HMG Advance Access published September 13, 2012

A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11

Afshan Siddiq[1]†, Fergus J. Couch[2,3]†, Gary K.Chen[4]†, Sara Lindström[5], Diana Eccles[6], Robert C. Millikan[7], Kyriaki Michailidou[8], Daniel O. Stram[4], Lars Beckmann[9], Suhn Kyong Rhie[4], Christine B. Ambrosone[10], Kristiina Aittomäki[11], Pilar Amiano[12], Carmel Apicella[13], Australian Breast Cancer Tissue Bank Investigators[14], Laura Baglietto[13,15], Elisa V. Bandera[16], Matthias W. Beckmann[17], Christine D. Berg[18], Leslie Bernstein[19], Carl Blomqvist[20], Hiltrud Brauch[21], Louise Brinton[22], Quang M. Bui[13], Julie E. Buring[23], Saundra S. Buys[24], Daniele Campa[25], Jane E. Carpenter[26], Daniel I. Chasman[27], Jenny Chang-Claude[28], Constance Chen[5], Françoise Clavel-Chapelon[29], Angela Cox[30] , Simon S. Cross[31], Kamila Czene[32], Sandra L. Deming[33], Robert B. Diasio[34], W. Ryan Diver[35], Alison M. Dunning[36], Lorraine Durcan[6], Arif B. Ekici[37], Peter A. Fasching[17,38], Familial Breast Cancer Study[39], Heather Spencer Feigelson[40], Laura Fejerman[41], Jonine D Figueroa[22], Olivia Fletcher[42], Dieter Flesch-Janys[43], Mia M. Gaudet[35], The GENICA Consortium[44], Susan M. Gerty[6], Jorge L. Rodriguez-Gil[45], Graham G. Giles[13,15], Carla H. van Gils[46], Andrew K. Godwin[47], Nikki Graham[6], Dario Greco[48], Per Hall[32], Susan E. Hankinson[23], Arndt Hartmann[49], Rebecca Hein[28,50], Judith Heinz[43], Robert N. Hoover[22], John L Hopper[13], Jennifer J. Hu[45], Scott Huntsman[51], Sue A. Ingles[4], Astrid Irwanto[52], Claudine Isaacs[53], Kevin B. Jacobs[22,54,55], Esther M. John[56], Christina Justenhoven[21], Rudolf Kaaks[28], Laurence N. Kolonel[57], Gerhard A. Coetzee[4,87], Mark Lathrop[58,59], Loic Le Marchand[57], Adam M. Lee[34], I-Min Lee[23], Timothy Lesnick[2], Peter Lichtner[60], Jianjun Liu[52], Eiliv Lund[61], Enes Makalic[13], Nicholas G. Martin[62], Catriona A McLean[63], Hanne Meijers-Heijboer[64], Alfons Meindl[65], Penelope Miron[66], Kristine R. Monroe[4], Grant W. Montgomery[62], Bertram Müller-Myhsok[67], Stefan Nickels[28], Sarah J. Nyante[22], Curtis Olswold[2], Kim Overvad[68], Domenico Palli[69], Daniel J Park[70], Julie R. Palmer[71], Harsh Pathak[47], Julian Peto[72], Paul Pharoah[36], Nazneen Rahman[39], Fernando Rivadeneira[73], Daniel F.

Schmidt[13], Rita K Schmutzler[74], Susan Slager[2], Melissa C. Southey[70], Kristen N. Stevens[2], Hans-Peter Sinn[75], Michael F. Press[76], Eric Ross[77], Elio Riboli[78], Paul M. Ridker[27], Fredrick R. Schumacher[4], Gianluca Severi[13,15], Isabel dos Santos Silva[72], Jennifer Stone[13], Malin Sund[79], William J. Tapper[6], Michael J. Thun[35], Ruth C. Travis[80], Clare Turnbull[39], Andre G. Uitterlinden[73], Quinten Waisfisz[64], Xianshu Wang[3], Zhaoming Wang[22,54], JoEllen Weaver[81], Rüdiger Schulz-Wendtland[82], Lynne R. Wilkens[57], David Van Den Berg[4], Wei Zheng[83], Regina G. Ziegler[22], Elad Ziv[51], Heli Nevanlinna[48], Douglas F. Easton[36], David J. Hunter[84,85], Brian E. Henderson[4], Stephen J. Chanock[22], Montserrat Garcia-Closas[86], Peter Kraft[5]†, Christopher A. Haiman[4]†, Celine M. Vachon[2]†*

[1] Department of Epidemiology and Biostatistics & Department of Genomics of Common Disease, School of Public Health, Imperial College London, United Kingdom

[2] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

[3] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

[4] Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA

[5] Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA, USA

[6] Faculty of Medicine, University of Southampton, Southampton, UK

[7] Department of Epidemiology, Gillings School of Global Public Health, and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

[8] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[9] Institute for Quality and Efficiency in Health Care, IQWiG, Cologne, Germany

[10] Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA

[11] Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland

[12] Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain

[13] Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, Melbourne School of Population Health, The University of Melbourne, Australia

[14] ABCTB, University of Sydney, NSW, Australia

[15] Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Australia & Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, The University of Melbourne, Australia

[16] The Cancer Institute of New Jersey, New Brunswick, NJ, USA

[17] Friedrich-Alexander University Erlangen-Nuremberg , University Hospital Erlangen, University Breast Center Franconia, Department of Gynecology and Obstetrics, Erlangen, Germany

[18] Early Detection Research Group, Division of Cancer Prevention, National Cancer Institute, Rockville, Maryland, USA

[19] Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, CA, USA

[20] Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland

[21] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University Tübingen, Germany

[22] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA

[23] Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[24]Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

[25] Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany

[26] Australian Breast Cancer Tissue Bank, University of Sydney at the Westmead Millennium Institute, Westmead, NSW, Australia

[27] Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[28]Division of Cancer Epidemiology, German Cancer Research Center, Deutsches Krebsforschungszentrum, Heidelberg, Germany

[29] INSERM UMR 1018, Team 9: Nutrition, Hormones et Sante´ desfemmes, Centre de Recherche en Epide´miologie et Sante´ des Populations, Hoˆpital Paul Brousse, Villejuif, France

[30] Institute for Cancer Studies, Department of Oncology , Faculty of Medicine, Dentistry & Health, University of Sheffield, UK

[31] Academic Unit of Pathology, Department of Neuroscience, Faculty of Medicine, Dentistry & Health, University of Sheffield, UK

[32] Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden

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

[34] Department of Pharmacology, Mayo Clinic, Rochester, MN, USA

[35] Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA

[36] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

[37] Friedrich-Alexander University Erlangen-Nuremberg, Institute of Human Genetics, Erlangen, Germany

[38] University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine, Division of hematology and Oncology, Los Angeles, CA, USA

[39] Section of Cancer Genetics, Institute of Cancer Research, Sutton, UK

[40] Institute for Health Research, Kaiser Permanente, Denver, CO, USA

[41] Division of General Internal Medicine and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California

[42] Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, UK

[43] Department of Cancer Epidemiology/Clinical Cancer Registry University Cancer Center Hamburg (UCCH) and Department of Medical Biometrics and Epidemiology University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[44] Gene Environment Interaction and Breast Cancer in Germany (GENICA): Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University Tübingen, Germany (HB, CJ);

Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany (Ute Hamann); Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany (Yon-Dschun Ko, Christian Baisch); Institute of Pathology, Medical Faculty of the University of Bonn, Germany (Hans-Peter Fischer); Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, Germany (Thomas Bruening, Beate Pesch, Sylvia Rabstein), Institute and Outpatient Clinic of Occupational Medicine, Saarland University Medical Center and Saarland University Faculty of Medicine, Homburg, Germany (Volker Harth)

[45] Sylvester Comprehensive Cancer Center and Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, FL, USA

[46] Julius Center, University Medical Center, Utrecht, The Netherlands

[47] Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Kansas City, KS, USA

[48] Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland

[49] Friedrich-Alexander University Erlangen-Nuremberg, Institute of Pathology, University Hospital Erlangen, Erlangen, Germany

[50] PMV Research Group at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany

[51] Division of General Internal Medicine, Department of Medicine, Institute for Human Genetics and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, USA

[52] Human Genetics Division, Genome Institute of Singapore, Singapore.

[53] Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC

[54] Core Genotyping Facility, SAIC-Frederick Inc., NCI-Frederick, Frederick, MD, USA

[55] Bioinformed Consulting Services, Gaithersburg, MD, USA

[56] Cancer Prevention Institute of California, Fremont, CA, USA, and Stanford University School of Medicine and Stanford Cancer Institute, Stanford, CA, USA

[57] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA

[58] Centre National de Genotypage, Evry, France.

[59] Fondation Jean Dausset – CEPH, Paris, France.

[60] Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[61] Institute of Community Medicine University of Tromsø, Tromsø,Norway

[62] QIMR GWAS Collective, Queensland Institute of Medical Research, Brisbane, Australia

[63] The Alfred Hospital, Melbourne, Australia

[64] Department of Clinical Genetics, VU University Medical Center, section Oncogenetics, Amsterdam, The Netherlands

[65] Clinic of Gynaecology and Obstetrics, Division for Gynaecological Tumor-Genetics, Technische Universität München, München, Germany

[66] Dana Farber Cancer Institute, Boston, MA, USA

[67] Max Planck Institute of Psychiatry, Munich, Germany

[68] Department of Cardiology, Center for Cardiovascular Research,Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark

[69] Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute, ISPO, Florence, Italy

[70] Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Australia

[71] Slone Epidemiology Center at Boston University, Boston, MA, USA

[72] Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London, UK.

[73] Department of Internal Medicine and Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

[74] Department of Obstetrics and Gynaecology, Division of Molecular Gynaeco-Oncology, University of Cologne, Germany

[75] Department of Pathology, University Hospital Heidelberg, Heidelberg, Germany

[76] Department of Pathology, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA

[77] Department of Biostatistics, Fox Chase Cancer Center, Philadelphia, PA, USA

[78] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom

[79] Department of Surgery, Umea° University, Umea°, Sweden

[80] Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

[81] Biosample Repository, Fox Chase Cancer Center, Philadelphia, PA, USA

[82] Friedrich-Alexander University Erlangen-Nuremberg, Institute of Diagnostic Radiology, University Hospital Erlangen, Erlangen, Germany

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

[84] Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

[85] Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA, USA

[86] Section of Epidemiology and Genetics, Institute of Cancer Research, Sutton, United Kingdom

[87] Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA, 90089

*To whom correspondence should be addressed: Celine M. Vachon, Mayo Clinic, 200 First St SW, Charlton 6-239, Rochester, MN 55905, (Tel): 507-284-9977, (Fax): 507-266-2478, E-mail: Vachon.Celine@mayo.edu or Christopher A. Haiman, Harlyne Norris Research Tower, 1450 Biggy Street, Room 1504, Los Angeles, CA 90033 USA, Email: Christopher.Haiman@med.usc.edu

† These authors contributed equally.

ABSTRACT

Genome-wide association studies (GWAS) of breast cancer defined by hormone receptor status have revealed loci contributing to susceptibility of estrogen receptor (ER)-negative subtypes. To identify additional genetic variants for ER-negative breast cancer we conducted the largest meta-analysis of ER-negative disease to date, comprising 4,754 ER-negative cases and 31,663 controls from three GWAS: NCI Breast and Prostate Cancer Cohort Consortium (BPC3) (2,188 ER-negative cases; 25,519 controls of European ancestry), Triple Negative Breast Cancer Consortium (TNBCC) (1,562 triple negative cases; 3,399 controls of European ancestry) and African American Breast Cancer Consortium (AABC) (1,004 ER-negative cases; 2,745 controls). We performed *in silico* replication of 86 SNPs at $P \leq 1x10^{-5}$ in an additional 11,209 breast cancer cases (946 with ER-negative disease) and 16,057 controls of Japanese, Latino and European ancestry. We identified two novel loci for breast cancer at 20q11 and 6q14. SNP rs2284378 at 20q11 was associated with ER-negative breast cancer (combined two stage OR=1.16; $P= 1.1x10^{-8}$) but showed a weaker association with overall breast cancer (OR=1.08, $P=1.3x10^{-6}$) based on 17,869 cases and 43,745 controls and no association with ER-positive disease (OR=1.01, P=0.67) based on 9,965 cases and 22,902 controls. Similarly, rs17530068 at 6q14 was associated with breast cancer (OR=1.12; $P= 1.1x10^{-9}$), and with both ER-positive (OR=1.09; $P=1.5x10^{-5}$) and ER-negative (OR=1.16, $P=2.5x10^{-7}$) disease. We also confirmed three known loci associated with ER-negative (19p13) and both ER-negative and ER-positive breast cancer (6q25 and 12p11). Our results highlight the value of large-scale collaborative studies to identify novel breast cancer risk loci.

INTRODUCTION

Breast cancer is a heterogeneous disease and has multiple histological and molecular subtypes, likely with distinct etiologies. Tumors that lack expression of the estrogen receptor (ER) tend to have more aggressive disease, higher histological grade, and lower survival rates (1). ER-negative breast cancer is more common in women of African ancestry, accounting for as much as 40% of cases in African American women compared with 15-20% in women of European ancestry. The etiologic heterogeneity between breast cancer subtypes is supported by different associations with ER-positive versus ER-negative disease for many of the known breast cancer risk factors (such as reproductive factors and BMI)(2). Tumors in women with *BRCA1* mutations are predominantly ER-negative, while tumors in *BRCA2* mutation carriers are predominantly ER-positive(3). Furthermore, genome-wide association studies have identified multiple common genetic variants more strongly associated with ER-positive than ER-negative breast cancer(4). Through collaborative efforts, we recently identified risk loci on 5p15 and 19p13 that are associated specifically with ER-negative and triple negative (TN) (ER-negative, progresterone (PR)-negative and HER2-negative) breast cancer(5-7).

In order to identify genetic loci associated with risk of ER-negative breast cancer, we conducted a meta-analysis of three GWAS of ER-negative breast cancer, comprising 4,754 cases and 31,663 controls with further replication in an additional 11,209 cases (946 with ER-negative disease) and 16,057 controls.

RESULTS

The meta-analysis included GWAS of ER-negative breast cancer (4,754 ER-negative cases and 31,663 controls) from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3) (2,188 ER-negative cases and 25,519 controls of European ancestry), the Triple Negative Breast Cancer Consortium (TNBCC) (1,562 triple negative cases and 3,399 controls of European ancestry) and the African American Breast Cancer Consortium (AABC) (1,004 ER-negative cases and 2,745 controls). **(Figure 1, Supplementary Table 1)**. We observed little evidence of over-inflation in the test statistics ($\lambda \leq 1.04$ for each study; λ=1.04 for meta-analysis) **(Supplementary Figure 1)**. A total of 86 SNPs were associated with ER-

negative breast cancer at $P \leq 10^{-5}$ **(Supplementary Table 2).** An *in silico* replication of the 86 SNPs was conducted using GWAS of European (BCAC combined), Latino (MEC-LAT, SFBCS/NC-BCFR) and Japanese (MEC-JPT) ancestry populations, totaling 11,209 breast cancer cases (946 with ER-negative disease) and 8,404 controls (Stage 2)**(Supplementary Table 1).**

Combining results for ER-negative breast cancer from stages 1 and 2, variants in three regions showed genome-wide significance [20q11-rs2284378, T allele: odds ratio, OR=1.16, $P = 1.1x10^{-8}$ (**Table 1**); 19p13-rs8100241, G allele: OR=1.14, $P=3.5x10^{-8}$; 6q25-rs9383938, T allele: OR=1.28, $P = 2.37 \times 10^{-10}$]. Variants at 6q25 have previously been associated with breast cancer risk(8), and variants at the 19p13 locus have been associated with ER-negative and TN breast cancer risk(5, 7). The rs2284378 variant at 20q11 is located in a region containing *RALY* (RNA binding protein, autoantigenic), *EIF2S2* (eukaryotic translation initiation factor 2, subunit 2 beta) and ~100kb upstream of *ASIP* (agouti signaling protein), and is in high linkage disequilibrium (r^2=0.96 and D'=1) with rs4911414, which has been associated with melanoma and basal cell carcinoma(9) (**Supplementary Figure 2).** The T allele at rs2284378 was associated with an increased ER-negative breast cancer risk (OR>1) in all racial/ethnic populations, except Japanese (OR=0.99) (**Table 1**). However this group had the smallest sample size. Furthermore, no significant evidence of heterogeneity was observed by race (*P*=0.28) or study (*P*=0.54) **(Table 1, Supplementary Table 3)**. When the study was extended to include all available breast cancer cases (ER-positive and ER-negative) and controls from the participating GWAS, rs2284378 showed a weaker association with overall breast cancer (OR=1.08, $P=1.3x10^{-6}$ based on 17,868 cases and 43,744 controls; **Table 1**) and no evidence for association with ER-positive disease (OR=1.01, *P*=0.67 based on 9,965 cases and 22,902 controls **(Supplementary Table 5)**. A case-only analysis of ER-negative versus ER-positive breast cancer indicated a highly significant difference in ORs by ER status ($P=1.3x10^{-4}$, **Supplementary Table 5**). Furthermore, rs2284378 appeared more strongly associated with triple negative (TN) breast cancer (OR=1.16; $P=6.4x10^{-3}$), than ER-negative, PR-negative, HER2-positive breast cancer (OR=1.07, *P*=0.41), although these differences were not statistically significant (case-only *P*=0.44) **(Supplementary Table 5)**.

Next, we examined the associations between all candidate loci from stage 1 (n=86 SNPs) and overall breast cancer risk using all available breast cancer cases and controls from the studies in stages 1 and 2 (**Figure 1**). We identified genome-wide statistically significant associations with variants at 6q25 (rs9383938, T allele: OR=1.20; P=8.7x10^{-14}), and a recently reported risk locus near the *PTHLH* gene at 12p11 (rs1975930, T allele: OR=1.22; P=1.4x10^{-13})(10). In addition, we observed genome wide significant associations with multiple variants in a gene-desert located at 6q14. Allele C of rs17530068 at 6q14 was associated with increased risk for overall breast cancer risk (OR=1.12; P=1.1x10^{-9}) (**Table 2, Supplementary Figure 3**, **Supplementary Table 4)** and both ER-positive (OR=1.09; P=1.5x10^{-5}) **(Supplementary Table 6)** and ER-negative (OR=1.16, P=2.5x10^{-7}) **(Table 2)** breast cancer. We observed no evidence of risk heterogeneity for rs17530068 by ER status (case-only analysis P=0.53) (**Supplementary Table 6**); study (P_{het}=0.16); or race/ethnicity (P_{het} =0.30) **(Table 2).** Furthermore, rs17530068 appeared more strongly associated with ER-negative, PR-negative, HER2-positive breast cancer (OR=1.26, P=8.0x10^{-3}), than TN breast cancer (OR=1.12, P=0.07), although these differences were not statistically significant (case-only P=0.17) **(Supplementary Table 6)**.

We also evaluated associations for 25 known breast cancer risk markers in European-ancestry women from our study (**Supplementary Table 7 and Supplementary Figure 4**). In our samples 8 of the 13 markers previously associated with both ER-negative and ER-positive disease or with ER-negative disease only (TERT and 19p13.1), were nominally significantly associated (P<0.05) with ER-negative disease. In contrast, none of the 10 markers previously associated with ER-positive disease only were associated with ER-negative disease. A risk score formed by summing the risk alleles at all 25 previously identified loci was significantly associated with ER-negative disease in our study (OR=1.06 (1.04-1.07); P=2.9 x10^{-14}). Risk scores for subsets of markers associated with ER-negative disease only (2 markers) or both ER-negative and ER-positive disease (11 markers) were also significantly associated with ER-negative disease (OR=1.22 (1.14-1.31), P=1.0 x10^{-8} and OR=1.08 (1.05-1.10), P=9.5 x10^{-12}, respectively). A risk score for the subset of loci previously associated with ER-positive disease only (10 markers) was not associated with risk of ER-negative disease (OR=1.02 (1.00-1.04), P=0.08). These score

results provide some confirmation of earlier results and an estimate of the effects of previously-identified breast cancer risk markers on risk of ER-negative disease.

DISCUSSION

We present results from the largest meta-analysis to date to specifically focus on ER-negative disease. We identify two novel loci for breast cancer: 20q11 associated with ER-negative and triple negative, but not ER-positive breast cancer, and 6q14 associated with both ER-positive and ER-negative breast cancer. In addition, we confirm three known regions previously associated with ER-negative (19p13) or ER-negative and ER-positive breast cancer (6q25 and 12p11). Correction for genomic control results in similar but attenuated findings for 20q11-rs2284378 (P_{GC}=2.4x10^{-8}) and 6q14-rs17530068 (P_{GC} =3.2x10^{-9}).

The novel association at 20q11 with ER-negative breast cancer spans the *ASIP, RALY* and *EIF2S2* genes. Agouti signaling protein (product of the *ASIP* gene) was first described to inhibit melanogenesis in human melanocytes in 1997(11). ASIP is a melanocortin 1 receptor (MC1R) ligand that antagonises the function of the transmembrane receptor(12). The variants we identified at 20q11 for breast cancer are highly correlated with variants previously associated with pigmentation traits as well as risk of both cutaneous melanoma and basal cell carcinoma(9), suggesting a possible biological link between these cancers. Further studies have confirmed the importance of the genetic variation spanning the *ASIP* locus, where a variant at 20q11 showed the strongest association with pigmentation and was implicated in a probable linkage disequilibrium (LD) with variants within an *ASIP* regulatory region(13). *EIF2S2* encodes eukaryotic translation initiation factor 2, subunit 2 beta, which is involved in early steps of protein synthesis by forming a ternary complex with GTP and initiator tRNA. The deletion of *Eif2s2* has been associated with suppression of testicular germ cell tumor incidence and recessive lethality in mice(14). The agouti-yellow (*AV*) deletion is a genetic modifier known to suppress testicular germ cell tumor susceptibility in mice and humans. The *AV* mutation deletes both *RALY* and *Eif2s2,* and induces the ectopic expression of *agouti,* all of which are potential testicular germ cell tumor-modifying variations

(14). Both *RALY* and *EIF2S2* are expressed in many tissues including mammary gland(15). SNP rs2284378 was not consistently associated with expression of *EIF2S2*, *RALY*, or *ASIP* in lymphocytes (11), adipocytes or skin cells(16)although there was marginal evidence for association between rs2284378 and *EIF2S2* expression in one study (16)(**Supplementary Table 8**). However, several SNPs in high linkage disequilibrium with SNP rs2284378 (r2>0.8) within a 1MB region were significantly associated with expression of nearby genes *EIF2S2* and *RALY*. Rs4911379 (r^2=0.96) is statistically significantly associated with *EIF2S2* expression in fibroblasts (P=3.6 x10^{-4}) (17)and SNPs rs761238 and rs761236 (r^2=0.85) are associated with *RALY* expression in lymphocytes (P=8.3x10^{-4})(16). An additional 13 SNPs (r^2>0.85) have been associated with expression of *RALY*, *GGTL3*, *DYNLRB1*, and *AK054906* in liver cells, monocytes and lymphoblastoid cell lines (**Supplementary Table 9**). In addition to expression, several enhancer as well as promoter regions defined by overlapping chromatin marks in human mammary epithelial cells were found at 20q11 (**Supplemental Figure 5**). SNPs in high LD with rs2284378 (r^2>0.7), such as rs4911395, rs4911396 and rs1007090, are located in the promoter region of RALY. SNPs rs6142101, rs6087557, and rs4911408 (r^2>0.7) are present in the promoter region of *EIF2S*2, and rs1054534, rs1555075, rs2268086, rs2268088, rs4911401, rs2284388, rs2284389 and rs932388 are located in predicted enhancer regions in introns of *RALY*. Thus, variants at 20q11 may influence expression of multiple genes in mammary epithelial cells, as has been seen in prostate cancer (18).

In contrast, rs17530068 at 6q14 is located in a gene desert with no evidence of an open/active regulatory region in human mammary epithelial cells (**Supplementary Figure 6**). The closest gene (~262kb), family with sequence similarity 46, member A (*FAM46A*/*C6orf37*), encodes a protein of unknown function. Five SNPs in this region in low linkage disequilibrium with SNP rs17530068 (r2<0.02) were associated with expression of *IBTK* in lymphoblastoid cell lines (**Supplementary Table 10).** Additional studies of both of these novel regions will be necessary to identify the underlying biologically relevant variant/s.

SNP rs17530068 at chromosome 6q14 was associated with overall breast cancer risk and showed no differential association depending on ER status. The association of SNP rs2284378 at 20q11, however, was stronger for ER-negative than ER-positive breast cancer. This finding underscores the importance of investigating genetic variants for specific subtypes of breast cancer, as this locus had not been previously identified in the many GWAS of breast cancer to date that did not focus on this specific breast cancer subtype. The etiology of ER-negative disease is largely unknown. Identifying new loci associated with ER-negative and TN breast cancer will continue to provide insight into the biological mechanisms underlying this more aggressive form of breast cancer, and could result in improvements in risk prediction and treatment.

MATERIALS AND METHODS

Study populations

Stage 1 included the studies of the NCI Breast and Prostate Cancer Cohort Consortium (BPC3), Triple Negative Breast Cancer Consortium (TNBCC) and African American Breast Cancer Consortium (AABC). The BPC3 study includes 2,188 ER-negative cases and 25,519 controls, AABC includes 3,153 cases (1,004 ER-negative) and 2,745 controls from 9 studies and TNBCC includes 1,562 cases and 3,399 controls from 15 studies **(Supplementary Table 1)**. Replication studies include 886 cases (84 ER-negative) and 830 controls from a GWAS of breast cancer in Japanese (MEC-JPT) women and 546 cases (112 ER-negative) and 558 controls from a GWAS of breast cancer in Latino (MEC-LAT) women in the Multiethnic Cohort (MEC), 992 (188 ER-negative) and 640 controls from the San Francisco Bay Area Breast Cancer Study (SFBCS) and the Northern California Breast Cancer Family Registry (NC-BCFR), and 8,785 (562 ER-negative) and 14,029 controls from eight combined GWAS of breast cancer from BCAC. All participants in these studies have provided written consent for the research and approval for the study was obtained from the ethical review board from all local institutions. A description of each participating study has been provided in supplementary material.

Stage 1 genotyping and quality control

Genotyping in AABC was conducted using the Illumina Human1M-Duo BeadChip. Of the 5,984 samples in the AABC Consortium (3,153 cases and 2,831 controls), we attempted genotyping of 5,932, removing samples (n=52) with DNA concentrations <20 ng/ul. Following genotyping, we removed samples based on the following exclusion criteria: 1) unknown replicates (≥98.9% genetically identical) that we were able to confirm, n=15); 2) unknown replicates pair or triplicate removed, n=14); 3) samples with call rates <95% after a second attempt (n=100); 4) samples with ≤ 5% African ancestry (n=36) (discussed below); and, 5) samples with <15% mean heterozygosity of SNPs in the X chromosome and/or similar mean allele intensities of SNPs on the X and Y chromosomes (n=6). In the analysis, we removed SNPs with <95% call rates (n=21,732) or minor allele frequencies (MAFs) <1% (n=80,193). The concordance rate for blinded duplicates was 99.95%. We also eliminated SNPs with genotyping concordance rates <98% based on the replicates (n=11,701). The final analysis dataset included 1,043,036 SNPs genotyped on 3,016 cases (988 ER-negative, 1520 ER-positive, and the remaining 508 cases with unknown ER status) and 2,745 controls, with an average SNP call rate of 99.7% and average sample call rate of 99.8%.

Genotyping for the TNBCC GWAS was conducted on 1,718 cases from 10 studies (ABCTB, BBCC, DFCI, FCCC, GENICA, MARIE, MCBCS, MCCS, POSH, SBCS) using the Illumina 660-Quad SNP array. In addition, a subset of MARIE cases (n=52) were genotyped using the Illumina CNV370 SNP array. HEBCS cases (n=85) were genotyped using the Illumina 550 SNP array and population allele and genotype frequencies on healthy population controls (n=222) were genotyped on Illumina 370 SNP array, and obtained from the NordicDB, a Nordic pool and portal for genome-wide control data(19) from the Finnish Genome Center. GWAS data for public controls (n=3,448) were generated using the following arrays: Illumina 660-Quad SNP array (QIMR), Illumina 550 SNP array (CGEMS), Illumina 550 SNP array (KORA), and Illumina 1.2M (WTCCC). These GWAS data were independently evaluated by an iterative QC process with the following exclusion criteria: minor allele frequency (MAF) <0.01, call rate <95%, HWE p-value $<1x10^{-7}$ among controls and sample call rate <98%. In total, we excluded previously unknown replicates (n=2) and samples with call rates <98% (n=83), samples that failed sex

check (n=10), cases identified as non-triple negative breast cancer (n=20) and related samples (n=27). We removed SNPs with <95% call rates or MAF <5%. Because a number of our samples were genotyped at different locations, we removed SNPs if there was a difference >0.10 between the study allele frequency and the median frequency across all studies. Eigensoft software which uses principle component analysis (PCA) was used to evaluate confounding due to population stratification. We removed 101 subjects that did not cluster with the CEU HapMap Phase 2 samples, and a further 179 controls were removed which overlapped with CGEMS/NHS controls in BPC3, resulting in 1,562 cases and 3,399 controls in the GWAS analyses.

BPC3 GWAS genotyping was conducted at three genotyping centers (NCI Core Genotyping Facility, USA; University of Southern California, USA; and Imperial College London, UK). Subjects from CPSII, EPIC, MEC, PLCO, and PBCS were genotyped using the Illumina Human 660k-Quad SNP array(Illumina, Inc), NHSI/NHSII and part of the PLCO study were genotyped previously using the Illumina Human 550 SNP array(Illumina, Inc) (20). SNPs were filtered and removed based on deviations from Hardy-Weinberg proportions in control subjects (p<10e-5), autosomal SNPs with MAF of less than 5% and completion rate less than 95%. Samples were excluded based on genotyping call rates less than 95% (n=195), samples with extreme heterozygosity were excluded from the analysis (n=35), sex discordance (n=3), unexpected duplicates and relatedness (n=6), Subjects with evidence of significant non-European ancestry and population structure were also excluded. Non-European ancestry was assessed utilizing a subset of unlinked, population informative SNPs (21). Individuals determined to have less than 80% European ancestry were excluded from future analyses (n=16). The average concordance rate of blinded duplicates was 99.95%. In order to resolve a more detailed population substructure, PCA was conducted using *struct.pca* module of GLU (http://code.google.com/p/glu-genetics/). PCA was only performed in subjects with over 80% European ancestry. Furthermore, 958 controls from NHS (CGEMS) were removed from BPC3 analyses due to overlap between TNBCC and BPC3 studies. The overall number of cases and controls after all exclusions which contributed to the stage 1 analysis were 1,998 cases and 2,305 controls.

The WHS cohort subjects in BPC3 were previously genotyped using the Human-Hap300 Duo-plus BeadChip (22). Among the final 23,294 individuals of verified European ancestry, genotypes for a total of 2,608,509 SNPs were imputed from the experimental genotypes and LD relationships implicit in the HapMap r. 22 CEU samples. WHS contributed 190 cases and 23,214 control subjects to stage 1. WHS was meta-analyzed with the remaining BPC3 studies contributing a total of 2,188 cases and 25,519 control subjects to stage 1 analysis.

SNP rs2284378 and rs17530068 were genotyped in all stage 1 studies.

Stage 2 genotyping and quality control

The San Francisco Bay Area Breast Cancer Study (SFBCS)(23)and the Northern California Breast Cancer Family Registry (NC-BCFR)(24) study samples were genotyped with the Affymetrix 6.0 array according to the manufacturer's instructions (https://www.affymetrix.com) in the laboratory of Esteban Gonzalez Burchard at UCSF. A total of 15 cases and 30 controls were excluded from the SFBCS and NC-BCFR sample set that had a genotyping call rate <95% or showed either known or cryptic relatedness. The final sample included in the analysis was 992 cases (188 ER-negative cases) and 640 controls. Imputation was conducted with the program BEAGLE, with all unrelated HapMap Phase II samples included as references (http://hapmap.ncbi.nlm.nih.gov).

GWAS of breast cancer in Latino (MEC-LAT) and Japanese (MEC-JPT) samples from the MEC were genotyped with the Illumina 660W array at USC. For MEC-LAT, we excluded 48 samples from the MEC that had a genotyping call rate of <95% and 34 that showed either known or cryptic relatedness. The final MEC-LAT sample included 546 (112 ER-negative) and 558 controls. With similar exclusions, the final MEC-JPT sample included 886 (84 ER-negative) and 830 controls.

The BCAC combined GWAS includes primary genotype data from eight breast cancer GWAS in populations of European ancestry (ABCFS, BBCS, , GC-HBOC, MARIE, HEBCS, SASBAC, UK2, DFBBCS). All studies were genotyped with various versions of Illumina arrays, except GC-HBOC which was performed with the Affymetrix 5.0 (cases) and 6.0 (controls) arrays. Standard QC was performed on all scans. Specifically, all individuals with low call rate (<95%), extreme high or low

heterozygosity ($P<10^{-5}$), and all individuals evaluated to be of non-European ancestry (>15% non-European component, by multidimensional scaling using the three Hapmap2 populations as a reference) were excluded. SNPs with call rate <95%; call rate <99% and MAF<5%, all SNPs with MAF<1%, and SNPs with genotype frequencies departing from Hardy-Weinberg equilibrium at $P<10^{-6}$ in controls or $P<10^{-12}$ in cases were also excluded. Data were imputed for ~2.6M SNPs for all scans using Mach v1.0 with HapMap version 2 CEU as a reference. BBCS and UK2 used the same control data (WTCCC2). These studies were imputed separately. For the combined analysis, the control set was divided randomly between the two studies, in proportion to the size of case series, to provide disjoint strata. Estimated per-allele ORs and standard errors were generated from the imputed genotypes using Probabel (25).

SNP rs2284378 and rs17530068 were genotyped in all stage 2 studies except SFBCS and NC-BCFR where they were imputed. Both SNPs were genotyped by TaqMan in 483 samples from these studies and genotype concordance versus imputed genotypes was 93.3% for rs2284378 and 94.9% for rs17530068.

Taqman gentoyping in BPC3 for SNP rs2284378 and SNP rs17530068

In BPC3, genotyping of SNP rs2284378 and rs17530068 was performed for all available breast cancer cases and controls by TaqMan in four laboratories (CPS-II and MEC at the University of Southern California; NHS and WHS at Harvard University; EPIC at the German Cancer Research Center in Heidelberg; and PLCO at the NCI/Core Genotyping Facility). All studies typed SNP rs17530068; however for SNP rs2284378, PLCO and CPS-II typed a proxy SNP rs6059651 (r^2 =1, D'=1). The concordance for the Taqman genotyping data with that generated from Illumina for stage 1 ER-negative cases and controls was 0.997 for rs17530068 and 0.986 for rs2284378 for CPS2, MEC, NHS, EPIC and PLCO. The genotype concordance versus imputed for WHS was 95% for rs2284378 and 97% for rs17530068

Statistical analysis

In AABC, we tested for gene dosage effects in models adjusted for age, study and eigenvectors 1-10. Odds ratios (OR) and 95% confidence intervals (95% CI) were estimated using unconditional logistic regression. In TNBCC, unconditional logistic regression was used to assess single SNP associations also assuming a log-additive model, adjusting for country and the first two principal components. In BPC3, unconditional logistic regression model was used to assess single SNP associations adjusting for age categories and the top 6 eigenvectors.

In both AABC and TNBCC, phased haplotype data from the founders of the CEU and YRI HapMap Phase 2 samples (build 21) were used to infer LD patterns in order to impute untyped markers. For BPC3, Hapmap Phase 2 (release 21) and Hapmap Phase 3 were used to impute untyped markers. For all studies, genome-wide imputation was carried out using the software MACH. Filtered from the analysis were SNPs with Rsq<0.3 and MAF <1%.

We conducted a fixed effect meta-analysis of AABC, TNBCC and BPC3 using the inverse variance weighted method. The number of SNPs available for meta-analysis from AABC, TNBCC and BPC3 in stage 1 were 3,055,415, 2,134,490 and 245,3207 respectively. The union of these three data sets was meta-analyzed using the program METAL. We conducted *in silico* replication of 86 SNP with p-values $\leq 10^{-5}$ in stage 1 in the stage 2 studies, and a meta-analysis of these SNPs from stage 1 and 2 for both ER- negative and overall breast cancer. P-values from our top two loci were corrected for genomic inflation (P_{GC}) using the lambda value from the overall meta-analysis. Testing for heterogeneity by study was evaluated using the Q-statistic. Case-only analyses were performed to test for differences in the association by tumor subtypes, study and race/ethnicity.

The association between risk scores of 25 previously-identified breast cancer risk alleles and risk of breast cancer in our samples was calculated using meta-regression, assuming the per-allele odds ratio was constant across the markers analyzed. This is equivalent to combining the summary log odds ratio estimates at independent loci using inverse-variance weighted meta-analysis. The overlap between subjects contributing to this study and those contributing to previous studies varied from marker to marker (e.g. the TNBCC contributed to the initial report on rs8170 (5) and the BPC3 and TNBCC

contributed to the initial report on the *TERT* locus (6).Thus, the results could be overestimates since some of the studies here contributed to the discovery of these 25 loci.

Functional analysis

Expression quantitative trait loci (eQTL) were assessed for all SNPs in the chromosome 6 and 20 loci using the GTEX database (http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi), University of Chicago eQTL Browser (http://eqtl.uchicago.edu) and Genevar (http://www.sanger.ac.uk/resources/software/genevar/) (26)

In an attempt to identify functionality at the two novel breast cancer risk loci, we used the open-source R/Bioconductor package FunciSNP version 0.99(27), which systematically integrates the 1,000 Genomes Project SNP data (April 2012 data release) with chromatin features of interest. For each of the two novel breast cancer markers, we analyzed all SNPs with an r^2 value > 0.5 with each index SNP in the 1,000 Genomes Project EUR populations in a 1MB window around each index variant. We assessed whether these SNPs were co-located with 12 different chromatin features generated by next-generation sequencing technologies, which capture open chromatin regions, promoters, and enhancers genome-wide in human mammary epithelial cells (HMEC) as well as known DNaseI hypersensitive locations, FAIRE-seq peaks, and CTCF binding sites from more than 100 different cell types, which were collected in ENCODE data(28). We utilized the UCSC Genome Browser (http://genome.ucsc.edu/) to illustrate the correlated SNPs, which overlap chromatin features as well as chromatin feature tracks (**Supplemental Figures 5-6**).

FUNDING

AABC was supported by a Department of Defense Breast Cancer Research Program Era of Hope Scholar Award to CAH [W81XWH-08-1-0383] and the Norris Foundation. Each of the participating **AABC** studies was supported by the following grants: MEC (National Institutes of Health grants R01-CA63464 and R37-CA54281); CARE (National Institute for Child Health and Development grant NO1-HD-3-3175), WCHS (U.S. Army Medical Research and Material Command (USAMRMC) grant DAMD-17-01-0-0334, the National Institutes of Health grant R01-CA100598, and the Breast Cancer Research Foundation, SFBCS (National Institutes of Health grant R01-CA77305 and United States Army Medical Research Program grant DAMD17-96-6071), NC-BCFR (National Institutes of Health grant U01-CA69417), CBCS (National Institutes of Health Specialized Program of Research Excellence in Breast Cancer, grant number P50-CA58223, and Center for Environmental Health and Susceptibility, National Institute of Environmental Health Sciences, National Institutes of Health, grant number P30-ES10126), PLCO (Intramural Research Program, National Cancer Institute, National Institutes of Health), and NBHS (National Institutes of Health grant R01-CA100374),WFBC (National Institutes of Health grant R01-CA73629). TheNC-BCFR is one of 6 sites participating in The Breast Cancer Family Registry (BCFR) which was supported by the National Cancer Institute, National Institutes of Health under RFA CA-06-503 and through cooperative agreements with members of the Breast Cancer Family Registry and Principal Investigators. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or the BCFR.

The **TNBCC** studies were supported by the following grants: MCBCS (National Institutes of Health Grants CA122340 and a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), and the Breast Cancer Research Foundation (BCRF); MARIE (Deutsche Krebshilfe e.V., grant number 70-2892-BR I, the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany grant 01KH0402); GENICA (Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8,

01KW9977/0, 01KW0114, and the Robert Bosch Foundation Stuttgart, Germany;MCCS (Australian NHMRC grants 209057, 251553 and 504711 and infrastructure provided by the Cancer Council Victoria); SBCS (Breast Cancer Campaign (grant 2004Nov49 to AC), and by Yorkshire Cancer Research core funding); DFCI (DFCI Breast Cancer SPORE NIH P50 CA089393); POSH (Cancer Research UK); DEMOKRITOS (Hellenic Cooperative Oncology Group research grant (HR R_BG/04) and the Greek General Secretary for Research and Technology (GSRT) Program, Research Excellence II, funded at 75% by the European Union); BBCC (Dr. Mildred Scheel Stiftung of the Deutsche Krebshilfe e.V.); BBCS (Cancer Research UK and Breakthrough Breast Cancer and NHS funding to the NIHR biomedical Research Centre and the National Cancer Research Network (NCRN); LMBC (European Union Framework Programme 6 Project LSHC-CT-2003-503297 (the Cancerdegradome) and by the 'Stichting tegen Kanker' (232-2008); OBCS (Finnish Cancer Foundation, the Sigrid Juselius Foundation, the Academy of Finland, the University of Finland, and Oulu University Hospital); HEBCS (Helsinki University Central Hospital Research Fund, Academy of Finland (132473), the Finnish Cancer Society, The Nordic Cancer Union and the Sigrid Juselius Foundation); FCCC (U01CA69631, 5U01CA113916, the University of Kansas Cancer Center and the Kansas Bioscience Authority Eminent Scholar Program); RPCI (RPCI DataBank and BioRepository (DBBR), a Cancer Center Support Grant Shared Resource (P30 CA016056-32); SKKDKFZS (Deutsches Krebsforschungszentrum); BIGGS (National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust); ABCTB (National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation); ABCS (Dutch Cancer Society grant number 2009-4363); KARBAC (The Stockholm Cancer Society).

BPC3 is supported by the US National Institutes of Health, National Cancer Institute under cooperative agreements U01-CA98233 (NHS, NHSII, WHS), U01-CA98710 (CPS2), U01-CA98216 (EPIC), U01-CA98758 (MEC) and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics (PLCO). The authors thank Drs. Christine Berg and

Philip Prorok, Division of Cancer Prevention, NCI, the screening center investigators and staff of the PLCO Cancer Screening Trial, Mr. Thomas Riley and staff at Information Management Services, Inc., and Ms. Barbara O'Brien and staff at Westat, Inc. for their contributions to the PLCO Cancer Screening Trial.

The WHS is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute and CA 047988 from the National Cancer Institute, the Donald W. Reynolds Foundation and the Fondation Leducq, with collaborative scientific support and funding for genotyping provided by Amgen.

The **UK2** GWAS was funded by Wellcome Trust and Cancer Research UK. The WTCCC was funded by the Wellcome Trust. **BCAC** is funded by CR-UK [C1287/A10118, C1287/A12014] and by the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS). Meetings of the BCAC have been funded by the European Union COST programme [BM0606]. The **ABCFS** study was supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia), and the National Cancer Institute, National Institutes of Health under RFA-CA-06-503 and through cooperative agreements with members of the Breast Cancer Family Registry (BCFR) and the Principle Investigators. The University of Melbourne (U01 CA69638) contributed data to this study. The content of this manuscript does not necessarily reflect the views or the policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government or the BCFR. We extend our thanks to the many women and their families that generously participated in the Australian Breast Cancer Family Study and consented to us accessing their pathology material. JLH is a National Health and Medical Research Council Australia Fellow. MCS is a National Health and Medical Research Council Senior Research Fellow. JLH and MCS are both group leaders of the Victoria Breast Cancer Research Consortium. The **BBCS** is funded by Cancer Research UK and Breakthrough Breast Cancer and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BBCS GWAS received funding from The Institut National de Cancer. The

DFBBCS GWAS was funded by The Netherlands Organisation for Scientific Research (NWO) as part of a ZonMw/VIDI grant number 91756341. We thank Muriel Adank for selecting the samples and Margreet Ausems, Christi van Asperen, Senno Verhoef, and Rogier van Oldenburg for providing samples from their Clinical Genetic centers. The **GC-HBOC** was supported by Deutsche Krebshilfe [107054], the Dietmar-Hopp Foundation, the Helmholtz society and the German Cancer Research Centre (DKFZ). The GC-HBOC GWAS was supported by the German Cancer Aid (grant no. 107352). The **MARIE** study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I], the Hamburg Cancer Society, the German Cancer Research Center and the genotype work in part by the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. MARIE would like to thank Tracy Slanger and Elke Mutschelknauss for their valuable contributions, and S. Behrens, R. Birr, M.Celik, U. Eilber, B. Kaspereit, N. Knese and K. Smit for their excellent technical assistance. The **SASBAC** study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. **CGEMS.** The Nurses' Health Studies are supported by NIH grants CA 65725, CA87969, CA49449, CA67262, CA50385 and 5UO1CA098233. The **HEBCS** study has been financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (132473), the Finnish Cancer Society, The Nordic Cancer Union and the Sigrid Juselius Foundation. The population allele and genotype frequencies were obtained from the data source funded by the Nordic Center of Excellence in Disease Genetics based on samples regionally selected from Finland, Sweden and Denmark. We thank Drs. Kirsimari Aaltonen, Päivi Heikkilä and Tuomas Heikkinen and RN Hanna Jäntti and Irja Erkkilä for their help with the HEBCS data and samples.

The biofeature analysis was supported by NIH grant CA109147.

The breast cancer **GWAS in Japanese and Latinos** in the MEC (MEC-LAT and MEC-JPT) were supported by NIH grants CA132839, CA54281 and CA63464. Genotyping of the Latino breast cancer cases and controls from SFBCS and NC-BCFR was supported by NIH grant **CA120120**..

ACKNOWLEDGEMENTS

We thank the women who volunteered to participate in each study. We also thank Madhavi Eranti, Andrea Holbrook, Paul Poznaik, and David Wong from the University of Southern California for their technical support. We would also like to acknowledge co-investigators from the WCHS study: Dana H. Bovbjerg (University of Pittsburgh), Lina Jandorf (Mount Sinai School of Medicine) and Gregory Ciupak, Warren Davis, Gary Zirpoli, Song Yao and Michelle Roberts from Roswell Park Cancer Institute.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

1 Parl, F.F., Schmidt, B.P., Dupont, W.D. and Wagner, R.K. (1984) Prognostic significance of estrogen receptor status in breast cancer in relation to tumor stage, axillary node metastasis, and histopathologic grading. *Cancer*, **54**, 2237-2242.

2 Yang, X.R., Chang-Claude, J., Goode, E.L., Couch, F.J., Nevanlinna, H., Milne, R.L., Gaudet, M., Schmidt, M.K., Broeks, A., Cox, A. *et al.* (2011) Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J. Natl. Cancer. Inst.*, **103**, 250-263.

3 Milne, R.L. and Antoniou, A.C. (2011) Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. *Ann. Oncol.*, **22 Suppl 1**, i11-17.

4 Broeks, A., Schmidt, M.K., Sherman, M.E., Couch, F.J., Hopper, J.L., Dite, G.S., Apicella, C., Smith, L.D., Hammet, F., Southey, M.C. *et al.* (2011) Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum. Mol. Genet.*, **20**, 3289-3303.

5 Antoniou, A.C., Wang, X., Fredericksen, Z.S., McGuffog, L., Tarrell, R., Sinilnikova, O.M., Healey, S., Morrison, J., Kartsonaki, C., Lesnick, T. *et al.* (2010) A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat. Genet.*, **42**, 885-892.

6 Haiman, C.A., Chen, G.K., Vachon, C.M., Canzian, F., Dunning, A., Millikan, R.C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C.B. *et al.* (2011) A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.*, **43**, 1210-1214.

7 Stevens, K.N., Vachon, C.M., Lee, A.M., Slager, S., Lesnick, T., Olswold, C., Fasching, P.A., Miron, P., Eccles, D., Carpenter, J.E. *et al.* (2011) Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res.*, **71**, 6240-6249.

8 Zheng, W., Long, J., Gao, Y.T., Li, C., Zheng, Y., Xiang, Y.B., Wen, W., Levy, S., Deming, S.L., Haines, J.L. *et al.* (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.*, **41**, 324-328.

9 Gudbjartsson, D.F., Sulem, P., Stacey, S.N., Goldstein, A.M., Rafnar, T., Sigurgeirsson, B., Benediktsdottir, K.R., Thorisdottir, K., Ragnarsson, R., Sveinsdottir, S.G. *et al.* (2008) ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat. Genet.*, **40**, 886-891.

10 Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C. *et al.* (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.*, **6**, e107.

11 Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M. *et al.* (2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.*, **8**, e1002639.

12 Scherer, D. and Kumar, R. (2010) Genetics of pigmentation in skin cancer--a review. *Mutat. Res.*, **705**, 141-153.

13 Barrett, J.H., Iles, M.M., Harland, M., Taylor, J.C., Aitken, J.F., Andresen, P.A., Akslen, L.A., Armstrong, B.K., Avril, M.F., Azizi, E. *et al.* (2011) Genome-wide association study identifies three new melanoma susceptibility loci. *Nat. Genet.*, **43**, 1108-1113.

14 Heaney, J.D., Michelson, M.V., Youngren, K.K., Lam, M.Y. and Nadeau, J.H. (2009) Deletion of eIF2beta suppresses testicular cancer incidence and causes recessive lethality in agouti-yellow mice. *Hum. Mol. Genet.*, **18**, 1395-1404.

15 Mosca, E., Alfieri, R., Merelli, I., Viti, F., Calabria, A. and Milanesi, L. (2010) A multilevel data integration resource for breast cancer study. *BMC Syst. Biol.*, **4**, 76.

16 Nica, A.C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K. *et al.* (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.*, **7**, e1002003.

17 Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M. *et al.* (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, **325**, 1246-1250.

18 Pomerantz, M.M., Shrestha, Y., Flavin, R.J., Regan, M.M., Penney, K.L., Mucci, L.A., Stampfer, M.J., Hunter, D.J., Chanock, S.J., Schafer, E.J. *et al.* (2010) Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. *PLoS Genet.*, **6**, e1001204.

19 Leu, M., Humphreys, K., Surakka, I., Rehnberg, E., Muilu, J., Rosenstrom, P., Almgren, P., Jaaskelainen, J., Lifton, R.P., Kyvik, K.O. *et al.* (2010) NordicDB: a Nordic pool and portal for genome-wide control data. *Eur. J. Hum. Genet.*, **18**, 1322-1326.

20 Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.*, **39**, 870-874.

21 Yu, K., Wang, Z., Li, Q., Wacholder, S., Hunter, D.J., Hoover, R.N., Chanock, S. and Thomas, G. (2008) Population substructure and control selection in genome-wide association studies. *PLoS One.*, **3**, e2551.

22 Ridker, P.M., Chasman, D.I., Zee, R.Y., Parker, A., Rose, L., Cook, N.R. and Buring, J.E. (2008) Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin. Chem.*, **54**, 249-255.

23 John, E.M., Schwartz, G.G., Koo, J., Wang, W. and Ingles, S.A. (2007) Sun exposure, vitamin D receptor gene polymorphisms, and breast cancer risk in a multiethnic population. *Am. J. Epidemiol.*, **166**, 1409-1419.

24 John, E.M., Miron, A., Gong, G., Phipps, A.I., Felberg, A., Li, F.P., West, D.W. and Whittemore, A.S. (2007) Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. *JAMA*, **298**, 2869-2876.

25 Aulchenko, Y.S., Struchalin, M.V. and van Duijn, C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, **11**, 134.

26 Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474-2476.

27 Coetzee, S.G., Rhie, S.K., Berman, B.P., Coetzee, G.A. and Noushmehr, H. (2012) FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic Acids Res.*, June 22 [Epub ahead of print].

28 Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43-49.

LEGEND

Figure 1. Multi-stage study design.

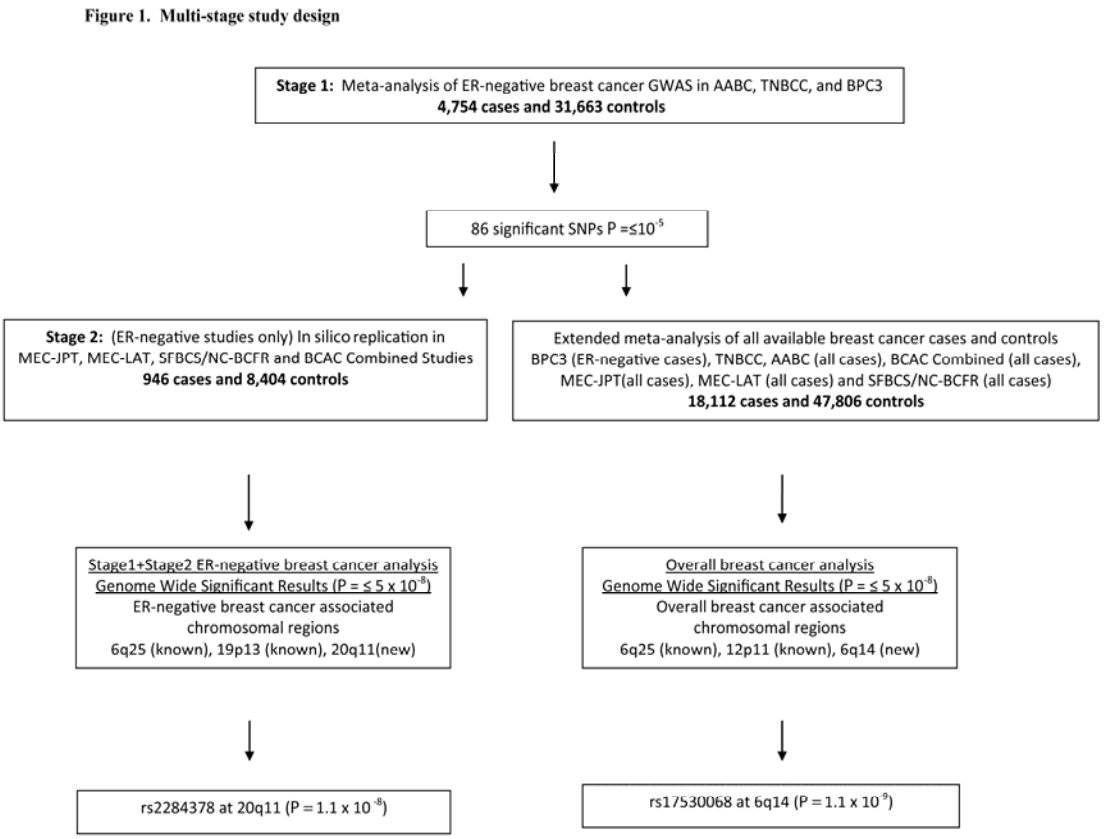

Table 1. Association of SNP rs2284378 (T/C) at chromosome 20q11 and breast cancer risk by study and race/ethnicity

Consortium/ Study	Race/ Ethnicity	Case/ control[a]	RAF (T allele)[b]	OR (95% CI)[c]	*P*-value[d]	$P_{Het\text{-}study}$/ $P_{Het\text{-}race}$[e]
Stage 1 ER-negative cases versus controls						
BPC3	European	2,188/25,519	0.31	1.14 (1.05-1.24)	0.0028	
TNBCC	European	1,478/3,345	0.33	1.18(1.07-1.30)	0.0010	
AABC	African	1,004/2,744	0.16	1.19 (1.03-1.37)	0.020	
Stage 1		**4,670/31,608**		**1.16 (1.09-1.23)**	**$6.5x10^{-7}$**	**0.85/0.76**
Stage 2 ER-negative cases versus controls						
BCAC Combined GWAS	European	562/6410	0.35	1.10 (0.96-1.25)	0.17	
MEC-JPT	Japanese	84/830	0.26	0.99 (0.68-1.44)	0.95	
MEC-LAT	Latino	112/553	0.29	1.27 (0.94-1.71)	0.13	
SFBCS/NC-BCFR	Latino	188/611	0.29	1.45 (1.13-1.87)	0.004	
Stage 2 (ER-negative)		946/8,404		1.16 (1.04-1.29)	0.0048	**0.98/0.12**
Stage 1+2 (ER-negative)		**5,616/40,012**		**1.16 (1.10-1.22)**	**$1.1x10^{-8}$**	**0.54/0.28**
All breast cancer cases versus controls						
AABC	African	3,016/2,745	0.16	1.06 (0.95-1.17)	0.30	
BCAC Combined GWAS	European	8,785/10,142	0.35	1.04 (0.99-1.09)	0.11	
MEC-JPT	Japanese	886/830	0.26	1.08 (0.91-1.24)	0.46	
MEC-LAT	Latino	546/553	0.29	1.24 (1.03-1.48)	0.021	
SFBCS/NC-BCFR	Latino	970/611	0.29	1.23 (1.05-1.44)	0.011	
Stage 2 (all cases)		14,202/14,880		1.06 (1.02-1.10)	0.0025	**0.14/0.073**
Stage 1+2 (all cases)		**17,869/43,745**		**1.08 (1.05-1.12)**	**$1.3x10^{-6}$**	**0.056/0.19**

[a]Number of cases and controls with genotype data for rs2284378. [b]Risk Allele Frequency (RAF) in controls. [c]Adjusted for age, study and principal components in AABC. Adjusted for age and country in TNBCC. Adjusted for age categories and top 6 eigenvectors in BPC3. Adjusted for age and top 10 eigenvectors in MEC-JPT, MEC-LAT and SFBCS/NC-BCFR studies. Combined analysis (Stage1, Stage2 and Stage 1+2) are from the meta-analysis. [d]*P* for trend (1-d.f.). [e]*P* for heterogeneity by study and race/ethnicity, respectively.

Table 2. Association of SNP rs17530068 (C/T) at chromosome 6q14 and breast cancer risk by study and race/ethnicity

Consortium/ Study	Race/ Ethnicity	Case/ control[a]	RAF (C allele)[b]	OR (95% CI)[c]	*P*-value[d]	$P_{Het\text{-}study}$/ $P_{Het\text{-}race}$[e]
Stage 1 ER-negative cases versus controls						
BPC3	European	2,188/25,519	0.24	1.23 (1.12-1.35)	$2.23x10^{-5}$	
TNBCC	European	1,478/3,345	0.24	1.13 (1.02-1.26)	0.023	
AABC	African	1,004/2,745	0.07	1.07 (0.86-1.34)	0.54	
Stage 1		**4,670/31,609**		**1.17 (1.09-1.26)**	**$3.5x10^{-6}$**	**0.37/0.41**
Stage 2 ER-negative cases versus controls						
BCAC combined GWAS	European	562/6,410	0.22	1.09 (0.95-1.25)	0.24	
MEC-JPT	Japanese	84/830	0.19	1.16 (0.79-1.71)	0.45	
MEC-LAT	Latino	112/553	0.23	1.06 (0.75-1.50)	0.73	
SFBCS/NC-BCFR	Latino	188/611	0.22	1.40 (1.07-1.84)	0.014	
Stage 2 (ER-negative)		946/8,404		1.14 (1.02-1.28)	0.022	**0.41/0.52**
Stage 1+2 (ER-negative)		**5,616/40,013**		**1.16 (1.10-1.23)**	**$2.5x10^{-7}$**	**0.54/0.78**
All breast cancer cases versus controls						
AABC	African	3,016/2,745	0.07	1.04 (0.89-1.21)	0.63	
BCAC combined GWAS	European	8,785/10,142	0.22	1.08 (1.02-1.14)	0.0021	
MEC-JPT	Japanese	886/830	0.19	1.13 (0.96-1.34)	0.14	
MEC-LAT	Latino	546/553	0.23	1.21 (0.99-1.47)	0.056	
SFBCS/NC-BCFR	Latino	970/611	0.22	1.27 (1.07-1.51)	0.006	
Stage 2 (all cases)		14,203/14,881		1.10 (1.05-1.15)	$1.8x10^{-5}$	**0.31/0.20**
Stage 1+2 (all cases)		**17,869/43,745**		**1.12 (1.08-1.16)**	**$1.1x10^{-9}$**	**0.16/0.30**

[a]Number of cases and controls with genotype data for rs17530068. [b]Risk Allele Frequency (RAF) in controls. [c]Adjusted for age, study and principal components in AABC. Adjusted for age and country in TNBCC. Adjusted for age categories and top 6 eigenvectors in BPC3. Adjusted for age and top 10 eigenvectors in MEC-JPT, MEC-LAT and SFBCS/NC-BCFR studies. Combined analysis (Stage1, Stage2 and Stage 1+2) are from the meta-analysis. [d]*P* for trend. [e]*P* for heterogeneity by study and race/ethnicity, respectively.

ABBREVIATIONS

ER=Estrogen Receptor

PR=Progesterone Receptor

SNP=Single nucleotide polymorphism

GWAS=Genome-wide Association Study

OR=Odds Ratio

BPC3=NCI Breast and Prostate Cancer Cohort Consortium

TNBCC=Triple Negative Breast Cancer Consortium

AABC=African American Breast Cancer Consortium