Title

Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer.

Journal

Carcinogenesis, 35(5)

ISSN

0143-3334

Authors

Purrington, Kristen S
Slager, Susan
Eccles, Diana
et al.

Publication Date

2014-05-01

DOI

10.1093/carcin/bgt404

Carcinogenesis vol.35 no.5 pp.1012–1019, 2014
doi:10.1093/carcin/bgt404
Advance Access publication December 9, 2013

Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer

Kristen S.Purrington[1], Susan Slager[1], Diana Eccles[2], Drakoulis Yannoukakos[3], Peter A.Fasching[4,5], Penelope Miron[6], Jane Carpenter[7], Jenny Chang-Claude[8], Nicholas G.Martin[9], Grant W.Montgomery[9], Vessela Kristensen[10,11], Hoda Anton-Culver[12], Paul Goodfellow[13], William J.Tapper[2], Sajjad Rafiq[2], Susan M.Gerty[2], Lorraine Durcan[2], Irene Konstantopoulou[3], Florentia Fostira[3], Athanassios Vratimos[3], Paraskevi Apostolou[3], Irene Konstanta[3], Vassiliki Kotoula[14], Sotiris Lakis[15], Meletios A.Dimopoulos[16], Dimosthenis Skarlos[17], Dimitrios Pectasides[18], George Fountzilas[19], Matthias W.Beckmann[5], Alexander Hein[5], Matthias Ruebner[5], Arif B.Ekici[20], Arndt Hartmann[21], Ruediger Schulz-Wendtland[22], Stefan P.Renner[5], Wolfgang Janni[23], Brigitte Rack[24], Christoph Scholz[23], Julia Neugebauer[24], Ulrich Andergassen[24], Michael P.Lux[5], Lothar Haeberle[5], Christine Clarke[25], Nirmala Pathmanathan[26], Anja Rudolph[8], Dieter Flesch-Janys[27], Stefan Nickels[8], Janet E.Olson[1], James N. Ingle[28], Curtis Olswold[1], Seth Slettedahl[1], Jeanette E.Eckel-Passow[1], S.Keith Anderson[1], Daniel W.Visscher[29], Victoria L. Cafourek[1], Hugues Sicotte[1], Naresh Prodduturi[1], Elisabete Weiderpass[30,31,32], Leslie Bernstein[33], Argyrios Ziogas[12], Jennifer Ivanovich[13], Graham G.Giles[34], Laura Baglietto[34], Melissa Southey[35], Veli-Matti Kosma[36], Hans-Peter Fischer[37], The GENICA Network[38,39,40,41,42,43], Malcom W.R.Reed[44], Simon S.Cross[45], Sandra Deming-Halverson[46], Martha Shrubsole[46], Qiuyin Cai[46], Xiao-Ou Shu[46], Mary Daly[47], JoEllen Weaver[48], Eric Ross[49], Jennifer Klemp[50], Priyanka Sharma[50], Diana Torres[43,51], Thomas Rüdiger[52], Heidrun Wölfing[52], Hans-Ulrich Ulmer[53], Asta Försti[54,55], Thaer Khoury[56], Shicha Kumar[57], Robert Pilarski[58], Charles L.Shapiro[59], Dario Greco[60], Päivi Heikkilä[61], Kristiina Aittomäki[62], Carl Blomqvist[63], Astrid Irwanto[64], Jianjun Liu[64], Vernon Shane Pankratz[1], Xianshu Wang[29], Gianluca Severi[34], Arto Mannermaa[36], Douglas Easton[65], Per Hall[66], Hiltrud Brauch[38], Angela Cox[44], Wei Zheng[46], Andrew K.Godwin[67], Ute Hamann[43], Christine Ambrosone[68], Amanda Ewart Toland[69], Heli Nevanlinna[60], Celine M.Vachon[1] and Fergus J.Couch[1,29,*]

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA, [2]Faculty of Medicine, University of Southampton SO17 1BJ, Southampton, UK, [3]Molecular Diagnostics Laboratory INRASTES, National Centre for Scientific Research "Demokritos", Athens 153 10, Greece, [4]Department of Medicine, Division of Hematology/Oncology, University of California at Los Angeles, David Geffen School of Medicine, Los Angeles, CA 90095, USA, [5]Department of Gynecology and Obstetrics, University Breast Center Franconia, University Hospital Erlangen; Friedrich-Alexander University Erlangen-Nuremberg, Erlangen D-91012, Germany, [6]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA, [7]Australian Breast Cancer Tissue Bank, University of Sydney at the Westmead Millennium Institute, Westmead, New South Wales, NSW 2145 Australia, [8]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, [9]QIMR GWAS Collective, Queensland Institute of Medical Research, Brisbane, Queensland, QLD 4006 Australia, [10]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Oslo 0316, Norway, [11]Faculty of Medicine (Faculty Division Ahus), Universitetet i Oslo, Oslo 0316, Norway, [12]Department of Epidemiology, University of California–Irvine, Irvine, CA 92697, USA, [13]Washington University School of Medicine, Barnes-Jewish Hospital and Siteman Cancer Center, St Louis, MO 63110, USA, [14]Department of Pathology, Laboratory of Molecular Oncology and [15]Laboratory of Molecular Oncology, Hellenic Foundation for Cancer Research, Aristotle University of Thessaloniki School of Medicine, Thessaloniki 54124, Greece, [16]Department of Clinical Therapeutics, "Alexandra" Hospital, University of Athens School of Medicine, Athens 115 27, Greece, [17]Second Department of Medical Oncology, "Metropolitan" Hospital, Athens 151 25, Greece, [18]Second Department of Internal Medicine, Oncology Section, "Hippokration" Hospital, University of Athens School of Medicine, Athens 115 27, Greece, [19]Department of Medical Oncology, "Papageorgiou" Hospital, Aristotle University of Thessaloniki School of Medicine, Thessaloniki 54124, Greece, [20]Institute of Human Genetics, [21]Institute of Pathology and [22] Institute of Diagnostic Radiology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen D-91012, Germany, [23]Department of Gynecology and Obstetrics, University Hospital Ulm, Ulm 89069, Germany, [24]Department of Gynecology and Obstetrics, University Hospital Ludwig Maximilians University, Campus Innenstadt, Munich 80539, Germany, [25]Westmead Institute for Cancer Research, Sydney Medical School Westmead, University of Sydney at the Westmead Millennium Institute, Westmead, New South Wales NSW 2145, Australia, [26]Westmead Breast Cancer Institute, Westmead Hospital, Westmead, New South Wales NSW 2145, Australia, [27]Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg 20246, Germany, [28]Department of Oncology and [29]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA, [30]Department of Community Medicine, University of Tromsø, Tromsø 9019, Norway, [31]Folkhälsan Research Cancer Centre, Helsinki 00250, Finland, [32]Cancer Registry of Norway, Oslo N-0304, Norway, [33]Department of Population Sciences, Division of Cancer Etiology, Beckman Research Institute, City of Hope, Duarte 91010, USA, [34]Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Victoria VIC 3053, Australia, [35]Department of Pathology, The University of Melbourne, Melbourne, Victoria VIC 3053, Australia, [36]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine; Biocenter Kuopio, Cancer Center of Eastern Finland, and Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, University of Eastern Finland, Kuopio 80130, Finland, [37]Department of Pathology, Medical Faculty University Bonn, Bonn 53127, Germany, [38]Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart and University of Tuebingen, Tuebingen 72074, Germany, 72074 [39]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn 53127, Germany, [40]Institute for Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany, [41]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum D-44789, Germany, [42]Institute of Pathology, Medical Faculty of the University of Bonn, Bonn 53127, Germany, [43]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, [44]Department of Oncology, Cancer Research UK/Yorkshire Cancer Research Sheffield Cancer Research Centre and [45]Department of Neuroscience, University of Sheffield, Sheffield S10 2TN, UK, [46]Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Division of Epidemiology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA, [47]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA 19111-2497, USA, [48]PennMed Biobank, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA, [49]Department of Biostatistics and Bioinformatics, Fox Chase Cancer Center, Philadelphia, PA 19111-2497, USA, [50]Department of Oncology/Hematology, University of Kansas Medical Center, Kansas City, KS 66160, USA, [51]Institute of Human Genetics, Pontificia University Javeriana, Bogota

Abbreviations: AUC, areas under the curve; CI, confidence interval; ER, estrogen receptor; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; HER2, human epidermal growth factor receptor-2; HMEC, human mammary epithelial cell; iCOGS, custom Illumina Infinium array; LD, linkage disequilibrium; OR, odds ratio; PRS, polygenic risk score; SNP, single nucleotide polymorphism; TN, triple-negative; TNBCC, Triple Negative Breast Cancer Consortium.

D.C. 11001000, Colombia, [52] Institute of Pathology, Städtisches Klinikum Karlsruhe, Karlsruhe 76133, Germany, [53]Frauenklinik der Stadtklinik Baden-Baden, Baden-Baden 76530, Germany, [54]Center for Primary Health Care Research, University of Lund, Malmö 223 63, Sweden, [55]Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, [56] Department of Pathology and [57]Department of Surgical Oncology, Roswell Park Cancer Institute, Buffalo, NY 14263, USA, [58]Department of Internal Medicine, Division of Human Genetics and [59]Department of Internal Medicine, Division of Medical Oncology, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA, [60]Department of Obstetrics and Gynecology, [61]Department of Pathology, [62]Department of Clinical Genetics and [63]Department of Oncology, Helsinki University Central Hospital, Helsinki 00014, Finland, [64]Human Genetics Division, Genome Institute of Singapore, Singapore 138672 Singapore, [65]Department of Public Health and Primary Care, Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB2 1TN, UK, [66]Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 171 77, Sweden, [67]Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66160, USA, [68]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY 14263, USA, [69]Departments of Internal Medicine and Molecular Virology, Immunology and Medical Genetics, Division of Human Cancer Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA

*To whom correspondence should be addressed. Stabile 2–42, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel: (507) 284 3623; Fax: (507) 538 1937;
Email: couch.fergus@mayo.edu

Triple-negative (TN) breast cancer is an aggressive subtype of breast cancer associated with a unique set of epidemiologic and genetic risk factors. We conducted a two-stage genome-wide association study of TN breast cancer (stage 1: 1529 TN cases, 3399 controls; stage 2: 2148 cases, 1309 controls) to identify loci that influence TN breast cancer risk. Variants in the 19p13.1 and *PTHLH* loci showed genome-wide significant associations ($P < 5 \times 10^{-8}$) in stage 1 and 2 combined. Results also suggested a substantial enrichment of significantly associated variants among the single nucleotide polymorphisms (SNPs) analyzed in stage 2. Variants from 25 of 74 known breast cancer susceptibility loci were also associated with risk of TN breast cancer ($P < 0.05$). Associations with TN breast cancer were confirmed for 10 loci (*LGR6*, *MDM4*, *CASP8*, 2q35, 2p24.1, *TERT*-rs10069690, *ESR1*, *TOX3*, 19p13.1, *RALY*), and we identified associations with TN breast cancer for 15 additional breast cancer loci ($P < 0.05$: *PEX14*, 2q24.1, 2q31.1, *ADAM29*, *EBF1*, *TCF7L2*, 11q13.1, 11q24.3, 12p13.1, *PTHLH*, *NTN4*, 12q24, *BRCA2*, *RAD51L1*-rs2588809, *MKL1*). Further, two SNPs independent of previously reported signals in *ESR1* [rs12525163 odds ratio (OR) = 1.15, $P = 4.9 \times 10^{-4}$] and 19p13.1 (rs1864112 OR = 0.84, $P = 1.8 \times 10^{-9}$) were associated with TN breast cancer. A polygenic risk score (PRS) for TN breast cancer based on known breast cancer risk variants showed a 4-fold difference in risk between the highest and lowest PRS quintiles (OR = 4.03, 95% confidence interval 3.46–4.70, $P = 4.8 \times 10^{-69}$). This translates to an absolute risk for TN breast cancer ranging from 0.8% to 3.4%, suggesting that genetic variation may be used for TN breast cancer risk prediction.

Introduction

Triple-negative (TN) breast cancer is a distinct histopathological subtype of breast cancer that accounts for approximately 15% of all invasive breast cancers (1,2). This disease subtype is defined by low or no expression of estrogen receptor (ER), progesterone receptor and human epidermal growth factor receptor-2 (HER2). In addition, TN tumors tend to be of higher histologic grade, more proliferative, and have medullary and metaplastic features (1,3). Women with TN tumors are more likely to be *BRCA1* mutation carriers, young or premenopausal, African American or Hispanic ethnicity, and experience higher rates of disease recurrence and progression, especially within the first 3 years following treatment, compared with other breast cancer subtypes (4). TN breast cancer is also associated with low socioeconomic status, an earlier age at menarche, higher body mass index during premenopausal years, higher parity and lower lifetime duration of breast feeding (1,5).

In addition to these epidemiologic factors, several common genetic variants have been established as risk factors for TN breast cancer (6). Among these, 19p13.1 (7), *TERT*-rs10069690 (8) and *MDM4* (9) are specific to TN breast cancer, such that these loci are not associated with risk of ER-positive or ER-negative, HER2-positive breast cancer. Four other loci (*RALY*/*EIF2S2*, *LGR6*, 2p24.1, *FTO*-rs11075995) associated with ER-negative but not ER-positive breast cancer (9,10) may also influence TN breast cancer risk. More recently, a large study by the Breast Cancer Association Consortium identified 46 additional common breast cancer susceptibility loci (11–13). Although 26 of these loci were associated with ER-negative as well as ER-positive breast cancer, the influence of the loci on TN breast cancer and other histopathological subtypes of breast cancer has not yet been assessed.

Given the substantial heterogeneity in genetic risk profiles for different breast cancer subtypes that we and others have demonstrated (14–17), we hypothesized that additional genetic variants for TN breast cancer remain to be identified. These may include variants that could not be detected by previous breast cancer genome-wide association studies (GWAS) conducted predominantly with ER-positive breast cancer cases, and perhaps a subset of the 42 breast cancer hits recently identified by the Breast Cancer Association Consortium. In addition, recent evidence has shown that risk loci are often complex and may contain multiple independent risk-associated variants that influence different subtypes of breast cancer (11–13). Here, we present results from a comprehensive analysis of genetic variants and TN breast cancer within the Triple Negative Breast Cancer Consortium (TNBCC), including a two-stage GWAS of TN breast cancer, examining the contributions of known breast cancer risk loci to TN breast cancer in terms of overall associations, independent signals and expression quantitative trait loci (eQTL), and estimating the cumulative effect of all common genetic risk factors on TN breast cancer risk.

Materials and methods

Ethics statement

Study participants were recruited under protocols approved by the institutional review board at each institution and all subjects provided written informed consent.

Study participants: TNBCC

TNBCC subjects included in this analysis were recruited by 22 studies in 7 different countries (Supplementary Table S1, available at *Carcinogenesis* Online). In addition, data from four publicly available control GWAS data sets (Wellcome Trust Case Control Consortium UK 1958 Birth Cohort, National Cancer Institute's Cancer Genetic Markers of Susceptibility project, Cooperative Health Research in the Region of Augsburg study and the Australian Twin Cohort study from the Queensland Institute of Medical Research; $n = 3180$) were utilized. These studies are described in more detail in Supplementary Material, available at *Carcinogenesis* Online, and have been described in detail elsewhere (8,10,14).

Pathology and tumor markers

A TN breast cancer case was defined as an individual with an ER-negative, progesterone-receptor-negative and HER2-negative (0 or 1 by immunohistochemical staining) breast cancer diagnosed after age 18 years. Criteria used for defining ER, progesterone receptor and HER2 status varied by study and have been described previously (8,10,14).

TN breast cancer GWAS

Stage 1 of the TNBCC GWAS has been described previously (8,10,14). Briefly, 1529 TN breast cancer cases and 3399 country-matched controls from 10 study sites were genotyped using the Illumina 660-Quad single nucleotide polymorphism (SNP) array, CNV370 SNP array and 550-Duo SNP array (10). GWAS data for public controls were generated using the Illumina 660-Quad (Queensland Institute of Medical Research), Illumina 550 (v1; National Cancer Institute's Cancer Genetic Markers of Susceptibility), Illumina 550

(Cooperative Health Research in the Region of Augsburg) and Illumina 1.2M (Wellcome Trust Case Control Consortium UK 1958 Birth Cohort). Genotype data from the various GWAS were independently evaluated by an iterative quality control process as described previously (10). Common SNP genotypes were imputed to HapMap phase 2 (release 21). Quantile–quantile plots showed no substantial evidence for cryptic population substructure or differential genotype calling between cases and controls. We excluded all SNPs with a minor allele frequency <0.05, imputation quality score <0.5 and effect size (β) with absolute value <0.3.

TN breast cancer iCOGS (stage 2) genotyping

The design of the custom Illumina Infinium array (iCOGS) array (211 155 SNPs) and genotyping methods has been described previously (11). Briefly, samples were genotyped as part of the COGS project using the iCOGS array at two genotyping centers (Mayo Clinic and Genome Quebec). In this analysis, 1263 cases and 1105 controls from the TNBCC were genotyped on the iCOGS array at the Mayo Clinic, and 885 cases and 204 controls were genotyped at Genome Quebec. A total of 4628 from the 6087 TNBCC GWAS SNPs proposed for the iCOGS array yielded high-quality genotype data. A total of 147 762 SNPs from the iCOGS array overlapped with the TNBCC stage 1 GWAS data.

Complementary DNA-mediated annealing, selection, extension and ligation expression data

Expression profiles were generated for a total of 702 TN tumors (Supplementary Table S2, available at *Carcinogenesis* Online) using the Illumina whole genome complementary DNA-mediated annealing, selection, extension and ligation assay (v4.0). Tumor samples were either whole 10 µm sections or 1 mm cores from formalin-fixed paraffin-embedded tumor blocks. Whole sections were macrodissected to select the tumor region on the slide, guided by a pathologist-read hematoxylin and eosin stained slide from the same block. RNA was extracted using the Roche High Pure RNA Isolation Kit (Indianapolis, IN). Samples were plated randomly by study on 96-well plates with two universal human reference samples and two duplicate tumor RNA samples. Complementary DNA-mediated annealing, selection, extension and ligation expression profiling was performed by the Mayo Clinic Medical Genome Facility Gene Expression Core (Rochester, MN).

Statistical analyses

SNP analyses. Estimated per-allele log odds ratios (OR) and standard errors were calculated using unconditional logistic regression of the allele counts (dosage for imputed data). Analyses were adjusted by country of origin and principal components as described previously (10). Analyses assumed a log-additive genetic model and *P*-values were based on the one degree-of-freedom Wald test.
Expression data. Raw intensity values for tumor samples were summarized using box plots. After $\log_2$-transformation of raw intensity values, a per-sample quality (stress) measure was calculated (18). Samples with stress >0.5, denoting a 2-fold change in the overall expression values after normalization, and replicates with the higher stress measure were excluded ($n = 34$). Log_2-transformed intensity values were median-quantile normalized. Probes with a *P*-value of detection >0.05 in all samples were excluded ($n = 713$) for a total of 28 664 probes analyzed. Samples were median-centered by 96-well plate to correct for batch effects. Tumors with ESR1 (ILMN_1678535) expression values more than 1.5 standard deviations from the median were excluded ($n = 72$). Of the 596 remaining TN tumors, 486 also had genotype data from the pooled GWAS and iCOGS data and were used in subsequent analyses.
eQTL analyses. Cis-associations between SNPs and probe expression, defined as probes within 1 Mb of the SNP of interest, were calculated for the 24 loci of interest (Table I). Associations were evaluated using a robust linear model to appropriately account for outliers in the expression data. For the 30 TN-associated SNPs reported in this study, *cis*-eQTL associations at $P < 0.05$ were considered significant. For all remaining SNPs, a false discovery rate was generated using 100 permutations and *cis*-eQTLs were excluded at a 10% false discovery rate threshold (equivalent to $P < 1.0 \times 10^{-3}$).
Polygenic risk score. Polygenic risk scores (PRS) were calculated using a leave-one-out cross-validation approach. Two scores were calculated, one using all known breast cancer risk SNPs and one using the 30 TN breast cancer-associated risk SNPs reported in this study. For the first model, a total of 74 SNPs were used (Supplementary Table S3, available at *Carcinogenesis* Online), including proxy SNPs ($R^2 > 0.8$) from three of seven loci (1p13.2, *RALY* and *MKL1*) missing genotype data for the original breast cancer risk SNPs. For the second model, only the 30 SNPs associated with TN risk were included. For each subject, TN ORs were estimated for each SNP after dropping that subject from the data set. The log OR for the tested allele for each SNP was multiplied by the number of tested alleles (0, 1 or 2) for the subject. The PRS for a subject was calculated as the sum across SNPs. Quintiles were

Table I. Known breast cancer susceptibility SNPs associated with TN breast cancer

SNP	G/I	Chr	Position	Locus	Allele	TN			ER-negative (9)		
						OR	95% CI	*P*-value	OR	95% CI	*P*-value
Previously reported TN associations											
rs6678914	G	1	200453799	*LGR6*	A	0.90	0.84–0.97	3.31×10^{-3}	0.91	0.88–0.94	1.4×10^{-8}
rs4245739	I	1	202785465	*MDM4*	C	1.19	1.11–1.29	4.00×10^{-6}	1.14	1.10–1.18	2.1×10^{-12}
rs13387042	G	2	217614077	2q35	G	0.93	0.87–1.00	0.049	0.95	0.92–0.98	0.002
rs12710696	I	2	19184284	2p24.1	A	1.11	1.04–1.19	3.51×10^{-3}	1.10	1.06–1.13	4.6×10^{-8}
rs10069690	I	5	1332790	*TERT*	A	1.24	1.14–1.34	1.43×10^{-7}	1.15	1.11–1.20	4.5×10^{-12}
rs2736108[a]	G	5	1350488	*TERT*	T	0.77	0.69–0.87	8.33×10^{-6}	0.89[b]	0.83–0.93	1.41×10^{-8}
rs3757318	G	6	151955806	*ESR1*	A	1.33	1.17–1.51	9.25×10^{-6}	1.22	1.15–1.30	2.5×10^{-11}
rs2046210	I	6	151990059	*ESR1*	A	1.16	1.08–1.24	5.26×10^{-5}	1.15	1.11–1.19	4.9×10^{-16}
rs3803662	G	16	51143842	*TOX3*	A	1.09	1.01–1.17	0.022	1.14	1.10–1.18	5.5×10^{-13}
rs8170	G	19	17250704	19p13.1	A	1.26	1.16–1.37	1.26×10^{-7}	1.15	1.11–1.20	9.3×10^{-13}
rs2363956	G	19	17255124	19p13.1	C	0.82	0.77–0.88	2.33×10^{-8}			
Newly identified TN associations											
rs616488	G	1	10488802	*PEX14*	G	0.91	0.85–0.98	9.73×10^{-3}	0.91	0.88–0.94	1.0×10^{-8}
rs4849887	G	2	120961592	2q14.2	A	0.89	0.79–1.00	0.041	0.93	0.88–0.99	0.013
rs2016394	G	2	172681217	2q31.1	A	1.10	1.03–1.18	6.90×10^{-3}	1.00	0.97–1.04	0.85
rs6828523	I	4	176083001	*ADAM29*	A	0.84	0.75–0.93	1.33×10^{-3}	0.99	0.95–1.04	0.77
rs1432679	G	5	158176661	*EBF1*	G	1.10	1.02–1.17	8.62×10^{-3}	1.08	1.04–1.11	6.7×10^{-6}
rs7904519	G	10	114763917	*TCF7L2*	G	1.12	1.05–1.20	9.95×10^{-4}	1.06	1.03–1.09	2.9×10^{-4}
rs3903072	I	11	65339642	11q13.1	A	0.92	0.86–0.99	0.024	0.97	0.94–1.00	0.027
rs11820646	I	11	128966381	11q24.3	A	0.92	0.86–0.98	0.016	0.94	0.91–0.97	2.3×10^{-4}
rs12422552	I	12	14305198	12p13.1	C	1.13	1.04–1.21	2.70×10^{-3}	1.05	1.02–1.09	0.005
rs10771399	I	12	28046347	*PTHLH*	G	0.72	0.64–0.80	1.55×10^{-8}	0.83	0.79–0.87	2.4×10^{-12}
rs17356907	G	12	94551890	*NTN4*	G	0.90	0.84–0.97	7.55×10^{-3}	0.92	0.89–0.96	9.3×10^{-6}
rs1292011	G	12	114320905	12q24	G	1.08	1.01–1.16	0.035	0.99	0.96–1.02	0.44
rs11571833	I	13	31870626	*BRCA2*	T	1.44	1.05–1.96	0.023	1.52	1.31–1.77	6.0×10^{-6}
rs2588809	I	14	67730181	*RAD51L1*	A	0.91	0.83–1.00	0.041	1.00	0.96–1.05	0.94
rs6001930[a]	G	22	39206180	*MLK1*	C	1.21	1.02–1.43	0.025	1.14	1.08–1.20	1.6×10^{-6}

[a]Genotyped in stage 2 only on the iCOGS platform (2148 cases, 1309 controls).
[b]ER-negative breast cancer risk results for rs2736108 from Bojesen *et al.* (12).

determined based on the distribution of the PRS in controls. ORs for TN breast cancer were calculated comparing each quintile to the median (third) quintile or the lowest (first) quintile as the reference.

Cumulative risk estimates of TN breast cancer in United States Caucasian women were calculated using a multistep approach. Both age-specific Surveillance, Epidemiology, and End Results Program (SEER) breast cancer incidence rates (http://seer.cancer.gov) and age-specific ratios of TN breast cancer to overall breast cancer from the California Cancer Registry were obtained (3). Age-specific incidence rates for TN breast cancer were estimated by multiplying the overall age-specific breast cancer incidence rates from SEER by the calculated proportion of TN breast cancer among all breast cancers within age groups from the California Cancer Registry. Finally, we estimated the cumulative risk of TN breast cancer by integrating these age-specific incidence rates for TN breast cancer. Changes in cumulative risk by PRS quintile were calculated using the OR estimates obtained as described previously. Quintile-specific cumulative risk estimates were calculated by multiplying cumulative risk estimates by both the OR for that quintile and the attributable risk for the PRS. Attributable risk for the PRS was calculated using the following formula, where the OR for each case was assigned according to the quintile to which that case belonged:

$$\text{Attributable risk} = 1 - \frac{\sum_{i=1}^{n\,\text{cases}} \text{OR}_i^{-1}}{n\,\text{cases}} \quad (1)$$

Discriminatory accuracy of the PRS was assessed using receiver operating characteristic curves and corresponding areas under the curve (AUC) and 95% confidence intervals (CIs), generated using the fitted probabilities of TN cases status from a logistic regression model using the PRS as a continuous predictor variable.

Results

TNBCC two-stage GWAS

Stage 1 of the TN GWAS (8,10,14) was comprised of 1529 TN cases and 3399 country-matched controls (Supplementary Table S1, available at *Carcinogenesis* Online). There was no evidence for genomic inflation ($\lambda = 1.04$) (10), and no SNPs achieved genome-wide significance ($P < 5 \times 10^{-8}$). Candidate SNPs were selected for stage 2 replication based on a log-additive trend-test of directly genotyped SNPs ($P < 0.01$). A total of 4785 SNPs were included in stage 2 on the iCOGS genotyping array (11) and genotyped on 2148 TN cases and 1309 country-matched controls from the TNBCC (Supplementary Table S1, available at *Carcinogenesis* Online). In stage 2 alone, no SNPs achieved significance after Bonferroni correction for 4785 tests. However, there was substantial enrichment when comparing the observed with the expected number of SNPs at various levels of significance. Specifically, there were 357 SNPs (7.4%) at $P < 0.05$ compared with the expected number of 240 SNPs (1.5-fold enrichment), 48 SNPs at $P < 5 \times 10^{-3}$ compared with 24 expected (2-fold enrichment) and 9 SNPs compared with 2.4 expected (3.75-fold enrichment) at $P < 5 \times 10^{-4}$.

A pooled analysis of the TNBCC GWAS and iCOGS data for a total of 3677 TN cases and 4708 controls was performed. SNPs in the 19p13.1 (rs2363956 OR = 0.82, $P = 2.33 \times 10^{-8}$) and *PTHLH* (rs10771399 OR = 0.72, $P = 1.55 \times 10^{-8}$) loci displayed genome-wide significant associations with TN breast cancer (Table I). SNPs in the 19p13.1 locus have previously been specifically associated with both TN breast cancer and *BRCA1*-related breast cancer. SNPs in the *PTHLH* locus have previously been associated with breast cancer (9), but this is the first report of an association with TN breast cancer. After Bonferroni correction for 4785 tests, an additional five SNPs in *MDM4*, *ESR1*, *PTHLH* and 19p13.1 were significantly associated with risk of TN breast cancer (Supplementary Table S4, available at *Carcinogenesis* Online). Known associations between TN breast cancer and variants in the *MDM4* and *ESR1* loci (7,9,14) were also confirmed. The 10 SNPs with the lowest *P*-values not located in known breast cancer loci are shown in Supplementary Table S5, available at *Carcinogenesis* Online.

Known breast cancer susceptibility loci

Next, we evaluated whether any known breast cancer susceptibility SNPs that were genotyped or imputed in the combined TNBCC data were associated with risk of TN breast cancer (Supplementary Tables S3 and S6, available at *Carcinogenesis* Online). Genotype data were available for 74 of the 78 known breast cancer risk SNPs (Supplementary Table S3, available at *Carcinogenesis* Online). Of these, a total of 26 SNPs were associated with risk of TN breast cancer at $P < 0.05$ (Table I). These included 11 SNPs in the 2q35, *LGR6*, *MDM4*, *TERT*, *ESR1*, *TOX3* and 19p13.1 loci that were previously associated with TN breast cancer. Of these, rs2588809 in the *RAD51L1* locus replaced rs999737 from earlier studies as the SNP most significantly associated with TN breast cancer (Table I). A further 15 SNPs at the *PEX14*, 2q14.2, 2q31.1, *ADAM29*, *EBF1*, *TCF7L2*, 11q13.1, 11q24.3, 12p13.1, *NTN4*, *PTHLH*, 12q24, *BRCA2* and *MLK1* loci showed associations with TN breast cancer risk, which have not previously been described (Table I). In contrast, SNPs in *CASP8*, *MAP3K1* and *LSP1*, which had been marginally associated with TN breast cancer in other studies (6), were not associated with TN disease in this combined analysis. Furthermore, the *FTO* locus that was recently associated with ER-negative disease (9) was not significantly associated with TN breast cancer in this study (rs11075995 OR = 1.08, 95% CI 1.00–1.17, $P = 0.065$).

Two of the TN breast cancer risk loci we identified contained additional SNPs with lower *P*-values for TN breast cancer than the reported SNP (*ESR1* and *PEX14*; Supplementary Table S7a, available at *Carcinogenesis* Online). In 1000 genomes data from Caucasians (19), these new SNPs were in high linkage disequilibrium (LD) with the originally reported SNPs suggesting that the additional SNPs better capture the associations with TN breast cancer. Additionally, although the reported SNP in the *CASP8* locus was not associated with TN breast cancer risk, another highly correlated SNP (rs3731711; $R^2 = 0.93$) was significantly associated with risk ($P = 1.0 \times 10^{-4}$) (Supplementary Table S7b, available at *Carcinogenesis* Online). Finally, a SNP in the *RALY* locus, for which the reported SNP was not genotyped in this study, was significantly associated with TN risk (rs6142050 $P = 3.8 \times 10^{-3}$; Supplementary Table S7c, available at *Carcinogenesis* Online). The *RALY* SNP was in high LD with the reported SNP in these regions.

To better understand the patterns of risk associated with genetic variation in these TN-associated loci, we looked for independent signals in each locus by adjusting each SNP in a 250 kb region for the SNP with the lowest *P*-value. We found evidence for additional independent associations in the 19p13.1 locus (Supplementary Figure S1, available at *Carcinogenesis* Online) and the *ESR1* locus (Supplementary Figure S2, available at *Carcinogenesis* Online). In a multivariable model for 19p13.1, including rs8100241 and rs1864112, both SNPs remained strongly associated with risk of TN breast cancer (Table II). The newly identified rs1864112 is not in LD with

Table II. Multiple independent SNPs in 19p13.1 and *ESR1*

Locus	SNP	Previously reported	Single-SNP analysis			Multiple SNP regression		
			OR	95% CI	*P*-value	OR	95% CI	*P*-value
19p13.1	rs8100241	Yes	0.82	0.77–0.88	1.8×10^{-8}	0.81	0.75–0.97	1.8×10^{-9}
	rs1864112	No	0.86	0.79–0.92	6.8×10^{-5}	0.84	0.78–0.90	5.5×10^{-6}
ESR1	rs9397437	Yes	1.42	1.25–1.61	8.9×10^{-8}	1.15	1.27–1.65	1.6×10^{-8}
	rs12525163	No	1.12	1.04–1.21	3.0×10^{-3}	1.15	1.06–1.24	4.9×10^{-4}

rs8100241 (R^2 = 0.025) or rs8170 (R^2 = 0.093). Using data from the ENCODE project (20), we found that rs1864112 is located in a region overlapping a DNaseI hypersensitivity site and promoter-associated histone mark (H3KMe1) site in primary human mammary epithelial cells (HMEC), indicating that this SNP may play a role in transcriptional regulation. In *ESR1*, both rs9397437 and rs12525163 were associated with TN risk, with the significance of the association for rs12525163 increasing in the multivariate model (Table II). This SNP is not in LD with either of the *ESR1* SNPs previously associated with breast cancer risk (rs9397437, R^2 = 0.005; rs2046210, R^2 = 0.021) and does not overlap with any DNaseI hypersensitivity, H3K4Me1 or H3K4Me3 sites. These data provide evidence for two novel TN risk SNPs in 19p13.1 and *ESR1*.

eQTL for TN risk loci

To better understand the potential biological mechanisms that underlie the associations between SNPs in the 25 loci (Tables I and II, Supplementary Table S7b and c, available at *Carcinogenesis* Online) and risk of TN breast cancer, we conducted an eQTL analysis. Genome-wide messenger RNA expression data were available for 578 TN cases from corresponding clinically defined TN breast tumors, of which 62 were excluded because of *ESR1* expression in the tumors (see Materials and methods), for a total of 516 TN cases included in the eQTL analysis (Supplementary Table S2, available at *Carcinogenesis* Online). We then examined each of the 30 SNPs present in the 25 TN loci of interest (Tables I and II, Supplementary Table S7b and c, available at *Carcinogenesis* Online) for associations with gene expression. We found evidence for 51 *cis*-associations with the 30 TN risk SNPs ($P < 0.05$) (Supplementary Table S8, available at *Carcinogenesis* Online), involving 46 genes in the 25 loci. Functional annotation of the eQTL SNPs by HaploReg (21) showed that eQTL SNPs were more likely located in normal mammary epithelial cell enhancer elements (HMEC: 9 observed versus 3.1 expected, $P = 3.6 \times 10^{-3}$) and DNase hypersensitivity sites (HMEC: 7 observed versus 1 expected, $P = 7.5 \times 10^{-5}$).

A recent study functionally annotated SNPs in high LD ($R^2 > 0.5$) with 71 known breast cancer risk SNPs (22) using histone modification ChIP-seq and DNaseI-seq data published as part of the ENCODE project (20), formaldehyde-assisted isolation of regulatory elements data and publically available eQTL data. Twenty-three of the 25 TN risk loci we describe here were included in this report (Supplementary Table S9, available at *Carcinogenesis* Online); among these, 8 (34.8% in TN versus 26.8% overall) had high-LD SNPs in transcription start site regions, 17 (73.9% in TN versus 77.5% overall) had high-LD SNPs in enhancers and 6 (26.1% in TN versus 22.5% overall) had high-LD SNPs in exons, suggesting a slight enhancement for TN risk SNPs in transcription start site regions. The vast majority of functional SNPs identified by Rhie *et al.* (22) were not genotyped or imputed in our data. The functional SNPs rs633800 and rs11227311 in the 11q13.1 locus were associated with *CTSW* expression, which we also observed with the correlated index SNP, rs3903072 (Supplementary Table S8, available at *Carcinogenesis* Online).

We next analyzed all other SNPs in the 25 TN risk loci for eQTLs (within 1 Mb flanking the top risk SNP) and identified 41 candidate *cis*-eQTLs in 14 TN risk loci, involving 35 unique SNPs and 26 unique genes, based on a 10% false discovery rate threshold (Supplementary Table S10, available at *Carcinogenesis* Online). The 35 eQTL SNPs were enriched in HMEC enhancers (6 observed versus 1.9 expected, P = 0.012) and mammary ductal adenocarcinoma DNase hypersensitivity sites (T47D: 2 observed versus 0.4 expected, P = 0.049). Notably, the *MDM4*, *TERT* and 19p13.1 TN-specific risk loci contained *cis*-eQTLs (Supplementary Table S10, available at *Carcinogenesis* Online). Among these 35 eQTL SNPs, 8 were associated with *CTSW* expression and were in low to moderate LD ($0.084 \leq R^2 \leq 0.516$) with synonomous exonic mutations (Supplementary Table S11, available at *Carcinogenesis* Online), SNPs in transcription start site regions (Supplementary Table S12, available at *Carcinogenesis* Online) and SNPs in enhancers (Supplementary Table S13, available at *Carcinogenesis* Online) identified by Rhie *et al.* (22). No other eQTL SNPs we identified were correlated with putative functional SNPs.

Sensitivity analysis

We conducted a sensitivity analysis of all 30 TN risk SNPs identified in this study (Tables I and II, Supplementary Table S7b and c, available at *Carcinogenesis* Online) to evaluate the influence of potential misclassification with respect to ER status. We first examined the 30 SNPs in 578 TN cases with expression data and 4638 country-matched controls. The ORs for these SNPs were very similar to the ORs observed in the overall TN analysis (Supplementary Table S14, available at *Carcinogenesis* Online), although the reduction in sample size produced some variability. We then repeated the analysis after excluding 62 TN cases because of *ESR1* expression in the tumors. All ORs were in the same direction and similar in magnitude for the majority of these SNPs, with the exception of 2q14.2 and *ADAM29* moving slightly closer toward the null. Although the numbers are low, the results further strengthen the evidence that these 30 SNPs are associated with TN breast cancer risk.

Polygenic risk score

These results provide strong evidence that at least 24 of the 74 known breast cancer susceptibility SNPs are individually associated with risk of TN breast cancer (Table I). We implemented a PRS to approximate the combined effect of these SNPs on risk of TN disease. The PRS was calculated using all reported SNPs in known breast cancer loci for which genotype data were available (n = 74; Supplementary Table S3, available at *Carcinogenesis* Online), both to avoid bias from data-driven SNP selection and to account for SNPs that may be associated with TN risk that did not achieve significance in this study due to limited study size. Compared with the median quintile, an individual in the first or second quintile of the PRS was 0.51-fold or 0.76-fold less likely to have TN breast cancer, respectively (Table III). In contrast, an individual in the fourth or fifth quintile of the PRS was 1.29-fold or 2.05-fold more likely to have TN breast cancer compared with subjects in the median quintile. Further, our data show that there is more than 4-fold difference in risk comparing those in the highest versus lowest quintiles (Supplementary Table S15, available at *Carcinogenesis* Online). The receiver operating characteristic curves for predicting TN breast cancer using the 74 SNP PRS produced an AUC of 0.64 (95% CI 0.63–0.65; Supplementary Figure S3, available at *Carcinogenesis* Online). Applying the PRS to the population-based cumulative risk (up to age 90 years) of TN breast cancer among Caucasian women, defined as approximately 1.8% (see Materials and methods), yielded an estimated cumulative risk of TN breast cancer

Table III. PRS for TN breast cancer

PRS quintile	74 SNPs				30 SNPs			
	Quintile definitions	OR	95% CI	*P*-value	Quintile definitions	OR	95% CI	*P*-value
1	PRS ≤ 0.24	0.51	0.43–0.60	9.9×10^{-16}	PRS ≤ −0.57	0.52	0.45–0.62	3.9×10^{-15}
2	0.24 < PRS ≤ 0.58	0.76	0.67–0.90	1.1×10^{-3}	−0.57 < PRS ≤ −0.26	0.75	0.65–0.87	1.6×10^{-4}
3	0.58 < PRS ≤ 0.86	1.00	—	—	−0.26 < PRS ≤ 0.039	1.00	—	—
4	0.86 < PRS ≤ 1.24	1.29	1.12–1.48	4.6×10^{-4}	0.039 < PRS ≤ 0.40	1.37	1.20–1.57	6.7×10^{-6}
5	1.24 < PRS	2.05	1.80–2.33	1.8×10^{-25}	0.40 < PRS	2.13	1.87–2.43	1.1×10^{-29}

of 3.4% for women in the highest PRS quintile and 0.8% for women in the lowest PRS quintile (Figure 1).

To better understand how the additional TN risk SNPs reported in this study contribute to cumulative risk beyond the 74 overall breast cancer variants, the PRS was recalculated using all 30 TN risk SNPs identified in this study (Tables I and II, Supplementary Table S7b and c, available at *Carcinogenesis* Online). Estimates were slightly stronger for each PRS quintile compared with the 74 SNP PRS (Table III), and the discriminatory accuracy of the 30 SNP PRS was comparable with the 74 SNP PRS (Supplementary Figure S3, available at *Carcinogenesis* Online). This suggests that the identification of additional TN risk loci may improve the stratification of cumulative risk estimates for TN breast cancer (Supplementary Figure S4, available at *Carcinogenesis* Online). These findings also suggest that additional prospective studies are needed in order to understand the implications of these genetic data for risk prediction of TN and other subtypes of breast cancer. Considering all known TN risk variants simultaneously is a significant step toward understanding how common genetic variants can be used for TN risk prediction, which will be enhanced by the incorporation of traditional epidemiologic risk factors in future studies.

Discussion

In this report, we present results from the first 2-stage GWAS of TN breast cancer in Caucasian women. Variants in the *PTHLH* and 19p13.1 loci showed genome-wide significant associations ($P < 5.0 \times 10^{-8}$) with TN disease (Tables I and II). Ten SNPs with near-genome associations with TN breast cancer (Supplementary Table S5, available at *Carcinogenesis* Online) warrant follow-up in larger studies of TN breast cancer. In addition, 26 of 74 known overall breast cancer risk SNPs were associated with TN breast cancer (Table I, Supplementary Table S6, available at *Carcinogenesis* Online). Specifically, this study confirmed TN associations with SNPs in 10 loci (*LGR6*, *MDM4*, *CASP8*, 2q35, 2p24.1, *TERT*-rs10069690, *ESR1*, *TOX3*, 19p13.1 and *RALY*) and identified TN associations with 15 other loci. Furthermore, two novel signals that are independent of previously known risk-associated SNPs were identified in the *ESR1* and 19p13.1 loci (Table II). Given the complexity of known breast cancer risk loci such as *CCND1* and *TERT* (12,13), further studies involving extensive fine mapping, haplotyping and functional characterization are needed for full understanding of the relationship between genetic variation in these loci and risk of TN breast cancer.

To gain some insight into whether the TN risk SNPs we identified have stronger effects for TN breast cancer compared with ER-negative breast cancer, we compared 25 of the SNPs in our combined analysis for which data were available from a recent ER-negative meta-analysis (9). As expected, stronger ORs were observed in our TN study compared with the ER-negative study for *MDM4*, *TERT* (rs10069690) and 19p13.1 (Table I), which have previously been shown to be TN-specific loci (7–9). In addition, stronger ORs were observed in our TN study for 2q14.2, *ESR1*, *TCF7L2*, 11q13.1, 12p13.1 and *PTHLH* in TN compared with the ER-negative study. Furthermore, four of the TN loci (2q31.1, *ADAM29*, 12q24 and *RAD51L1*-rs2588809) had no reported association with ER-negative breast cancer. Studies that directly compare ER-negative, non-TN to TN breast cancer are required to determine whether any of these loci are TN specific.

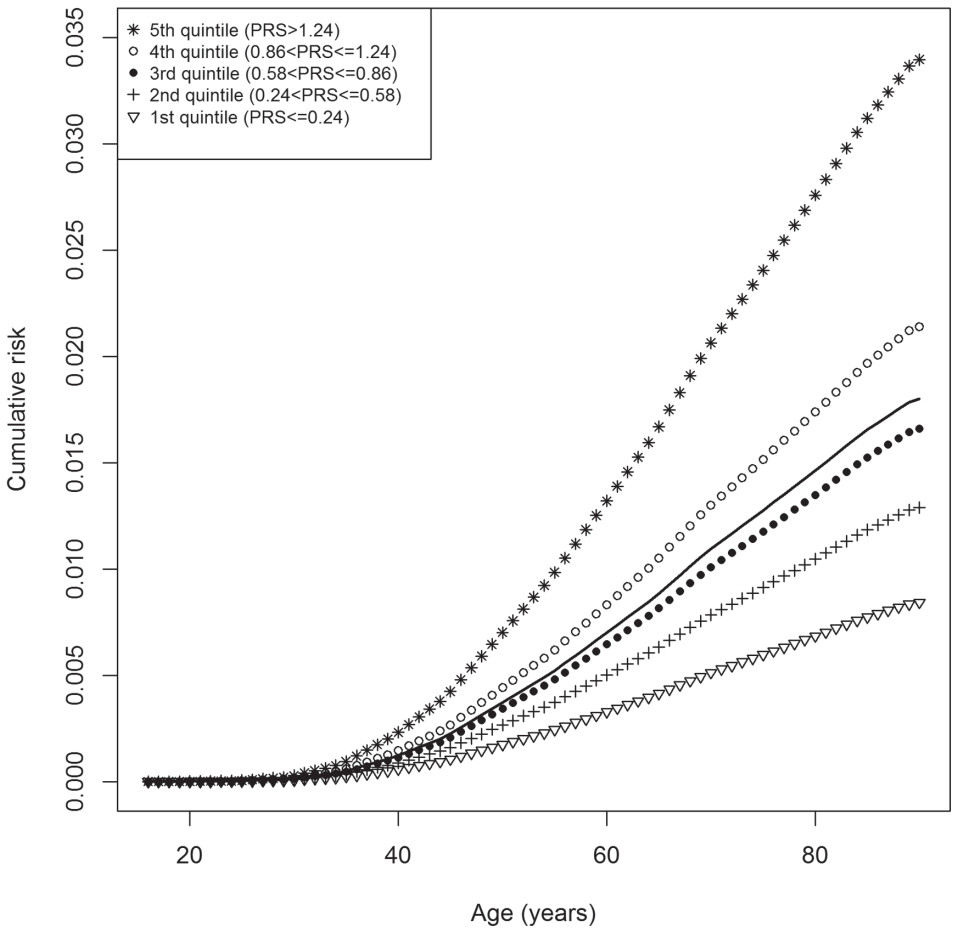

Fig. 1. Cumulative incidence of TN breast cancer stratified by 74 SNP PRS. The effect of the 74 SNP PRS on cumulative risk of TN breast cancer among Caucasian women, stratified by PRS quintile, is shown. The population-based cumulative risk curve is shown as a solid black line, and the first through fifth quintile-specific cumulative risk estimates are shown as indicated by labels.

In addition, we have provided evidence for SNP-mediated regulation of gene expression in these TN risk loci through *cis*-eQTL analyses involving over 500 TN breast tumors. Many of the 27 TN risk SNPs (Supplementary Table S8, available at *Carcinogenesis* Online) and an additional 35 SNPs in the TN risk loci (Supplementary Table S10, available at *Carcinogenesis* Online) that were associated with gene expression were located in transcriptional enhancers and DNase hypersensitivity sites in normal mammary epithelial cell lines, suggesting direct effects on gene transcription. Several interesting candidate genes were identified as *cis*-eQTLs. *PTHLH*, which encodes parathyroid hormone-like hormone, influences mammary gland development through regulation of epithelial to mesenchymal cellular interactions, is involved in lactation and is expressed in 60% of breast cancers (23–25). *IGFBP2* (insulin-like growth factor binding protein 2) in the 2q35 locus displays elevated expression in breast tumors and promotes the growth and survival of breast epithelial cells though regulation of the ER-α (26,27). *TBX3* in the 12q24 locus encodes T-box 3, a transcription factor involved in developmental regulation that is overexpressed in breast tumors (28) and can induce mammary stem-like cells and mammary gland hyperplasia in mice (29). Although the *cis*-eQTL results suggest mechanisms by which certain loci influence TN breast cancer risk, additional functional validation of these SNP–gene expression relationships in breast cancer cell lines is needed.

Beyond etiology, the identification of 30 TN risk SNPs provides an opportunity to better understand how genetic variation may inform TN breast cancer risk prediction. As we have shown through our PRS, where we observed a 4-fold difference in risk between the highest and lowest PRS quintiles of the TN breast cancer population, it may be possible to identify women who are substantially above or below population level risk of TN breast cancer. Our PRS had better discriminatory accuracy (AUC = 0.64) compared with that of the Gail model applied in the Women's Health Initiative (overall AUC = 0.58, 95% CI 0.58–0.62; ER-negative AUC = 0.50, 95% CI 0.45–0.54) (30). It is also likely that the inclusion of additional TN breast cancer risk SNPs will further stratify these women with respect to cumulative incidence of TN breast cancer. It will also be important to combine these TN risk SNPs with known epidemiologic risk factors such as parity, age at menarche, body mass index during premenopausal years and duration of breast feeding (1,5) to understand the cumulative influence on TN breast cancer risk. An important limitation of this study was that the PRS was applied to the study population from which the TN breast cancer risk estimates were derived. Although our cross-validation approach mitigates potential bias arising from this approach, it will be important to develop a risk model with these SNPs and validate the model in an independent study population. Overall, the findings provide strong evidence that integration of SNPs into predictive models will have a substantial impact on our ability to identify women at elevated risk of TN breast cancer.

Supplementary material

Supplementary Materials, Tables 1–15 and Figures 1–4 can be found at http://carcin.oxfordjournals.org/

Funding

Breast Cancer Research Foundation, National Institutes of Health (R01 CA128978), National Cancer Institute specialized program of research excellence (SPORE) in Breast Cancer (P50 CA116201); Mayo Cancer Genetic Epidemiology Training Grant (R25 CA092049). MCBCS was supported by the David and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation; The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany (01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114), the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany; OSU studies were supported by the OSU Comprehensive Cancer Center and the Stefanie Spielman Fund; SBCS was supported by Yorkshire Cancer Research Programme (S305PA); SKKDKFZS is supported by the DKFZ. The Demokritos has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund; The University of Kansas Cancer Center's (P30 CA168524) Biospecimen Repository Core Facility. A.K.G. was funded by National Institutes of Health (5U01CA113916 and R01CA14032), a grant from the Kansas Bioscience Authority Eminent Scholar Program and by the Chancellors Distinguished Chair in Biomedical Sciences Professorship; The Australian Twin Cohort Study was supported by National Health and Medical Research Council of Australia; The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the Academy of Finland and by the strategic funding of the University of Eastern Finland; The HEBCS study has been financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (132473), the Finnish Cancer Society and the Sigrid Juselius Foundation. The population allele and genotype frequencies were obtained from the data source funded by the Nodic Center of Excellence in Disease Genetics based on samples regionally selected from Finland, Sweden and Denmark.

Acknowledgements

We acknowledge the support of the Mayo Clinic Genotyping Core and Mayo Clinic Expression Core SKKDKFZS. We are grateful to all the patients for their participation. We thank the physicians, other hospital staff and research assistants who contributed to the patient recruitment, data collection and sample preparation. KBCP thank Eija Myöhänen and Helena Kemiläinen for technical assistance. HEBCS thank research nurses Hanna Jäntti and Irja Erkkilä for their help with the patient data and samples and Drs Ari Ristimäki, Tuomas Heikkinen, Mira Heinonen and Laura Hautala for their help with the tumor marker and pathology information, and gratefully acknowledges the Finnish Cancer Registry for the cancer data.

Conflict of Interest Statement: None declared.

References

1. Foulkes,W.D. *et al.* (2010) Triple-negative breast cancer. *N. Engl. J. Med.*, **363**, 1938–1948.
2. American Cancer Society. (2012) *Breast Cancer Facts & Figures, 2011–2012*. American Cancer Society, Atlanta, GA.
3. Bauer,K.R. *et al.* (2007) Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer*, **109**, 1721–1728.
4. Nofech-Mozes,S. *et al.* (2009) Patterns of recurrence in the basal and non-basal subtypes of triple-negative breast cancers. *Breast Cancer Res. Treat.*, **118**, 131–137.
5. Millikan,R.C. *et al.* (2008) Epidemiology of basal-like breast cancer. *Breast Cancer Res. Treat.*, **109**, 123–139.
6. Stevens,K.N. *et al.* (2013) Genetic susceptibility to triple-negative breast cancer. *Cancer Res.*, **73**, 2025–2030.
7. Stevens,K.N. *et al.* (2012) 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. *Cancer Res.*, **72**, 1795–1803.
8. Haiman,C.A. *et al.* (2011) A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.*, **43**, 1210–1214.
9. Garcia-Closas,M. *et al.* (2013) Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.*, **45**, 392–398.
10. Siddiq,A. *et al.* (2012) A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum. Mol. Genet.*, **21**, 5373–5384.
11. Michailidou,K. *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, **45**, 353–361.

12. Bojesen,S.E. *et al.* (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.*, **45**, 371–384.
13. French,J.D. *et al.* (2013) Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am. J. Hum. Genet.*, **92**, 489–503.
14. Stevens,K.N. *et al.* (2011) Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res.*, **71**, 6240–6249.
15. Figueroa,J.D. *et al.* (2011) Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Hum. Mol. Genet.*, **20**, 4693–4706.
16. Broeks,A. *et al.* (2011) Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum. Mol. Genet.*, **20**, 3289–3303.
17. Milne,R.L. *et al.* (2011) Confirmation of 5p12 as a susceptibility locus for progesterone-receptor-positive, lower grade breast cancer. *Cancer Epidemiol. Biomarkers Prev.*, **20**, 2222–2231.
18. Mahoney,D.W. *et al.* (2013) Quality assessment metrics for whole genome gene expression profiling of paraffin embedded samples. *BMC Res. Notes*, **6**, 33.
19. Johnson,A.D. *et al*. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.
20. Consortium,E.P., *et al*. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
21. Ward,L.D. *et al.* (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–D934.
22. Rhie,S.K. *et al.* (2013) Comprehensive functional annotation of seventy-one breast cancer risk loci. *PLoS One*, **8**, e63925.
23. Fleming,N.I. *et al*. (2009) Parathyroid hormone-related protein protects against mammary tumor emergence and is associated with monocyte infiltration in ductal carcinoma in situ. *Cancer Res*., **69**, 7473–7479.
24. Sowers,M.F. *et al*. (1996) Elevated parathyroid hormone-related peptide associated with lactation and bone density loss. *JAMA*, **276**, 549–554.
25. Southby,J. *et al*. (1990) Immunohistochemical localization of parathyroid hormone-related protein in human breast cancer. *Cancer Res*., **50**, 7710–7716.
26. Foulstone,E.J. *et al*. (2013) Insulin-like growth factor binding protein 2 (IGFBP-2) promotes growth and survival of breast epithelial cells: novel regulation of the estrogen receptor. *Endocrinology*, **154**, 1780–1793.
27. Busund,L.T. *et al.* (2005) Significant expression of IGFBP2 in breast cancer compared with benign lesions. *J. Clin. Pathol.*, **58**, 361–366.
28. Yarosh,W. *et al*. (2008) TBX3 is overexpressed in breast cancer and represses p14 ARF by interacting with histone deacetylases. *Cancer Res*., **68**, 693–699.
29. Liu,J. *et al*. (2011) TBX3 over-expression causes mammary gland hyperplasia and increases mammary stem-like cells in an inducible transgenic mouse model. *BMC Dev. Biol*., **11**, 65.
30. Chlebowski,R.T. *et al.* (2007) Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J. Natl Cancer Inst.*, **99**, 1695–1705.

Received August, 5, 2013; revised November 4, 2013; accepted November 27, 2013