Supplementary Information

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

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], 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 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 DP 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 BL 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*]

Affiliations:
[1]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA
[2]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
[3]Genomic Epidemiology Group, German Cancer Research Center (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, NC, USA
[6]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
[7]Department of Medicine, Roswell Park Cancer Institute, Buffalo, NY, USA
[8]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA
[9]Cancer Epidemiology Centre, The Cancer Council Victoria & Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, The University of Melbourne, Melbourne, Victoria, Australia
[10]Department of Translational Oncology, Westmead Hospital, Western Sydney Local Health Network, Westmead, NSW, Australia
[11]The Cancer Institute of New Jersey, New Brunswick, NJ, USA

Nature Genetics: doi:10.1038/ng.985

[12] Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg , Erlangen, Germany
[13]Division of Cancer Prevention, US National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[14]Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, CA, 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, TN, USA
[17]International Epidemiology Institute, Rockville, MD, 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, MA, USA
[21]Department of Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA
[22]Australian Breast Cancer Tissue Bank, University of Sydney at the Westmead Millennium Institute, Westmead, NSW, Australia
[23]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany
[24]Division of Cancer Epidemiology and Genetics, US National Cancer Institute, National Institutes of Health Bethesda, MD, 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, MN, USA
[28]Department of Clinical Therapeutics, “Alexandra” Hospital, University of Athens School of Medicine, Athens, Greece
[29]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA
[30]Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany
[31]Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK
[32] Institute of Human Genetics, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
[33] Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA
[34]Kaiser Permanente Colorado, Denver, CO, 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, Deutsches Krebsforschungszentrum (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, KS, USA
[42]Washington University School of Medicine, Barnes-Jewish Hospital and Siteman Cancer Center, St. Louis, MO, USA

Nature Genetics: doi:10.1038/ng.985

[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, MA, USA.
[45]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA.
[46] Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany
[47]Sylvester Comprehensive Cancer Center and Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, FL, USA
[48]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA, USA
[49]Human Genetics Division, Genome Institute of Singapore, Singapore.
[50]Cancer Prevention Institute of California, Fremont, CA
[51]Stanford University School of Medicine and Stanford Cancer Center, Stanford, CA, 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, German Cancer Research Center (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, HI, USA
[57]Department of Pathology, Imaging Centre, Kuopio University Hospital, Kuopio, Finland.
[58]Department of Surgical Oncology, Roswell Park Cancer Institute, Buffalo, NY, 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, Australia
[65]Dana Farber Cancer Institute, Boston, MA, USA
[66]Slone Epidemiology Center at Boston University, Boston MA, USA
[67]Department of Medical Oncology, Fox Chase Cancer Center, Philadelphia, PA, 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, NC, 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, CA
[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, PA, USA
[74]Department of Biostatistics, Fox Chase Cancer Center, Philadelphia, PA, 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

Nature Genetics: doi:10.1038/ng.985

[79]Genetic Epidemiology Laboratory, Department of Pathology, The University of 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 Institute – 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, MA, USA

Nature Genetics: doi:10.1038/ng.985

Contents

Nature Genetics: doi:10.1038/ng.985

Supplementary Table 1. Participating breast cancer studies.

				Samples in the 2 GWAS		Samples genotyped for rs10069690	
Consortium	Study Abbreviation	Full Name	Country	Cases	Controls	Cases	Controls
AABC	CARE	The Los Angeles component of The Women's Contraceptive and Reproductive Experiences Study	USA	380	224		
	CBCS	The Carolina Breast Cancer Study	USA	656	608		
	MEC	Multiethnic Cohort	USA	734	1003		
	NBHS	The Nashville Breast Health Study	USA	310	186		
	NC-BCFR	The Northern California Breast Cancer Family Registry	USA	440	53		
	PLCO	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	USA	64	133		
	SFBCS	The San Francisco Bay Area Breast Cancer Study	USA	172	231		
	WCHS	The Women's Circle of Health Study	USA	272	240		
	WFBC	Wake Forest University Breast Cancer Study	USA	125	153		
			TOTAL	3153	2831		
TNBCC	ABCS	Amsterdam Breast Cancer Study	Netherlands			67[a]	
	ABCTB	Australian Breast Cancer Tissue Bank	Australia	166		162	
	BBCC	Bavarian Breast Cancer Cases and Controls	Germany	240		325	
	BBCS	British Breast Cancer Study	UK			58	58[a]
	BIGGS	Breast Cancer In Galway Genetic Study	Ireland			38	86
	CGEMS	Cancer Genetic Markers of Susceptibility	USA		1142		
	Demokritos	Hellenic Cooperative Oncology Group	Greece			281	91
	DFCI	Harvard Breast Cancer SPORE Blood Repository	USA	303		304	
	FCCC	Fox Chase Cancer Center	USA	148		159	159
	GENICA	Gene Environment Interaction and Breast Cancer in Germany	Germany	60		65	66
	HEBCS	Helsinki Breast Cancer Study	Finland	85	222		
	KARBAC	Karolinska Breast Cancer Study	Sweden			27	26
	KBCP	Kuopio Breast Cancer Project	Finland			36	
	KORA	Cooperative Health Research in the Region of Augsburg	Germany		226		
	LMBC	Leuven Multidisciplinary Breast Centre	Germany			88	95
	MARIE	Mammary Carcinoma Risk Factor Investigation	Germany	205		231	248
	MCBCS	Mayo Clinic Breast Cancer Study	USA	153		152	155
	MCCS	Melbourne Collaborative Cohort Study	Australia	41		58	66
	NBHS	Nashville Breast Health Study	USA			123	119

Nature Genetics: doi:10.1038/ng.985

	OBCS	Oulu Breast Cancer Study	Finland			68	96
	POSH	Prospective Study of Outcomes in Sporadic Versus Hereditary Breast Cancer	UK	274		273	
	QIMR	Australian Twin Cohort study from the Queensland Institute of Medical Research	Australia		659		
	RPCI	Roswell Park Cancer Institute	USA			142	143
	SBCS	Sheffield Breast Cancer Study	UK	43		47	54
	SKKDKFZ	Städtisches Klinikum Karlsruhe and Deutsches Krebsforschungszentrum Breast Cancer Study	Germany			167	170
	WASHU	Washington University	USA			92	
	WTCCC	Wellcome Trust Case Control Consortium	UK		1421		
			TOTAL	1718	3670	2963	1632
BPC3	CPS-II	Cancer Prevention Study II Nutrition Cohort				583	791
	EPIC	European Prospective Investigation into Cancer and Nutrition	Europe			2533	3382
	MCCS	Melbourne Collaborative Cohort Study	Australia			688	766
	MEC	Multiethnic Cohort	USA			527	561
	NHS	The Nurses' Health Study	USA			1974	2572
	NHSII	The Nurses' Health Study II	USA			587	1176
	PLCO	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	USA			799	1013
	WHS	The Women's Health Study	USA			674	674
			TOTAL			8365	10935
SEARCH	SEARCH	Studies of Epidemiology and Risk Factors in Cancer Heredity	UK			6182	5966

[a]TNBCC samples used for re-genotyping of rs10069690. ABCS was not included in the analysis as no county-specific controls were available.

Nature Genetics: doi:10.1038/ng.985

Supplementary Table 2. The association of rs10069690 and ER negative breast cancer risk by study/country.

Consortium/Study	No. Cases / No. Controls with genotype data	Risk Allele Frequency	OR (95% CI)[a]	P-value	P_{Het}
AABC/CARE	129/214	0.56	1.55(1.11-2.16)	0.010	
AABC/CBCS	316/588	0.61	1.23(1.00-1.50)	0.051	
AABC/MEC	176/990	0.56	1.41(1.11-1.79)	0.0047	
AABC/NBHS	65/182	0.56	1.62(1.04-2.51)	0.032	
AABC/NC-BCFR	121/50	0.54	1.45(0.90-2.35)	0.13	
AABC/PLCO	14/116	0.59	1.00(0.45-2.23)	0.99	
AABC/SFBC	50/220	0.58	1.19(0.74-1.93)	0.47	
AABC/WCHS	88/239	0.58	1.39(0.94-2.05)	0.10	
AABC/WFBC	43/144	0.60	1.05(0.61-1.81)	0.86	
Total	1002/2743				0.86
TNBCC[b]/USA	940/565	0.30	1.19(1.02-1.40)	0.029	
TNBCC/Australia	180/64	0.26	1.10(0.69-1.74)	0.69	
TNBCC/UK	376/112	0.31	1.02(0.70-1.49)	0.92	
TNBCC/Finland	102/96	0.25	1.29(0.84-1.98)	0.25	
TNBCC/Germany	844/571	0.26	1.17(0.99-1.38)	0.071	
TNBCC/Greece	281/88	0.21	1.44(0.95-2.19)	0.086	
TNBCC/Ireland	35/80	0.30	0.83(0.41-1.65)	0.59	
TNBCC/Sweden	27/26	0.23	1.70(0.70-4.12)	0.24	
Total	2785/1602				0.85
BPC3/CPS2	35/787	0.26	0.87(0.50-1.53)	0.63	
BPC3/EPIC	498/3236	0.25	1.15(0.99-1.34)	0.066	
BPC3/MCCS	128/721	0.27	1.01(0.74-1.37)	0.96	
BPC3/MEC	77/536	0.27	0.89(0.60-1.34)	0.59	
BPC3/NHS	287/2448	0.25	1.01(0.83-1.23)	0.90	
BPC3/NHS2	92/1138	0.28	1.10(0.79-1.52)	0.59	
BPC3/PLCO	76/887	0.25	1.33(0.93-1.90)	0.12	
BPC3/WHS	96/644	0.26	1.14(0.83-1.57)	0.41	
Total	1289/10397				0.37
SEARCH	933/5966	0.26	1.21(1.09-1.36)	$6.9x10^{-4}$	
			All 4 studies		0.070

[a]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. [b]rs10069690 was directly genotyped in TNBCC

Nature Genetics: doi:10.1038/ng.985

Supplementary Table 3. The association of rs10069690 with tumor subtype by age.

ER- cases				
Age Group	**No. Cases / No. Controls with genotype data[a]**	**OR (95% CI)[b]**	**P-value**	**P_{int}[c]**
<50	2158/4218	1.32(1.20-1.45)	$1.4x10^{-8}$	
50-<60	2020/7303	1.20(1.10-1.31)	$3.8x10^{-5}$	
60-<70	1305/6469	1.10(0.99-1.22)	0.082	
≥70	416/2671	1.01(0.83-1.22)	0.93	0.035
ER-/PR-/HER2- cases				
Age Group	**No. Cases / No. Controls with genotype data[a]**	**OR (95% CI)[b]**	**P-value**	**P_{Int}[c]**
<50	1431/3901	1.48(1.30-1.68)	$1.9x10^{-9}$	
50-<60	1146/6709	1.21(1.07-1.37)	$2.5x10^{-3}$	
60-<70	735/5966	1.14(0.99-1.33)	0.078	
≥70	286/2495	1.04(0.81-1.35)	0.74	$3.2x10^{-3}$

[a]Numbers do not match those in Tables 1 or 2 as cases or controls were removed for any given study if not both observed in an age group category. [b]Results combined by meta-analysis. 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. [c]Test for interaction (1 df) between age (continuous) and genotype (trend).

Nature Genetics: doi:10.1038/ng.985

Supplementary Table 4. Criteria used to define ER, PR, and HER2 status by study site.

Study	ER	PR	HER2
ABCS	Positive=>10% cells stained on TMA (Neomarkers, 1D5 and 6F11 clones)	Positive =>10% cells stained on TMA (ImmunoLogic, PR-1 clone)	Positive=Score 3+ on TMA (NeoMarkers, 3B5 and 23 clones)
ABCTB	Positive = staining of any intensity in >1% of cells	Positive = staining of any intensity in >1% of cells	Single probe for HER2 SISH. Positive if >6 copies of HER2 gene per cell. Equivocal if between 4 and 6 cpc. If equivocal Cep17:HER2 ratio performed Score >2.2 = Pos (N.B. some of the older cases were done by FISH)
BBCC	Positive = > 9% of the cells stained positive; 1D5; 1:200; monoclonal mouse IgG 1k; Dako, Denmark; whole sections of FFPE.	Positive = >9% of the cells stained positive; PgR 636; 1:200; monoclonal mouse IgG; Dako, Denmark; whole sections of FFPE.	Positive= DAKO Score 3+ or 2+ and FISH positive; whole sections of FFPE.
BBCS	Data extracted from clinical notes. Quick (Allred) score (intensity & proportion). Negative=Quick score 0-2; Positive=Quick score 3+	Data extracted from clinical notes. Quick (Allred) score (intensity & proportion). Negative=Quick score 0-2; Positive=Quick score 3+	Data extracted from clinical notes. IHC scoring method: Positive=scores 3+; Negative=scores 0,1+; Borderline: 2+. Fish ratio: Positive >2.0; Negative <2.0
BIGGS	Data from hospital pathology reports: Positive=Allred score (intensity*percentage)= 3-8 (score range 0-8	Data from hospital pathology reports:Positive=Allred score (intensity*percentage)= 3-8 (score range 0-8	Data from hospital pathology reports: Negative=No staining, IHC Score 1 or 2+FISH negative; Positive=IHC Score 3 or 2+FISH positive (ASCO guidelines J jClin Path 2007: 25:1:118-145)
CARE	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals	HER2 expression status was determined by immunohistochemistry using the 10H8 monoclonal antibody to assess HER2 membrane protein immunostaining. No immunostaining (0) or weak (1+) membrane

Nature Genetics: doi:10.1038/ng.985

			immunostaining was considered low HER2 expression (HER2−). Moderate (2+) or strong membrane immunostaining (3+) was considered HER2 overexpression (HER2+).
CBCS	Antibody: SP1 Vendor: Ventana 5% or more nuclei positive staining	Antibody: Y85 Vendor: Ventana 5% or more nuclei positive staining	Antibody: CB11 Vendor: Biogenex 10% or more cells showing membrane or cytoplasmic plus membrane staining with weak or greater intensity
CPS-II	ER status data were extracted directly from pathology reports collected from hospitals throughout the U.S. If no pathology report was available, then for some cases ER status data were collected from tumor registries.	PR status data were extracted directly from pathology reports collected from hospitals throughout the U.S. If no pathology report was available, then for some cases PR status data were collected from tumor registries.	HER2 status data were extracted directly from pathology reports collected from hospitals throughout the U.S.
DEMOKRITOS	Positive=>1% immunoreactive nuclei (central pathology review on TMA for samples initially evaluated before 2004 (30% of the samples), clone 6F11, Leica BioSystems, Newcastle, UK, or abstracted from medical records for newer samples)	Positive=>1% immunoreactive nuclei (central pathology review on TMA for samples initially evaluated before 2004 (30% of the samples), clone 1A6, Leica BioSystems, Newcastle, UK, or abstracted from medical records for newer samples)	Positive=Score 3+ on TMA or Score 2+ and CISH/FISH positive (central pathology review on TMA for samples initially evaluated before 2004 (30% of the samples), clone PL, Dako, Glostrup, Denmark, or abstracted from medical records for newer samples)
DFCI	Negative= <1% of cells staining; Low-positive=1-10% of cells staining; Positive = >10% of cells staining (Dako, 1D5)	Negative= <1% of cells staining; Low-positive=1-10% of cells staining; Positive = >10% of cells staining (Dako, PgR 636))	Positive=3+ membrane staining (Dako A0485) or FISH amplified ratio >=2.0; Negative=0 or 1+ staining, or 2+ IHC and FISH not amplified ratio < 2.0
EPIC	Different methods used in different subcohorts, including: - Percentage of stained cells (positive =	Different methods used in different subcohorts, including: - Percentage of stained cells (positive =	Different methods used in different subcohorts, including: - Hercept test (range 0-3+, where 0 is no

Nature Genetics: doi:10.1038/ng.985

	≥10% of cells staining) - Femtomoles of ER/PR per milligram of cytosol protein (positive = ≥20fmol/mg) - Plus score (range: ?, +, ++ or +++; positive if at least +) - H-score (range 0-300, obtained by multiplying percentages of cell staining at each intensity category by the weighted intensity of staining; positive = H≥10) - Allred/quick score (range 0-8, obtained by adding the percentage of nuclei staining score to the intensity of staining score; positive = score≥4) - Immunoreactive score (range 0-12, obtained by multiplying the percentage of stained cells score by the intensity of staining score; positive = score≥3)	≥10% of cells staining) - Femtomoles of ER/PR per milligram of cytosol protein (positive = ≥20fmol/mg) - Plus score (range: ?, +, ++ or +++; positive if at least +) - H-score (range 0-300, obtained by multiplying percentages of cell staining at each intensity category by the weighted intensity of staining; positive = H≥10) - Allred/quick score (range 0-8, obtained by adding the percentage of nuclei staining score to the intensity of staining score; positive = score≥4) - Immunoreactive score (range 0-12, obtained by multiplying the percentage of stained cells score by the intensity of staining score; positive = score≥3)	staining and 3+ is complete and intense staining in >10% of the invasive tumor cells; positive = score of 2+ or 3+) - ECD (c-erb 2 levels measured in serum; when centres provides ECD scores, the variable was coded as unknown because of the high proportion of false positives of this test (30%))
FCCC	Positive=Any nuclear staining on whole sections (Novocastra, 6F11/2 clone)	Positive=Any nuclear staining on whole sections (Dako, PgR 636 clone)	Positive=Complete strong cytoplasmic staining in >30% tumor cells on whole sections (Dako, HercepTest™)
GENICA	Positive=Number of cells x intensity (german immuno reactive score) 3-12 positive on whole sections (Dako, 1D5 clone)	Positive=Number of cells x intensity (german immuno reactive score) 3-12 positive on whole sections (Dako, PgR 636 clone)	Positive=Score 2+ on whole sections (Dako, HercepTest™)
HEBCS	Positive=>10% cells stained (Novocastra), abstracted from medical records	Positive=>10% cells stained (Dako), abstracted from medical records	Positive= Score 2+/ CISH-result; 0-1=neg, 2-3=pos / combined if no CISH result: IHC 0-1=neg, 3=pos on TMA (IHC Novocastra Zymed, NCL-BC11 ErbB2 probe))

Nature Genetics: doi:10.1038/ng.985

KARBAC	Positive= ≥0.05 fmol/µg DNA (quantitative method, cytosol assay) or ≥10% stained (IHC). Abstracted from medical records	Positive= ≥0.05 fmol/µg DNA (quantitative method, cytosol assay) or ≥10% stained (IHC). Abstracted from medical records	Negative= 0 or 1+ with IHC, 2+ with IHC and FISH-negative. Positive= 2+ or 3+ with IHC and FISH-positive
KBCP	Positive=Intensity score (0.1,2,3)*percentage score (0,1,2,3)= 3-6 (score range 0-6) from whole sections (Abbot, ER_ICA kit); abstracted from medical records	Positive=Intensity score (0.1,2,3)*percentage score (0,1,2,3)= 3-6 (score range 0-6) from whole sections (Abbot, PR_ICA kit); abstracted from medical records	Data source: Hospital registry.Intensity scores: 0 = no staining, 1=weak, 2=moderate, 3=strong; Scoring of % cells stained: 0-10 %=negative, 10-30 %=1, 30-60=2, Over 60 %=3; Reference: Pellikainen J. et al. Eur J. Cancer 2004;40:1485-1495
LMBC	Negative=Quick score 0-2, corresponding to no or weak stainnig; Positive=Quick score 3+, corresponding to moderate to strong staining	Negative=Quick score 0-2, corresponding to no or weak stainnig; Positive=Quick score 3+, corresponding to moderate to strong staining	Negative=No staining, Score 1 or 2+FISH negative; Positive=Score2,3+FISH positive
MARIE	Positive=>10% tumor nucei stained with intensity score (0,1,2,3,) >1 (Dako, ID5 clone); abstracted from medical records	Positive=>10% tumor nucei stained with intensity score (0,1,2,3,) >1 (Dako, PgR 636 clone); abstracted from medical records	Positive=Score 3+ in >30% stained tumor cells or FISH amplified (Dako A0485, cerB2 clone); abstracted from medical records
MCBCS	Positive=Any nuclear staining on whole sections (Novocastra, 6F11/2 clone)	Positive=Any nuclear staining on whole sections (Dako, PgR 636 clone)	Positive=Complete strong cytoplasmic staining in >30% tumor cells on whole sections (Dako, HercepTest™)
MCCS	Positive=Nuclei positive with intensity score (0,1,2,3) >=1 on whole sections (Neomarkers RM9101, SP1 clone); abstracted from medical records	Positive=Nuclei positive with intensity score (0,1,2,3) >=1 on whole sections (Dako M3569, PgR 636 clone); abstracted from medical records	Positive=Score 2+ on whole sections (Dako A0485, CerB2 clone))
MEC	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals

Nature Genetics: doi:10.1038/ng.985

NBHS	Data extracted from clinical and pathology reports using Allred score or equivalent scoring system based on staining intensity and percent of cells positive. Positive = Score of 3-8. Negative = Score of 0-2.	Data extracted from clinical and pathology reports using Allred score or equivalent scoring system based on staining intensity and percent of cells positive. Positive = Score of 3-8. Negative = Score of 0-2.	Data extracted from clinical and pathology reports. Positive for amplification= (3+ with IHC, or FISH positive). Negative for amplification= (0, 1+ with IHC or FISH negative) 2+ staining with IHC considered equivocal and was usually associated with FISH assessment. FISH results supplanted IHC result when IHC indeterminate.
NC-BCFR	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals	Data were available from SEER tumor registry files and were extracted directly from pathology reports collected from hospitals
NHS/NHSII	Data were extracted from medical records where available and TMA otherwise. For TMA, >=1% of cells staining is positive.	Data were extracted from medical records where available and TMA otherwise. For TMA >=1% of cells staining is positive.	Data were extracted from medical records where available and TMA otherwise. Medical records data came from both IHC and FISH, and the TMA are all based on IHC. 0,1=negative; 2,3=positive
OBCS	Defined by nuclear immunostaining on FFPE tissue sections 0 = negative staining 1 = positive staining: >2% of tumour cells with nuclear staining (DAKO, monoclonal, clone 1D5)	Defined by nuclear immunostaining on FFPE tissue sections 0 = negative staining 1 = positive staining: >2% of tumour cells with nuclear staining (DAKO, monoclonal, clone PgR636)	Defined as membranous immunostaining on FFPE tissue sections 0 = completely negative 1 = positive staining: faint, moderate or strong membranous positivity (DAKO, polyclonal)
PLCO	Data were extracted from medical records when available. Qualitative	Data were extracted from medical records when available. Qualitative data	Data were extracted from medical records when available. Qualitative data from FISH

Nature Genetics: doi:10.1038/ng.985

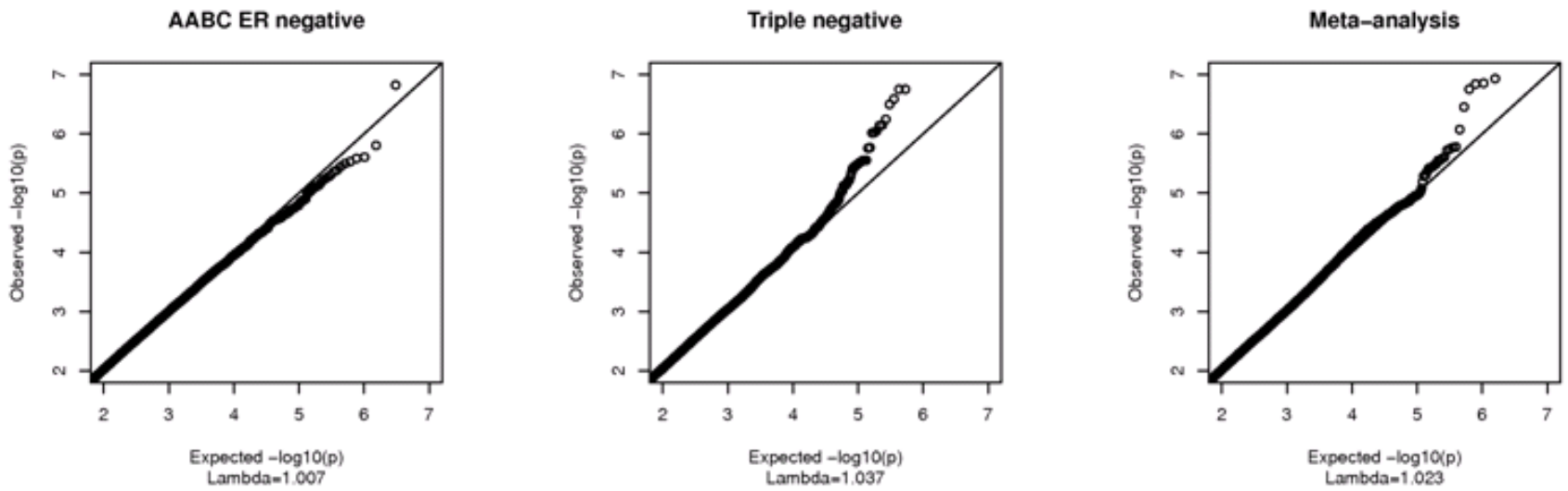
	results (positive/negative) were accepted as reported. For quantitative results, ≥ 10% of cells staining was positive, 1-9% was low positive, and 0% was negative.	(positive/negative) were accepted as reported. For quantitative results, ≥ 10% of cells staining was positive, 1-9% was low positive, and 0% was negative.	(positive/negative) were accepted as reported. For FISH ratio data, > 0 to < 1.8 was negative, 1.8-2.2 was low positive, and > 2.2 was positive. For IHC data, 0 or 1+ was negative, 2+ was equivocal, and 3+ was positive.
POSH	Data abstracted from clincal pathology report - where scored using Allred or equivalent system scores of <3 treated as negative.	Data abstracted from clincal pathology report - where scored using Allred or equivalent system scores of <3 treated as negative.	Data abstracted from clincal pathology report - most scored by IHC, negative score = +1 or 0 and positive score = +3. FISH for borderline (IHC score=2+). Only negative scores accepted.
RPCI	Positive=Allred score (intensity*percentage)= 3-8 (score range 0-8)	Positive=Allred score (intensity*percentage)= 3-8 (score range 0-8)	Positive=Score 3+ in >30% stained tumor cells or FISH amplified
SBCS	Positive=Intensity score (0,1,2,3) * percentage of cells stained (0-100%) >=50 (total score range 0-300) on TMA (Vector, 6F11/2 clone); abstracted from medical records	Positive=Allred score (intensity*percentage)= 3-8 (score range 0-8) on TMA (Vector, 1A6 clone)	Positive=Score 2+ on TMA (Dako, HercepTest™K5204)
SEARCH	Positive=Allred score (intensity*percentage)= 3-8 (score range 0-8) on TMA (Novocastra, 6F11/2 clone); abstracted from medical records	Positive=Allred score (intensity*percentage)= 3-8 (score range 0-8) on TMA (Dako, PgR 636 clone); abstracted from medical records	Positive=Score 2+ on TMA (Dako, HercepTest™K5204); abstracted from medical records
SFBCS	Positive = if >=1% of cells stain positive; data obtained from the SEER registry records	Positive = if >=1% of cells stain positive; data obtained from the SEER registry records	n/a

Nature Genetics: doi:10.1038/ng.985

SKKDKFZS	Negative: Remmele Score <3. If the Remmele Score was not available, the status was based on the biochemical analysis with <20 fmol/mg protein being negative.	Negative: Remmele Score <3. If the Remmele Score was not available, the status was based on the biochemical analysis with <20 fmol/mg protein being negative.	Negative: Score 0, 1; positive: Score 2, 3 on whole sections (Dako A0485, CerB2 clone) if no FISH
WASHU	Positive=Any nuclear staining on whole sections (Novocastra, 6F11/2 clone)	Positive=Any nuclear staining on whole sections (Dako, PgR 636 clone)	Positive=Complete strong cytoplasmic staining (3+) in >30% tumor cells on whole sections (Dako, HercepTest™) or 2+ and FISH positive
WCHS	Data abstracted from pathology reports from numerous hospitals. Scoring was based on clinical testing procedures (utilized for treatment decisions) at hospitals in NY/NJ.	Data abstracted from pathology reports from numerous hospitals. Scoring was based on clinical testing procedures (utilized for treatment decisions) at hospitals in NY/NJ.	Data abstracted from pathology reports from numerous hospitals. Scoring was based on clinical testing procedures (utilized for treatment decisions) at hospitals in NY/NJ.
WFBC	Positive=>10% cells stained (Clone 6F11), abstracted from pathology reports	Positive=>10% cells stained (Clone 16), abstracted from pathology reports	Negative: Score 0 or 1+; positive: Score 2+ or 3+ on whole sections (c-erbB-2-Dako Herceptest), abstracted from pathology reports
WHS	Data abstracted from physician review of medical records collected after self-report of breast cancer. Results coded into 5 categories: assay not performed, positive (≥10fmol/mg protein), negative, borderline, not found in medical record.	Data abstracted from physician review of medical records collected after self-report of breast cancer. Results coded into 5 categories: assay not performed, positive (≥10fmol/mg protein), negative, borderline, not found in medical record.	n/a

Nature Genetics: doi:10.1038/ng.985

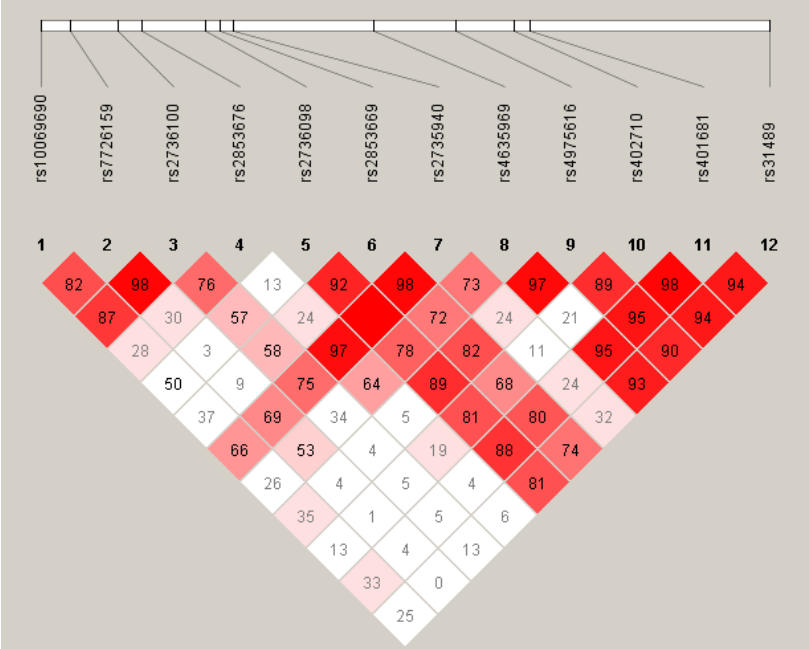
Supplementary Figure 1. Quantile-quantile plots for AABC, TNBCC and the meta-analysis of AABC and TNBCC.



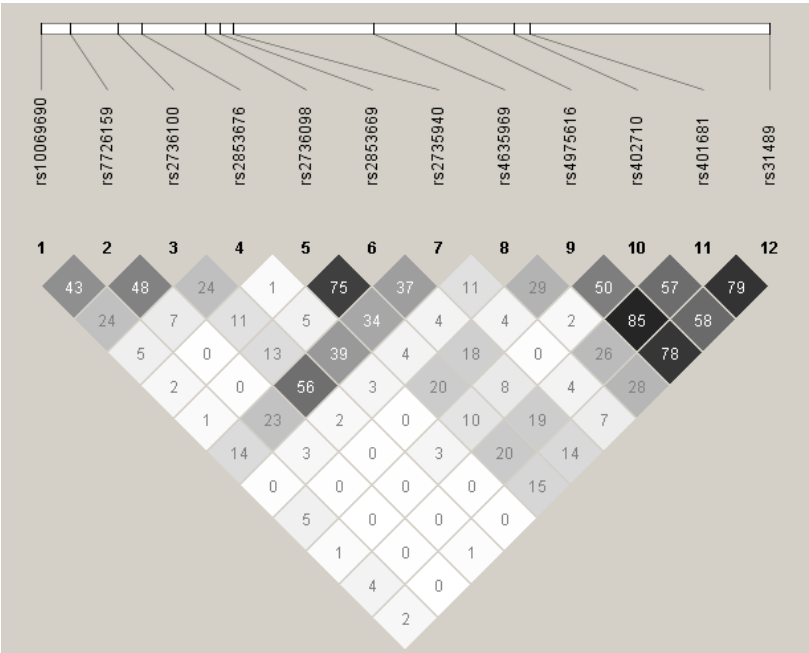
Nature Genetics: doi:10.1038/ng.985

Supplementary Figure 2. Correlations of cancer risk SNPs at 5p15 in populations of European and African ancestry from the 1000 Genomes Project.

European ancestry, D'

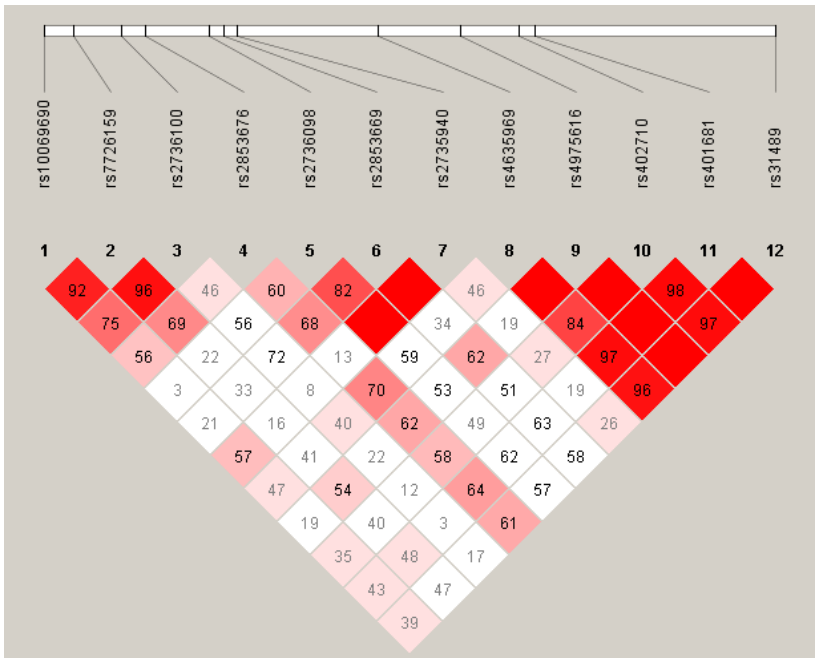


European ancestry, r^2

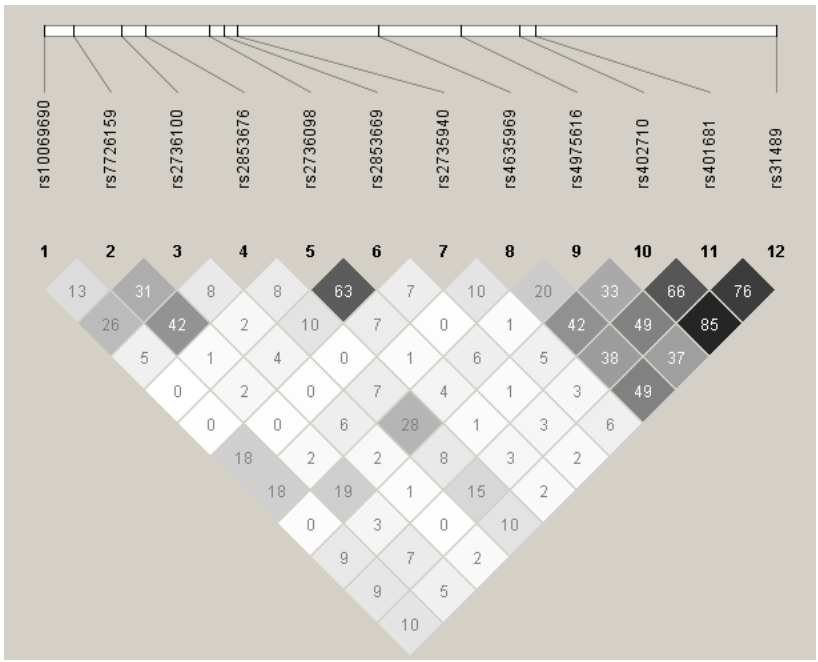


Nature Genetics: doi:10.1038/ng.985

African ancestry, D'



African ancestry, r^2



Nature Genetics: doi:10.1038/ng.985

Supplementary Note

Study Populations

Stage 1 included the studies of the African American Breast Cancer Consortium (AABC) and the Triple Negative Breast Cancer Consortium (TNBCC). The replication studies include the NCI Breast and Prostate Cancer Cohort Consortium (BPC3) and Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH). All participants in these studies have provided written informed consent for the research and approval for the study was obtained from the ethical review board from all local institutions. Below is a description of each study.

The African American Breast Cancer Consortium (AABC)

The Multiethnic Cohort Study (MEC): The Multiethnic Cohort Study is a population-based prospective cohort study (n=215,251) that was initiated between 1993 and 1996 and includes subjects from various ethnic groups – African Americans and Latinos primarily from Californian (great Los Angeles area) and Native Hawaiians, Japanese-Americans, and European Americans primarily from Hawaii. State drivers' license files were the primary sources used to identify study subjects in Hawaii and California. Additionally, in Hawaii, state voter's registration files were used, and, in California, Health Care Financing Administration (HCFA) files were used to identify additional African American study subjects. In the cohort, incident cancer cases are identified annually through cohort linkage to population-based cancer Surveillance, Epidemiology, and End Results (SEER) registries in Hawaii and Los Angeles County as well as to the California State cancer registry. Information on estrogen receptor status was also obtained through these registries. Blood sample collection in the MEC began in 1994 and targeted incident breast cancer cases and a random sample of study participants to serve as controls for genetic analyses. Subjects were frequency matched on age at blood draw and race/ethnicity. Through December, 31 2007, a nested breast cancer case-control study in the MEC included 556 African American cases (544 invasive and 12 in situ) and 1,003 African American controls. An additional 178 African American breast cancer cases (ages: 50-84 years) diagnosed between June 1, 2006 and December 31, 2007 in Los Angeles County (but outside of the MEC) were included in the study.

The Los Angeles component of The Women's Contraceptive and Reproductive Experiences (CARE) Study: The NICHD Women's CARE Study is a large multi-center population-based case-control study that was designed to examine the effects of oral contraceptive use on invasive breast cancer risk among African American women and white women ages 35-64 years in five U.S. locations. Cases from Los Angeles County were identified through the National Cancer Institute's local Surveillance, Epidemiology, and End Results (SEER) registry using rapid-reporting techniques and were diagnosed from July 1, 1994 through April 30, 1998. Information about tumor pathology is obtained through the registry. Controls were sampled by random-digit dialing from the same population and time period. In stage 1, 380 African American cases and 224 African American controls were genotyped.

The Women's Circle of Health Study (WCHS): The WCHS is an ongoing case-control study of breast cancer among women of European descent and African American women. Breast cancers are ascertained from hospitals in the NYC boroughs (Manhattan, the Bronx, Brooklyn and Queens) and in seven counties in New Jersey (Bergen, Essex, Hudson, Mercer, Middlesex, Passaic, and Union). In New Jersey, cases are identified through the New Jersey State Cancer Registry in collaboration with researchers at Cancer Epidemiology Services (CES) of the New Jersey Department of Health and Senior Services (NJDHSS). Pathology information, including tumor receptor status is collected from hospital pathology records. Eligible cases included women with invasive breast cancer between 20 and 74 years of age. Controls

Nature Genetics: doi:10.1038/ng.985

were identified through random digit dialing and are frequency matched to cases by 5-year age groups and race. The WCHS contributed 272 invasive African American cases and 240 African American controls.

The San Francisco Bay Area Breast Cancer Study (SFBCS): The SFBCS is a population-based case-control study of invasive breast cancer in Hispanic, African American and non-Hispanic White women conducted between 1995 and 2003 in the San Francisco Bay Area. Women newly diagnosed with breast cancer were identified through the Greater Bay Area Cancer Registry which ascertains all incident cancers as part of the Surveillance, Epidemiology, and End Results (SEER) program and the California Cancer Registry. African American cases, ages 35-79 years, were diagnosed between April 1, 1995 and April 30, 1999. Information on tumor pathology was obtained through the cancer registry. Controls were identified by random-digit dialing conducted from 1996 and 2001 in the same geographic area (i.e., 5 counties of the San Francisco Bay area), and frequency matched to cases on five-year age group and race/ethnicity. Included from this study were 172 invasive African American cases and 231 African American controls.

The Northern California Breast Cancer Family Registry (NC-BCFR): The NC-BCFR is a population-based family study conducted in the Greater San Francisco Bay Area, and is one of 6 sites participating in the international Breast Cancer Family Registry (BCFR). Newly diagnosed breast cancer cases were identified through the Greater Bay Area Cancer Registry and enriched with early-onset and familial breast cancer cases. African American breast cancer cases in NC-BCFR were diagnosed after January 1, 1995 and between the ages of 18 and 64 years. Information ontumor pathology was obtained through the cancer registry. Controls were identified through random-digit dialing conducted from 1999-2000 and were frequency-matched to cases by five-year age group and race/ethnicity. Genotyping was conducted for 440 invasive African American cases and 53 African American controls.

The Carolina Breast Cancer Study (CBCS): The CBCS is a population-based case-control study conducted between 1993 and 2001 in 24 counties of central and eastern North Carolina. Cases were identified by rapid case ascertainment system in cooperation with the North Carolina Central Cancer Registry. Information on tumor pathology was obtained through the cancer registry. Controls were selected from the North Carolina Division of Motor Vehicle and United States Health Care Financing Administration beneficiary lists and frequency-matched to cases on age and race. Participants' ages ranged from 20 to 74 years. DNA samples were provided from 656 African American cases with invasive breast cancer and 608 African American controls.

The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) Cohort: PLCO, coordinated by the U.S. National Cancer Institute (NCI) in 10 U.S. centers, enrolled during 1993 - 2001 approximately 155,000 men and women, aged 55-74 years, in a randomized trial to determine the efficacy of screening for these four cancers. Approximately 39,000 women were assigned to each arm of the trial: an intervention arm and a control arm in which women received their regular medical care. At entry, demographic, medical, and cancer risk factor information was collected through self-administered questionnaires. Sequential blood samples, the first obtained at baseline, were collected from participants assigned to the screening arm. Buccal cells were collected once, ~3 years after study entry, from the participants assigned to the control arm. All incident cancers are ascertained through annual questionnaires mailed to the participants. Cancers and vital status are also identified through the National Death Registry, state cancer registries, physician reports, and next-of-kin reports. Hospital records are used to confirm cancer diagnoses; for all confirmed breast cancer cases, hormone receptor status and other tumor characteristics are abstracted from hospital pathology reports. A total of 121 Black non-Hispanic women with no history of breast cancer were diagnosed with invasive breast cancer in both arms of the trial by December, 2008; 72 had DNA available for genotyping and provided

Nature Genetics: doi:10.1038/ng.985

informed consent. Using incidence density sampling, two controls, free of breast cancer at the age of diagnosis of the case, were matched to each case on race, study arm, date at cohort entry, and age at cohort entry. A total of 64 African American invasive breast cancer cases and 133 African American controls contributed to this study.

The Nashville Breast Health Study (NBHS): The NBHS is a population-based case-control study of breast cancer conducted in Tennessee. The study was initiated in 2001 to recruit patients with invasive breast cancer or ductal carcinoma in situ between the ages of 25 and 75 years. Cases were identified from participating hospitals in and around the Nashville Metropolitan area as well as from the Tennessee Cancer Registry (TCR). Diagnosis and tumor pathology were confirmed via medical record abstraction and ascertainment from the TCR. Controls were recruited through random digit dialing. NBHS contributed 310 African American cases (57 in situ), and 186 African American controls.

Wake Forest University Breast Cancer Study (WFBC): African American breast cancer cases and controls in WFBC were recruited at Wake Forest University Health Sciences from November 1998 through December 2008. Information on tumor pathology was obtained from the medical records. Controls were recruited from the patient population receiving routine mammography at the Breast Screening and Diagnostic Center. Age range of participants was 30-86 years. WFBC contributed 125 cases (116 invasive and 9 in situ) and 153 controls to the analysis.

The Triple-Negative Breast Cancer Consortium (TNBCC)

TNBCC was composed of 23 studies providing cases alone or cases and matched controls and four public control datasets. Triple negative status of cases was defined by clinical pathology report.

Amsterdam Breast Cancer Study (ABCS): Breast cancer cases were obtained from a Netherlands Cancer Institute – Antoni van Leeuwenhoek hospital -based consecutive case series of operable cases with invasive mammary carcinoma aged <50 years period 1995-2002 and enriched with familial breast cancer cases of all ages, counseled from 1995 to 2007. DNA samples from 67 triple negative breast cancer cases were genotyped for rs10069690.

Australia Breast Cancer Tissue Bank (ABCTB): Breast cancer cases were collected from six hospitals in New South Wales, Australia: Royal Prince Alfred Hospital, Westmead Hospital, Royal North Shore Hospital, St. Vincent's Hospital, Hunter Area Hospitals, and Port Macquarie beginning in 2006. A total of 166 triple negative breast cancer cases from the ABCTB tumor bank were genotyped in the Stage 1 GWAS, and 162 of these cases were re-genotyped for rs10069690.

Bavarian Breast Cancer Cases and Controls (BBCC): This is a consecutive series of cases with invasive breast cancer recruited at the University Breast Centre, Franconia in Northern Bavaria, Germany from 2002-2006. Cases were between 22-96 years of age. Controls were population-based unaffected women from the same geographical area. A total of 240 triple negative breast cancer cases were genotyped in the Stage 1 GWAS, and 325 cases were re-genotyped for rs100696902 of which 102 were not included in Stage 1 genotyping.

British Breast Cancer Study (BBCS): Cases were identified from English & Scottish Cancer Registries and oncological departments, including all breast cancer cases who developed a first primary before age 65 in 1971 or later and who subsequently developed a second primary cancer and unilateral breast cancer cases diagnosed before age 70 in 1971 or later. Cases were between 24-76 years of age. Population-based controls were identified as a friend, sister-in-law, daughter-in-law or other non-blood relative of

Nature Genetics: doi:10.1038/ng.985

cases. A total of 58 triple negative breast cancer cases and 58 controls from this study of cancer registry and National Cancer Research network (NCRN) based cases were genotyped for rs10069690.

Breast Cancer in Galway Genetic Study (BIGGS): Cases were unselected breast cancers recruited from University College Hospital Galway and surrounding hospitals in the West of Ireland since 2001 with an age range of 24-90 years. Controls were women > 60 years with no personal history of any cancer and no family history of breast or ovarian cancer, identified from retirement groups in the West of Ireland (same catchment area as cases) during the period 2001-2008. A total of 38 triple negative cases (ascertained from hospital pathology reports) and 86 controls from this study of hospital-based breast cancer cases and population-based controls were genotyped for rs10069690.

Cancer Genetic Markers of Susceptibility (CGEMS): The Nurses' Health Study (NHS) is a longitudinal study of 121,700 women enrolled in 1976. The CGEMS nested case-control study is derived from 32,826 participants who provided a blood sample between 1989 and 1990 and were free of diagnosed breast cancer at blood collection and followed for incident disease until June 1, 2004. Controls were not diagnosed with breast cancer during follow-up, and were matched to cases based on age at diagnosis, blood collection variables (time of day, season, and year of blood collection, as well as recent (<3 months) use of postmenopausal hormones), ethnicity (all cases and controls are self-reported Caucasians), and menopausal status (all cases were postmenopausal at diagnosis). Genotype data from a total of 1,142 controls were included in Stage 1 of this study.

DEMOKRITOS: Cases were enrolled from 1997 until 2010 in several major hospitals covering most geographical areas of Greece, such as Athens metropolitan area, Thessaloniki, Ioannina, Patras, and Crete (Chania), in collaboration with the Hellenic Cooperative Oncology Group (HECOG). Cases had an age range of 20-87 years. Controls were population-based unaffected women of the same age range. A total of 281 triple negative breast cancer cases from an unselected breast tumor series and 91 regional controls were genotyped for rs10069690 in this study.

Dana Farber Cancer Institute (DFCI): Cases were obtained from an unselected series of breast tumors patients from the Dana Farber Cancer Institute. DNA from residual bloods from 303 triple negative breast cancer patients were genotyped in the Stage 1 GWAS and DNA from 304 triple negative cases were genotyped for rs10069690.

Fox Chase Cancer Center (FCCC): Cases were between 28-80 years of age at diagnosis and DNA was obtained from peripheral blood samples. Comprehensive clinical data including histology, staging, treatment and outcomes was provided for all cases. Controls were healthy females with no personal cancer history matched geographically and by gender, race and age. A total of 148 triple negative breast cancer cases from the Fox Chase Cancer Center were genotyped in the Stage 1 GWAS. 159 triple negative cases, of which 148 were also genotyped in Stage 1, and 159 unaffected matched controls were genotyped for rs10069690.

Gene Environment Interaction and Breast Cancer in Germany (GENICA): GENICA institutions and investigators are Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University Tübingen, Germany (Hiltrud Brauch, Christina Justenhoven); 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, Volker Harth, Sylvia Rabstein).

Nature Genetics: doi:10.1038/ng.985

GENICA is a population-based case-control study of breast cancer in the Greater Bonn area of Germany. Cases were incident breast cancer cases enrolled between 2000 and 2004 from the Greater Bonn area (by of the hospitals within the study region), all of which were enrolled within 6 months of diagnosis. Cases were between 23-80 years of age. Controls were selected from population registries from 31 communities in the greater Bonn area and matched to cases in 5-year age classes between 2001 and 2004. 60 triple negative cases were genotyped in Stage 1. A total of 65 triple negative cases, of which 59 were included in Stage 1, and 66 controls were genotyped for rs10069690.

Helsinki Breast Cancer Study (HEBCS): Cases from this hospital-based case-control study in Southern Finland were consecutive breast cancer cases from the 1) Department of Oncology, Helsinki University Central Hospital 1997-8 and 2000, 2) consecutive cases from the Department of Surgery, Helsinki University Central Hospital 2001 – 2004, or 3) Familial breast cancer patients from the Helsinki University Central Hospital, Departments of Oncology and Clinical Genetics (from 1995). Cases were between 22 and 96 years of age. The population allele and genotype frequencies were obtained from the Finnish Genome Centre on 221 healthy population controls in the NordicDB, a Nordic pool and portal for genome-wide control data. A total of 85 triple negative breast cancer cases and 222 controls from this hospital-based case-control study were genotyped in Stage 1.

Karolinska Breast Cancer Study (KARBAC): This Swedish case-control study consisting of population- and hospital-based cases and geographically matched controls. Cases were collected consecutively from the Department of Oncology, Huddinge & Söder Hospital, Stockholm between 1998-2000. Controls were blood donors of mixed gender from the same geographical region. Excess material was received from all blood donors over a three month period in 2004 (approximately 3000) and DNA was extracted from a random sample of 1500 subjects. A total of 27 triple negative cases and 26 controls were genotyped for rs10069690.

Kuopio Breast Cancer Project (KBCP): Cases in this hospital-based prospective clinical cohort were women in the cohort seen at Kuopio University Hospital between 1990 and 1995 because of breast lump, mammographic abnormality, or other breast symptom and who were found to have breast cancer. A total of 36 triple negative breast cancers were genotyped for rs10069690.

Cooperative Health Research in the Region of Augsburg (KORA): In total, four population based health surveys have been conducted between 1984 and 2000 with 18,000 participants between the age of 25 to 74 years, and a biological specimen bank was established in order to enable the researchers to perform epidemiologic research with respect to molecular and genetic factors. The KORA study center conducts regular follow-up investigations and has collected a wealth of information on sociodemography, general medical history, environmental factors, smoking, nutrition, alcohol consumption, and various laboratory parameters. Follow-up activities include address inquiry for all participants (incl. assessment of vital status and cause of death), postal questionnaires focusing on chronic diseases, and complete follow-up studies with interviews and physical examination. Genotype data from 226 controls from this cohort study were included in Stage 1.

Leuven Multidisciplinary Breast Centre (LMBC): Cases from this hospital-based case control study in Leuven, Belgium included all patients diagnosed with breast cancer and seen in the Multidisciplinary Breast Center in Leuven (Gasthuisberg) since June 2007 plus retrospective collection of cases diagnosed since 2000. Healthy controls (blood donors) were collected at the Red Cross located in Gasthuisberg hospital (Oct-2007-March 2008). A total of 88 triple negative cases and 95 controls were genotyped for rs10069690.

Nature Genetics: doi:10.1038/ng.985

Mammary Carcinoma Risk Factor Investigation (MARIE): This is a population-based case-control study of breast cancer in Northern and Southern Germany. Cases from this study were incident and prevalent cases diagnosed from 2001-2005 in the study region of Hamburg in Northern Germany and from 2002-2005 in the study region of Rhein-Neckar-Karlsruhe in Southern Germany. Controls were randomly drawn from population registries and frequency matched by birth year and study region to the case. Controls were recruited from 2002 to 2006. 205 triple negative breast cancer cases were genotyped in Stage 1. A total of 231 triple negative cases, of which 132 were in Stage 1, and 248 controls were genotyped for rs10069690.

Mayo Clinic Breast Cancer Study (MCBCS): This is a clinic-based breast cancer case-control study at the Mayo Clinic. Subjects were enrolled between February 1, 2001 and June 30, 2005. Cases were comprised of Caucasian women with primary invasive breast cancer ascertained with 6 months of diagnosis. Controls were comprised of Caucasian women visiting the Mayo Clinic for general medical exams in the Department of Internal Medicine with no prior history of cancer. Controls were frequency matched to cases on region of residence, race, and 5-year age group. A total of 153 triple negative breast cancer cases were genotyped in Stage 1. Genotyping of rs10069690 included 152 triple negative cases and 155 controls.

Melbourne Collaborative Cohort Study (MCCS): This is a prospective cohort study. Incident cases of breast cancer were diagnosed within the Melbourne Collaborative Cohort Study in Melbourne, Australia during the follow-up from baseline (1990-1994) to 2008 of the 24469 participating women, and controls were randomly sampled from the initial cohort among members not diagnosed with breast cancer at the end of follow-up. A total of 41 triple negative breast cancer cases were genotyped in Stage 1. Re-genotyping of rs10069690 was performed for 58 triple negative cases, of which 38 were included in Stage 1, and 66 controls.

The Nashville Breast Health Study (NBHS): This population-based case-control study is described for BPC3. A total of 123 Caucasian triple negative breast cancer cases and 119 Caucasian controls were genotyped for rs10069690.

Oulu Breast Cancer Study (OBCS): This is a Finnish hospital-based case-control study. Cases were consecutive incident cases diagnosed at the Oulu University Hospital between 2000 and 2004. Controls were healthy, consecutive, anonymous, female Finnish Red-Cross blood donors recruited in 2002 from the same geographical region in Northern Finland. A total of 68 triple negative breast cancer cases and 96 controls were genotyped for rs10069690.

Prospective Study of Outcomes in Sporadic Versus Hereditary Breast Cancer (POSH): Cases from this prospective cohort study in the United Kingdom were aged 40 or younger at breast cancer diagnosis, recruited across the UK, and diagnosed between January 2000 and December 2007. A total of 274 cases of triple negative breast cancer were genotyped in Stage 1, and 273 of these cases were also re-genotyped for rs10069690.

Australian Twin Cohort study from the Queensland Institute of Medical Research (QIMR): Australian controls were unselected parents of adolescent twins taking part in studies of melanoma risk factors and cognition. Both mothers and fathers were collected but only mothers were used in this analysis. No phenotype data were collected for parents but a blood DNA sample was collected and typed using the Illumina 610K array. All controls were of northern European ancestry, mainly British Isles. Phenotypic and genotypic data collection was approved by the Queensland Institute of Medical Research (QIMR)

Nature Genetics: doi:10.1038/ng.985

Ethics Committee and informed consent was obtained from all participants. Genotype data from 659 controls were provided for Stage 1.

RPCI: The Data Bank and Biorepository (DBBR) at Roswell Park Cancer Institute (RPCI), is a comprehensive data and sample bank containing pretreatment biospecimens that are rigorously collected and processed, with comprehensive clinical and epidemiologic data. Briefly, patients newly diagnosed with cancer at RPCI are invited to participate during their initial visit with the surgical oncologist. After consent, blood samples are collected (prior to any treatment for breast cancer, including surgery) in phlebotomy when specimens for clinical measures are drawn, transported to the laboratory through a pneumatic tube system, and processed within one hour of blood draw. Specimens are maintained in liquid nitrogen until analysis. The average time interval between the time of diagnosis and the time of blood draw for the women in our study was 27 days. Data on tumor characteristics, including ER, PR and HER2 status are obtained from the Pathology Resource Network and matched to cases by an honest broker by medical record numbers. Healthy controls are identified from family members, friends or visitors (n=35, 6%) of patients with cancer other than breast, employee volunteers, and women recruited from community events such as Susan G. Komen Foundation sponsored Western New York Race for the Cure. Data and specimens from 142 women with triple negative breast cancer and 143 healthy controls were genotyped for rs10069690.

Sheffield Breast Cancer Study (SBCS): This is a hospital-based case-control study of breast cancer. The study consists of women with pathologically confirmed breast cancer recruited from surgical outpatient clinics at the Royal Hallamshire Hospital, Sheffield, 1998 – 2005 and unselected women attending the Sheffield Mammography Screening Service between Sep 2000 - Aug 2004 if their mammograms showed no evidence of a breast lesion. Cases are a mixture of prevalent and incident disease. A total of 43 triple negative breast cancer cases were genotyped in Stage 1. Genotyping of rs10069690 included 47 triple negative cases, of which 43 were included in Stage 1, and 54 controls.

Städtisches Klinikum Karlsruhe and Deutsches Krebsforschungszentrum Breast Cancer Study (SKKDKFZS): This breast cancer case cohort study consists of women with pathologically confirmed breast cancer recruited at the Städtisches Klinikum Karlruhe, Karlsruhe, Germany from 1993 - 2005. Cases were between 21-93 years of age. Controls were from an unselected series of unaffected women from the same geographical area. A total of 167 triple negative breast cancer cases and 170 controls were genotyped for rs10069690.

Wellcome Trust Case Control Consortium (WTCCC): The 1958 Birth Cohort (also known as the National Child Development Study) includes all births in England, Wales and Scotland, during one week in 1958. From an original sample of over 17,000 births, survivors were followed up at ages 7, 11, 16, 23, 33 and 42 yrs. In a biomedical examination at 44-45 yrs, 9,377 cohort members were visited at home providing 7,692 blood samples with consent for future Epstein–Barr virus (EBV)-transformed cell lines. DNA samples extracted from 1,500 cell lines of self-reported white ethnicity and representative of gender and each geographical region were selected for use as controls. Genotype data from 1,421 controls were included in Stage 1.

Washington University Young Women's Breast Cancer Study (WASHU): This breast cancer case cohort study consists of women with pathologically confirmed breast cancer identified through the Young Women's Breast Cancer Program at Washington University Siteman Cancer Center. A total of 92 triple negative breast cancer cases were genotyped for rs10069690.

The NCI Breast and Prostate Cancer Consort Consortium (BPC3)

Nature Genetics: doi:10.1038/ng.985

The following BPC3 studies were included in the present analysis: CPS-II, EPIC, NHS, NHSII, WHS, MCCS, PLCO and MEC 8,313 cases (1,308 ER negative and ER positive 5,017) and 10,879 controls

Cancer Prevention Study II Nutrition Cohort (CPS-II): The Cancer Prevention Study II Nutrition Cohort (CPS-II) was established in 1992 by the American Cancer Society; the cohort includes over 86,000 men and 97,000 women from 21 U.S. states who completed a mailed questionnaire in 1992. At baseline, the cohort was 97% white and the median age of participants was 63 (range: 40-92). Starting in 1997, follow-up questionnaires have been sent to surviving cohort members every other year to update exposure information and to ascertain occurrence of new cases of cancer; a >90% response rate has been achieved for each follow-up questionnaire. Incident cancers are verified through medical records, state cancer registries, or death certificates. From 1998 - 2001, blood samples were collected from a subgroup of 39,376 cohort members. To further supplement the DNA resources, during 2000 - 2001, buccal cell samples were collected by mail from an additional 70,004 cohort members. ER status was assessed through medical records, and state cancer registries. A total of 583 cases (35 ER negative, 433 ER positive) and 791 controls or European ancestry were genotyped for rs1006960.

European Prospective Investigation into Cancer and Nutrition (EPIC): The European Prospective Investigation into Cancer and Nutrition (EPIC) is a prospective study designed to investigate both genetic and non-genetic risk factors for different forms of cancer. Approximately 500,000 individuals (~ 366,500 women) in EPIC were recruited between 1992 and 2000, from 23 centers in 10 European countries. Overall approximately 400,000 subjects also provided a blood sample at recruitment. Case subjects were selected through follow-up among women who developed breast cancer after blood collection. Tumor pathology information, include ER status is determined using laboratory quantification methods such as cell staining, immunohistochemistry etc . The present study includes 2,533 breast cancer cases (522 ER negative and 1,107 ER positive), and 3,382 controls from all 10 participating countries: France, Norway, Denmark, Germany, Greece, Italy, the Netherlands, Spain, Sweden and the United Kingdom (UK).

The Harvard Cohorts: NHS, NHSII and WHS: The Nurses' Health Study (NHS) cohort was initiated in 1976, when 121,700 US registered nurses aged 30 to 55 returned an initial questionnaire. The NHS breast cancer case-control study is nested within a subcohort of 32,826 women that donated blood during 1989 and 1990 and followed until 2006 for incident disease until 2006. In 1989, 116,609 US additional registered nurses returned an initial questionnaire (NHSII). The NHSII breast cancer case-control study is nested within a subcohort of 29,611 women that donated blood during 1993 to 1995 and followed until 2005. Medical records were used to confirm the diagnoses in women who reported a diagnosis of breast cancer on the biennial questionnaires for both NHS and NHSII. Control subjects were matched to cases based on age, menopausal status, recent hormone replacement therapy, and blood-draw specific variables (such as date and time of day). A total of 1974 cases (256 ER negative and 1163 ER positive) and 2572 controls from NHS and 587 cases (86 ER negative and 317 ER positive) and 1176 controls from NHSII were genotyped for rs10069690. The Women's Health Study (WHS) began in 1992-1994 as a ~10 year trial of vitamin E in primary prevention of cancer among female healthcare professionals and included 28,263 women who provided consent for blood-based analyses. As in the NHS/NHS II populations, cancer diagnoses were reported by questionnaire and confirmed by review of medical records. In the WHS, 674 cases (98 ER negative and 540 ER positive) and 674 controls were genotyped for rs10069690. In all studies, pathology reports were reviewed to obtain information on ER and PR status.

Nature Genetics: doi:10.1038/ng.985

Melbourne Collaborative Cohort Study (MCCS): The MCCS is a prospective cohort study of more than 41,000 women and men living in Melbourne, Australia, aged 40 to 69 years at baseline (1990-1994). In this report we include 688 breast cancer cases (136 ER negative and 451 ER positive) and 766 controls. Incident cases of breast cancer were participants in the MCCS diagnosed with breast cancer during the follow-up from baseline to 2008 and ascertained through linkage with the Victorian Cancer Registry and population-based registries from other States in Australia. Controls were a random sample of the initial cohort of 24,469 women.

Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) Cohort: Details of the PLCO cohort study are provided above. SNP rs1006960 was genotyped for 799 invasive breast cancer cases (98 ER negative and 622 ER positive) and 1013 controls of White non-Hispanic ancestry in PLCO.

The Multiethnic Cohort (MEC): Details of the MEC are provided above. SNP rs10069690 was genotyped for 527 breast cancer cases (77 ER negative and 384 ER positive) and 561 controls of European ancestry in the MEC.

Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH)

SEARCH is a series of ongoing population-based studies in Eastern England, with cases ascertained through the Eastern Cancer Registration and Information Centre (ECRIC), and covering the counties of Cambridgeshire, Norfolk, Suffolk, Bedfordshire, Hertfordshire and Essex. SEARCH-breast began recruitment in 1996 and has recruited over 10,000 patients. Eligible patients include women aged 18-69 at diagnosis with invasive breast cancer. Approximately 66% of eligible patients are enrolled in the study. Study participants were asked to provide a blood sample for DNA analysis and to complete a comprehensive epidemiological questionnaire. Basic pathology data including grade and ER/PR/HER2 status are available for the majority of patients. For this study we genotyped 6,182 patients (933 ER negative and 3,434 ER positive) and 5,966 controls. The control samples (for 4,552 patients) were randomly selected from the Norfolk component of EPIC (European Prospective Investigation of Cancer), in the same region from which the cases have been recruited, or controls from SEARCH recruited through general practices that recruited cases.

Nature Genetics: doi:10.1038/ng.985

Acknowledgements

The participating AABC studies are 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 P50-CA58223, and Center for Environmental Health and Susceptibility, National Institute of Environmental Health Sciences, National Institutes of Health grant P30-ES10126), NBHS (National Institutes of Health grant R01-CA100374) and WFBC (National Institutes of Health grant R01-CA73629). PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute. The Breast Cancer Family Registry (BCFR) was supported by the National Cancer Institute, National Institutes of Health under RFA CA-95-011 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), grants from the Komen Foundation for the Cure and the Breast Cancer Research Foundation (BCRF); MARIE (Deutsche Krebshilfe e.V., grant 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 (by the 'Stichting tegen Kanker' (232-2008 and 196-2010); 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, and the Sigrid Juselius Foundation); FCCC (SPORE P-50CA83638, U01CA69631, 5U01CA113916, and the Eileen Stein Jacoby Fund); 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 and Oxford Biomedical Research Centre); ABCTB (National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation); WASHU (Celebrate Fitness); ABCS (Dutch Cancer Society grant number 2009-4363); KARBAC (The Stockholm Cancer Society); KBCP (The Finnish Cancer Society; the Academy of Finland (grant number 127220); EVO Research Fund (grant number 5654113 and 5501) of Kuopio University Hospital; and the strategic

Nature Genetics: doi:10.1038/ng.985

funding of the University of Eastern Finland). KBCP also thanks Mrs. Helena Kemiläinen, Mrs Aija Parkkinen and Mrs. Eija Myöhänen for their skillful technical assistance

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) and U01-CA98758 (MEC) and through the Intramural Research Program of the NCI Division of Cancer Epidemiology and Genetics (PLCO). The authors thank Dr. 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 MCCS was supported by Australian NHMRC grants 209057, 251553 and 504711 and infrastructure provided by the Cancer Council Victoria. We acknowledge the contribution of other MCCS investigators: Professor John L. Hopper and Professor Dallas R. English.

SEARCH is supported by Cancer Research UK grants C1287/A10118, C1287/A5260, C8197/A10123 and C490/A10124. DFE is a Principal Research Fellow of Cancer Research UK. AMD is supported by the Joseph Mitchell Fund.

We thank the women who volunteered to participate in each study. We also thank Madhavi Eranti, 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.

Nature Genetics: doi:10.1038/ng.985