$) or in the combined dataset ($I^2 = 19.7\%$, $P = .23$). Survival curves for ER-negative tumors by rs4778137 genotype for the SEARCH hypothesis-generating set, the BCAC validation set, and the combined dataset are provided in Figure 3. We next examined the association among patients with ER-negative tumors in the combined dataset after adjustment for age at diagnosis, stage, grade, and study; we found that the strength of the association was similar to that in the unadjusted analysis (HR of death per rare allele carried = 0.79, 95% CI = 0.68 to 0.92, $P = .0023$) (Table 5). We found no evidence that study-specific minor allele frequencies of rs4771837 tracked with weighted study-specific annual mortality rates (pairwise correlation = −0.41 and $P = .13$), excluding substantial bias from population stratification for this SNP. No associations were observed among patients with ER-positive tumors between SNP rs4778137 and overall survival in analyses of the combined dataset.

We could not validate the association between SNP rs6626269 and overall survival that we observed in the SEARCH hypothesis-generating dataset. Statistically significant heterogeneity for the effect of this SNP was observed between the SEARCH hypothesis-generating dataset and the BCAC validation dataset ($I^2 = 45.2\%$ and $P = .039$).

Discussion

We report an association between the SNP rs4778137 (C>G) that is located in the *OCA2* gene and overall survival for patients with ER-negative tumors, with the rare G allele of this SNP being associated with better prognosis for this group. The strength of the association was not reduced after adjusting for other prognostic factors, including age at diagnosis, TNM stage, and histopathologic grade, indicating that SNP rs4778137 was independently associated with overall survival among patients with ER-negative breast cancer in the combined SEARCH hypothesis-generating and BCAC validation datasets.

For this study, we evaluated associations between 10 621 SNPs that were genotyped in patients with invasive breast cancer from the second stage of a genome-wide association study of breast cancer susceptibility (SEARCH study) and all-cause mortality after a diagnosis of breast cancer. We used data from 15 international BCAC studies of patients with invasive breast cancer (14 096 patients) for the validation dataset. Major strengths of this study included the number of SNPs evaluated, the large and homogeneous sample in the hypothesis-generating dataset (SEARCH study), and the large validation sample.

The SNP rs4778137 was originally chosen as a tagging SNP for part of a genome-wide panel of 266 722 SNPs in a genome-wide association study of breast cancer susceptibility and is unlikely to be a functional variant. However, this SNP is located on chromosome 15 within the *OCA2* gene and close to the *HERC2* gene; both genes have been associated with skin pigmentation and hair and iris color (48–52), and genetic variation in this region of chromosome 15 has been associated with risk of melanoma (53,54). Because of the association with skin pigmentation and eye color, which could vary across our study populations, we evaluated our results for possible bias from population stratification. We found no evidence that study-specific rs4778137 minor allele frequencies tracked with study annual mortality rates, indicating that population stratification bias is not a threat to the validity of our conclusions (47). Furthermore, rs4778137 was only weakly correlated ($r^2 = .046$) with rs12913832, the SNP that was most strongly associated with blue eye color (D. Duffy, Genetic Epidemiology Laboratory, The Queensland Institute of Medical Research, Brisbane, Australia, personal communication).

The *OCA2* gene encodes a 12-transmembrane domain protein of unknown function that is localized to lysosomes when expressed in nonpigment cells (55). The OCA2 protein is not expressed in normal breast tissue (56), but overexpression of this protein in ductal breast carcinomas, compared with normal

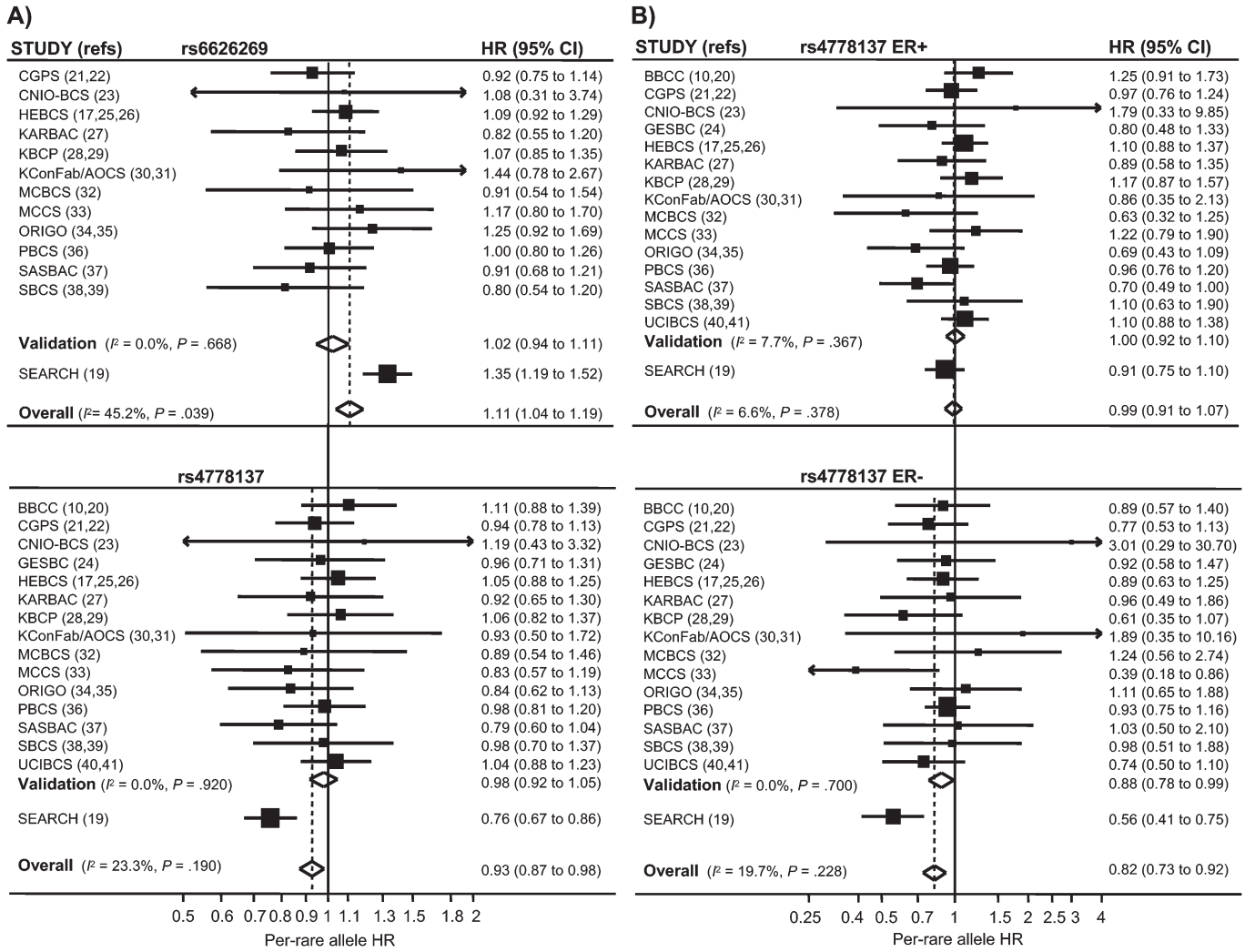

Figure 2. Associations between single-nucleotide polymorphism rs6626269 or rs4778137 and overall survival. Per-allele hazard ratios (HRs) and 95% confidence intervals (CIs) are presented. **A**) Nonstratified analyses for rs6626269 and rs4778137. **B**) Analyses for rs4778137 stratified by estrogen receptor (ER) status (ER positive [ER+] and ER negative [ER−]). Data are shown for analyses with the hypothesis-generating set (Studies of Epidemiology and Risk factors in Cancer Heredity [SEARCH]), individual studies in the validation dataset, the entire validation set, and a combined dataset (overall) containing data from the hypothesis-generating set and the validation set. Supplementary Table 1 (available online) presents descriptions of all studies. **Squares** = study-specific hazard ratios; area of the each square = inverse of the variance of the estimate; **horizontal lines** = 95% CIs; **diamonds** = summary hazard ratio estimates encompassing 95% CIs; **dotted vertical line** = combined summary hazard ratio. Hazard ratio statistical significance assessed by a trend test with 1 *df*. Study heterogeneity was assessed by use of the I^2 statistic. All statistical tests were two-sided. BBCC = Bavarian Breast Cancer Cases and Controls; CGPS = Copenhagen Breast Cancer Study and Copenhagen General Population Study; CNIO-BCS = Spanish National Cancer Centre Breast Cancer Study; GESBC = Genetic Epidemiology Study of Breast Cancer by Age 50; HEBCS = Helsinki Breast Cancer Study; KARBAC = Karolinska Breast Cancer Study; KBCP = Kuopio Breast Cancer Project; kConFab/AOCS = The Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer/Australian Ovarian Cancer Study; MCBCS = Mayo Clinic Breast Cancer Study; MCCS = Melbourne Collaborative Cohort Study; ORIGO = Leiden University Medical Centre Breast Cancer Study; PBCS = Polish Breast Cancer Study; SASBAC = Singapore and Swedish Breast Cancer Study; SBCS = Sheffield Breast Cancer Study; UCIBCS = University of California Irvine Breast Cancer Study.

breast tissue, has been reported (57). To study its function in more detail, researchers have expressed the p protein (a mouse *OCA2* homolog, pink-eyed dilution gene) in the yeast *Saccharomyces cerevisiae* by use of yeast expression plasmids and found that expression of this gene leads to higher sensitivity to a number of toxic compounds, as shown by the concentration of compound that inhibited colony formation by 50% (IC_{50} value) (58). Similarly, cultured murine melanocytes expressing a functional *p* gene were found to be more sensitive to these compounds, as well as the cytotoxic agents, cisplatin and doxorubicin, when assessed with a cell viability assay. Cisplatin and doxorubicin are detoxified in mammalian cells as glutathione conjugates, and intracellular glutathione expression decreased by 50% in *p*-expressing *S cerevisiae*, suggesting that the pink-eyed dilution protein can modulate intracellular glutathione metabolism (58). Doxorubicin and other anthracyclines are commonly used to treat breast cancer, particularly ER-negative breast cancer. Therefore, given the results of this study for SNP rs4778137, additional studies are warranted to investigate whether SNP rs4778137 is associated with sensitivity to anthracycline treatment

Table 4. Associations between single-nucleotide polymorphisms (SNPs), rs6626269 or rs4778137, and overall survival: Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH), validation, and combined datasets*

SNP	Gene (chromosomal band and position)†	Alleles‡	MAF§	SEARCH dataset			Validation dataset			Combined dataset		
				No. of patients	Per-allele HR (95% CI)	*P* for trend	No. of patients	Per-allele HR‖ (95% CI)	*P* for trend	No. of patients	Per-allele HR‖ (95% CI)	*P* for trend
rs6626269	*FMR1* (Xq27.3 and chrX:146673425)	A, G	0.21	3752	1.35 (1.19 to 1.52)	2.2×10^{-6}	10960	1.02 (0.94 to 1.11)	.64	14712	1.11 (1.04 to 1.19)	.0025
rs4778137	*OCA2* (15q13.1 and chr15:28327835)											
Total		C, G	0.30	3755	0.76 (0.67 to 0.86)	1.9×10^{-5}	13712	0.98 (0.92 to 1.05)	.54	17467	0.93 (0.87 to 0.98)	.011
ER positive				1978	0.91 (0.75 to 1.10)	.32	8865	1.01 (0.92 to 1.10)	.92	10843	0.99 (0.91 to 1.10)	.75
ER negative				498	0.56 (0.41 to 0.75)	9.2×10^{-5}	2656	0.88 (0.78 to 0.99)	.030	3154	0.82 (0.73 to 0.92)	5.0×10^{-4}

* Statistical significance based on a 1 *df* trend test. All statistical tests were two-sided. CI = confidence interval; ER = estrogen receptor; HR = hazard ratio; MAF = minor allele frequency.

† Gene name, chromosomal band, and chromosomal position based on dbSNP building 130 (ftp.ncbi.nih.gov/snp).

‡ Common allele and rare allele.

§ The MAF was based on SEARCH population data.

‖ Fixed-effects meta-analysis of study-specific hazard ratios.

among patients with ER-negative breast cancer. Alternatively, mutations in the *HERC2* gene have been associated with reduced growth, male sterility, and female semisterility (59), and it is possible that SNP rs4778137 may be associated with the expression or function of the *HERC2* gene. Further fine mapping will be required to explore this area for associations with breast cancer prognosis and to identify the additional variant(s) for additional study.

The SNPs that we analyzed in this study were from a set selected for the second stage of a multistage genome-wide association study of breast cancer susceptibility (18). Most of the SNPs in that study were expected to provide false-positive data (for susceptibility), but we believe that it is reasonable to hypothesize that SNPs that are associated with susceptibility may also be associated with tumor behavior and prognosis. The SNPs from that study, therefore, may constitute a large set of rather weak candidate prognostic variants. The main advantage of using the SEARCH dataset was the ease with which we could evaluate a large number of previously unexplored common germline genetic polymorphisms for an association with overall survival. However, it is important to note that our dataset was composed of only 10621 SNPs, which do not provide genome-wide coverage; therefore, there were areas of the genome that we could not assess for an association with overall survival. Further work, including a genome-wide scan of breast cancer prognosis, is required to identify more common germline variants that may be associated with overall survival.

This study had some limitations. Our dataset of 10621 SNPs does not represent a set of SNPs that captures common genetic variation across the entire genome. The observed association between rs4778137 and overall survival among patients with ER-negative tumors did not reach nominal genome-wide statistical significance ($P = 5.0 \times 10^{-4}$). A false-positive association caused by confounding cannot be ruled out, although it is unlikely that rs4778137 genotype would be associated with possible confounders, such as socioeconomic status or comorbidity. Treatment data were too limited to assess possible interaction with specific adjuvant therapies. It is also possible that the effect of rs4778137 is limited to specific subtypes of ER-negative disease such as the HER2-expressing tumors or the basal tumors, but data on other markers were not available to investigate this issue. The power of this study to detect association was modest, and it is possible that we missed some valid associations. For example, in the SEARCH dataset, at a nominal genome-wide statistical significance of 5×10^{-8}, the power to detect a minor allele with a frequency of 0.30 that confers a per-allele HR = 1.5 was 89%, but, if the effect size associated with the minor allele is weaker (eg, HR = 1.3), the power to detect an association is reduced to 13%. In addition, genetic effects may be restricted to specific patient subsets (eg, those based on clinicopathological features or treatment groups), further reducing power.

In summary, we identified a polymorphism in the *OCA2* gene located at 15q13.1, rs4778137 (C>G), that appears to be associated with all-cause mortality among patients with ER-negative breast cancer. We used a large validation dataset from 15 European, North American, and Australian BCAC studies of breast cancer to validate results in the SEARCH hypothesis-generating set. The effect size of the association between this

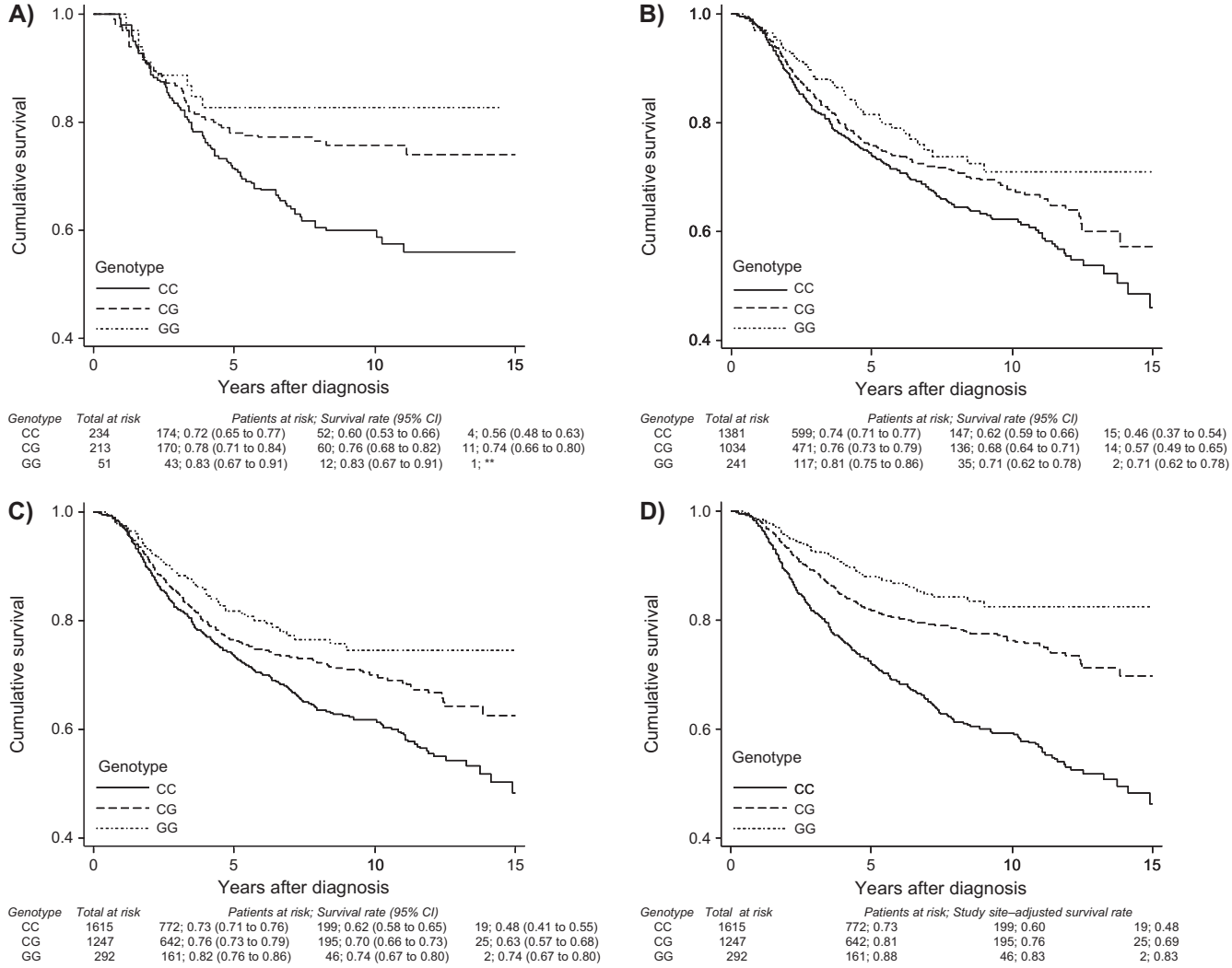

Figure 3. Cumulative overall survival among patients with estrogen receptor (ER)–negative breast cancer by genotype of the single-nucleotide polymorphism rs4778137. A Kaplan–Meier analysis was used. Total patients at risk in the analysis and number of patients at risk and overall survival rates with 95% confidence intervals (95% CI) for years 5, 10, and 15 after breast cancer diagnosis are presented. **A**) Cumulative survival for patients in the hypothesis-generating dataset from the Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) study. **B**) Cumulative survival for patients in the validation set. **C**) Cumulative survival for patients in the combined hypothesis-generating and validation datasets. **D**) Predicted cumulative survival adjusted for study site for patients in the combined hypothesis-generating and validation datasets, adjusted to the baseline hazard function of the SEARCH study. All four analyses were stratified by rs4778137 genotype (C = common; G = rare). ** = patient numbers were too few to estimate a survival rate at this time point.

SNP and overall survival, if a true association, was small, and so this SNP is unlikely to have clinical utility by itself. However, this effect compounded over a 10-year period for an appreciable overall survival difference. For example, among patients with ER-negative tumors, the study-adjusted, predicted overall survival rate was 88% at 5 years after diagnosis and 83% at 10 years after diagnosis for patients with two copies of the protective (rare, G) allele of rs4778137 compared with a study-adjusted, predicted overall survival rate of 73% at 5 years and 60% at 10 years for patients with two copies of the common (C) allele (Figure 3, D). Furthermore, the discriminatory accuracy of prognostication tools such as AdjuvantOnline! (http://www.adjuvantonline.com/), which is based on established prognostic factors, might be improved by incorporation of markers with associated effects of this magnitude. Because these tools are widely used to select patients most likely to benefit from adjuvant chemotherapy, even markers with small effects may have utility and so further validation of our findings, particularly in clinical trial populations, is warranted. In addition, the observed association may provide clues to the underlying tumor biology that is involved in the response to chemotherapy and might ultimately improve patient treatment and outcome. Germline variants in many genes are likely to affect many steps in cancer development and progression, as well as treatment suitability (currently based on assessment of functional status, performance status, life expectancy, and existing comorbidities), tolerance, and response. Future work will need to focus on a variety of specific outcomes and subgroups, as well as breast cancer–specific survival, to elucidate the complex relationships among tumors, hosts, and treatment effects.

Table 5. Associations between single-nucleotide polymorphisms (SNPs), rs6626269 or rs4778137, and overall survival in adjusted analyses: Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH), validation, and combined datasets*

SNP	Gene (chromosomal band and position)†	Allele‡	MAF§	SEARCH dataset			Validation dataset			Combined dataset		
				No. of patients	Per-allele HR (95% CI)	*P* for trend	No. of patients	Per-allele HR‖ (95% CI)	*P* for trend	No. of patients	Per-allele HR‖ (95% CI)	*P* for trend
rs6626269	*FMR1* (Xq27.3 and chrX:146673425)	A, G	0.21	2201	1.49 (1.25 to 1.77)	5.9×10^{-6}	6113	0.99 (0.88 to 1.11)	.83	8314	1.11 (1.01 to 1.22)	.03
rs4778137	*OCA2* (15q13.1 and chr15:28327835)											
Total		C, G	0.30	2205	0.72 (0.61 to 0.86)	2.8×10^{-4}	7855	0.95 (0.87 to 1.05)	.31	10060	0.9 (0.83 to 0.97)	.009
ER-positive				1762	0.83 (0.67 to 1.03)	.10	6246	0.97 (0.87 to 1.08)	.52	8008	0.94 (0.85 to 1.03)	.2
ER-negative				443	0.54 (0.40 to 0.74)	1.4×10^{-4}	1609	0.9 (0.76 to 1.07)	.25	2052	0.79 (0.68 to 0.92)	.002

* Statistical significance based on 1 *df* trend test. All statistical tests were two-sided. CI = confidence interval; ER = estrogen receptor; HR = hazard ratio; MAF = minor allele frequency.

† Gene name, chromosomal band, and chromosomal position based on dbSNP building 130 (ftp.ncbi.nih.gov/snp).

‡ Common allele and rare allele.

§ This value is based on SEARCH population data.

‖ Cox proportional hazards models adjusted for age at diagnosis and stratified by study stage, grade, and ER status.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin.* 2005;55(2):74–108.
2. Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med.* 2000;124(7):966–978.
3. Udler M, Pharoah PD. Germline genetic variation and breast cancer survival: prognostic and therapeutic implications. *Future Oncol.* 2007; 3(5):491–495.
4. Zubor P, Dokus K, Kajo K, Danko J. Integration of polymorphic data into the clinical management of breast cancer. *Lancet Oncol.* 2008;9(12): 1121–1122.
5. Lifsted T, Le Voyer T, Williams M, et al. Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int J Cancer.* 1998;77(4):640–644.
6. Hsieh SM, Lintell NA, Hunter KW. Germline polymorphisms are potential metastasis risk and prognosis markers in breast cancer. *Breast Dis.* 2006;26(1):157–162.
7. Hartman M, Lindstrom L, Dickman PW, Adami HO, Hall P, Czene K. Is breast cancer prognosis inherited? *Breast Cancer Res.* 2007;9(3):R39.
8. Abraham JE, Harrington P, Driver KE, et al. Common polymorphisms in the prostaglandin pathway genes and their association with breast cancer susceptibility and survival. *Clin Cancer Res.* 2009;15(6):2181–2191.
9. Azzato EM, Driver KE, Lesueur F, et al. Effects of common germline genetic variation in cell cycle control genes on breast cancer survival: results from a population-based cohort. *Breast Cancer Res.* 2008; 10(3):R47.
10. Fasching PA, Loehberg CR, Strissel PL, et al. Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients. *Breast Cancer Res Treat.* 2008;112(1): 89–98.
11. Goode EL, Dunning AM, Kuschel B, et al. Effect of germ-line genetic variation on breast cancer survival in a population-based study. *Cancer Res.* 2002;62(11):3052–3057.
12. Udler M, Azzato EM, Healey CS, et al. Common germline polymorphisms in COMT, CYP19A1, ESR1, PGR, SULT1E1 and STS and survival after a diagnosis of breast cancer. *Int J Cancer.* 2009;125(11): 2687–2696.
13. Udler M, Maia AT, Cebrian A, et al. Common germline genetic variation in antioxidant defense genes and survival after diagnosis of breast cancer. *J Clin Oncol.* 2007;25(21):3015–3023.
14. Hughes S, Agbaje O, Bowen RL, et al. Matrix metalloproteinase single-nucleotide polymorphisms and haplotypes predict breast cancer progression. *Clin Cancer Res.* 2007;13(22, pt 1):6673–6680.
15. Ambrosone CB, Ahn J, Singh KK, et al. Polymorphisms in genes related to oxidative stress (MPO, MnSOD, CAT) and survival after treatment for breast cancer. *Cancer Res.* 2005;65(3):1105–1111.
16. Ambrosone CB, Sweeney C, Coles BF, et al. Polymorphisms in glutathione S-transferases (GSTM1 and GSTT1) and survival after treatment for breast cancer. *Cancer Res.* 2001;61(19):7130–7135.
17. Fagerholm R, Hofstetter B, Tommiska J, et al. NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nat Genet.* 2008;40(7):844–853.
18. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007;447(7148):1087–1093.
19. Lesueur F, Pharoah PD, Laing S, et al. Allelic association of the human homologue of the mouse modifier Ptprj with breast cancer. *Hum Mol Genet.* 2005;14(16):2349–2356.
20. Schrauder M, Frank S, Strissel PL, et al. Single nucleotide polymorphism D1853N of the ATM gene may alter the risk for breast cancer. *J Cancer Res Clin Oncol.* 2008;134(8):873–882.
21. Bojesen SE, Tybjaerg-Hansen A, Axelsson CK, Nordestgaard BG. No association of breast cancer risk with integrin beta3 (ITGB3) Leu33Pro genotype. *Br J Cancer.* 2005;93(1):167–171.
22. Weischer M, Bojesen SE, Tybjaerg-Hansen A, Axelsson CK, Nordestgaard BG. Increased risk of breast cancer associated with CHEK2*1100delC. *J Clin Oncol.* 2007;25(1):57–63.

23. Milne RL, Ribas G, Gonzalez-Neira A, et al. ERCC4 associated with breast cancer risk: a two-stage case-control study using high-throughput genotyping. *Cancer Res.* 2006;66(19):9420–9427.
24. 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(8):687–695.
25. Kilpivaara O, Bartkova J, Eerola H, et al. Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients. *Int J Cancer.* 2005;113(4): 575–580.
26. Syrjakoski K, Vahteristo P, Eerola H, et al. Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients. *J Natl Cancer Inst.* 2000;92(18):1529–1531.
27. Margolin S, Werelius B, Fornander T, Lindblom A. BRCA1 mutations in a population-based study of breast cancer in Stockholm County. *Genet Test.* 2004;8(2):127–132.
28. Hartikainen JM, Tuhkanen H, Kataja V, et al. An autosome-wide scan for linkage disequilibrium-based association in sporadic breast cancer cases in eastern Finland: three candidate regions found. *Cancer Epidemiol Biomarkers Prev.* 2005;14(1):75–80.
29. Hartikainen JM, Tuhkanen H, Kataja V, et al. Refinement of the 22q12-q13 breast cancer–associated region: evidence of TMPRSS6 as a candidate gene in an eastern Finnish population. *Clin Cancer Res.* 2006;12(5):1454–1462.
30. Beesley J, Jordan SJ, Spurdle AB, et al. Association between single-nucleotide polymorphisms in hormone metabolism and DNA repair genes and epithelial ovarian cancer: results from two Australian studies and an additional validation set. *Cancer Epidemiol Biomarkers Prev.* 2007; 16(12):2557–2565.
31. Mann GJ, Thorne H, Balleine RL, et al. Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the kConFab familial breast cancer resource. *Breast Cancer Res.* 2006;8(1):R12.
32. Olson JE, Ingle JN, Ma CX, et al. A comprehensive examination of CYP19 variation and risk of breast cancer using two haplotype-tagging approaches. *Breast Cancer Res Treat.* 2007;102(2):237–247.
33. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ.* 2002;156:69–70.
34. de Bock GH, Schutte M, Krol-Warmerdam EM, et al. Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant. *J Med Genet.* 2004;41(10):731–735.
35. Huijts PE, Vreeswijk MP, Kroeze-Jansema KH, et al. Clinical correlates of low-risk variants in FGFR2, TNRC9, MAP3K1, LSP1 and 8q24 in a Dutch cohort of incident breast cancer cases. *Breast Cancer Res.* 2007;9(6):R78.
36. Garcia-Closas M, Egan KM, Newcomb PA, et al. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. *Hum Genet.* 2006;119(4):376–388.
37. Wedren S, Lovmar L, Humphreys K, et al. Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study. *Breast Cancer Res.* 2004;6(4):R437–R449.
38. MacPherson G, Healey CS, Teare MD, et al. Association of a common variant of the CASP8 gene with reduced risk of breast cancer. *J Natl Cancer Inst.* 2004;96(24):1866–1869.
39. Rafii S, O'Regan P, Xinarianos G, et al. A potential role for the XRCC2 R188H polymorphic site in DNA-damage repair and breast cancer. *Hum Mol Genet.* 2002;11(12):1433–1438.
40. Anton-Culver H, Cohen PF, Gildea ME, Ziogas A. Characteristics of BRCA1 mutations in a population-based case series of breast and ovarian cancer. *Eur J Cancer.* 2000;36(10):1200–1208.
41. Ziogas A, Gildea M, Cohen P, et al. Cancer risk estimates for family members of a population-based family registry for breast and ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2000;9(1):103–111.
42. Sobin LH, Wittekind C. *TNM Classification of Malignant Tumours.* 5th ed. New York, NY: Wiley-Liss; 1997.
43. Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol.* 1999;17(5):1474–1481.
44. Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer.* 1999;80(suppl 1):95–103.
45. Azzato EM, Greenberg D, Shah M, et al. Prevalent cases in observational studies of cancer survival: do they bias hazard ratio estimates? *Br J Cancer.* 2009;100(11):1806–1811.
46. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21(11):1539–1558.
47. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol Biomarkers Prev.* 2002;11(6):513–520.
48. Branicki W, Brudnik U, Wojas-Pelc A. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. *Ann Hum Genet.* 2009;73(2):160–170.
49. Han J, Kraft P, Nan H, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* 2008;4(5):e1000074.
50. Kayser M, Liu F, Janssens AC, et al. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet.* 2008;82(2):411–423.
51. Sturm RA, Duffy DL, Zhao ZZ, et al. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am J Hum Genet.* 2008;82(2):424–431.
52. Sulem P, Gudbjartsson DF, Stacey SN, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet.* 2007;39(12): 1443–1452.
53. Fernandez LP, Milne RL, Pita G, et al. Pigmentation-related genes and their implication in malignant melanoma susceptibility. *Exp Dermatol.* 2009; 18(7):634–642.
54. Jannot AS, Meziani R, Bertrand G, et al. Allele variations in the OCA2 gene (pink-eyed-dilution locus) are associated with genetic susceptibility to melanoma. *Eur J Hum Genet.* 2005;13(8):913–920.
55. Sitaram A, Piccirillo R, Palmisano I, et al. Localization to mature melanosomes by virtue of cytoplasmic dileucine motifs is required for human OCA2 function. *Mol Biol Cell.* 2009;20(5):1464–1477.
56. Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics.* 2008; 9(271):271.
57. Richardson AL, Wang ZC, De Nicolo A, et al. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell.* 2006;9(2):121–132.
58. Staleva L, Manga P, Orlow SJ. Pink-eyed dilution protein modulates arsenic sensitivity and intracellular glutathione metabolism. *Mol Biol Cell.* 2002;13(12):4206–4220.
59. Lehman AL, Nakatsu Y, Ching A, et al. A very large protein with diverse functional motifs is deficient in rjs (runty, jerky, sterile) mice. *Proc Natl Acad Sci U S A.* 1998;95(16):9436–9441.

Funding

Cancer Research UK (C1287 and A7497 to D.F.E.). E.M.A. is supported by the National Cancer Institute and the National Institutes of Health (NIH)–Cambridge Graduate Partnership Program. Bavarian Breast Cancer Cases and Controls (BBCC) was supported by the ELAN Fonds of the University of Erlangen, and P.A.F. is funded by Dr Mildred Scheel Stiftung and Deutsche Krebshilfe e.V. CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Copenhagen University Hospital, Herlev Hospital. Spanish National Cancer Centre Breast Cancer Study (CNIO-BCS) was supported by the Genome Spain Foundation, the *Red Temática de Investigación Cooperativa en Cáncer*, and grants from the *Asociación Española Contra el Cáncer* and the *Fondo de Investigación Sanitario* (PI081120 to J.Ben., PI081583 to R.L.M.]. Genetic Epidemiology Study of Breast Cancer by Age 50 (GESBC) was supported by the Deutsche Krebshilfe e.V. (70492) and GESBC genotyping in part by the state of Baden-Württemberg through the Medical Faculty of the University of Ulm (P.685). Helsinki Breast Cancer Study (HEBCS) has been financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (110663), the Finnish Cancer Society, and the Sigrid Juselius Foundation. Kuopio Breast Cancer Project (KBCP) is supported by grants from erityisvaltionosuus (EVO)

funds of Kuopio University Hospital, EVO funds of Vaasa Central Hospital, the Finnish Cancer Foundation, and the Academy of Finland. The Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer is supported by grants from the National Breast Cancer Foundation and the National Health and Medical Research Council (NHMRC) (145684, 288704, and 454508) and by the Queensland Cancer Fund; the Cancer Councils of New South Wales, Victoria, Tasmania, and South Australia; and the Cancer Foundation of Western Australia. Mayo Clinic Breast Cancer Study (MCBCS) was supported by grants from the NIH (P50 CA116201 and R01 CA122340). Melbourne Collaborative Cohort Study (MCCS) was supported by NHMRC (209057, 251533, 396414, and 504711). Infrastructure support for the MCCS recruitment and follow-up is provided by the Cancer Council Victoria, whereas cohort recruitment was partly funded by VicHealth. Genotyping was in part supported by the Prostate Cancer Foundation of Australia. Leiden University Medical Centre Breast Cancer Study (ORIGO) was supported by the Dutch Cancer Society. Polish Breast Cancer Study was funded by intramural research funds of the National Cancer Institute, Department of Health and Human Services. SASBAC was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US NIH, and the Susan G. Komen Breast Cancer Foundation. Sheffield Breast Cancer Study (SBCS) was supported by Yorkshire Cancer Research and the Breast Cancer Campaign. University of California Irvine Breast Cancer Study (UCIBCS) is supported by the NIH, National Cancer Institute (CA-58860), and the Lon V. Smith Foundation (LVS-18840).

Notes

The authors have full responsibility for design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript.

P. A. Fasching is conducting research sponsored by Novartis and AstraZeneca. N.E.C. is a Senior Investigator in the National Cancer Institute. G.C.-T. is a NHMRC Fellow. M.C.S. is a NHMRC Senior Research Fellow.

Spanish National Cancer Centre Breast Cancer Study (CNIO-BCS) thanks Pilar Zamora from La Paz University Hospital, Madrid, and Anna González-Neira, Charo Alonso, Tais Moreno, and Guillermo Pita from the Spanish National Cancer Research Centre (CNIO). Genetic Epidemiology Study of Breast Cancer by Age 50 (GESBC) thanks Ursula Eilber and Tanya Koehler for competent technical assistance. Helsinki Breast Cancer Study (HEBCS) thanks Dr Kirsimari Aaltonen and R. N. Hanna Jäntti for their help with the patient data and gratefully acknowledges the Finnish Cancer Registry for the cancer data. Karolinska Breast Cancer Study (KARBAC) thanks the Cancer Society in Stockholm, the Nilsson Ehle Foundation, and the Swedish Society of Medicine. Kuopio Breast Cancer Project (KBCP) is grateful to Eija Myöhänen for her assistance. The Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab) thanks David Duffy for helpful comments on the manuscript; Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow-Up Study for their contributions to this resource; and the many families that contribute to kConFab. Leiden University Medical Centre Breast Cancer Study (ORIGO) thanks C. J. van Asperen, P. E. A. Huijts, E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. Polish Breast Cancer Study (PBCS) thanks Drs Neonila Szeszenia-Dabrowska and Beata Peplonska of the Nofer Institute of Occupational Medicine (Lodz, Poland); Witold Zatonski of the Department of Cancer Epidemiology and Prevention, the M. Sklodowska-Curie Cancer Center and Institute of Oncology (Warsaw, Poland); Louise Brinton and Stephen Chanock from the Division of Cancer Epidemiology and Genetics of the National Cancer Institute; and Pei Chao from Information Management Services (Silver Spring, MD, USA) for their valuable contributions to the study. Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) thanks the women who have taken part in the study, the SEARCH study team and the consultants and general practitioners throughout East Anglia for their help in recruiting patients, and the staff of the Eastern Cancer Registration and Information Centre and the Thames Cancer Registry for providing outcome and clinical data. Sheffield Breast Cancer Study (SBCS) thanks Helen Cramp, Sabapathy Balasubramanian, Sue Higham, and Dan Connley for assistance with subject recruitment and data collection and Simon Cross for assistance with histopathology.

Affiliations of authors: Department of Oncology, Strangeways Research Laboratory, University of Cambridge, Cambridge, UK (EMA, JT, PDPP); Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD (EMA, NEC); Bavarian Breast Cancer Cases and Controls (BBCC), University Breast Center (PAF, MWB) and BBCC, Institute of Diagnostic Radiology (RS-W), University Hospital Erlangen, Erlangen, Germany; BBCC, Institute of Human Genetics, Friedrich Alexander University Erlangen Nuremberg, Erlangen, Germany (ABE); BBCC, Division of Hematology and Oncology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA (PAF); Copenhagen Breast Cancer Study and Copenhagen General Population Study (CGPS), Department of Clinical Biochemistry and CGPS, Department of Breast Surgery, Herlev University Hospital, University of Copenhagen, Copenhagen, Denmark (SEB, BGN, HF); Spanish National Cancer Centre Breast Cancer Study (CNIO-BCS), Genetic and Molecular Epidemiology Group (RLM) and CNIO-BCS, Human Genetics Group (JBen), Spanish National Cancer Research Centre (CNIO), Madrid, Spain; CNIO-BCS, El Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Spain (JBen); CNIO-BCS, Hospital Monte Naranco, Oviedo, Spain (JIA); CNIO-BCS, Hospital Universitario Central de Asturias (HUCA-Oviedo) (PM), Oviedo, Spain; Genetic Epidemiology Study of Breast Cancer by Age 50 (GESBC), Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany (JC-C, RH); GESBC, Department of Obstetrics and Gynecology, University of Ulm, Ulm, Germany (SW-G); Helsinki Breast Cancer Study (HEBCS), Department of Obstetrics and Gynecology (HN, TH), HEBCS, Department of Clinical Genetics (KA), and HEBCS, Department of Oncology (CB), Helsinki University Central Hospital, Helsinki, Finland; Karolinska Breast Cancer Study (KARBAC), Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden (SM, AM); Kuopio Breast Cancer Project (KBCP), Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Kuopio (AM, V-MK, VK), Kuopio, Finland; KBCP, Department of Pathology (AM, V-MK) and KBCP, Department of Oncology (AM, V-MK, VK), Kuopio University Hospital, Kuopio, Finland; KBCP, Department of Oncology, Vaasa Central Hospital, Vaasa, Finland (VK); Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia (kConFab); Queensland Institute of Medical Research, Brisbane, Australia (GC-T, JBee, XC); Mayo Clinic Breast Cancer Study (MCBCS), Department of Laboratory Medicine and Pathology (FJC, XW) and MCBCS, Department of Health Sciences Research (FJC, JEO, ZSF), Mayo Clinic, Rochester, MN; Melbourne Collaborative Cohort Study (MCCS), Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Australia (GGG, GS, LB); MCCS, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia (MCS); Leiden University Medical Centre Breast Cancer Study (ORIGO), Department of Human Genetics (PD), ORIGO, Department of Pathology (PD), and ORIGO, Department of Surgical Oncology (RAEMT), Leiden University Medical Centre, Leiden, the Netherlands; ORIGO, Department of Medical Oncology, Rotterdam Family Cancer Clinic, Erasmus MC-Daniel den Hoed Cancer Center, Rotterdam, the Netherlands (CS); Polish Breast Cancer Study (PBCS), Hormonal and Reproductive Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD (MG-C, MES, KLB); PBCS, Department of Cancer Epidemiology and Prevention, The M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland (JL); Singapore and Swedish Breast Cancer Study (SASBAC), Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden (KC, PH); Sheffield Breast Cancer Study (SBCS), Institute for Cancer Studies (AC, GCE, IWB), Department of Oncology, University of Sheffield, Sheffield, UK; SBCS, Academic Unit of Surgical Oncology (MWRR), University of Sheffield, Sheffield, UK; Eastern Cancer Registration and Information Centre, Unit C-Magog Court, Shelford Bottom, Cambridge, UK (DG); University of California Irvine Breast Cancer Study (UCIBCS), Department of Epidemiology, University of California Irvine, Irvine, CA (AZ, HA-C); Cancer Research UK Genetic Epidemiology Unit, Department of Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Cambridge, UK (DFE, MH).