A common variant at the *TERT-CLPTM1L* locus is associated with estrogen receptor–negative breast cancer

© 2011 Nature America, Inc. All rights reserved.

Estrogen receptor (ER)-negative breast cancer shows a higher incidence in women of African ancestry compared to women of European ancestry. In search of common risk alleles for ER-negative breast cancer, we combined genome-wide association study (GWAS) data from women of African ancestry (1,004 ER-negative cases and 2,745 controls) and European ancestry (1,718 ER-negative cases and 3,670 controls), with replication testing conducted in an additional 2,292 ER-negative cases and 16,901 controls of European ancestry. We identified a common risk variant for ER-negative breast cancer at the *TERT-CLPTM1L* locus on chromosome 5p15 (rs10069690: per-allele odds ratio (OR) = 1.18 per allele, $P = 1.0 \times 10^{-10}$). The variant was also significantly associated with triple-negative (ER-negative, progesterone receptor (PR)-negative and human epidermal growth factor-2 (HER2)-negative) breast cancer (OR = 1.25, $P = 1.1 \times 10^{-9}$), particularly in younger women (<50 years of age) (OR = 1.48, $P = 1.9 \times 10^{-9}$). Our results identify a genetic locus associated with estrogen receptor negative breast cancer subtypes in multiple populations.

Compared to women of European ancestry, women of African descent are more likely to be diagnosed with ER-negative breast cancer[1]. ER-negative tumors and triple-negative tumors are observed at even higher rates among African women currently residing in Africa[2], suggesting a genetic component to the high risk of ER-negative phenotypes in women of African descent. Similarly, ER-negative breast cancers and triple-negative breast cancers are also the predominant histological subtypes in women with germline mutations in *BRCA1* (ref. 3). The enrichment for ER-negative disease in this genetically predisposed population also suggests the existence of additional genetic factors that contribute to the risk of ER-negative disease. Support for the presence of these factors was recently provided by a GWAS of breast cancer in *BRCA1* mutation carriers, in which a common risk variant for ER-negative breast cancer on chromosome 19p13 was identified that also was significantly associated with ER-negative and triple-negative disease in the general population[4].

To search for genetic risk factors for ER-negative breast cancer phenotypes, we combined results from a GWAS of breast cancer in African-American women (African American Breast Cancer Consortium (AABC): 3,016 cases (1,004 with ER-negative disease) and 2,745 controls) with results from a GWAS of triple-negative breast cancer in women of European ancestry (Triple-Negative Breast Cancer Consortium (TNBCC): 1,718 cases and 3,670 controls). Genotyping in AABC was conducted with the Illumina Infinium 1M Duo. In TNBCC, cases were genotyped with the Illumina 660W array, a subset of cases from the Mammary Carcinoma Risk Factor Investigation (MARIE) component were genotyped using the Illumina CNV370 SNP array, and cases and controls from the Helsinki Breast Cancer Study (HEBCS) component were genotyped using the Illumina 550-Duo SNP array. Genotypes of TNBCC cases were compared with GWAS data for publicly available controls (Online Methods). Both studies imputed genotypes for common SNPs in phase 2 HapMap populations (release 21) (**Supplementary Table 1** and Online Methods). A total of 3,154,485 SNPs, genotyped and imputed, were analyzed in stage 1 of the meta-analysis.

We observed little evidence of inflation in the test statistics in AABC ($\lambda = 1.01$) or TNBCC ($\lambda = 1.04$) or in the meta-analysis of the two GWAS ($\lambda = 1.02$; **Supplementary Fig. 1**). In the combined results, only SNP rs10069690 (NCBI36/hg18, chr5:1,332,790) located in intron 4 of the *TERT* gene (encoding telomerase reverse transcriptase) at chromosome 5p15 showed a genome-wide significant association with ER-negative breast cancer (AABC: OR per allele = 1.32, $P = 1.3 \times 10^{-6}$; TNBCC: OR = 1.25, $P = 1.2 \times 10^{-3}$; combined OR = 1.29,

Table 1 Association of rs10069690 at 5p15 and ER-negative breast cancer risk

Stage	Consortium or study	Cases/controls[a]	RAF[b] T allele	Heterozygotes OR (95% CI)[c]	Homozygotes OR (95% CI)[c]	Per-allele OR (95% CI)[c]	*P* value (1-d.f.)[d]
1	AABC	1,002/2,743	0.57	1.32 (1.05–1.67)	1.74 (1.37–2.21)	1.32 (1.18–1.48)	1.3×10^{-6}
1	TNBCC	2,785/1,602	0.27	1.10 (0.97–1.26)	1.53 (1.21–1.95)	1.18 (1.07–1.30)	1.0×10^{-3}
2	BPC3	1,289/10,397	0.26	1.08 (0.96–1.22)	1.19 (0.95–1.49)	1.09 (0.99–1.19)	0.077
2	SEARCH	933/5,966	0.26	1.23 (1.06–1.43)	1.44 (1.10–1.89)	1.21 (1.09–1.36)	6.9×10^{-4}
Combined		6,009/20,708		1.15 (1.06–1.23)	1.46 (1.29–1.64)	1.18 (1.13–1.25)	1.0×10^{-10}

[a]Number of cases and controls with genotype data for rs10069690. All subjects were directly genotyped. [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, study and country (European Prospective Investigation into Cancer and Nutrition (EPIC) only) in BPC3. Adjusted for age in SEARCH. Combined results are from the meta-analysis. [d]*P* for trend (one degree of freedom (1-d.f.)).

A full list of authors and affiliations appears at the end of the paper.

Received 3 May; accepted 28 September; published online 30 October 2011; doi:10.1038/ng.985

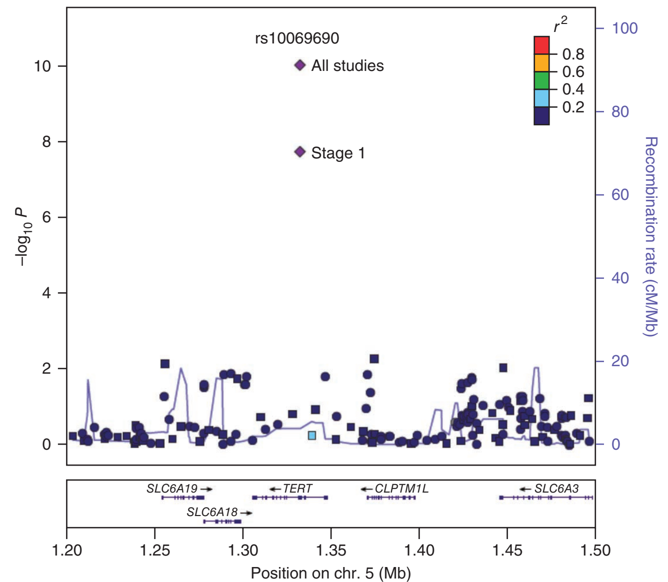

Figure 1 A regional plot of the $-\log_{10}$ P values for SNPs at the chromosome 5p15 risk locus from the meta-analysis of the AABC and TNBCC stage 1 studies. SNP rs10069690 is designated with the purple diamonds. The colors depict the strength of the correlation (r^2) between SNP rs10069690 and the SNPs tested in the region. The correlation is estimated using 1000 Genomes Project (1KGP) data for the HapMap CEU population (June 2010). Squares are SNPs that were genotyped in AABC and TNBCC. Circles are SNPs that were genotyped in one study and imputed in the other or imputed in both studies. The blue line indicates the recombination rates in centimorgans (cM) per megabase (Mb). Also shown are the SNP Build 36 coordinates and genes in the region.

$P = 1.0 \times 10^{-8}$). Whereas SNP rs10069690 was genotyped in AABC, it was imputed in TNBCC ($R^2 = 0.55$). To verify the imputed genotypes and the significance of the association in TNBCC, we re-genotyped rs10069690 in available DNA samples from 2,963 TNBCC cases and 1,632 study-specific TNBCC controls (Online Methods). Although the overlapping samples between the TNBCC GWAS and the re-genotyping study showed that the quality of imputation for rs10069690 in the GWAS was poor (Online Methods), the association with ER-negative breast cancer for rs10069690 remained statistically significant in the larger re-genotyped TNBCC sample (OR = 1.18, $P = 1.0 \times 10^{-3}$; **Table 1** and **Fig. 1**) and in the new combined results for AABC and the re-genotyped TNBCC sample (OR = 1.24, $P = 1.6 \times 10^{-8}$).

To further confirm the association at 5p15, we genotyped SNP rs10069690 in women of European ancestry, which included 8,365 cases (1,359 ER negative) and 10,935 controls from the US National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3) and 6,182 cases (933 ER negative) and 5,966 controls from Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH). Evidence for replication was observed for rs10069690 and ER-negative breast cancer in both studies (BPC3: OR = 1.09, $P = 0.077$; SEARCH: OR = 1.21, $P = 6.9 \times 10^{-4}$; **Table 1**).

In combining the results across all studies (6,009 ER-negative cases and 20,708 controls with genotype data), rs10069690 was significantly associated with an increased risk of ER-negative breast cancer (OR = 1.18, 95% confidence interval (CI), 1.13–1.25; $P = 1.0 \times 10^{-10}$; **Table 1**). The risk for heterozygote and homozygote carriers was 1.15 (95% CI, 1.06–1.23) and 1.46 (95% CI, 1.29–1.64), respectively. We observed little evidence of heterogeneity for the reported association for this variant by study or country in AABC (test for heterogeneity, $p_{\text{het}} = 0.86$), TNBCC ($p_{\text{het}} = 0.85$) or BPC3 ($p_{\text{het}} = 0.37$; **Supplementary Table 2**).

In an analysis of ER-positive cases, rs10069690 was only weakly associated with risk in African Americans (AABC: 1,558 ER-positive cases and 2,743 controls with genotype data, OR = 1.08, $P = 0.10$) and in women of European ancestry (BPC3: 4,890 ER-positive cases and 10,397 controls, OR = 1.03, $P = 0.31$; SEARCH: 3,534 ER positive cases and 5,966 controls, OR = 1.03, $P = 0.37$; combined for all populations: OR = 1.04, $P = 0.06$, $p_{\text{het}} = 0.64$). The statistical power to detect an OR of 1.18 (observed for ER-negative disease) for ER-positive disease was >99% in the combined sample (9,982 cases and 19,106 controls), assuming the risk allele frequency of 0.26 in people of European decent. This result suggests that the association with breast cancer might be specific for ER-negative subtypes (P value for case-only test of ER negative versus ER positive = 1.7×10^{-4}).

We further stratified the cases by HER2 status to assess whether this region may be a risk locus for triple-negative disease. In AABC, BPC3 and SEARCH the association with rs10069690 was greater for triple-negative tumors than for ER-negative, PR-negative, HER2-positive tumors (**Table 2**), and, in combining all studies, including TNBCC, the association with rs10069690 was significantly greater for triple-negative disease (3,707 triple-negative cases and 19,728 controls with genotype data, OR = 1.25, $P = 1.1 \times 10^{-9}$; 376 ER-negative, PR-negative, HER2-positive cases and 18,126 controls, OR = 1.03, $P = 0.71$; P value for case-only test = 0.010). The association with rs10069690 was also observed to be significantly greater for ER-negative and triple-negative disease at younger ages (<50 years: ER negative,

Table 2 Association of rs10069690 at 5p15 stratified by HER2 status

Consortium or study	Subtype	Cases/controls[a]	Heterozygotes OR (95% CI)[b]	Homozygotes OR (95% CI)[b]	Per-allele OR (95% CI)[b]	P value (1-d.f.)[c]	Case-only P
AABC[d]	ER−PR−HER2−	440/2,407	1.35 (0.97–1.89)	1.78 (1.27–2.49)	1.33 (1.14–1.55)	3.0×10^{-4}	0.19
	ER−PR−HER2+	115/2,407	1.83 (0.99–3.40)	1.59 (0.82–3.05)	1.15 (0.86–1.52)	0.34	
TNBCC	ER−PR−HER2−	2,785/1,602	1.10 (0.97–1.26)	1.53 (1.21–1.95)	1.18 (1.07–1.30)	1.0×10^{-3}	–
BPC3[e]	ER−PR−HER2−	300/9,753	1.19 (0.93–1.52)	1.64 (1.10–2.46)	1.25 (1.04–1.49)	0.015	0.13
	ER−PR−HER2+	198/9,753	0.99 (0.73–1.33)	0.95 (0.53–1.70)	0.98 (0.78–1.23)	0.87	
SEARCH	ER−PR−HER2−	182/5,966	1.42 (1.03–1.95)	2.41 (1.47–3.95)	1.51 (1.20–1.89)	4.2×10^{-4}	0.058
	ER−PR−HER2+	63/5,966	1.31 (0.79–2.16)	0.27 (0.04–1.95)	0.97 (0.64–1.46)	0.88	
Combined	ER−PR−HER2−	3,707/19,728[f]	1.17 (1.06–1.30)	1.69 (1.43–1.99)	1.25 (1.16–1.34)	1.1×10^{-9}	0.010
	ER−PR−HER2+	376/18,126	1.15 (0.91–1.46)	1.11 (0.73–1.70)	1.03 (0.88–1.21)	0.71	

[a]Number of cases and controls with genotype data for rs10069690. All subjects were directly genotyped. [b]Adjusted for age, study and principal components in AABC. Adjusted for age and country in TNBCC. Adjusted for age, study and country (EPIC only) in BPC3. Adjusted for age in SEARCH. Combined results are from the meta-analysis. [c]P for trend (1-d.f.). [d]Excludes San Francisco Bay Area Breast Cancer Study (SFBCS) and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO), as HER2 data were not available. [e]Excludes WHS , as HER2 data were not available. [f]Includes TNBCC. Without TNBCC: 922 ER−PR−HER2− cases and 18,126 controls; OR per allele = 1.33 (1.20–1.48), $P = 6.3 \times 10^{-8}$; heterozygotes: OR = 1.29 (1.09–1.53); homozygotes: OR = 1.85 (1.47–2.33).

© 2011 Nature America, Inc. All rights reserved.
npg

OR = 1.32, $P = 1.4 \times 10^{-8}$; triple negative, OR = 1.48, $P = 1.9 \times 10^{-9}$; P for interaction with age = 0.035 and 3.2×10^{-3}, respectively; **Supplementary Table 3**). We found no significant association with rs1006960 among ER- and PR-positive cases when stratified by HER2 status (513 triple-positive cases and 18,126 controls, OR = 1.09, $P = 0.21$; 2,808 ER-positive, PR-positive, HER2-negative cases and 18,126 controls, OR = 1.04, $P = 0.29$), which suggests the association may be limited to triple-negative disease and not all HER2-negative tumors.

Similar to 8q24 (refs. 5–7) and 11q13 (refs. 8–10), the *TERT-CLPTM1L* locus harbors multiple risk variants for different cancers (reviewed in ref. 11). SNP rs10069690 is modestly correlated ($r^2 = 0.13$–0.43 in 1000 Genomes Project populations of European and African ancestry, **Supplementary Fig. 2**) with variants found for serous ovarian cancer (rs7726159), glioma (rs2736100) and lung cancer (rs2736100, rs2735940)[12–14]. Aside from risk variant rs2853676 found for glioma[14], which we found to be associated with risk in TNBCC ($P = 0.014$, $r^2 = 0.05$ with rs10069690), none of the known risk variants identified for other cancers in the *TERT-CLPTM1L* region was significantly associated with breast cancer risk in TNBCC or AABC. Although rs7726159 was not tested in AABC or TNBCC (as it is not on the Illumina arrays or in HapMap), it is noteworthy that the first common risk variant identified for ER-negative breast cancer, at chromosome 19p13, is also associated with risk for serous ovarian cancer[15]. The *TERT* gene encodes the catalytic subunit of telomerase, which controls telomere length, a process linked with genomic instability and implicated in tumorigenesis. Sequencing of the coding exons of *TERT* in 96 African-American women (Online Methods) did not reveal a coding variant strongly correlated with rs10069690. The *TERT* locus may highlight another biological process common to the pathogenesis of ER-negative breast cancer subtypes and serous ovarian cancer that is also shared with other cancers.

Identification of the variant directly responsible for the association will be required to fully address the extent to which this locus contributes to the greater incidence of ER-negative and triple-negative tumors in women of African ancestry. However, it is notable that the risk allele frequency of rs10069690 is greater in African American women (frequency, 0.57) than in women of European ancestry (frequency, 0.26). If this variant is an equally good surrogate for the biologically functional allele in each population, then this locus may be responsible for a 15% (95% CI, 10–20%) higher incidence rate of ER-negative or triple-negative breast cancer in women of African compared to European ancestry (Online Methods). Larger studies with well-characterized tumor pathology information will be needed to determine whether the association we observed applies to all ER-negative disease or just the triple-negative subtype. Our findings provide further support for the presence of genetic susceptibility to ER-negative breast cancer subtypes and demonstrate the importance of discovery efforts in multiple populations.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported by a US Department of Defense Breast Cancer Research Program Era of Hope Scholar Award to C.A.H. (W81XWH-08-1-0383), the Norris Foundation, the Mayo Clinic College of Medicine, Komen Foundation for the Cure, the Breast Cancer Research Foundation and US National Institutes of Health grants CA128978, CA122340 and CA148065. Study specific acknowledgments are listed in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Conceived of and designed the experiments: C.A.H. and F.J.C. Performed the experiments and analyzed the data: C.A.H., L.C.P., D.V.D.B., X.S., G.K.C., A. Holbrook, P.W., F.C., D.O.S., X.W., T.L., C.O., K.N.S., A.M.L., L.Y.X., S.L.S. and C.M.V. Contributed reagents, materials, analysis tools or comments on the manuscript: C.A.H., C.M.V., A.D., R.C.M., X.W., F.A., S.A., C.B.A., L. Baglietto, R.B., E.V.B., M.W.B., C.D.B., L. Bernstein, C.B., W.J.B., H.B., J.E.B., L.A.C., J.E.C., J.C.-C., S.J.C., D.I.C., C.L.C., A.C., S.S.C., S.L.D., R.B.D., A.M.D., W.R.D., T.D., L.D., D.E., C.K.E., A.B.E., P.A.F., H.S.F., D.F.-J., F.F., A.F., G.F., S.M.G., G.G.G., A.K.G., P.G., N.G., D.G., U.H., S.E.H., A. Hartmann, R.H., J.H., R.N.H., J.J.H., D.J.H., S.A.I., A.I., J.I., E.M.J., N.J., A.J.-V., R.K., Y.-D.K., L.N.K., I.K., V.-M.K., S.K., D.L., A.M.L., L.L.M., T.L., J.L., S.L., A.M., S.M., N.G.M., P.M., G.W.M., H.N., S. Nickels, S. Nyante, C.O., J. Palmer, H.P., D.P., C.M.P., J. Peto, P.D.P.P., L.C.P., M.F.P., K.P., T.R.R., J.L.R.-G., L.R., E.R., T.R., I.d.S.S., E.S., M.K.S., R.S.-W., F.S., G.S., X.S., L.B.S., H.-P.S., K.N.S., M.C.S., W.J.T., I.T., F.B.L.H., E.W., J.W., H.W., R.W., D.Y., W.Z., R.G.Z., A.S., S.L.S., D.O.S., D.E., P.K., B.E.H. and F.J.C. Wrote the paper: C.A.H. and F.J.C.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Carey, L.A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J. Am. Med. Assoc.* **295**, 2492–2502 (2006).
2. Huo, D. *et al.* Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. *J. Clin. Oncol.* **27**, 4515–4521 (2009).
3. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874 (2001).
4. Antoniou, A.C. *et al.* 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 (2010).
5. Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
6. Haiman, C.A. *et al.* A common genetic risk factor for colorectal and prostate cancer. *Nat. Genet.* **39**, 954–956 (2007).
7. Kiemeney, L.A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* **40**, 1307–1312 (2008).
8. Purdue, M.P. *et al.* Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat. Genet.* **43**, 60–65 (2011).
9. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
10. Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42**, 504–507 (2010).
11. Baird, D.M. Variation at the TERT locus and predisposition for cancer. *Expert Rev. Mol. Med.* **12**, e16 (2010).
12. Johnatty, S.E. *et al.* Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and TERT, a cancer susceptibility "hot-spot". *PLoS Genet.* **6**, e1001016 (2010).
13. McKay, J.D. *et al.* Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.* **40**, 1404–1406 (2008).
14. Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.* **41**, 899–904 (2009).
15. Bolton, K.L. *et al.* Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat. Genet.* **42**, 880–884 (2010).

Christopher A Haiman[1], Gary K Chen[1], Celine M Vachon[2], Federico Canzian[3], Alison Dunning[4], Robert C Millikan[5], Xianshu Wang[6], Foluso Ademuyiwa[7], Shahana Ahmed[4], Christine B Ambrosone[8], Laura Baglietto[9], Rosemary Balleine[10], Elisa V Bandera[11], Matthias W Beckmann[12], Christine D Berg[13], Leslie Bernstein[14], Carl Blomqvist[15], William J Blot[16,17], Hiltrud Brauch[18,19], Julie E Buring[20], Lisa A Carey[21], Jane E Carpenter[22], Jenny Chang-Claude[23], Stephen J Chanock[24], Daniel I Chasman[20], Christine L Clarke[22],

© 2011 Nature America, Inc. All rights reserved.

Angela Cox[25], Simon S Cross[26], Sandra L Deming[16], Robert B Diasio[27], Athanasios M Dimopoulos[28], W Ryan Driver[29], Thomas Dünnebier[30], Lorraine Durcan[31], Diana Eccles[31], Christopher K Edlund[1], Arif B Ekici[32], Peter A Fasching[12,33], Heather S Feigelson[34], Dieter Flesch-Janys[35], Florentia Fostira[36], Asta Försti[37,38], George Fountzilas[39], Susan M Gerty[31], The Gene Environment Interaction and Breast Cancer in Germany (GENICA) Consortium[40], Graham G Giles[9], Andrew K Godwin[41], Paul Goodfellow[42], Nikki Graham[31], Dario Greco[43], Ute Hamann[30], Susan E Hankinson[44,45], Arndt Hartmann[46], Rebecca Hein[23], Judith Heinz[35], Andrea Holbrook[1], Robert N Hoover[24], Jennifer J Hu[47], David J Hunter[45,48], Sue A Ingles[1], Astrid Irwanto[49], Jennifer Ivanovich[42], Esther M John[50,51], Nicola Johnson[52], Arja Jukkola-Vuorinen[53], Rudolf Kaaks[54], Yon-Dschun Ko[55], Laurence N Kolonel[56], Irene Konstantopoulou[36], Veli-Matti Kosma[57], Swati Kulkarni[58], Diether Lambrechts[59,60], Adam M Lee[27], Loïc Le Marchand[56], Timothy Lesnick[2], Jianjun Liu[49], Sara Lindstrom[45,48], Arto Mannermaa[61,62], Sara Margolin[63], Nicholas G Martin[64], Penelope Miron[65], Grant W Montgomery[64], Heli Nevanlinna[43], Stephan Nickels[23], Sarah Nyante[5], Curtis Olswold[2], Julie Palmer[66], Harsh Pathak[67], Dimitrios Pectasides[68], Charles M Perou[69], Julian Peto[70], Paul D P Pharoah[4], Loreall C Pooler[1], Michael F Press[71], Katri Pylkäs[72], Timothy R Rebbeck[73], Jorge L Rodriguez-Gil[47], Lynn Rosenberg[66], Eric Ross[74], Thomas Rüdiger[75], Isabel dos Santos Silva[70], Elinor Sawyer[76], Marjanka K Schmidt[77], Rüdiger Schulz-Wendtland[46], Fredrick Schumacher[1], Gianluca Severi[9], Xin Sheng[1], Lisa B Signorello[16,17], Hans-Peter Sinn[78], Kristen N Stevens[2], Melissa C Southey[79], William J Tapper[31], Ian Tomlinson[80], Frans B L Hogervorst[81], Els Wauters[59,60], JoEllen Weaver[67], Hans Wildiers[82], Robert Winqvist[72], David Van Den Berg[1], Peggy Wan[1], Lucy Y Xia[1], Drakoulis Yannoukakos[36], Wei Zheng[16], Regina G Ziegler[24], Afshan Siddiq[83], Susan L Slager[2], Daniel O Stram[1], Douglas Easton[4], Peter Kraft[45,48,84], Brian E Henderson[1] & Fergus J Couch[2,6]

© 2011 Nature America, Inc. All rights reserved.

[1]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA. [2]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. [3]Genomic Epidemiology Group, DKFZ, Heidelberg, Germany. [4]Centre for Cancer Genetic Epidemiology, Strangeways Laboratory, Worts Causeway, Cambridge, UK. [5]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, USA. [6]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. [7]Department of Medicine, Roswell Park Cancer Institute, Buffalo, New York, USA. [8]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York, USA. [9]Cancer Epidemiology Centre, The Cancer Council Victoria & Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, The University of Melbourne, Victoria, Australia. [10]Department of Translational Oncology, Westmead Hospital, Western Sydney Local Health Network, Westmead, New South Wales, Australia. [11]The Cancer Institute of New Jersey, New Brunswick, New Jersey, USA. [12]Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. [13]Division of Cancer Prevention, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [14]Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, California, USA. [15]Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland. [16]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [17]International Epidemiology Institute, Rockville, Maryland, USA. [18]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany. [19]University of Tübingen, Tübingen, Germany. [20]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [21]Department of Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, USA. [22]Australian Breast Cancer Tissue Bank, University of Sydney at the Westmead Millennium Institute, Westmead, New South Wales, Australia. [23]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany. [24]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [25]Institute for Cancer Studies, Department of Oncology, Faculty of Medicine, Dentistry & Health, University of Sheffield, Sheffield, UK. [26]Academic Unit of Pathology, Department of Neuroscience, Faculty of Medicine, Dentistry & Health, University of Sheffield, Sheffield, UK. [27]Department of Pharmacology, Mayo Clinic, Rochester, Minnesota, USA. [28]Department of Clinical Therapeutics, "Alexandra" Hospital, University of Athens School of Medicine, Athens, Greece. [29]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA. [30]Molecular Genetics of Breast Cancer, DKFZ, Heidelberg, Germany. [31]Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK. [32]Institute of Human Genetics, Friedrich-Alexander University of Erlangen-Nuremberg, Erlangen, Germany. [33]Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California–Los Angeles, Los Angeles, California, USA. [34]Kaiser Permanente Colorado, Denver, Colorado, USA. [35]Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany. [36]Molecular Diagnostics Laboratory Institute of Radioisotopes and Radiodiagnostic Products, National Centre for Scientific Research "Demokritos", Athens, Greece. [37]Division of Molecular Genetic Epidemiology, DKFZ, Heidelberg, Germany. [38]Center for Primary Health Care Research, University of Lund, Malmö, Sweden. [39]Department of Medical Oncology, Aristotle University of Thessaloniki, Papageorgiou Hospital, Thessaloniki, Greece. [40]A full list of members is provided in the **Supplementary Note**. [41]Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Lawrence, Kansas, USA. [42]Washington University School of Medicine, Barnes-Jewish Hospital and Siteman Cancer Center, St. Louis, Missouri, USA. [43]Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland. [44]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [45]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [46]Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander University of Erlangen-Nuremberg, Erlangen, Germany. [47]Sylvester Comprehensive Cancer Center and Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, Florida, USA. [48]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [49]Human Genetics Division, Genome Institute of Singapore, Singapore. [50]Cancer Prevention Institute of California, Fremont, California. [51]Stanford University School of Medicine and Stanford Cancer Center, Stanford, California, USA. [52]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK. [53]Department of Oncology, Oulu University Hospital, University of Oulu, Oulu, Finland. [54]Division of Cancer Epidemiology, DKFZ, Heidelberg, Germany. [55]Department of Internal Medicine, Evangelische Kliniken Johanniter- und Waldkrankenhaus Bonn gGmbH, Bonn, Germany. [56]Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, Hawaii, USA. [57]Department of Pathology, Imaging Centre, Kuopio University Hospital, Kuopio, Finland. [58]Department of Surgical Oncology, Roswell Park Cancer Institute, Buffalo, New York, USA. [59]Vesalius Research Center, Vlaams Instituut voor Biotechnologie, Leuven, Belgium. [60]Vesalius Research Center, University of Leuven, Leuven, Belgium. [61]Institute of Clinical Medicine, Department of Pathology, University of Eastern Finland Biocenter Kuopio, Kuopio, Finland. [62]Department of Pathology, Imaging Centre, Kuopio University Hospital, Kuopio, Finland. [63]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden. [64]Queensland Institute of Medical Research (QIMR) Genome-Wide Association Study Collective, Brisbane, Queensland, Australia. [65]Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [66]Slone Epidemiology Center at Boston University, Boston, Massachusetts, USA. [67]Department of Medical Oncology, Fox Chase Cancer Center, Philadelphia,

Pennsylvania, USA. [68]Department of Internal Medicine, Oncology Section, "Hippokration" Hospital, Athens, Greece. [69]Departments of Genetics and Pathology, Lineberger Comprehensive Cancer Center, The University of North Carolina, Chapel Hill, North Carolina, USA. [70]Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. [71]Department of Pathology, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California. [72]Laboratory of Cancer Genetics, Department of Clinical Genetics and Biocenter Oulu, University of Oulu, Oulu University Hospital, Oulu, Finland. [73]University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. [74]Department of Biostatistics, Fox Chase Cancer Center, Philadelphia, Pennsylvania, USA. [75]Institute of Pathology, Städtisches Klinikum Karlsruhe, Karlsruhe, Germany. [76]National Institute for Health Research Comprehensive Biomedical Research Centre, Guy's & St. Thomas' National Health Service Foundation Trust, London, UK. [77]Division of Experimental Therapy and Molecular Pathology and Division of Epidemiology, Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. [78]Department of Pathology, University Hospital Heidelberg, Heidelberg, Germany. [79]Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. [80]Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. [81]Family Cancer Clinic, Netherlands Cancer Institut–Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. [82]Multidisciplinary Breast Center, University Hospital Gasthuisberg, Leuven, Belgium. [83]Imperial College, London, UK. [84]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. Correspondence should be addressed to C.A.H. (haiman@usc.edu) or F.J.C. (couch.fergus@mayo.edu).

© 2011 Nature America, Inc. All rights reserved.

npg

© 2011 Nature America, Inc. All rights reserved.

ONLINE METHODS

Study populations. Stage 1 included the studies of the AABC and the TNBCC. AABC includes 3,153 breast cancer cases (1,017 ER negative and 1,608 ER positive) and 2,831 controls from 9 studies (**Supplementary Table 1**). TNBCC is composed of 2,963 triple-negative breast cancer cases and 1,632 controls from 22 studies, GWAS genotype data from an additional 85 triple-negative breast cancer cases and 222 controls from HEBCS, and public GWAS genotype data from 3,448 controls from Cancer Genetic Markers of Susceptibility (CGEMS), Wellcome Trust Case-Control Consortium (WTCCC), KORA and QIMR (**Supplementary Table 1**). Replication studies include 8,365 breast cancer cases (1,359 ER negative and 5,255 ER positive) and 10,935 controls of the BPC3 and 6,182 breast cancer cases (933 ER negative and 3,434 ER positive) and 5,966 controls of the SEARCH. All participants in these studies have provided written informed consent for the research, and approval for the study was obtained from the ethics review boards at all the local institutions. A description of each participating study is provided in the **Supplementary Note**. Details regarding the measurement and collection of ER, PR and HER2 data for each study are provided in **Supplementary Table 4**.

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/μl. Following genotyping, we removed samples on the basis of the following exclusion criteria: (i) unknown replicates (≥98.9% genetically identical, n = 29); (ii) samples with call rates <95% after a second attempt (n = 100); (iii) samples with ≤5% African ancestry (n = 36) (discussed below); and (iv) samples with <15% mean heterozygosity of SNPs on 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 data set included 1,043,036 SNPs genotyped on 3,016 cases (988 ER negative, 1,520 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%. The call rate for rs10069690 was very high in stage 1 (99.9%) and similar in cases (99.9%) and controls (99.9%). We also re-genotyped rs10069690 using TaqMan in 1,456 of the stage 1 samples; the concordance was 99.8%.

Genotyping for the TNBCC GWAS was conducted on 1,577 cases from ten studies (Australian Breast Cancer Tissue Bank (ABCTB), Bavarian Breast Cancer Cases and Controls (BBCC), Dana-Farber Cancer Institute, Fox Chase Cancer Center, GENICA, MARIE, Melbourne Collaborative Cohort Study (MCBCS), Prospective Study of Outcomes in Sporadic Versus Hereditary Breast Cancer (POSH), Sheffield Breast Cancer Study (SBCS)) using the Illumina 660-Quad SNP array. In addition, a set of MARIE cases (n = 56) were genotyped using the Illumina CNV370 SNP array. HEBCS cases (n = 85) were genotyped using the Illumina 550-Duo SNP array, bringing the total number of cases to 1,718. Population allele and genotype frequencies on healthy population controls (n = 222) genotyped on Illumina HumanHap 370CNV in the NordicDB, a Nordic pool and portal for genome-wide control data, were obtained from the Finnish Genome Center. GWAS data for public controls (n = 3,448) were generated using the following arrays: Illumina 660-Quad (QIMR), Illumina 550(v1) (CGEMS), Illumina 550 (KORA) and Illumina 1.2M (WTCCC). The combined total number of controls was 3,670. These GWAS data were independently evaluated by an iterative quality control process with the following exclusion criteria: MAF <0.01, call rate <95%, Hardy-Weinberg equilibrium (HWE) P value $< 1 \times 10^{-7}$ among controls and sample call rate <98%. In total, we excluded cases failing in the genotyping process (n = 5), 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 of >0.10 between the study allele frequency and the median frequency across all studies. Eigensoft was used to evaluate confounding due to population stratification. We removed 101 subjects that did not cluster with the CEU HapMap phase 2 samples, resulting in 1,562 cases and 3,578 controls in the GWAS analyses.

Re-genotyping of rs10069690 on 2,963 TNBCC cases and 1,632 study-specific controls was conducted using a single multiplex on the iPLEX Mass Array platform (Sequenom). We removed 31 cases from MCCS that were part of the MCCS replication sample in BPC3. SNPs and samples evaluated on the iPLEX were excluded on the basis of the following criteria: SNP call rate was <97%, HWE P value < 0.001 among controls and sample call rate <95% (for the overall experiment). The final data set of 2,849 cases and 1,602 controls for rs10069690 had a SNP call rate >99% and HWE P value of 0.53 in controls. The concordance rate, on the basis of blinded duplicates, was 100%. The concordance of the imputed (R^2 = 0.55) versus the genotyped data was 70%.

Replication genotyping. In BPC3, genotyping of rs10069690 was performed by TaqMan in five laboratories (Cancer Prevention Study II Nutrition Cohort (CPS2) and Multiethnic Cohort (MEC) at the University of Southern California; the Nurses' Health Study (NHS) and the Women's Health Study (WHS) at Harvard University; EPIC at the German Cancer Research Center in Heidelberg; MCCS at Melbourne University and PLCO at the NCI Core Genotyping Facility). Genotyping in SEARCH was performed by TaqMan at Cambridge University. Genotype call rates were >92% in cases and controls, and concordance of blinded duplicates was ≥99.5% in all studies. The P value for HWE in controls was >0.01 in all studies except WHS (P = 0.007).

DNA sequencing. Bi-directional sequencing of the 15 coding exons of *TERT* was performed in 96 African-American women using the ABI 3730xl DNA Analyzer (Applied Biosystems). Sequencing purification was performed using DyeDX 96 columns (Qiagen) following their standard protocol, and PolyPhred was used for analyzing sequence traces (http://droog.gs.washington.edu/polyphred/). More than 95% of samples were sequenced for each exon except for exon 15 (n = 74) and 16 (n = 86). Exon 1 could not be sequenced, as well as 112bp (9%) of exon 2, because of high GC content.

Statistical analysis. In AABC, we tested for gene dosage effects through a one-degree-of-freedom likelihood ratio test in models adjusted for age, study and genetic ancestry eigenvectors 1–10. OR and 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. For the analyses of the iPLEX genotyping data on rs10069690, unconditional logistic regression was used assuming a log-additive model and adjusting for age and country.

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 linkage disequilibrium patterns in order to impute untyped markers. For both studies, genome-wide imputation was carried out using the software MACH. Filtered from the analysis were SNPs with $R^2 < 0.3$.

We conducted a fixed-effect meta-analysis of AABC and TNBCC using the inverse variance weighted method. The number of SNPs available for meta-analysis from AABC and TNBCC was 3,055,415 and 2,134,490 respectively. The union of these two data sets (3,154,485 SNPs) was meta-analyzed using the program METAL.

SNP rs10069690 was analyzed in BPC3 and SEARCH using logistic regression controlling for age and study or country (BPC3 only). The meta-analysis of rs10069690 from AABC, TNBCC, BPC3 and SEARCH was conducted using the inverse variance weighted method. 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.

We estimated the relative risk in African-ancestry women compared to women of European descent that could plausibly be attributable to the association with rs10069690. The calculation of the attributable racial/ethnic ratio (ARR) is ARR = $\sum_{i=0}^{2} f_A OR^i / \sum_{i=0}^{2} f_E OR^i$, where $f_A(i)$ is the probability in the African American women of carrying i = 0, 1 or 2 copies of the risk variant and $f_E(i)$ is the same probability for European women. The per-allele OR is for triple-negative disease from the meta-analysis (1.25), and both a log linear model for risk and Hardy-Weinberg equilibrium for the alleles (in both populations) is assumed. A confidence interval for the ARR is calculated from the confidence interval for the OR in the meta-analysis.

doi:10.1038/ng.985