Identification of novel susceptibility loci and genes for breast cancer risk: A transcriptome-wide association study of 229,000 women of European descent

Lang Wu[1,160], Wei Shi[2,160], Jirong Long[1], Xingyi Guo[1], Kyriaki Michailidou[3,4], Jonathan Beesley[2], Manjeet K. Bolla[3], Xiao-Ou Shu[1], Yingchang Lu[1], Qiuyin Cai[1], Fares Al-Ejeh[2], Esdy Rozali[2], Qin Wang[3], Joe Dennis[3], Bingshan Li[151], Chenjie Zeng[1], Helian Feng[5,6], Alexander Gusev[153, 154, 155], Richard T. Barfield[5], Irene L. Andrulis[7,8], Hoda Anton-Culver[9], Volker Arndt[10], Kristan J. Aronson[11], Paul L. Auer[12,13], Myrto Barrdahl[14], Caroline Baynes[15], Matthias W. Beckmann[16], Javier Benitez[17,18], Marina Bermisheva[19,20], Carl Blomqvist[21,159], Natalia V. Bogdanova[20,22,23], Stig E. Bojesen[24-26], Hiltrud Brauch[27-29], Hermann Brenner[10,29,30], Louise Brinton[31], Per Broberg[32], Sara Y. Brucker[33], Barbara Burwinkel[34,35], Trinidad Caldés[36], Federico Canzian[37], Brian D. Carter[38], J. Esteban Castelao[39], Jenny Chang-Claude[14,40], Xiaoqing Chen[2], Ting-Yuan David Cheng[41], Hans Christiansen[22], Christine L. Clarke[42], NBCS Collaborators[43-120,44-124,45], Margriet Collée[46], Sten Cornelissen[47], Fergus J. Couch[48], David Cox[49,50], Angela Cox[51], Simon S. Cross[52], Julie M. Cunningham[48], Kamila Czene[53], Mary B. Daly[54], Peter Devilee[55,56], Kimberly F. Doheny[57], Thilo Dörk[20], Isabel dos-Santos-Silva[58], Martine Dumont[59], Miriam Dwek[60], Diana M. Eccles[61], Ursula Eilber[14], A. Heather Eliassen[6,62], Christoph Engel[63], Mikael Eriksson[53], Laura Fachal[15], Peter A. Fasching[16,64], Jonine Figueroa[31,65], Dieter Flesch-Janys[66,67], Olivia Fletcher[68], Henrik Flyger[69], Lin Fritschi[70], Marike Gabrielson[53], Manuela Gago-Dominguez[71,72], Susan M. Gapstur[38], Montserrat García-Closas[31], Mia M. Gaudet[38], Maya Ghoussaini[15], Graham G. Giles[73,74], Mark S. Goldberg[75,76], David E. Goldgar[77], Anna González-Neira[17], Pascal Guénel[78], Eric Hahnen[79-81], Christopher A. Haiman[82], Niclas Håkansson[83], Per Hall[53], Emily Hallberg[84], Ute Hamann[85], Patricia Harrington[15], Alexander Hein[16], Belynda Hicks[86], Peter Hillemanns[20], Antoinette Hollestelle[87], Robert N. Hoover[31], John L. Hopper[74], Guanmengqian Huang[85], Keith Humphreys[53], David J. Hunter[6,158], Anna Jakubowska[88], Wolfgang Janni[89], Esther M. John[90-92], Nichola Johnson[68], Kristine Jones[86], Michael E. Jones[93], Audrey Jung[14], Rudolf Kaaks[14], Michael J. Kerin[94], Elza Khusnutdinova[19,95], Veli-Matti Kosma[96-98], Vessela N. Kristensen[99-101], Diether Lambrechts[102,103], Loic Le Marchand[104], Jingmei Li[157], Sara Lindström[5,105], Jolanta Lissowska[106], Wing-Yee Lo[27,28], Sibylle Loibl[107], Jan Lubinski[88], Craig Luccarini[15], Michael P. Lux[16], Robert J. MacInnis[73,74], Tom Maishman[61,108], Ivana Maleva Kostovska[20,109], Arto Mannermaa[96-98], JoAnn E. Manson[6,110], Sara Margolin[111], Dimitrios Mavroudis[112], Hanne Meijers-Heijboer[152], Alfons Meindl[113], Usha Menon[114], Jeffery Meyer[48], Anna Marie Mulligan[115,116], Susan L. Neuhausen[117], Heli Nevanlinna[118], Patrick Neven[119], Sune F. Nielsen[24,25], Børge G. Nordestgaard[24-26], Olufunmilayo I. Olopade[120], Janet E. Olson[84], Håkan Olsson[32], Paolo Peterlongo[121], Julian Peto[58], Dijana Plaseska-Karanfilska[109], Ross Prentice[12], Nadege Presneau[60], Katri Pylkäs[122,123], Brigitte Rack[89], Paolo Radice[125], Nazneen Rahman[126], Gad Rennert[127], Hedy S. Rennert[127], Valerie Rhenius[15], Atocha Romero[36,128], Jane Romm[57], Anja Rudolph[14], Emmanouil Saloustros[129], Dale P. Sandler[130], Elinor J. Sawyer[131], Marjanka K. Schmidt[47,132], Rita K. Schmutzler[79-81], Andreas Schneeweiss[34,133], Rodney J. Scott[134,135], Christopher Scott[84], Sheila Seal[126], Mitul Shah[15], Martha J. Shrubsole[1], Ann Smeets[119], Melissa C. Southey[136], John J. Spinelli[137,138], Jennifer Stone[139,140], Harald Surowy[34,35], Anthony J. Swerdlow[93,141], Rulla M. Tamimi[5,6,62], William Tapper[61], Jack A. Taylor[130,142], Mary Beth Terry[143], Daniel C. Tessier[144], Abigail Thomas[84], Kathrin Thöne[67], Rob A.E.M. Tollenaar[145], Diana Torres[85,146], Thérèse Truong[78], Michael Untch[147], Celine Vachon[84], David Van Den Berg[82], Daniel Vincent[144], Quinten Waisfisz[152], Clarice R. Weinberg[148], Camilla Wendt[111], Alice S. Whittemore[91,92], Hans Wildiers[119], Walter C.

Willett[6,62,156], Robert Winqvist[122,123], Alicja Wolk[83], Lucy Xia[82], Xiaohong R. Yang[31], Argyrios Ziogas[9], Elad Ziv[149], kConFab/AOCS Investigators[150], Alison M. Dunning[15], Paul D.P. Pharoah[3,15], Jacques Simard[59], Roger L. Milne[73,74], Stacey L. Edwards[2], Peter Kraft[5,6], Douglas F. Easton[3,15], Georgia Chenevix-Trench[2]*, Wei Zheng[1]*

***Corresponding Authors:** Wei Zheng, MD, PhD, Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, 2525 West End Ave, Suite 800, Nashville, Tennessee, 37203, USA. Email: wei.zheng@vanderbilt.edu and Georgia Chenevix-Trench, PhD, Cancer Division, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston 4006, Australia. Email: Georgia.Trench@qimrberghofer.edu.au

Key words: eQTL, genetics, breast cancer, gene expression, GWAS, susceptibility

1. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA.
2. Cancer Division, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
3. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
4. Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus.
5. Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
6. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
7. Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada.
8. Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.
9. Department of Epidemiology, University of California Irvine, Irvine, CA, USA.
10. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.
11. Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada.
12. Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
13. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA.
14. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
15. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.
16. Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany.

17. Human Cancer Genetics Program, Spanish National Cancer Research Centre, Madrid, Spain.
18. Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia, Spain.
19. Institute of Biochemistry and Genetics, Ufa Scientific Center of Russian Academy of Sciences, Ufa, Russia.
20. Gynaecology Research Unit, Hannover Medical School, Hannover, Germany.
21. Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland.
22. Department of Radiation Oncology, Hannover Medical School, Hannover, Germany.
23. N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus.
24. Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.
25. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.
26. Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
27. Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany.
28. University of Tübingen, Tübingen, Germany.
29. German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.
30. Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.
31. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA.
32. Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden.
33. Department of Gynecology and Obstetrics, University of Tübingen, Tübingen, Germany.
34. Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany.
35. Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg, Germany.
36. Medical Oncology Department, CIBERONC Hospital Clínico San Carlos, Madrid, Spain.
37. Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany.
38. Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA.
39. Oncology and Genetics Unit, Instituto de Investigacion Biomedica (IBI) Orense-Pontevedra-Vigo, Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain.
40. University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.
41. Department of Epidemiology, University of Florida, Gainesville, FL, USA.
42. Westmead Institute for Medical Research, University of Sydney, Sydney, Australia.
43. Department of Oncology, Haukeland University Hospital, Bergen, Norway.
44. National Advisory Unit on Late Effects after Cancer Treatment, Oslo University Hospital Radiumhospitalet, Oslo, Norway.
45. Oslo University Hospital, Oslo, Norway.

46. Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands.
47. Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.
48. Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.
49. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK.
50. INSERM U1052, Cancer Research Center of Lyon, Lyon, France.
51. Sheffield Institute for Nucleic Acids, Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK.
52. Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK.
53. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
54. Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA.
55. Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands.
56. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.
57. Center for Inherited Disease Research (CIDR), Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
58. Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.
59. Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, QC, Canada.
60. Department of Biomedical Sciences, Faculty of Science and Technology, University of Westminster, London, UK.
61. Cancer Sciences Academic Unit, Faculty of Medicine, University of Southampton, Southampton, UK.
62. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
63. Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany.
64. David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA.
65. Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh Medical School, Edinburgh, UK.
66. Institute for Medical Biometrics and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.
67. Department of Cancer Epidemiology, Clinical Cancer Registry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.
68. The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK.
69. Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.
70. School of Public Health, Curtin University, Perth, Australia.

71. Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago De Compostela, Spain.
72. Moores Cancer Center, University of California San Diego, La Jolla, CA, USA.
73. Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Australia.
74. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia.
75. Department of Medicine, McGill University, Montréal, QC, Canada.
76. Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, Canada.
77. Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA.
78. Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France.
79. Center for Hereditary Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany.
80. Center for Integrated Oncology (CIO), University Hospital of Cologne, Cologne, Germany.
81. Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany.
82. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
83. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
84. Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.
85. Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany.
86. Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD, USA.
87. Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands.
88. Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland.
89. Department of Gynecology and Obstetrics, University Hospital Ulm, Ulm, Germany.
90. Department of Epidemiology, Cancer Prevention Institute of California, Fremont, CA, USA.
91. Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford, CA, USA.
92. Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA.
93. Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK.
94. School of Medicine, National University of Ireland, Galway, Ireland.
95. Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia.
96. Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland.

97. Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland.
98. Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland.
99. Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway.
100. Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway.
101. Department of Clinical Molecular Biology, Oslo University Hospital, University of Oslo, Oslo, Norway.
102. VIB KULeuven Center for Cancer Biology, VIB, Leuven, Belgium.
103. Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium.
104. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA.
105. Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA.
106. Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Institute - Oncology Center, Warsaw, Poland.
107. German Breast Group, GmbH, Neu Isenburg, Germany.
108. Southampton Clinical Trials Unit, Faculty of Medicine , University of Southampton, Southampton, UK.
109. Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov" , Macedonian Academy of Sciences and Arts, Skopje, Republic of Macedonia.
110. Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
111. Department of Oncology - Pathology, Karolinska Institutet, Stockholm, Sweden.
112. Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece.
113. Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany.
114. Gynaecological Cancer Research Centre, Women’s Cancer, Institute for Women's Health, University College London, London, UK.
115. Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada.
116. Laboratory Medicine Program, University Health Network, Toronto, ON, Canada.
117. Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA.
118. Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland.
119. Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium.
120. Center for Clinical Cancer Genetics and Global Health, The University of Chicago, Chicago, IL, USA.
121. IFOM, The FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy.
122. Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland.

123. Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland.
124. Department of Gynecology and Obstetrics, Ludwig-Maximilians University of Munich, Munich, Germany.
125. Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS (Istituto Di Ricovero e Cura a Carattere Scientifico) Istituto Nazionale dei Tumori (INT), Milan, Italy.
126. Section of Cancer Genetics, The Institute of Cancer Research, London, UK.
127. Clalit National Cancer Control Center, Haifa, Israel.
128. Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain.
129. Hereditary Cancer Clinic, University Hospital of Heraklion, Heraklion, Greece.
130. Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.
131. Research Oncology, Guy's Hospital, King's College London, London, UK.
132. Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands.
133. National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany.
134. Division of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, Australia.
135. Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Callaghan, Australia.
136. Department of Pathology, The University of Melbourne, Melbourne, Australia.
137. Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada.
138. School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada.
139. The Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and University of Western Australia, Perth, Australia.
140. Department of Obstetrics and Gynaecology, University of Melbourne and the Royal Women's Hospital, Melbourne, Australia.
141. Division of Breast Cancer Research, The Institute of Cancer Research, London, UK.
142. Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.
143. Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA.
144. McGill University and Génome Québec Innovation Centre, Montréal, QC, Canada.
145. Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands.
146. Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota, Colombia.
147. Department of Gynecology and Obstetrics, Helios Clinics Berlin-Buch, Berlin, Germany.
148. Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.
149. Department of Medicine, Institute for Human Genetics, UCSF Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA.
150. Peter MacCallum Cancer Center, Melbourne, Australia.
151. Department of Molecular Physiology & Biophysics, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA.

152. Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands.
153. Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA.
154. Department of Medicine, Harvard Medical School, Boston, MA.
155. Division of Genetics, Brigham and Women's Hospital, Boston, MA.
156. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA.
157. Human Genetics, Genome Institute of Singapore, Singapore, Singapore.
158. Nuffield Department of Population Health, University of Oxford, Big Data Institute, Old Road Campus, Oxford OX3 7LF, UK.
159. Department of Oncology University of Örebro, Örebro, Sweden.
160. Lang Wu and Wei Shi are joint co-first authors.

Abstract:

Breast cancer risk variants identified in genome-wide association studies explain only a small fraction of familial relative risk, and genes responsible for these associations remain largely unknown. To identify novel risk loci and likely causal genes, we performed a transcriptome-wide association study evaluating associations of genetically predicted gene expression with breast cancer risk in 122,977 cases and 105,974 controls of European ancestry. We used data from 67 subjects included in the Genotype-Tissue Expression Project to establish genetic models to predict gene expression in breast tissue and evaluated model performance using data from 86 subjects included in The Cancer Genome Atlas. Of the 8,597 genes evaluated, significant associations were identified for 48 at a Bonferroni-corrected threshold of $P < 5.82\times10^{-6}$, including 14 genes at loci not yet reported for breast cancer risk. We silenced 13 genes and showed an effect for 11 on cell proliferation and/or colony forming efficiency. Our study provides new insights into breast cancer genetics and biology.

Breast cancer is the most commonly diagnosed malignancy among women in many countries[1]. Genetic factors play an important role in breast cancer etiology. Multiple high- and moderate-penetrance genes, including *BRCA1*, *BRCA2*, *PALB2*, *CHEK2* and *ATM*, have been identified as contributors to familial breast cancer[2,3]. However, deleterious germline mutations in these genes are rare, thus accounting for only a small fraction of breast cancer cases in the general population[4,5]. Since 2007, genome-wide association studies (GWAS) have identified approximately 180 genetic loci harboring common, low-penetrance variants for breast cancer[6-13], but these more common variants explain less than 20% of familial relative risk[7].

A large proportion of disease-associated risk variants identified by GWAS are located in non-protein coding or intergenic regions and are not in linkage disequilibrium (LD) with any nonsynonymous coding single nucleotide polymorphisms (SNPs)[14]. Many of these susceptibility variants are located in gene regulatory elements[15,16], and it has therefore been hypothesized that most of the GWAS-identified associations may be driven by the regulatory function of risk variants on the expression levels of nearby genes. For breast cancer, recent studies have shown that GWAS-identified associations at 1p34, 1p36, 2q35, 5p12, 5p15.33, 5q11.2, 5q14, 6q25, 7q22, 9q31.2, 10q21.3, 10q26.13, 11p15, 11q13.3, 15q26.1, 19p13 and 19q13.31 are likely due to the effect of risk variants at these loci on regulating the expression of either nearby or more distal genes: *CITED4*, *KLHDC7A*, *IGFBP5*, *FGF10*/*MRPS30*, *TERT*, *MAP3K1*, *ATP6AP1L*, *RMND1*, *RASA4*/*PRKRIP1*, *KLF4*, *NRBF2*, *FGFR2*, *PIDD1*, *CCND1*, *RCCD1*, *ABHD8*, and *ZNF404*[7,9,10,13,17-22]. However, for the large majority of the GWAS-identified breast cancer risk loci, the genes responsible for the associations remain unknown.

Several recent studies have reported that regulatory variants may account for a large proportion of disease heritability not yet discovered through GWAS[23-25]. Many of these variants may have a small effect size, and thus are difficult to identify in individual SNP-based GWAS studies, even with a very large sample size. Applying gene-based approaches that aggregate the effects of multiple variants into a single testing unit may increase study power to identify novel disease-associated loci. Transcriptome-wide association studies (TWAS) systematically investigate across the transcriptome the association of genetically predicted gene expression with disease risk, providing an effective approach to identify novel susceptibility genes[26-29]. Instead of testing millions of SNPs in GWAS, TWAS evaluate the association of predicted expression for selected genes, thus greatly reducing the burden of multiple comparisons in statistical inference. Recently, Hoffman et al performed a TWAS including 15,440 cases and 31,159 controls and reported significant associations for five genes with breast cancer risk[30]. However, the sample size of that study was relatively small and several reported associations were not statistically significant after Bonferroni correction. Herein, we report results from a larger TWAS of breast cancer that used the MetaXcan method[26] to analyze summary statistics data from 122,977 cases and 105,974 controls of European descent from the Breast Cancer Association Consortium (BCAC).

Results

Gene expression prediction models

The overall study design is shown in **Supplementary Figure 1**. We used transcriptome and high-density genotyping data from 67 women of European descent included in the Genotype-Tissue Expression (GTEx) project to build genetic models to predict RNA expression levels for

each of the genes expressed in normal breast tissues, by applying the elastic net method (α=0.5) with ten-fold cross-validation. Genetically regulated expression was estimated for each gene using variants within a 2 MB window flanking the respective gene boundaries, inclusive. SNPs with a minor allele frequency of at least 0.05 and included in the HapMap Phase 2 subset were used for model building. Of the models built for 12,696 genes, 9,109 showed a prediction performance (R^2) of at least 0.01 ($\geq$10% correlation between predicted and observed expression). For genes for which the expression could not be predicted well using this approach, we built models using only SNPs located in the promoter or enhancer regions, as predicted using three breast cell lines in the Roadmap Epigenomics Project/Encyclopedia of DNA Elements Project. This approach leverages information from functional genomics and reduces the number of variants for variable selection, and therefore potentially improving statistical power. This enabled us to build genetic models for additional 3,715 genes with $R^2 \geq 0.01$. **Supplementary Table 1** provides detailed information regarding the performance threshold and types of models built in this study. Overall, genes that were predicted with $R^2 \geq 0.01$ in GTEx data were also predicted well in The Cancer Genome Atlas (TCGA) tumor-adjacent normal tissue data (correlation coefficient of 0.55 for R^2 in two datasets; **Supplementary Figure 2**). Based on model performance in GTEx and TCGA, we prioritized 8,597 genes for analyses of the associations between predicted gene expression and breast cancer risk using the following criteria: 1) genes with a model prediction R^2 of at least 0.01 in the GTEx set (10% correlation) and a Spearman's correlation coefficient of $\geq$0.1 in the external validation experiment using TCGA data, 2) genes with a prediction R^2 of at least 0.09 (30% correlation) in the GTEx set regardless of their performance in the TCGA set, 3) genes with a prediction R^2 of at least 0.01 in

the GTEx set (10% correlation) that could not be evaluated in the TCGA set because of a lack of data.

Association analyses of predicted gene expression with breast cancer risk

Using the MetaXcan method[26], we performed association analyses to evaluate predicted gene expression and breast cancer risk using the meta-analysis summary statistics of individual genetic variants generated for 122,977 breast cancer cases and 105,974 controls of European ancestry included in BCAC. For the majority of the tested genes, most of the SNPs selected for prediction models were used for the association analyses (e.g., ≥95% predicting SNPs used for 83.8% of the tested genes, and ≥80% predicting SNPs used for 95.6% of the tested genes). Lambda 1,000 ($\lambda_{1,000}$), a standardized estimate of the genomic inflation scaling to a study of 1,000 cases and 1,000 controls, was 1.004 in our study (Quantile-quantile (QQ) plot presented in **Supplementary Figure 3 (A)**). Of the 8,597 genes evaluated in this study, we identified 179 genes whose predicted expression was associated with breast cancer risk at $P<1.05\times10^{-3}$, a FDR-corrected significance level (**Figure 1**, **Supplementary Table 2**). Of these, 48 showed a significant association at the Bonferroni-corrected threshold of $P\leq5.82\times10^{-6}$ (**Figure 1**, **Tables 1-3**), including 14 genes located at 11 loci that are 500 kb away from any of the risk variants identified in previous GWAS of breast cancer risk (**Table 1**). An association between lower predicted expression and increased breast cancer risk was detected for *LRRC3B* (3p24.1), *SPATA18* (4q12), *UBD* (6p22.1), *MIR31HG* (9p21.3), *RIC8A* (11p15.5), *B3GNT1* (11q13.2), *GALNT16* (14q24.1) and *MAN2C1* and *CTD-2323K18.1* (15q24.2). Conversely, an association between higher predicted expression and increased breast cancer risk was identified for *ZSWIM5* (1p34.1), *KLHDC10* (7q32.2), *RP11-867G23.10* (11q13.2), *RP11-218M22.1* (12p13.33) and

454 *PLEKHD1* (14q24.1). The remaining 34 significantly associated genes are all located at breast
455 cancer susceptibility loci identified in previous GWAS (**Tables 2-3**). Among them, 23 have not
456 yet been previously implicated as genes responsible for association signals with breast cancer
457 risk identified at these loci through expression quantitative trait loci (eQTL) and/or functional
458 studies, and do not harbor GWAS or fine-mapping identified risk variants (**Table 2**), while the
459 other eleven (*KLHDC7A*[7], *ALS2CR12*[31], *CASP8*[31,32], *ATG10*[9], *SNX32*[33], *STXBP4*[34,35] , *ZNF404*[8],
460 *ATP6AP1L*[9], *RMND1*[17], *L3MBTL3*[6], and *RCCD1*[10]) had been reported as potential causal genes
461 at breast cancer susceptibility loci or harbor GWAS or fine-mapping identified risk variants
462 (**Table 3**). Except for *RP11-73O6.3* and *L3MBTL3*, there was no evidence of heterogeneity in
463 the gene-expression association (I^2<0.2) across the iCOGS, OncoArray, and GWAS datasets
464 included in our analyses (**Supplementary Table 3**). Overall, through our agnostic search, we
465 identified 37 novel susceptibility genes for breast cancer, including 21 protein-coding genes, 15
466 long non-coding RNAs (lncRNAs) and a processed transcript, and confirmed eleven genes
467 known to potentially play a role in breast cancer susceptibility.
468
469 To determine whether the associations between predicted gene expression and breast cancer risk
470 were independent of the association signals identified in previous GWAS, we performed
471 conditional analyses adjusting for the GWAS-identified risk SNPs closest to the TWAS-
472 identified gene (**Supplementary Table 4**)[36]. We found that the associations for 11 genes
473 (*LRRC3B*, *SPATA18*, *KLHDC10*, *MIR31HG*, *RIC8A*, *B3GNT1*, *RP11-218M22.1*, *MAN2C1*,
474 *CTD-2323K18.1* **(Table 1)**, *ALK*, *CTD-3051D23.1* **(Table 2)**) remained statistically significant
475 at $P<5.82\times10^{-6}$ (**Tables 1-3**). This suggests the expression of these genes may be associated with
476 breast cancer risk independent of the GWAS-identified risk variant(s). For nine of the genes

(*SPATA18*, *KLHDC10*, *MIR31HG*, *RIC8A*, *RP11-218M22.1*, *MAN2C1*, *CTD-2323K18.1* **(Table 1)**, *ALK*, and *CTD-3051D23.1* **(Table 2)**), the significance level of the association remained essentially unchanged, suggesting these associations may be entirely independent of GWAS-identified association signals.

Of the 131 genes showing a significant association at *P* values between 5.82×10^{-6} and 1.05×10^{-3} (significant after FDR-correction but not Bonferroni-correction), 38 are located at GWAS-identified breast cancer risk loci (± 500 kb of the index SNPs) (**Table 4**). Except for *RP11-400F19.8*, there was no evidence of heterogeneity in TWAS association ($I^2<0.2$) across the iCOGS, OncoArray, and GWAS studies (**Supplementary Table 3**). After adjusting for the index SNPs, breast cancer associations for *MTHFD1L*, *PVT1*, *RP11-123K19.1*, *FES*, *RP11-400F19.8*, *CTD-2538G9.5*, and *CTD-3216D2.5* remained significant at $p \leq 1.05\times10^{-3}$, again suggesting that the association of these genes with breast cancer risk may be independent of the GWAS-identified association signals (**Table 4**).

For 41 of the 48 associated genes that reached the Bonferroni-corrected significant level, we obtained individual-level data from subjects included in the iCOGS (n=84,740) and OncoArray (n=112,133) datasets, which was 86% of the subjects included in the analysis using summary statistics (**Supplementary Table 5**). The results from the analysis using individual-level data were very similar to those described above using MetaXcan analyses (Pearson correlation of z-scores was 0.991 for iCOGS data and 0.994 for OncoArray data), although not all associations reached the Bonferroni-corrected significant level, possibly due to a smaller sample size (**Supplementary Table 5**). Conditional analyses using individual level data also revealed

consistent results compared with analyses using summary data. We found that for several genes within the same genomic region, their predicted expression levels were correlated with each other (**Tables 1-3**). The associations between predicted expression of *PLEKHD1* and *ZSWIM5* and breast cancer risk were largely influenced by their corresponding closest risk variants identified in GWAS, although these risk variants are >500 kb away from these genes (**Table 1**). There were significant correlation of rs999737 and rs1707302 with genetically predicted expression of *PLEKHD1* (r = -0.47 in the OncoArray dataset and -0.48 in the iCOGS dataset) and *ZSWIM5* (r = 0.50 in the OncoArray dataset and 0.51 in the iCOGS dataset), respectively.

INQUISIT algorithm scores for the identified genes

For the 48 associated genes after Bonferroni correction, we assessed their integrated expression quantitative trait and *in silico* prediction of GWAS target (INQUISIT) scores[7] to assess whether there are other lines of evidence beyond the scope of eQTL for supporting our TWAS-identified genes as candidate target genes at GWAS-identified loci. The detailed methodology for INQUISIT scores have been described elsewhere[7]. In brief, a score for each gene-SNP pair is calculated across categories representing potential regulatory mechanisms - distal or proximal gene regulation (promoter). Features contributing to the score are based on functionally important genomic annotations such as chromatin interactions, transcription factor binding, and eQTLs. Compared with evidence from eQTL only, INQUISIT scores incorporate additional lines of evidence, including distal regulations. The INQUISIT scores for our identified genes are shown in **Supplementary Table 6**. Except for *UBD* with a very low score in the distal regulation category (0.05), none of the genes at novel loci (**Table 1**) showed evidence to be potential target genes for any of the GWAS-identified breast cancer susceptibility loci. This is interesting and

within the expectation since these genes may represent novel association signals. There was evidence suggesting that *RP11-439A17.7*, *NUDT17*, *ANKRD34A*, *BTN3A2*, *AP006621.6*, *RPLP2*, *LRRC37A2*, *LRRC37A*, *KANSL1-AS1*, *CRHR1* and *HAPLN4* listed in Table 2, and all eleven genes listed in Table 3, may be target genes for risk variants identified in GWAS at these loci (**Supplementary Table 6**). For *NUDT17*, *ANKRD34A*, *RPLP2*, *LRRC37A2*, *LRRC37A*, *KANSL1-AS1*, *CRHR1*, *HAPLN4*, *KLHDC7A*, *ALS2CR12*, *CASP8*, *ATG10*, *ATP6AP1L*, *L3MBTL3*, *RMND1*, *SNX32*, *RCCD1*, *STXBP4* and *ZNF404,* the INQUISIT scores were not derived only from eQTL data, providing orthogonal support for these loci. For these loci, the associations of candidate causal SNPs with breast cancer risk may be mediated through these genes. This is in general consistent with the findings from the conditional analyses described above.

Pathway enrichment analyses

Ingenuity Pathway Analysis (IPA)[37] suggested potential enrichment of cancer-related functions for the significantly associated protein-coding genes identified in this study (**Supplementary Table 7**). The top canonical pathways identified in these analyses included apoptosis related pathways (Granzyme B signaling (p=0.024) and cytotoxic T lymphocyte-mediated apoptosis of target cells (p=0.046)), immune system pathway (inflammasome pathway (p=0.030)), and tumoricidal function of hepatic natural killer cells (p=0.036). The identified pathways are largely consistent with findings in previous studies[7]. For the significantly associated lncRNAs identified in this study, pathway analysis of their highly co-expressed protein-coding genes also revealed potential over-representation of cancer related functions (**Supplementary Table 7**).

Knockdown of predicted risk-associated genes in breast cells

To assess the function of genes whose high levels of predicted expression were associated with increased breast cancer risk, we selected 13 genes for knockdown experiments in breast cells: *ZSWIM5*, *KLHDC10*, *RP11-218M22.1* and *PLEKHD1* (**Table 1**), *UBLCP1*, *AP006621.6*, *RP11-467J12.4*, *CTD-3032H12.1* and *RP11-15A1.7* (**Table 2**), and *ALS2CR12*, *RMND1*, *STXBP4* and *ZNF404* (**Table 3**). As negative controls, we selected *B2M*, *ARHGDIA* and *ZAP70* using the following criteria: 1) at least 2 MB from any known breast cancer risk locus; 2) not an essential gene in breast cancer[38,39]; and 3) not predicted to be a target gene in INQUISIT. In addition, as positive controls, we included in the experiments *PIDD1* (**Table 4**)[7], *NRBF2*[20] and *ABHD8*[22], which have been functionally validated as the target genes at breast cancer risk loci. We performed quantitative PCR (qPCR) on a panel of three 'normal' mammary epithelial and 15 breast cancer cell lines to analyze their expression level (**Supplementary Figure 4 and Supplementary Table 8**). All 19 genes were expressed in the normal mammary epithelial line 184A1[40] and the luminal breast cancer cell lines, MCF7 and T47D, so we used these cell lines for the proliferation assay, and MCF7 for the colony formation assay[41]. We also evaluated *SNX32*, *ALK* and *BTN3A2* by qPCR, but they were not expressed in T47D and MCF7 cells; therefore they were not evaluated further. It was difficult to design siRNAs against *RP11-867G23.1* and *RP11-53O19.1* because they both have multiple transcripts with limited, GC-rich regions in common. We did not include *RPLP2* because it is already known to be an essential gene for breast cancer survival[42]. Knockdown of the 19 tested genes was achieved by small short interfering RNA (siRNA) (**Supplementary Table 9**) and the knockdown efficiency was calculated in 184A1, MCF7 and T47D for each siRNA pair. Robust knockdown of the gene of

interests (GOI) was validated by qPCR with the majority of the siRNAs (**Supplementary Figure 5**).

To evaluate the survival and proliferation ability of cells following gene interruption, we used an IncuCyte to quantify cell proliferation in real time and quantified the corrected proliferation of cells with knocking down of GOI in comparison to that of cells with non-target control (NTC) siRNA). As expected, knockdown of the three negative control genes (*B2M*, *ARHGDIA* and *ZAP70*) did not significantly change cell proliferation in any of the three cell lines (**Figure 2A, Supplementary Figure 6**). However, with the exception of *UBLCP1, RMND1* and *STXBP4,* knockdown of all other genes (11 TWAS-identified genes along with two known genes, *ABHD8* and *NRBF2*) resulted in significantly decreased cell proliferation in 184A1 normal breast cells, with *KLHDC10, PLEKHD1, RP11-218M22.1, AP006621.6, ZNF404, RP11-467J12.4, CTD-3032H12.1* and *STXBP4* showing a similar effect in one or both cancer cell lines. Down-regulation of three lncRNAs (*RP11-218M22.1*, *RP11-467J12.4* and *CTD-3032H12.1)* resulted in significant reduction in cell proliferation in all three cell lines. We also evaluated the effect of inhibition of these genes on colony forming ability in MCF7 cells. Knockdown of the three negative control genes did not significantly affect colony forming efficiency (CFE). By contrast, knockdown of *PIDD1, RP11-15A1.7, RP11-218M22.1, AP006621.6, ZNF404, RP11-467J12.4* and *CTD-3032H12.1* resulted in significantly decreased colony forming efficiency in MCF7 cells compared to the NTC (**Figure 2B, Supplementary Figure 7).**

Discussion

This is the largest study to systematically evaluate associations of genetically predicted gene

591 expression across the human transcriptome with breast cancer risk. We identified 179 genes
592 showing a significant association at the FDR-corrected significance level. Of these, 48 showed a
593 significant association at the Bonferroni-corrected threshold, including 14 genes at genomic loci
594 that have not previously been implicated for breast cancer risk. Of the 34 genes we identified that
595 are located at known risk loci, 23 have not previously been shown to be the targets of GWAS-
596 identified risk SNPs at corresponding loci and not harbor any risk SNPs. Our study provides
597 substantial new information to improve the understanding of genetics and etiology for breast
598 cancer, the most common malignancy among women in most countries.
599
600 It is possible that TWAS-identified genes may be associated with breast cancer risk through their
601 correlation with disease causal genes. To determine the potential functional significance of
602 TWAS-identified genes and provide evidence for causal inference, we knocked down 13 genes
603 for which high predicted levels of expression were associated with an increased breast cancer
604 risk, in one normal and two breast cancer cell lines, and measured the effect on proliferation and
605 colony forming efficiency. Although there was some variation between cell lines, knockdown of
606 11 of the 13 genes showed an effect in at least one cell line, particularly on proliferation in
607 184A1 normal breast cells; the effects were strongest and most consistent for the lncRNAs,
608 *RP11-218M22.1*, *RP11-467J12.4* and *CTD-3032H12.1*. The observation of a more consistent
609 effect in the normal breast cell line compared with the cancer cell lines is not surprising as cancer
610 cell lines have increased capacity to handle gene interference through mutations which enhance
611 cell survival. Rewiring of pathways and compensatory mechanisms is a hallmark of cancer.
612 Knockdown of *PIDD1, NRBF2* and *ABHD8*¸ for which breast cancer risk associated haplotypes
613 have been shown to be associated with increased expression in reporter assays[7,20,22], affected

either proliferation or colony forming efficiency, supporting the results from this study. Knockdown of *UBLCP1* and *RMND1* did not affect proliferation or colony formation but they could mediate breast cancer risk through other mechanisms.

Some of the genes with strong functional evidence from our study have been reported to have important roles in carcinogenesis. For example, *RP11-467J12.4* (PR-lncRNA-1) is a p53-regulated lncRNA that modulates gene expression in response to DNA damage downstream of p53[43]. *STXBP4* encodes Syntaxin binding protein 4, a scaffold protein that can stabilise and prevent degradation of an isoform of p63, a member of the p53 tumor suppressor family[44]. *KLHDC10* encodes a member of the Kelch superfamily that can activate apoptosis signal-regulating kinase 1, contributing to oxidative stress-induced cell death[45]. Notably, another member of this superfamily, *KLHDC7A*, has recently been identified as the target gene at the 1p36 breast cancer risk locus[7].

SNX32, ALK and *BTN3A2* are also likely susceptibility genes for breast cancer risk. However, their low or absent expression in our chosen breast cell lines prevented further functional analysis. *SNX32* (Sorting Nexin 32) is not well characterized, but *ALK* (Anaplastic lymphoma kinase) copy number gain and overexpression have been reported in aggressive and metastatic breast cancers[46]. Therapeutic targeting of ALK rearrangement has significantly improved survival in advanced ALK-positive lung cancer[47], making it an attractive target for breast and other cancers. *BTN3A2* is a member of the B7/butyrophilin-like group of Ig superfamily receptors modulating the function of T-lymphocytes. While the exact role of *BTN3A2* remains

636 unknown, over-expression of this gene in epithelial ovarian cancer is associated with higher
637 infiltrating immune cells and a better prognosis[48].
638
639 Our analyses identified multiple genes with reduced expression levels associated with increased
640 breast cancer risk. Among them, *LRRC3B* and *CASP8* are putative tumor suppressors in multiple
641 cancers, including breast cancer. Leucine-rich repeat-containing 3B (*LRRC3B)* is a putative
642 LRR-containing transmembrane protein, which is frequently inactivated via promoter
643 hypermethylation leading to inhibition of cancer cell growth, proliferation, and invasion[49].
644 *CASP8* encodes a member of the cysteine-aspartic acid protease family, which play a central role
645 in cell apoptosis. Previous studies have suggested that caspase-8 may act as a tumor suppressor
646 in certain types of lung cancer and neuroblastoma, although this function has not yet been
647 demonstrated in breast cancer. Notably, several large association studies have identified SNPs at
648 the 2q33/*CASP8* locus associated with increased breast cancer risk[31,50]. Consistent with our data,
649 eQTL analyses showed that the risk alleles for breast cancer were associated with reduced
650 *CASP8* mRNA levels in both peripheral blood lymphocytes and normal breast tissue[31].
651
652 For seven of the genes listed in Tables 1 and 2, we found some evidence from studies using
653 tumor tissues, *in vitro* or *in vivo* experiments linking them to cancer risk (**Supplementary Table**
654 **10)**, although their association with breast cancer has not been previously demonstrated in human
655 studies. For five of them, including *LRRC3B*, *SPATA18*, *RIC8A*, *ALK* and *CRHR1*, previous *in*
656 *vitro* and *in vivo* experiments and human tissue studies showed a consistent direction of the
657 association as demonstrated in our studies. For two other genes (*UBD* and *MIR31HG*), however,
658 results from previous studies were inconsistent, reporting both potential promoting and inhibiting

effects on breast cancer development. Future studies are needed to evaluate functions of these genes.

We included a large number of cases and controls in this study, providing strong statistical power for the association analysis. This large sample size enabled us to identify a large number of candidate breast cancer susceptibility genes, much larger than the number identified in a TWAS study with a sample size of about 20% of ours[30]. The previous study included subjects of different races, which could affect the results as linkage disequilibrium (LD) patterns differ by races. Of the five genes reported in that smaller TWAS that showed a suggestive association with breast cancer risk, the association for the *RCCD1* gene was replicated in our study **(Table 3)**. The other four genes (*ANKLE1*, *DHODH*, *ACAP1* and *LRRC25)* were not evaluated in our study because of unsatisfactory performance of our breast specific models for these genes which were built using the GTEx reference dataset including only female European descendants. In our study, the expression prediction model for *ANKLE1* has a marginal performance in predicting gene expression (R^2=0.013 in the GTEx). The model, however, did not perform well in the TCGA data. For *ACAP1* and *LRRC25,* previous results for suggestive associations were based on blood tissue models.

A substantial proportion of SNPs included in the OncoArray and iCOGS were selected from breast cancer GWAS and fine-mapping analyses, and thus these arrays were enriched for association signals with breast cancer risk. As a result, the overall λ value for the BCAC association analyses of individual variants is 1.26 after adjusting for population stratifications (QQ plot in **Supplementary Figure 3 (B)**)[7]. The λ value for the associations of the ~257,000

SNPs included in the gene expression prediction models of the 8,597 genes tested in our association analysis is 1.40 (QQ plot in **Supplementary Figure 3 (C)**). This higher λ value is perhaps expected because of a potential further enrichment of breast cancer associated signals in the set of SNPs selected to predict gene expression. There could be additional gain of power (and thus a higher λ value) in TWAS as it aggregates the effect of multiple SNPs to predict gene expression and use genes as the unit for association analyses. The lambda (λ) for our associated analyses of 8,597 genes was 1.51 (QQ plot presented in **Supplementary Figure 3 (A)**) likely due to the potential enrichment and power gain discussed above as well as our large sample size, and the highly polygenic nature of the disease[7,51]. Interestingly, high λ values were also found in recent large studies of other polygenic traits, such as body mass index (BMI) ($\lambda = 1.99$) and height ($\lambda = 2.7$)[52,53]. The $\lambda_{1,000}$, a standardized estimate of the genomic inflation scaling to a study of 1,000 cases and 1,000 controls, is 1.004 in our study.

The statistical power of our study is very large to detect associations for genes with a relatively high cis-heritability (h^2) (**Supplementary Figure 8**). For example, our study has 80% statistical power to detect an association with breast cancer risk at $P<5.82\times10^{-6}$ with an OR of 1.07 or higher per one standard deviation increase (or decrease) in the expression level of genes with an h^2 of 0.1 or higher. One limitation of our study is the small sample size for building gene expression prediction models, which may have affected the precision of model parameter estimates. The prediction performance (R^2) for several of the genes identified in our study was not optimal, and thus additional research is needed to confirm our findings. We expect that models built with a larger sample size (and thus with more stable estimates of model parameters) will identify additional association signals. We used samples from women of European origin in

model building, given differences in gene expression patterns between males and females and in genetic architecture across ethnicities[54]. We also used gene expression data of tumor-adjacent normal tissue samples from European descendants in TCGA as an external validation step to prioritize genes for association analyses. Given potential somatic alterations in tumor-adjacent normal tissues, we retained all models showing a prediction performance (R^2) of at least 0.09 in GTEx, regardless of their performance in TCGA. Not all genes have a significant hereditary component in expression regulation, and thus these genes could not be investigated in our study. For example, previous studies have provided strong evidence to support a significant role of the *TERT*, *ESR1*, *CCND1*, *IGFBP5*, *TET2* and *MRPS30* genes in the etiology of breast cancer. However, expression of these genes cannot be predicted well using the data from female European descendants included in the GTEx and thus they were not included in our association analyses. **Supplementary Table 11** summarizes the performance of prediction models and association results for breast cancer target genes reported previously at GWAS-identified loci.

In summary, our study has identified multiple gene candidates that can be further functionally characterized. By evaluating the associations of predicted gene expression levels with breast cancer risk, we provided evidence for the direction of the association for the identified genes. The silencing experiments we performed suggest that many of the genes identified by TWAS are likely to mediate risk of breast cancer by affecting proliferation or colony forming efficiency, two of the hallmarks of cancer. Further investigation of genes identified in our study will provide additional insight into the biology and genetics of breast cancer.

Methods

Building of gene expression prediction models

We used transcriptome and high-density genotyping data from the Genotype-Tissue Expression (GTEx) study to establish prediction models for genes expressed in normal breast tissues. Details of the GTEx have been described elsewhere[55]. Genomic DNA samples obtained from study subjects included in the GTEx were genotyped using Illumina OMNI 5M or 2.5M SNP Array and RNA samples from 51 tissue sites were sequenced to generate transcriptome profiling data. Genotype data were processed according to the GTEx protocol (http://www.gtexportal.org/home/documentationPage). SNPs with a call rate < 98%, with differential missingness between the two array experiments (5M/2.5M Arrays), with Hardy-Weinberg equilibrium p-value < 10^{-6} (among subjects of European ancestry), or showing batch effects were excluded. One Klinefelter individual, three related individuals, and a chromosome 17 trisomy individual were also excluded. The genotype data were imputed to the Haplotype Reference Consortium reference panel[56] using Minimac3 for imputation and SHAPEIT for prephasing[57,58]. SNPs with high imputation quality ($r^2 \geq 0.8$), minor allele frequency (MAF) ≥ 0.05, and included in the HapMap Phase 2 version, were used to build expression prediction models. For gene expression data, we used Reads Per Kilobase per Million (RPKM) units from RNA-SeQC[59]. Genes with a median expression level of 0 RPKM across samples were removed, and the RPKM values of each gene were log2 transformed. We performed quantile normalization to bring the expression profile of each sample to the same scale, and performed inverse quantile normalization for each gene to map each set of expression values to a standard normal. We adjusted for the top ten principal components (PCs) derived from genotype data and the top 15 probabilistic estimation of expression residuals (PEER) factors to correct for batch effects and experimental confounders in model building[60]. Genetic and transcriptome data from 67 female

751 subjects of European descent without a prior breast cancer diagnosis were used to build gene
752 expression prediction models for this study.
753
754 We built an expression prediction model for each gene by using the elastic net method as
755 implemented in the glmnet R package, with α=0.5, as recommended by Gamazon et al[27]. The
756 genetically regulated expression for each gene was estimated by including variants within a 2
757 MB window flanking the respective gene boundaries, inclusive. Expression prediction models
758 were built for protein coding genes, long non-coding RNAs (lncRNAs), microRNAs (miRNAs),
759 processed transcripts, immunoglobulin genes, and T cell receptor genes, according to categories
760 described in the Gencode V19 annotation file (http://www.gencodegenes.org/releases/19.html).
761 Pseudogenes were not included in the present study because of potential concerns of inaccurate
762 calling[61]. Ten-fold cross-validation was used to validate the models internally. Prediction R^2
763 values (the square of the correlation between predicted and observed expression) were generated
764 to estimate the prediction performance of each of the gene prediction models established.
765
766 For genes that cannot be predicted well using the above approach, we built models using only
767 SNPs located in predicted promoter or enhancer regions in breast cell lines. This approach
768 reduces the number of variants for model building, and thus potentially improves model
769 accuracy, by increasing the ratio of sample size to effective degrees of freedom.
770 SNP-level annotation data in three breast cell lines, namely, Breast Myoepithelial Primary Cells
771 (E027), Breast variant Human Mammary Epithelial Cells (vHMEC) (E028), and HMEC
772 Mammary Epithelial Primary Cells (E119) in the Roadmap Epigenomics Project/Encyclopedia
773 of DNA Elements Project[16], were downloaded from

774 http://archive.broadinstitute.org/mammals/haploreg/data/ (Version 4.0, assessed on December 6,
775 2016). SNPs in regions classified as promoters (TssA, TssAFlnk), enhancers (Enh, EnhG), or
776 regions with both promoter and enhancer signatures (ExFlnk) according to the core 15 chromatin
777 state model[16] in at least one of the cell lines were retained as input SNPs for model building.
778
779 **Evaluating performance of gene expression prediction models using The Cancer Genome**
780 **Atlas (TCGA) data**
781 To assess further the validity of the models, we performed external validation using data
782 generated in tumor-adjacent normal breast tissue samples obtained from 86 European-ancestry
783 female breast cancer patients included in the TCGA. Genotype data were imputed using the same
784 approach as described for GTEx data. Expression data were processed and normalized using a
785 similar approach as described above. The predicted expression level for each gene was calculated
786 using the model established using GTEx data and then compared with the observed level of that
787 gene using the Spearman's correlation.
788
789 **Evaluating statistical power for association tests**
790 We conducted a simulation analysis to assess the power of our TWAS analysis. Specifically, we
791 set the number of cases and controls to be 122,977 and 105,974, respectively, and generated the
792 gene expression levels from the empirical distribution of predicted gene expression levels in the
793 BCAC. We calculated statistical power at $P<5.82\times10^{-6}$ (the significance level used in our
794 TWAS) according to cis-heritability (h^2) which we aim to capture using gene expression
795 prediction models (R^2). The results based on 1000 replicates are summarized in **Supplementary**
796 **Figure 8**. Based on the power calculation, our TWAS analysis has 80% power to detect a

797 minimum odds ratio of 1.11, 1.07, 1.05, 1.04, or 1.03 for breast cancer risk per one standard
798 deviation increase (or decrease) in the expression level of a gene whose cis-heritability is 5%,
799 10%, 20%, 40%, or 60%, respectively.
800
801 **Association analyses of predicted gene expression with breast cancer risk**
802 We used the following criteria to select genes for the association analysis: 1) with a model
803 prediction R^2 of ≥ 0.01 in GTEx and a Spearman's correlation coefficient of ≥ 0.1 in TCGA, 2)
804 with a prediction R^2 of ≥ 0.09 in GTEx regardless of the performance in TCGA, 3) with a
805 prediction R^2 of ≥ 0.01 in GTEx but unable to be evaluated in TCGA. The second group of genes
806 was selected because some gene expression levels might have changed in TCGA tumor-adjacent
807 normal tissues, and thus it is anticipated that some genes may show low prediction performance
808 in TCGA data due to the influence of tumor growth[62,63]. Overall, a total of 8,597 genes met the
809 criteria and were evaluated for their expression-trait associations.
810
811 To identify novel breast cancer susceptibility loci and genes, the MetaXcan method, as described
812 elsewhere, was used for the association analyses[26]. Briefly, the formula:

813 $$Z_g \approx \sum_{l \in \mathrm{Model}_g} w_{lg} \frac{\hat{\sigma}_l}{\hat{\sigma}_g} \frac{\hat{\beta}_l}{\mathrm{se}(\hat{\beta}_l)}$$

814 was used to estimate the Z-score of the association between predicted expression and breast
815 cancer risk. Here w_{lg} is the weight of SNP l for predicting the expression of gene g, $\hat{\beta}_l$and
816 $\mathrm{se}(\hat{\beta}_l)$ are the GWAS association regression coefficient and its standard error for SNP l, and $\hat{\sigma}_l$
817 and $\hat{\sigma}_g$ are the estimated variances of SNP l and the predicted expression of gene g respectively.
818 Therefore, the weights for predicting gene expression, GWAS summary statistics results, and

819 correlations between model predicting SNPs are the input variables for the MetaXcan analyses.
820 For this study we estimated correlations between SNPs included in the prediction models using
821 the phase 3, 1000 Genomes Project data focusing on European population.
822
823 For the association analysis, we used the summary statistics data of genetic variants associated
824 with breast cancer risk generated in 122,977 breast cancer patients and 105,974 controls of
825 European ancestry from the Breast Cancer Association Consortium (BCAC). The details of the
826 BCAC have been described elsewhere[7,9,13,64,65]. Briefly, 46,785 breast cancer cases and 42,892
827 controls of European ancestry were genotyped using a custom Illumina iSelect genotyping array
828 (iCOGS) containing ~211,155 variants. A further 61,282 cases and 45,494 controls of European
829 ancestry were genotyped using the OncoArray including 570,000 SNPs
830 (http://epi.grants.cancer.gov/oncoarray/). Also included in this analysis were data from nine
831 GWAS studies including 14,910 breast cancer cases and 17,588 controls of European ancestry.
832 Genotype data from iCOGS, OncoArray and GWAS were imputed using the October 2014
833 release of the 1000 Genomes Project data as reference. Genetic association results for breast
834 cancer risk were combined using inverse variance fixed effect meta-analyses[7]. For our study,
835 only SNPs with imputation $r^2 \geq 0.3$ were used. All participating BCAC studies were approved by
836 their appropriate ethics review boards. This study was approved by the BCAC Data Access
837 Coordination Committee.
838
839 Lambda 1,000 ($\lambda_{1,000}$) was calculated to represent a standardized estimate of the genomic
840 inflation scaling to a study of 1,000 cases and 1,000 controls, using the following formula:
841 $\lambda_{1,000}=1+(\lambda_{obs}-1) \times (1/n_{cases}+1/n_{controls})/(1/1{,}000_{cases}+1/1{,}000_{controls})$[66,67]. We used a Bonferroni

842 corrected *p* threshold of 5.82×10^{-6} (0.05/8,597) to determine a statistically significant association
843 for the primary analyses. To identify additional gene candidates at previously identified
844 susceptibility loci, we also used a false discovery rate (FDR) corrected *p* threshold of 1.05×10^{-3}
845 (FDR $\leq$ 0.05) to determine a significant association. Associated genes with an expression of >0.1
846 RPKM in less than 10 individuals in GTEx data were excluded as the corresponding prediction
847 models may not be stable.
848
849 To determine whether the predicted expression-trait associations were independent of the top
850 signals identified in previous GWAS, we performed GCTA-COJO analyses developed by Yang
851 et al[36] to calculate association betas and standard errors of variants with breast cancer risk after
852 adjusting for the index SNPs of interest. We then re-ran the MetaXcan analyses using the
853 association statistics after conditioning on the index SNPs. This information was used to
854 determine whether the detected expression-trait associations remained significant after adjusting
855 for the index SNPs.
856
857 For 41 identified associated genes at the Bonferroni-corrected threshold, we also performed
858 analyses using individual level data in iCOGS (n=84,740) and OncoArray (n=112,133) datasets.
859 We generated predicted gene expression using predicting SNPs, and then assessed the
860 association between predicted gene expression and breast cancer risk adjusting for study and
861 nine principal components in iCOGS dataset, and country and the first ten principal components
862 in OncoArray dataset. Conditional analyses adjusting for index SNPs were performed to assess
863 potential influence of reported index SNPs on the association between predicted gene expression
864 and breast cancer risk. Furthermore, we evaluated whether the predicted expression levels of

genes within a same genomic region were correlated with each other by using the OncoArray data.

INQUISIT algorithm scores for TWAS-identified genes

To evaluate whether there are additional lines of evidence supporting the identified genes as putative target genes of GWAS identified risk SNPs beyond the scope of eQTL, we assessed their INQUISIT algorithm scores, which have been described elsewhere[7]. Briefly, this approach evaluates chromatin interactions between distal and proximal regulatory transcription-factor binding sites and the promoters at the risk regions using Hi-C data generated in HMECs[68] and Chromatin Interaction Analysis by Paired End Tag (ChiA-PET) in MCF7 cells. This could detect genome-wide interactions brought about by, or associated with, CCCTC-binding factor (CTCF), DNA polymerase II (POL2), and Estrogen Receptor (ER), all involved in transcriptional regulation[68]. Annotation of predicted target genes used the Integrated Method for Predicting Enhancer Targets (IM-PET)[69], the Predicting Specific Tissue Interactions of Genes and Enhancers (PreSTIGE) algorithm[70], Hnisz[71] and FANTOM[72]. Features contributing to the scores are based on functionally important genomic annotations such as chromatin interactions, transcription factor binding, and eQTLs. The detailed information for the INQUISIT pipeline and scoring strategy has been included in a previous publication[7]. In brief, besides assigning integral points according to different features, we also set up-weighting and down-weighting criteria according to breast cancer driver genes, topologically associated domain (TAD) boundaries, and gene expression levels in relevant breast cell lines. Scores in the distal regulation category range from 0-7, and in the promoter category from 0-4. A score of "none" represents that no evidence was found for regulation of the corresponding gene.

888

889 **Functional enrichment analysis using Ingenuity Pathway Analysis (IPA)**

890 We performed functional enrichment analysis for the identified protein-coding genes reaching
891 Bonferroni corrected association threshold. To assess potential functionality of the identified
892 lncRNAs, we examined their co-expressed protein-coding genes determined using expression
893 data of normal breast tissue of European females in GTEx. Spearman's correlations between
894 protein-coding genes and identified lncRNAs of ≥ 0.4 or ≤ -0.4 were used to indicate a high co-
895 expression. Canonical pathways, top associated diseases and biofunctions, and top networks
896 associated with genes of interest were estimated using IPA software[37].

897

898 **Gene expression in breast cell lines**

899 Total RNA was isolated from 18 cell lines (**Supplementary Table 8**) using the RNeasy Mini Kit
900 (Qiagen). cDNA was synthesized using the SuperScript III (Invitrogen) and amplified using the
901 Platinum SYBR Green qPCR SuperMix-UDG cocktail (Invitrogen). Two or three primer pairs
902 were used for each gene and the mRNA levels for each sample was measured in technical
903 triplicates for each primer set. The primer sequences are listed in **Supplementary Table 12**.
904 Experiments were performed using an ABI ViiA(TM) 7 System (Applied Biosystems), and data
905 processing was performed using ABI QuantStudio™ Software V1.1 (Applied Biosystems). The
906 average of Ct from all the primer pairs for each gene was used to calculate ΔC_T. The relative
907 quantitation of each mRNA normalizing to that in 184A1 was performed using the comparative
908 Ct method ($\Delta\Delta C_T$) and summarized in **Supplementary Figure 4**.

909

910 **Short interfering RNA (siRNA) silencing**

MCF7 and T47D cells were reverse-transfected with siRNAs targeting genes of interest (GOI) or a non-targeting control siRNA (consi; Shanghai Genepharma) with RNAiMAX (Invitrogen) according to the manufacturer's protocol. Verification of siRNA knockdown of gene expression by qPCR was performed 36 hours after transfection.

Proliferation and colony formation assays

For proliferation assays, MCF7 and T47D cells were trypsinized at 16 hours post-transfection and seeded into 24 well plates to achieve ~10% confluency. Phase-contrast images were collected with IncuCyte ZOOM (Essen Bioscience) for seven days. Duplicate samples were assessed for each GOI siRNA transfected cells along with non-target control si (NTCsi) treated cells in the same plate. 184A1 cells were reverse-transfected in 96 well plates to achieve 50% confluence at 8 hours after transfection. Two independent experiments were carried out for all siRNAs in all three cell lines. Each cell proliferation time-course was normalized to the baseline confluency and analyzed in GraphPad Prism. The area under the curve was calculated for each concentration (n=4) and used to calculate corrected proliferation (Corrected proliferation % = 100 +/- (relative proliferation in indicated siRNA - proliferation in NTC siRNA) / knockdown efficiency ("+" if the GOI promotes proliferation and "-" if it inhibits proliferation)). For each gene, results from two siRNAs in two independent experiments were averaged and summarized in **Figure 2** and **Supplementary Figure 6**. For colony formation assays; the same number of GOI siRNA transfected MCF7 cells was seeded in 6 well plates at 16 hours after transfection to assay colony forming efficiency at two weeks. All siRNA-treated cells were seeded in duplicate. Colonies (defined to consist of at least 50 cells) were fixed with methanol, stained with crystal violet (0.5% w/v), scanned and counted using ImageJ as batch analysis by a self-defined plug-in

Macro. Correct CFE % = 100 +/- (relative CFE in indicated siRNA - CFE in NTC siRNA) / knockdown efficiency ("+" if the GOI promotes CF and "-" if it inhibits CF). For each gene, results from two siRNAs in two independent experiments were averaged and summarized in **Figure 2** and **Supplementary Figure 7**.

Data availability

The GTEx data are publicly available via dbGaP (www.ncbi.nlm.nih.gov/gap; dbGaP Study Accession: phs000424.v6.p1). TCGA data are publicly available via National Cancer Institute's Genomic Data Commons Data Portal (https://gdc.cancer.gov/). Most of the BCAC data used in this study are or will be publicly available via dbGAP. Data from some BCAC studies are not publicly available due to restraints imposed by the ethics committees of individual studies; requests for further data can be made to the BCAC (http://bcac.ccge.medschl.cam.ac.uk/) Data Access Coordination Committee.

Code availability

The computer codes used in our study are available upon reasonable request.

Acknowledgements

The authors thank Jing He, Wanqing Wen, Ayush Giri, and Todd Edwards of Vanderbilt Epidemiology Center and Rao Tao of Department of Biostatistics, Vanderbilt University Medical Center for their help with the data analysis of this study. The authors also would like to thank all the individuals for their participation in the parent studies and all the researchers, clinicians, technicians and administrative staff for their contribution to the studies. We are also grateful to

957 Hae Kyung Im of University of Chicago for her help. The data analyses were conducted using
958 the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University.
959 This project at Vanderbilt University Medical Center was supported in part by grants
960 R01CA158473 and R01CA148677 from the U.S. National Institutes of Health as well as funds
961 from Anne Potter Wilson endowment. Dr. Wu was supported by the Vanderbilt Molecular and
962 Genetic Epidemiology of Cancer (MAGEC) training program funded by the US National Cancer
963 Institute grant R25 CA160056 (PI: X.-O. Shu). Genotyping of the OncoArray was principally
964 funded from three sources: the PERSPECTIVE project, funded from the Government of Canada
965 through Genome Canada and the Canadian Institutes of Health Research, the *Ministère de*
966 *l'Économie, de la Science et de l'Innovation du Québec* through Genome Québec, and the
967 Quebec Breast Cancer Foundation; the NCI Genetic Associations and Mechanisms in Oncology
968 (GAME-ON) initiative and Discovery, Biology and Risk of Inherited Variants in Breast Cancer
969 (DRIVE) project (NIH Grants U19 CA148065 and X01HG007492); and Cancer Research UK
970 (C1287/A10118 and. C1287/A16563). **BCAC** is funded by Cancer Research UK
971 [C1287/A16563], by the European Community's Seventh Framework Programme under grant
972 agreement 223175 (HEALTH-F2-2009-223175) (COGS) and by the European Union's Horizon
973 2020 Research and Innovation Programme under grant agreements 633784 (B-CAST) and
974 634935 (BRIDGES). Genotyping of the iCOGS array was funded by the European Union
975 (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710), the Canadian Institutes of
976 Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program, and the
977 Ministry of Economic Development, Innovation and Export Trade of Quebec – grant # PSR-
978 SIIRI-701. Combining the GWAS data was supported in part by The National Institute of Health
979 (NIH) Cancer Post-Cancer GWAS initiative grant U19 CA 148065 (DRIVE, part of the GAME-

ON initiative). A full description of funding and acknowledgments for BCAC studies are included in the Acknowledgments for BCAC studies section of the **Supplementary Material**.

Author Contributions

W.Z. and J.L. conceived the study. L.W. contributed to the study design, and performed statistical analyses. L.W., W.Z. and G.C.-T. wrote the manuscript with significant contributions from W.S., J.L., X.G., and S.L.E.. W.S. performed the *in vitro* experiments. G.C.-T. directed the *in vitro* experiments. X.G. contributed to the model building and pathway analyses. J.B. contributed to the bioinformatics analyses. F.A.-E., E.R., and S.L.E. contributed to the *in vitro* experiments. Y. L. and C. Z. contributed to the model building. K.M., M.K.B., X.-O.S., Q.W., J.D., B.L., C.Z., H.F., A.G., R.T.B., A.M.D., P.D.P.P., J.S., R.L.M., P.K., and D.F.E, contributed to manuscript revision, statistical analyses and/or BCAC data management. I.L.A., H.A.-C., V.A., K.J.A., P.L.A., M. Barrdahl, C.B., M.W.B., J.B., M. Bermisheva, C.B., N.V.B., S.E.B., H. Brauch, H. Brenner, L.B., P.B., S.Y.B., B.B., Q.C., T.C., F.C., B.D.C., J.E.C., J.C.-C., X.C., T.-Y.D.C., H.C., C.L.C., NBCS Collaborators, M.C., S.C., F.J.C., D.C., A.C., S.S.C., J.M.C., K.C., M.B.D., P.D., K.F.D., T.D., I.d.S.S., M. Dumont, M. Dwek, D.M.E., U.E., H.E., C.E., M.E., L.F., P.A.F., J.F., D.F.-J., O.F., H.F., L.F., M. Gabrielson, M.G.-D., S.M.G., M.G.-C., M.M.G., M. Ghoussaini, G.G.G., M.S.G., D.E.G., A.G.-N., P.G., E. Hahnen, C.A.H., N.H., P. Hall, E. Hallberg, U.H., P. Harrington, A. Hein, B.H., P. Hillemanns, A. Hollestelle, R.N.H., J.L.H., G.H., K.H., D.J.H., A.J., W.J., E.M.J., N.J., K.J., M.E.J., A. Jung, R.K., M.J.K., E.K., V.-M.K., V.N.K., D.L., L.L.M., J. Li, S.L., J. Lissowska, W.-Y.L., S.Loibl, J.L., C.L., M.P.L., R.J.M., T.M., I.M.K., A. Mannermaa, J.E.M., S.M., D.M., H.M.-H., A. Meindl, U.M., J.M., A.M.M., S.L.N., H.N., P.N., S.F.N., B.G.N., O.I.O., J.E.O., H.O., P.P., J.P., D.P.-K., R.P., N.P., K.P.,

B.R., P.R., N.R., G.R., H.S.R., V.R., A. Romero, J.R., A. Rudolph, E.S., D.P.S, E.J.S., M.K.S., R.K.S., A.S., R.J.S., C. Scott, S.S., M.S., M.J.S., A.S., M.C.S., J.J.S., J.S., H.S., A.J.S., R.T., W.T., J.A.T., M.B.T., D.C.T., A.T., K.T., R.A.E.M.T., D.T., T.T., M.U., C.V., D.V.D.B., D.V., Q.W., C.R.W., C.W., A.S.W., H.W., W.C.W., R.W., A.W., L.X., X.R.Y., A.Z., E.Z., kConFab/AOCS Investigators contributed to the collection of the data and biological samples for the original BCAC studies. All authors have reviewed and approved the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

1. Kamangar, F., Dores, G.M. & Anderson, W.F. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol* **24**, 2137-50 (2006).
2. Beggs, A.D. & Hodgson, S.V. Genomics and breast cancer: the different levels of inherited susceptibility. *Eur J Hum Genet* **17**, 855-6 (2009).
3. Southey, M.C. *et al.* PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. *J Med Genet* (2016).
4. Nathanson, K.L., Wooster, R. & Weber, B.L. Breast cancer genetics: what we know and what we need. *Nat Med* **7**, 552-6 (2001).
5. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. *Br J Cancer* **83**, 1301-8 (2000).
6. Milne, R.L. *et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet* **49**, 1767-1778 (2017).
7. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
8. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-61, 361e1-2 (2013).
9. Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* **47**, 373-80 (2015).
10. Cai, Q. *et al.* Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet* **46**, 886-90 (2014).
11. Zheng, W. *et al.* Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum Mol Genet* **22**, 2539-50 (2013).
12. Zhang, B., Beeghly-Fadiel, A., Long, J. & Zheng, W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol* **12**, 477-88 (2011).
13. French, J.D. *et al.* 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 (2013).
14. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
15. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
16. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
17. Dunning, A.M. *et al.* Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nat Genet* **48**, 374-86 (2016).
18. Ghoussaini, M. *et al.* Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun* **4**, 4999 (2014).

19. Li, Q. *et al.* Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* **152**, 633-41 (2013).
20. Darabi, H. *et al.* Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate NRBF2 Expression. *Am J Hum Genet* **97**, 22-34 (2015).
21. Glubb, D.M. *et al.* Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. *Am J Hum Genet* **96**, 5-20 (2015).
22. Lawrenson, K. *et al.* Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. *Nat Commun* **7**, 12675 (2016).
23. Lee, D. *et al.* A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet* **47**, 955-61 (2015).
24. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
25. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* **95**, 535-52 (2014).
26. Barbeira, A.N. *et al*. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *bioRxiv* (2017).
27. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).
28. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
29. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
30. Hoffman, J.D. *et al.* Cis-eQTL-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk. *PLoS Genet* **13**, e1006690 (2017).
31. Lin, W.Y. *et al.* Identification and characterization of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. *Hum Mol Genet* **24**, 285-98 (2015).
32. Camp, N.J. *et al.* Discordant Haplotype Sequencing Identifies Functional Variants at the 2q33 Breast Cancer Risk Locus. *Cancer Res* **76**, 1916-25 (2016).
33. Li, Q. *et al.* Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. *Hum Mol Genet* **23**, 5294-302 (2014).
34. Caswell, J.L. *et al.* Multiple breast cancer risk variants are associated with differential transcript isoform expression in tumors. *Hum Mol Genet* **24**, 7421-31 (2015).
35. Darabi, H. *et al.* Fine scale mapping of the 17q22 breast cancer locus using dense SNPs, genotyped within the Collaborative Oncological Gene-Environment Study (COGs). *Sci Rep* **6**, 32512 (2016).
36. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
37. Kramer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-30 (2014).
38. Koh, J.L. *et al.* COLT-Cancer: functional genetic screening resource for essential genes in human cancer cell lines. *Nucleic Acids Res* **40**, D957-63 (2012).
39. Marcotte, R. *et al.* Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov* **2**, 172-89 (2012).

40. Walen, K.H. & Stampfer, M.R. Chromosome analyses of human mammary epithelial cells at stages of chemical-induced transformation progression to immortality. *Cancer Genet Cytogenet* **37**, 249-61 (1989).
41. Treszezamsky, A.D. *et al.* BRCA1- and BRCA2-deficient cells are sensitive to etoposide-induced DNA double-strand breaks via topoisomerase II. *Cancer Res* **67**, 7078-81 (2007).
42. Marcotte, R. *et al.* Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov* **2**, 172-189 (2012).
43. Sanchez, Y. *et al.* Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature. *Nat Commun* **5**, 5812 (2014).
44. Li, Y., Peart, M.J. & Prives, C. Stxbp4 regulates DeltaNp63 stability by suppression of RACK1-dependent degradation. *Mol Cell Biol* **29**, 3953-63 (2009).
45. Sekine, Y. *et al.* The Kelch repeat protein KLHDC10 regulates oxidative stress-induced ASK1 activation by suppressing PP5. *Mol Cell* **48**, 692-704 (2012).
46. Kim, M.H. *et al.* Anaplastic lymphoma kinase gene copy number gain in inflammatory breast cancer (IBC): prevalence, clinicopathologic features and prognostic implication. *PLoS One* **10**, e0120320 (2015).
47. Crizotinib versus Chemotherapy in Advanced ALK-Positive Lung Cancer. *N Engl J Med* **373**, 1582 (2015).
48. Le Page, C. *et al.* BTN3A2 expression in epithelial ovarian cancer is associated with higher tumor infiltrating T cells and a better prognosis. *PLoS One* **7**, e38541 (2012).
49. Kan, L. *et al.* LRRC3B is downregulated in non-small-cell lung cancer and inhibits cancer cell proliferation and invasion. *Tumour Biol* **37**, 1113-20 (2016).
50. Cox, A. *et al.* A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**, 352-8 (2007).
51. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).
52. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190 (2017).
53. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* **50**, 26-41 (2018).
54. Mele, M. *et al.* Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660-5 (2015).
55. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
56. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
57. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
58. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
59. DeLuca, D.S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530-2 (2012).

60. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
61. Guo, X., Lin, M., Rockowitz, S., Lachman, H.M. & Zheng, D. Characterization of human pseudogene-derived non-coding RNAs for functional potential. *PLoS One* **9**, e93972 (2014).
62. Casbas-Hernandez, P. *et al.* Tumor intrinsic subtype is reflected in cancer-adjacent tissue. *Cancer Epidemiol Biomarkers Prev* **24**, 406-14 (2015).
63. Huang, X., Stern, D.F. & Zhao, H. Transcriptional Profiles from Paired Normal Samples Offer Complementary Information on Cancer Patient Survival--Evidence from TCGA Pan-Cancer Data. *Sci Rep* **6**, 20567 (2016).
64. Ghoussaini, M. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* **44**, 312-8 (2012).
65. Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* **45**, 392-8, 398e1-2 (2013).
66. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
67. Freedman, M.L. *et al.* Assessing the impact of population stratification on genetic association studies. *Nat Genet* **36**, 388-93 (2004).
68. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
69. He, B., Chen, C., Teng, L. & Tan, K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A* **111**, E2191-9 (2014).
70. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* **24**, 1-13 (2014).
71. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
72. Consortium, F. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-70 (2014).

Figure Legends

Figure 1. Manhattan plot of association results from the breast cancer transcriptome-wide association study. The red line represents $P = 5.82 \times 10^{-6}$. The blue line represents $P = 1.00 \times 10^{-3}$.

Figure 2. Heat maps of proliferation and colony formation efficiency in breast cells. (A) 184A1, MCF7 or T47D cells were transfected with indicated siRNAs over seven days and phase-contrast images collected using an IncuCyte ZOOM. Each cell proliferation time-course was normalized to the baseline confluency and analyzed using GraphPad Prism. Corrected proliferation % = 100 +/- (relative proliferation in indicated siRNA - proliferation in control siRNA (consi))/knockdown efficiency. **(B)** MCF7 cells were transfected with indicated siRNAs, then reseeded after 16 hours for colony formation (CF) assay. At day 14, colonies were fixed with methanol, stained with crystal violet, scanned and batch analyzed by ImageJ. Corrected CF efficiency (CFE) % = 100 +/- (relative CFE in indicated siRNA - CFE in control siRNA (consi))/knockdown efficiency. Error bars, SD (*N*=2). *P*-values were determined by one-way ANOVA followed by Dunnett's multiple comparisons test: **P*-value < 0.05. NTC: non-target control.

Table 1. Fourteen expression-trait associations for genes located at genomic loci at least 500 kb away from any GWAS-identified breast cancer risk variants

Region	Gene[a]	Type[b]	Z score	*P* value[c]	R^{2c}	Closest risk SNP[d]	Distance to the closest risk SNP (kb)	*P* value after adjusting for adjacent risk SNPs[e]
1p34.1	***ZSWIM5***	Protein	5.26	1.43×10^{-7}	0.17	rs1707302	829	0.006
3p24.1	*LRRC3B*	Protein	-9.57	1.11×10^{-21}	0.17	rs653465	591	1.60×10^{-6}
4q12	*SPATA18*	Protein	-4.62	3.86×10^{-6}	0.11	rs6815814	14,101	3.98×10^{-6}
6p22.1	*UBD*	Protein	-4.87	1.10×10^{-6}	0.13	rs9257408	597	0.94
7q32.2	***KLHDC10***	Protein	5.21	1.92×10^{-7}	0.14	rs4593472	892	2.90×10^{-7}
9p21.3	*MIR31HG*	lncRNA	-5.02	5.22×10^{-7}	0.12	rs1011970	502	1.23×10^{-7}
11p15.5	*RIC8A*	Protein	-5.27	1.40×10^{-7}	0.15	rs6597981	588	4.95×10^{-6}
11q13.2	*B3GNT1*	Protein	-5.85	4.88×10^{-9}	0.09	rs3903072	530	3.50×10^{-6}
11q13.2	*RP11-867G23.10*	transcript	4.71	2.49×10^{-6}	0.03	rs3903072	594	2.61×10^{-4}
12p13.33	***RP11-218M22.1***	lncRNA	5.02	5.27×10^{-7}	0.19	rs12422552	13,641	5.17×10^{-7}
14q24.1	*GALNT16*	Protein	-8.27	1.38×10^{-16}	0.04	rs999737	691	8.57×10^{-4}
14q24.1	***PLEKHD1***	Protein	7.50	6.55×10^{-14}	0.02	rs999737	917	0.12
15q24.2	*MAN2C1* [f]	Protein	-5.32	1.02×10^{-7}	0.39	rs2290203	15,851	9.56×10^{-8}
15q24.2	*CTD-2323K18.1* [f]	lncRNA	-4.65	3.27×10^{-6}	0.07	rs2290203	15,619	3.16×10^{-6}

[a] Genes that were siRNA-silenced for functional assays are bolded; SNPs used to predict gene expression are listed in the Supplementary Table 13

[b] Protein: protein coding genes; lncRNA: long non-coding RNAs; transcript: processed transcript

[c] *P* value: derived from association analyses; associations with $p \leq 5.82 \times 10^{-6}$ considered statistically significant based on Bonferroni correction of 8,597 tests (0.05/8,597); R^2: prediction performance (R^2) derived using GTEx data.

[d] Risk SNPs identified in previous GWAS or fine-mapping studies. The risk SNP closest to the gene is presented. A full list of all risk SNPs, and their distances to the genes are presented in the **Supplementary Table 4**

[e] Use of COJO method[36]

[f] Predicted expression of *MAN2C1* and *CTD-2323K18.1* was correlated (spearman R=0.76)

Table 2. Twenty-three expression-trait associations for genes located at genomic loci within 500 kb of any previous GWAS-identified breast cancer risk variants but not yet implicated as target genes of risk variants[#]

Region	Gene[a]	Type[b]	Z score	*P* value[c]	R^{2c}	Closest risk SNP[d]	Distance to the closest risk SNP (kb)	*P* value after adjusting for adjacent risk SNPs[e]
1p11.2	*RP11-439A17.7*	lncRNA	-5.34	9.07×10^{-8}	0.22	rs11249433	442	0.02
1q21.1	*NUDT17*	Protein	-6.27	3.58×10^{-10}	0.01	rs12405132	56	0.08
1q21.1	*ANKRD34A*	Protein	-5.05	4.42×10^{-7}	0.01	rs12405132	169	4.28×10^{-5}
2p23.1-2p23.2	*ALK*	Protein	4.67	3.06×10^{-6}	0.06	rs4577244	295	2.70×10^{-6}
3p21.31	*PRSS46*	Protein	-5.83	5.68×10^{-9}	0.13	rs6796502	89	0.002
3q12.2	*RP11-114I8.4*	lncRNA	-5.84	5.19×10^{-9}	0.02	rs9833888	356	0.09
5p12	*RP11-53O19.1*	lncRNA	10.38	2.94×10^{-25}	0.03	rs10941679	39	7.46×10^{-4}
5q33.3	***UBLCP1***	Protein	5.93	3.04×10^{-9}	0.07	rs1432679	446	0.37
5q33.3	*RP11-32D16.1*	lncRNA	-5.41	6.37×10^{-8}	0.09	rs1432679	283	1.32×10^{-4}
6p22.2	*BTN3A2*	Protein	4.61	3.97×10^{-6}	0.28	rs71557345	229	0.72
6q23.1	*RP11-73O6.3* [f]	lncRNA	-6.61	3.74×10^{-11}	0.11	rs6569648	105	0.41
11p15.5	***AP006621.6*** [g]	lncRNA	5.61	2.01×10^{-8}	0.34	rs6597981	21	0.52
11p15.5	*RPLP2* [g]	Protein	4.64	3.46×10^{-6}	0.27	rs6597981	7	0.51
14q32.33	*CTD-3051D23.1*	lncRNA	-5.06	4.21×10^{-7}	0.05	rs10623258	97	7.05×10^{-7}
16q12.2	***RP11-467J12.4***	lncRNA	8.04	9.02×10^{-16}	0.23	rs3112612	434	0.79
16q12.2	***CTD-3032H12.1***	lncRNA	4.92	8.58×10^{-7}	0.03	rs28539243	290	0.006
17q21.31	*LRRC37A* [g]	Protein	-5.89	3.85×10^{-9}	0.43	rs2532263	118	0.79
17q21.31	*KANSL1-AS1* [g]	lncRNA	-5.58	2.44×10^{-8}	0.62	rs2532263	18	0.95
17q21.31	*CRHR1* [g]	Protein	-5.29	1.22×10^{-7}	0.22	rs2532263	339	0.99
17q21.31	*LINC00671*	lncRNA	-5.85	4.95×10^{-9}	0.07	rs72826962	190	0.26
17q21.31	*LRRC37A2*	Protein	-5.77	7.93×10^{-9}	0.46	rs2532263	336	0.93
19p13.11	*HAPLN4*	Protein	-7.13	9.88×10^{-13}	0.02	rs2965183	172	0.22
19q13.31	***RP11-15A1.7*** [h]	lncRNA	5.45	5.06×10^{-8}	0.02	rs3760982	215	0.28

[#] not yet reported from eQTL and/or functional studies as target genes of GWAS-identified risk variants and not harbor GWAS or fine-mapping identified risk variants

[a] Genes that were siRNA-silenced for functional assays are bolded; SNPs used to predict gene expression are listed in the Supplementary Table 13

[b] Protein: protein coding genes; lncRNA: long non-coding RNAs
[c] *P* value: nominal *P* value from association analysis; the threshold after Bonferroni correction of 8,597 tests ($0.05/8,597=5.82\times10^{-6}$) was used; R^2: prediction performance (R^2) derived using GTEx data
[d] Risk SNPs identified in previous GWAS or fine-mapping studies. The risk SNP closest to the gene is presented. A full list of all risk SNPs, and their distances to the genes are presented in the **Supplementary Table 4**
[e] Use of COJO method[36]; all index SNPs in the corresponding region were adjusted in the conditional analyses
[f] Predicted expression of *RP11-73O6.3* and *L3MBTL3* was correlated (spearman R=0.88)
[g] Predicted expression of *AP006621.6* and *RPLP2* was correlated; predicted expression of *LRRC37A*, *KANSL1-AS1*, *and CRHR1* was correlated (spearman R>0.1)
[h] Predicted expression of *RP11-15A1.7* and *ZNF404* was correlated (spearman R=0.64)

Table 3. Eleven expression-trait associations for genes previously reported as potential target genes of GWAS-identified breast cancer risk variants or genes harboring risk variants

Region	Gene[a]	Type[b]	Z score	*P* value[c]	R^{2c}	Closest risk SNP[d]	Distance to the closest risk SNP (kb)	*P* value after adjusting for adjacent risk SNPs[e]	Association direction reported previously[f]	Reference
1p36.13	*KLHDC7A*	Protein	-5.67	1.40×10^{-8}	0.04	rs2992756	0.085	0.06	-	[7]
2q33.1	***ALS2CR12***	Protein	6.70	2.11×10^{-11}	0.10	rs1830298	intron of the gene	0.17	NA	[31]
2q33.1	*CASP8*	Protein	-8.05	8.51×10^{-16}	0.22	rs3769821	intron of the gene	0.16	-	[31,32]
5q14.1	*ATG10*	Protein	-6.65	2.85×10^{-11}	0.51	rs7707921	intron of the gene	0.21	NA	[9]
5q14.2	*ATP6AP1L*	Protein	-4.98	6.32×10^{-7}	0.63	rs7707921	37	0.98	NA	[9]
6q23.1	*L3MBTL3* [g]	Protein	-6.69	2.27×10^{-11}	0.10	rs6569648	208	0.44	NA	[6]
6q25.1	***RMND1***	Protein	4.76	1.95×10^{-6}	0.13	rs3757322	169	1.11×10^{-4}	mixed	[17]
11q13.1	*SNX32*	Protein	4.70	2.60×10^{-6}	0.19	rs3903072	18	0.17	NA	[33]
15q26.1	*RCCD1*	Protein	-7.18	7.23×10^{-13}	0.13	rs2290203	6	1.66×10^{-4}	-	[10]
17q22	***STXBP4***	Protein	6.69	2.21×10^{-11}	0.03	rs6504950	intron of the gene	0.90	+ in GTEx	[34,35]
19q13.31	***ZNF404*** [h]	Protein	7.42	1.15×10^{-13}	0.15	rs3760982	90	0.005	NA	[8]

[a] Genes that were siRNA silenced for functional assays are bolded; SNPs used to predict gene expression are listed in the Supplementary Table 13

[b] Protein: protein coding genes; lncRNA: long non-coding RNAs; NA: not available

[c] *P* value: nominal *P* value from association analysis; the threshold after Bonferroni correction of 8,597 tests ($0.05/8,597=5.82\times10^{-6}$) was used; R^2: prediction performance (R^2) derived using GTEx data .

[d] Risk SNPs identified in previous GWAS or fine-mapping studies. The risk SNP closest to the gene is presented. A full list of all risk SNPs, and their distances to the genes are presented in the **Supplementary Table 4**

[e] Use of COJO method[36]; all index SNPs in the corresponding region were adjusted for the conditional analyses

[f] -: inverse association; +: positive association; mixed: both inverse and positive associations reported; NA: not available

[g] Predicted expression of *L3MBTL3* and *RP11-73O6.3* was correlated (spearman R=0.88)

[h] Predicted expression of *ZNF404* and *RP11-15A1.7* was correlated (spearman R=0.64)

Table 4. Genes at GWAS-identified breast cancer risk loci (± 500kb of the index SNPs) whose predicted expression levels were associated with breast cancer risk at *p*-values between 5.82×10^{-6} and 1.05×10^{-3} (FDR corrected *p*-value≤0.05)

Region	Gene	Type[a]	Z score	*P* value[b]	R^{2b}	Closest risk SNP[c]	Distance to the closest risk SNP (kb)	*P* value after adjusting for adjacent risk SNPs[d]
1p34.1	*UQCRH*	Protein	-3.90	9.51×10^{-5}	0.12	rs1707302	168	0.06
1p22.3	*LMO4*	Protein	-3.76	1.73×10^{-4}	0.09	rs12118297	15	0.002
2p23.3	*DNAJC27-AS1*	lncRNA	3.84	1.24×10^{-4}	0.03	rs6725517	65	0.13
4p14	*KLHL5*	Protein	3.52	4.35×10^{-4}	0.13	rs6815814	230	0.03
5q11.2	*AC008391.1*	miRNA	-4.03	5.60×10^{-5}	0.13	rs16886113	242	0.76
6p22.1	*HCG14*	lncRNA	-3.47	5.19×10^{-4}	0.11	rs9257408	61	0.03
6p22.2	*TRNAI2*	miRNA	-3.71	2.09×10^{-4}	0.02	rs71557345	307	0.007
6q25.1	*MTHFD1L*	Protein	3.85	1.17×10^{-4}	0.10	rs3757318	491	2.36×10^{-4}
8q24.21	*PVT1*	transcript	3.85	1.20×10^{-4}	0.03	rs11780156	81	1.09×10^{-4}
9q33.3	*RP11-123K19.1*	lncRNA	-4.10	4.05×10^{-5}	0.05	rs10760444	20	1.26×10^{-4}
10q25.2	*RP11-57H14.3*	lncRNA	3.42	6.16×10^{-4}	0.08	rs7904519	108	0.002
10q26.13	*RP11-500G22.2*	lncRNA	4.48	7.54×10^{-6}	0.15	rs2981582	336	0.91
11p15.5	*PTDSS2*	Protein	-3.47	5.16×10^{-4}	0.04	rs6597981	312	0.02
11p15.5	*AP006621.5*	Protein	4.35	1.37×10^{-5}	0.51	rs6597981	19	0.01
11p15.5	*PIDD1*	Protein	4.24	2.28×10^{-5}	0.45	rs6597981	intron of the gene	0.12
11p15.5	*MRPL23-AS1*	lncRNA	-3.86	1.12×10^{-4}	0.10	rs3817198	95	0.06
11q13.1-11q13.2	*PACS1*	Protein	-3.59	3.36×10^{-4}	0.06	rs3903072	255	0.001
12p11.22	*RP11-860B13.1*	lncRNA	3.46	5.42×10^{-4}	0.17	rs10771399	221	0.86
13q22.1	*KLF5*	Protein	-4.08	4.44×10^{-5}	0.22	rs6562760	306	NA
14q24.1	*CTD-2566J3.1*	lncRNA	-3.84	1.22×10^{-4}	0.04	rs2588809	64	0.55
14q32.33	*C14orf79*	Protein	4.37	1.22×10^{-5}	0.11	rs10623258	240	0.91
15q26.1	*FES*	Protein	4.37	1.26×10^{-5}	0.21	rs2290203	73	3.04×10^{-6}
16q12.2	*BBS2*	Protein	3.97	7.23×10^{-5}	0.26	rs2432539	80	0.36
16q12.2	*CRNDE*	lncRNA	3.28	1.05×10^{-3}	0.02	rs28539243	271	0.69
16q24.2	*RP11-482M8.1*	lncRNA	3.32	9.16×10^{-4}	0.02	rs4496150	441	0.19

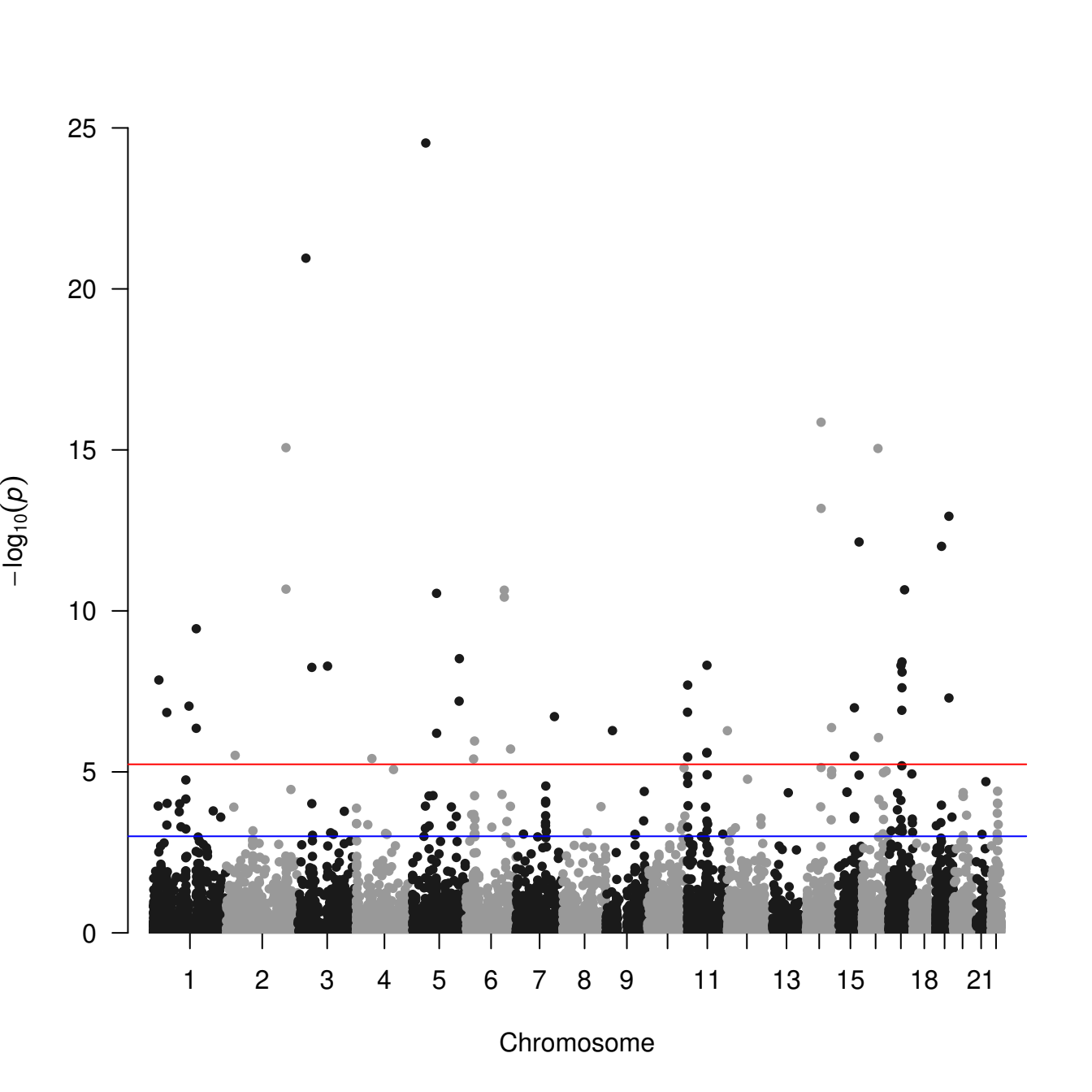
17q11.2	*GOSR1*	Protein	3.79	1.51×10^{-4}	0.10	rs146699004	376	0.04
17q21.2	*ATP6V0A1*	Protein	3.61	3.02×10^{-4}	0.03	rs72826962	162	0.01
17q21.2	*RP11-400F19.8*	transcript	-3.96	7.65×10^{-5}	0.01	rs72826962	122	6.62×10^{-4}
17q21.31	*RP11-105N13.4*	transcript	-4.51	6.46×10^{-6}	0.02	rs2532263	359	NA
17q25.3	*CBX8*	Protein	4.38	1.16×10^{-5}	0.05	rs745570	6	0.99
19p13.11	*CTD-2538G9.5*	lncRNA	3.56	3.76×10^{-4}	0.01	rs8170	432	4.38×10^{-4}
19p13.11	*HOMER3*	Protein	-3.87	1.08×10^{-4}	0.10	rs4808801	469	0.18
20q11.22	*CTD-3216D2.5*	lncRNA	4.03	5.60×10^{-5}	0.16	rs2284378	281	9.24×10^{-4}
22q13.1	*TRIOBP*	Protein	3.34	8.34×10^{-4}	0.07	rs738321	396	0.003
22q13.1	*RP5-1039K5.13*	lncRNA	3.73	1.93×10^{-4}	0.01	rs738321	99	0.053
22q13.1	*CBY1*	Protein	3.91	9.34×10^{-5}	0.05	chr22:39359355	289	0.06
22q13.1	*APOBEC3A*	Protein	-4.11	3.98×10^{-5}	0.07	chr22:39359355	0.2	0.02
22q13.2	*RP1-85F18.6*	lncRNA	3.52	4.28×10^{-4}	0.12	rs73161324	460	0.72

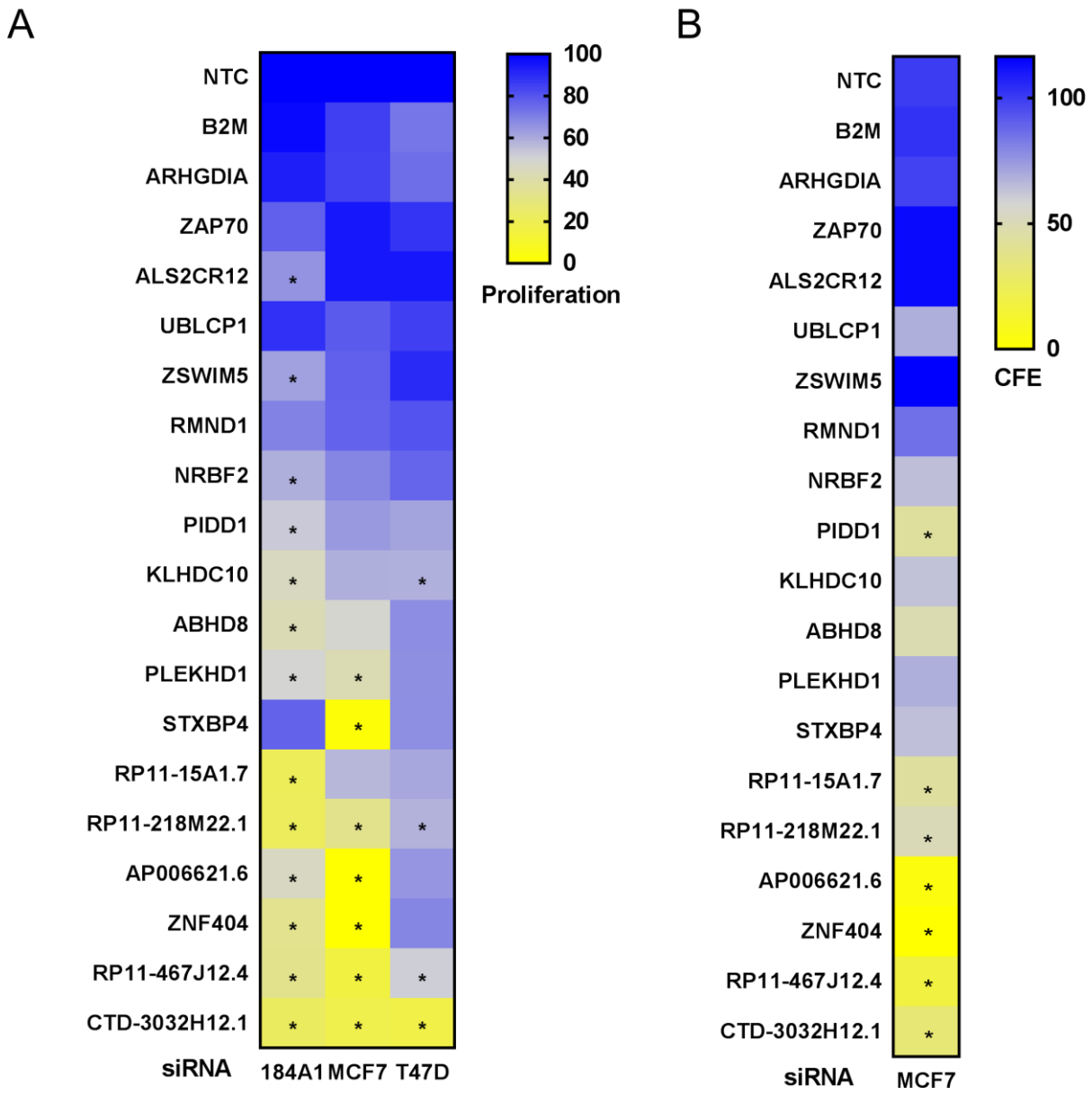
[a] Protein: protein coding genes; lncRNA: long non-coding RNAs; transcript: processed transcript

[b] *P* value: nominal *P* value from association analysis; R^2: prediction performance derived using GTEx data.

[c] Risk SNPs identified in previous GWAS or fine-mapping studies. The risk SNP closest to the gene is presented. A full list of all risk SNPs, and their distances to the genes are presented in the **Supplementary Table 4**

[d] Use of COJO method[36]; all index SNPs in the corresponding region were adjusted for the conditional analyses


25
20
15
10
5
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
11
13
15
18
21
Chromosome


A
NTC
B2M
ARHGDIA
ZAP70
ALS2CR12
UBLCP1
ZSWIM5
RMND1
NRBF2
PIDD1
KLHDC10
ABHD8
PLEKHD1
STXBP4
RP11-15A1.7
RP11-218M22.1
AP006621.6
ZNF404
RP11-467J12.4
CTD-3032H12.1
siRNA
184A1 MCF7 T47D
100
80
60
40
20
0
Proliferation
B
NTC
B2M
ARHGDIA
ZAP70
ALS2CR12
UBLCP1
ZSWIM5
RMND1
NRBF2
PIDD1
KLHDC10
ABHD8
PLEKHD1
STXBP4
RP11-15A1.7
RP11-218M22.1
AP006621.6
ZNF404
RP11-467J12.4
CTD-3032H12.1
siRNA
MCF7
100
50
0
CFE

Supplementary Material

Identification of novel susceptibility loci and genes for breast cancer risk: A transcriptome-wide association study of 229,000 women of European descent

Lang Wu[1,160], Wei Shi[2,160], Jirong Long[1], Xingyi Guo[1], Kyriaki Michailidou[3,4], Jonathan Beesley[2], Manjeet K. Bolla[3], Xiao-Ou Shu[1], Yingchang Lu[1], Qiuyin Cai[1], Fares Al-Ejeh[2], Esdy Rozali[2], Qin Wang[3], Joe Dennis[3], Bingshan Li[151], Chenjie Zeng[1], Helian Feng[5,6], Alexander Gusev[153, 154, 155], Richard T. Barfield[5], Irene L. Andrulis[7,8], Hoda Anton-Culver[9], Volker Arndt[10], Kristan J. Aronson[11], Paul L. Auer[12,13], Myrto Barrdahl[14], Caroline Baynes[15], Matthias W. Beckmann[16], Javier Benitez[17,18], Marina Bermisheva[19,20], Carl Blomqvist[21,159], Natalia V. Bogdanova[20,22,23], Stig E. Bojesen[24-26], Hiltrud Brauch[27-29], Hermann Brenner[10,29,30], Louise Brinton[31], Per Broberg[32], Sara Y. Brucker[33], Barbara Burwinkel[34,35], Trinidad Caldés[36], Federico Canzian[37], Brian D. Carter[38], J. Esteban Castelao[39], Jenny Chang-Claude[14,40], Xiaoqing Chen[2], Ting-Yuan David Cheng[41], Hans Christiansen[22], Christine L. Clarke[42], NBCS Collaborators[43-120,44-124,45], Margriet Collée[46], Sten Cornelissen[47], Fergus J. Couch[48], David Cox[49,50], Angela Cox[51], Simon S. Cross[52], Julie M. Cunningham[48], Kamila Czene[53], Mary B. Daly[54], Peter Devilee[55,56], Kimberly F. Doheny[57], Thilo Dörk[20], Isabel dos-Santos-Silva[58], Martine Dumont[59], Miriam Dwek[60], Diana M. Eccles[61], Ursula Eilber[14], A. Heather Eliassen[6,62], Christoph Engel[63], Mikael Eriksson[53], Laura Fachal[15], Peter A. Fasching[16,64], Jonine Figueroa[31,65], Dieter Flesch-Janys[66,67], Olivia Fletcher[68], Henrik Flyger[69], Lin Fritschi[70], Marike Gabrielson[53], Manuela Gago-Dominguez[71,72], Susan M. Gapstur[38], Montserrat García-Closas[31], Mia M. Gaudet[38], Maya Ghoussaini[15], Graham G. Giles[73,74], Mark S. Goldberg[75,76], David E. Goldgar[77], Anna González-Neira[17], Pascal Guénel[78], Eric Hahnen[79-81], Christopher A. Haiman[82], Niclas Håkansson[83], Per Hall[53], Emily Hallberg[84], Ute Hamann[85], Patricia Harrington[15], Alexander Hein[16], Belynda Hicks[86], Peter Hillemanns[20], Antoinette Hollestelle[87], Robert N. Hoover[31], John L. Hopper[74], Guanmengqian Huang[85], Keith Humphreys[53], David J. Hunter[6,158], Anna Jakubowska[88], Wolfgang Janni[89], Esther M. John[90-92], Nichola Johnson[68], Kristine Jones[86], Michael E. Jones[93], Audrey Jung[14], Rudolf Kaaks[14], Michael J. Kerin[94], Elza Khusnutdinova[19,95], Veli-Matti Kosma[96-98], Vessela N. Kristensen[99-101], Diether Lambrechts[102,103], Loic Le Marchand[104], Jingmei Li[157], Sara Lindström[5,105], Jolanta Lissowska[106], Wing-Yee Lo[27,28], Sibylle Loibl[107], Jan Lubinski[88], Craig Luccarini[15], Michael P. Lux[16], Robert J. MacInnis[73,74], Tom Maishman[61,108], Ivana Maleva Kostovska[20,109], Arto Mannermaa[96-98], JoAnn E. Manson[6,110], Sara Margolin[111], Dimitrios Mavroudis[112], Hanne Meijers-Heijboer[152], Alfons Meindl[113], Usha Menon[114], Jeffery Meyer[48], Anna Marie Mulligan[115,116], Susan L. Neuhausen[117], Heli Nevanlinna[118], Patrick Neven[119], Sune F. Nielsen[24,25], Børge G. Nordestgaard[24-26], Olufunmilayo I. Olopade[120], Janet E. Olson[84], Håkan Olsson[32], Paolo Peterlongo[121], Julian Peto[58], Dijana Plaseska-Karanfilska[109], Ross Prentice[12], Nadege Presneau[60], Katri Pylkäs[122,123], Brigitte Rack[89], Paolo Radice[125], Nazneen Rahman[126], Gad Rennert[127], Hedy S. Rennert[127], Valerie Rhenius[15], Atocha Romero[36,128], Jane Romm[57], Anja Rudolph[14], Emmanouil Saloustros[129], Dale P. Sandler[130], Elinor J. Sawyer[131], Marjanka K. Schmidt[47,132], Rita K. Schmutzler[79-81], Andreas Schneeweiss[34,133], Rodney J. Scott[134,135], Christopher Scott[84], Sheila Seal[126], Mitul Shah[15], Martha J. Shrubsole[1], Ann Smeets[119], Melissa C. Southey[136], John J. Spinelli[137,138], Jennifer Stone[139,140], Harald

Surowy[34,35], Anthony J. Swerdlow[93,141], Rulla M. Tamimi[5,6,62], William Tapper[61], Jack A. Taylor[130,142], Mary Beth Terry[143], Daniel C. Tessier[144], Abigail Thomas[84], Kathrin Thöne[67], Rob A.E.M. Tollenaar[145], Diana Torres[85,146], Thérèse Truong[78], Michael Untch[147], Celine Vachon[84], David Van Den Berg[82], Daniel Vincent[144], Quinten Waisfisz[152], Clarice R. Weinberg[148], Camilla Wendt[111], Alice S. Whittemore[91,92], Hans Wildiers[119], Walter C. Willett[6,62,156], Robert Winqvist[122,123], Alicja Wolk[83], Lucy Xia[82], Xiaohong R. Yang[31], Argyrios Ziogas[9], Elad Ziv[149], kConFab/AOCS Investigators[150], Alison M. Dunning[15], Paul D.P. Pharoah[3,15], Jacques Simard[59], Roger L. Milne[73,74], Stacey L. Edwards[2], Peter Kraft[5,6], Douglas F. Easton[3,15], Georgia Chenevix-Trench[2]*, Wei Zheng[1]*

Supplementary Tables 1-13

Supplementary Figures 1-8

Supplementary Note

Supplementary Table 1. Internal performance of gene expression prediction models built using GTEx data

Prediction performance (R^2)	**All**	**Protein**	**lncRNAs**	**miRNAs**	**Others***
Number of genes	15,148	10,483	4,277	68	320
0.01[#]	12,824	8,874	3,628	57	265
0.04	7,655	5,244	2,200	38	173
0.09	3,818	2,601	1,106	19	92
0.16	1,573	1,035	479	8	51

Protein: protein coding genes; lncRNAs: long non-coding RNAs; miRNAs: microRNAs

* Including processed transcripts, immunoglobulin genes, and T cell receptor genes

[#] The R^2 of 0.01 is the internal prediction performance threshold according to which the prediction models were retained for external evaluation in the TCGA data

Supplementary Table 2. Genes with predicted expression levels associated with breast cancer risk at $p<1.05 \times 10^{-3}$ (the significance level with false discovery rate correction)

Gene	Type*	Z score	*P* value	R^2	No. of predicting variants used	No. of predicting variants in model	Proportion of predicting variants used (%)
RP11-53O19.1	lncRNA	10.38	2.94×10^{-25}	0.03	8	15	53
LRRC3B	Protein	-9.57	1.11×10^{-21}	0.17	46	46	100
GALNT16	Protein	-8.27	1.38×10^{-16}	0.04	53	53	100
CASP8	Protein	-8.05	8.51×10^{-16}	0.22	15	15	100
RP11-467J12.4	lncRNA	8.04	9.02×10^{-16}	0.23	142	142	100
PLEKHD1	Protein	7.50	6.55×10^{-14}	0.02	6	6	100
ZNF404	Protein	7.42	1.15×10^{-13}	0.15	32	32	100
RCCD1	Protein	-7.18	7.23×10^{-13}	0.13	22	22	100
HAPLN4	Protein	-7.13	9.88×10^{-13}	0.02	53	53	100
ALS2CR12	Protein	6.70	2.11×10^{-11}	0.10	4	4	100
STXBP4	Protein	6.69	2.21×10^{-11}	0.03	58	60	97
L3MBTL3	Protein	-6.69	2.27×10^{-11}	0.10	5	5	100
ATG10	Protein	-6.65	2.85×10^{-11}	0.51	57	61	93
RP11-73O6.3	lncRNA	-6.61	3.74×10^{-11}	0.11	26	26	100
NUDT17	Protein	-6.27	3.58×10^{-10}	0.01	7	7	100
UBLCP1	Protein	5.93	3.04×10^{-9}	0.07	2	3	67
LRRC37A	Protein	-5.89	3.85×10^{-9}	0.43	31	32	97
B3GNT1	Protein	-5.85	4.88×10^{-9}	0.09	26	27	96
LINC00671	lncRNA	-5.85	4.95×10^{-9}	0.07	1	1	100
RP11-114I8.4	lncRNA	-5.84	5.19×10^{-9}	0.02	14	14	100
PRSS46	Protein	-5.83	5.68×10^{-9}	0.13	45	46	98
LRRC37A2	Protein	-5.77	7.93×10^{-9}	0.46	120	121	99
KLHDC7A	Protein	-5.67	1.40×10^{-8}	0.04	15	15	100
AP006621.6	lncRNA	5.61	2.01×10^{-8}	0.34	41	41	100
KANSL1-AS1	lncRNA	-5.58	2.44×10^{-8}	0.62	70	72	97

RP11-15A1.7	lncRNA	5.45	5.06×10^{-8}	0.02	2	2	100
RP11-32D16.1	lncRNA	-5.41	6.37×10^{-8}	0.09	44	46	96
RP11-439A17.7	lncRNA	-5.34	9.07×10^{-8}	0.22	93	94	99
MAN2C1	Protein	-5.32	1.02×10^{-7}	0.39	27	27	100
CRHR1	Protein	-5.29	1.22×10^{-7}	0.22	31	31	100
RIC8A	Protein	-5.27	1.40×10^{-7}	0.15	15	15	100
ZSWIM5	Protein	5.26	1.43×10^{-7}	0.17	67	67	100
KLHDC10	Protein	5.21	1.92×10^{-7}	0.14	52	53	98
CTD-3051D23.1	lncRNA	-5.06	4.21×10^{-7}	0.05	25	26	96
ANKRD34A	Protein	-5.05	4.42×10^{-7}	0.01	1	1	100
MIR31HG	lncRNA	-5.02	5.22×10^{-7}	0.12	1	1	100
RP11-218M22.1	lncRNA	5.02	5.27×10^{-7}	0.19	47	48	98
ATP6AP1L	Protein	-4.98	6.32×10^{-7}	0.63	64	67	96
CTD-3032H12.1	lncRNA	4.92	8.58×10^{-7}	0.03	23	23	100
UBD	Protein	-4.87	1.10×10^{-6}	0.13	31	31	100
RMND1	Protein	4.76	1.95×10^{-6}	0.13	91	91	100
RP11-867G23.10	transcript	4.71	2.49×10^{-6}	0.03	5	5	100
SNX32	Protein	4.70	2.60×10^{-6}	0.19	17	17	100
ALK	Protein	4.67	3.06×10^{-6}	0.06	47	48	98
CTD-2323K18.1	transcript	-4.65	3.27×10^{-6}	0.07	23	23	100
RPLP2	Protein	4.64	3.46×10^{-6}	0.27	45	45	100
SPATA18	Protein	-4.62	3.86×10^{-6}	0.11	43	43	100
BTN3A2	Protein	4.61	3.97×10^{-6}	0.28	66	66	100
RP11-105N13.4	transcript	-4.51	6.46×10^{-6}	0.02	15	16	94
SLC39A9	Protein	-4.48	7.32×10^{-6}	0.03	24	24	100
RP11-500G22.2	lncRNA	4.48	7.54×10^{-6}	0.15	8	8	100
FAT4	Protein	4.45	8.44×10^{-6}	0.06	42	54	78
CRIP2	Protein	4.44	9.14×10^{-6}	0.03	12	12	100
RP11-432I5.1	lncRNA	4.40	1.06×10^{-5}	0.03	11	14	79
CBX8	Protein	4.38	1.16×10^{-5}	0.05	12	12	100
C14orf79	Protein	4.37	1.22×10^{-5}	0.11	5	5	100
RHOD	Protein	4.37	1.23×10^{-5}	0.03	24	24	100

FES	Protein	4.37	1.26×10^{-5}	0.21	23	23	100
AP006621.5	Protein	4.35	1.37×10^{-5}	0.51	46	46	100
NUP107	Protein	4.30	1.69×10^{-5}	0.14	4	4	100
GSTM4	Protein	-4.29	1.78×10^{-5}	0.06	9	9	100
YBEY	Protein	4.26	2.01×10^{-5}	0.40	27	27	100
PIDD1	Protein	4.24	2.28×10^{-5}	0.45	61	61	100
RP11-126L15.4	lncRNA	-4.19	2.74×10^{-5}	0.05	59	59	100
AC010136.2	lncRNA	-4.14	3.52×10^{-5}	0.21	1	1	100
APOBEC3A	Protein	-4.11	3.98×10^{-5}	0.07	33	33	100
RP11-123K19.1	lncRNA	-4.10	4.05×10^{-5}	0.05	21	21	100
GABPB1-AS1	transcript	4.10	4.21×10^{-5}	0.45	28	28	100
CTD-3110H11.1	lncRNA	4.09	4.31×10^{-5}	0.53	25	26	96
EDEM2	Protein	4.09	4.39×10^{-5}	0.03	58	59	98
KLF5	Protein	-4.08	4.44×10^{-5}	0.22	30	30	100
HSF2	Protein	-4.05	5.02×10^{-5}	0.04	45	45	100
SMN2	Protein	-4.04	5.44×10^{-5}	0.19	33	34	97
XXbac-BPG170G13.32	lncRNA	4.03	5.50×10^{-5}	0.14	50	56	89
AC008391.1	miRNA	-4.03	5.60×10^{-5}	0.13	7	7	100
CTD-3216D2.5	lncRNA	4.03	5.60×10^{-5}	0.16	57	57	100
CPNE1	Protein	-4.02	5.80×10^{-5}	0.33	36	36	100
GSTM3	Protein	-3.98	6.95×10^{-5}	0.18	23	23	100
BBS2	Protein	3.97	7.23×10^{-5}	0.26	20	20	100
RP11-400F19.8	transcript	-3.96	7.65×10^{-5}	0.01	22	26	85
PILRA	Protein	3.94	8.16×10^{-5}	0.54	25	25	100
STAG3L5P-PVRIG2P-PILRB	transcript	3.91	9.27×10^{-5}	0.32	42	43	98
CBY1	Protein	3.91	9.34×10^{-5}	0.05	19	21	90
UQCRH	Protein	-3.90	9.51×10^{-5}	0.12	35	35	100
ALS2CL	Protein	-3.90	9.69×10^{-5}	0.23	1	3	33
ATF4	Protein	-3.90	9.74×10^{-5}	0.11	95	97	98
CCBL2	Protein	3.90	9.78×10^{-5}	0.01	13	17	76
HOMER3	Protein	-3.87	1.08×10^{-4}	0.10	16	16	100

CMTR2	Protein	-3.86	1.11×10^{-4}	0.01	22	38	58
MRPL23-AS1	lncRNA	-3.86	1.12×10^{-4}	0.10	13	14	93
ARHGEF19	Protein	-3.86	1.15×10^{-4}	0.13	95	96	99
NNT-AS1	lncRNA	3.86	1.15×10^{-4}	0.06	40	40	100
MTHFD1L	Protein	3.85	1.17×10^{-4}	0.10	24	24	100
PVT1	transcript	3.85	1.20×10^{-4}	0.03	14	17	82
CTD-2566J3.1	lncRNA	-3.84	1.22×10^{-4}	0.04	16	16	100
PDLIM4	Protein	-3.84	1.22×10^{-4}	0.08	42	43	98
MYRF	Protein	3.84	1.24×10^{-4}	0.01	10	10	100
DNAJC27-AS1	lncRNA	3.84	1.24×10^{-4}	0.03	22	22	100
ATP5I	Protein	-3.82	1.34×10^{-4}	0.02	9	9	100
GOSR1	Protein	3.79	1.51×10^{-4}	0.10	13	13	100
RP11-335O13.7	lncRNA	-3.77	1.63×10^{-4}	0.08	34	34	100
RP11-550I24.2	transcript	-3.76	1.67×10^{-4}	0.05	61	61	100
LMO4	Protein	-3.76	1.73×10^{-4}	0.09	1	1	100
RP5-1039K5.13	lncRNA	3.73	1.93×10^{-4}	0.01	37	38	97
TRNAI2	miRNA	-3.71	2.09×10^{-4}	0.02	12	12	100
RP4-625H18.2	lncRNA	-3.70	2.12×10^{-4}	0.02	5	5	100
ZNF334	Protein	-3.69	2.22×10^{-4}	0.12	55	55	100
PILRB	Protein	3.68	2.29×10^{-4}	0.30	70	71	99
METTL10	Protein	-3.68	2.35×10^{-4}	0.17	25	25	100
SH3TC2	Protein	3.67	2.42×10^{-4}	0.09	42	43	98
CTD-2026K11.3	lncRNA	3.67	2.46×10^{-4}	0.01	20	20	100
CTD-2026K11.2	lncRNA	3.66	2.52×10^{-4}	0.12	109	130	84
TMC4	Protein	3.66	2.54×10^{-4}	0.21	6	6	100
RP5-1139B12.4	lncRNA	-3.66	2.55×10^{-4}	0.17	47	47	100
TBX5	Protein	3.64	2.73×10^{-4}	0.11	85	85	100
SNUPN	Protein	-3.63	2.86×10^{-4}	0.03	4	4	100
RP11-1055B8.4	lncRNA	3.62	2.92×10^{-4}	0.20	5	5	100
PSORS1C2	Protein	3.62	2.96×10^{-4}	0.41	29	32	91
IST1	Protein	3.62	3.00×10^{-4}	0.01	18	18	100
ATP6V0A1	Protein	3.61	3.02×10^{-4}	0.03	98	99	99

KLC1	Protein	-3.61	3.08×10^{-4}	0.07	37	37	100
GPR144	Protein	3.59	3.31×10^{-4}	0.12	53	75	71
PACS1	Protein	-3.59	3.36×10^{-4}	0.06	49	49	100
ECT2L	Protein	3.58	3.47×10^{-4}	0.14	3	3	100
CTD-2538G9.5	lncRNA	3.56	3.76×10^{-4}	0.01	7	7	100
AZGP1	Protein	-3.55	3.79×10^{-4}	0.03	5	5	100
OXLD1	Protein	3.55	3.86×10^{-4}	0.15	31	31	100
CPLX1	Protein	-3.54	4.03×10^{-4}	0.05	17	17	100
DGKQ	Protein	3.54	4.06×10^{-4}	0.25	85	85	100
RP11-757G1.6	lncRNA	3.53	4.17×10^{-4}	0.19	33	33	100
CTA-109P11.4	lncRNA	-3.52	4.26×10^{-4}	0.10	10	10	100
RP1-85F18.6	lncRNA	3.52	4.28×10^{-4}	0.12	88	88	100
TBX5-AS1	lncRNA	3.52	4.31×10^{-4}	0.09	55	61	90
KLHL5	Protein	3.52	4.35×10^{-4}	0.13	106	109	97
MUTYH	Protein	3.51	4.47×10^{-4}	0.04	12	12	100
TRIM4	Protein	-3.50	4.64×10^{-4}	0.43	72	74	97
MIR1909	miRNA	3.50	4.68×10^{-4}	0.04	33	34	97
SLC22A5	Protein	-3.50	4.72×10^{-4}	0.19	28	28	100
CCDC18	Protein	-3.48	5.08×10^{-4}	0.38	94	94	100
PTDSS2	Protein	-3.47	5.16×10^{-4}	0.04	31	31	100
HCG14	lncRNA	-3.47	5.19×10^{-4}	0.11	2	2	100
SMIM8	Protein	3.47	5.20×10^{-4}	0.06	20	20	100
MAP3K14-AS1	lncRNA	-3.46	5.31×10^{-4}	0.04	3	3	100
FAM149B1	Protein	-3.46	5.35×10^{-4}	0.03	12	12	100
RP11-860B13.1	lncRNA	3.46	5.42×10^{-4}	0.17	14	14	100
PAIP1	Protein	-3.45	5.67×10^{-4}	0.02	2	2	100
GSTM5	Protein	-3.44	5.92×10^{-4}	0.28	20	20	100
RP11-57H14.3	lncRNA	3.42	6.16×10^{-4}	0.08	2	2	100
BRMS1	Protein	-3.40	6.62×10^{-4}	0.05	7	7	100
KDM6B	Protein	-3.40	6.73×10^{-4}	0.07	36	52	69
IGKV2D-24	IG_gene	-3.40	6.74×10^{-4}	0.02	1	1	100
RP11-174G6.5	lncRNA	3.39	7.00×10^{-4}	0.05	26	27	96

POLR2J	Protein	-3.39	7.01×10^{-4}	0.28	86	86	100
RP11-580I16.2	lncRNA	3.38	7.17×10^{-4}	0.04	4	4	100
RP13-20L14.1	lncRNA	-3.37	7.52×10^{-4}	0.02	8	9	89
RP11-553A10.1	Protein	3.36	7.76×10^{-4}	0.03	31	33	94
RP11-363E6.3	lncRNA	-3.36	7.83×10^{-4}	0.05	37	37	100
TSPAN5	Protein	-3.35	8.11×10^{-4}	0.04	12	12	100
PSORS1C1	Protein	3.34	8.28×10^{-4}	0.35	17	20	85
TRIOBP	Protein	3.34	8.34×10^{-4}	0.07	22	23	96
CLEC18A	Protein	-3.34	8.37×10^{-4}	0.43	32	32	100
DFNA5	Protein	-3.33	8.55×10^{-4}	0.19	28	28	100
TMEM136	Protein	3.33	8.56×10^{-4}	0.07	68	78	87
C9orf3	Protein	3.33	8.64×10^{-4}	0.03	23	26	88
GPR156	Protein	3.33	8.67×10^{-4}	0.19	69	71	97
IL10RB-AS1	lncRNA	-3.33	8.68×10^{-4}	0.17	91	92	99
BDH2	Protein	-3.33	8.72×10^{-4}	0.23	41	41	100
ZNF165	Protein	3.33	8.76×10^{-4}	0.06	17	17	100
LINC00092	lncRNA	-3.32	9.03×10^{-4}	0.08	43	43	100
RP11-482M8.1	lncRNA	3.32	9.16×10^{-4}	0.02	37	37	100
USP19	Protein	-3.31	9.28×10^{-4}	0.02	5	6	83
MMP24	Protein	-3.31	9.40×10^{-4}	0.13	2	2	100
CTD-2196P11.2	lncRNA	3.29	1.01×10^{-3}	0.04	28	29	97
NR1H3	Protein	3.29	1.01×10^{-3}	0.17	52	53	98
FLOT1	Protein	-3.28	1.03×10^{-3}	0.10	60	63	95
BAZ1B	Protein	-3.28	1.04×10^{-3}	0.14	63	63	100
AHI1	Protein	3.28	1.05×10^{-3}	0.23	13	14	93
CRNDE	lncRNA	3.28	1.05×10^{-3}	0.02	22	25	88
AL450992.2	lncRNA	-3.28	1.05×10^{-3}	0.03	6	6	100

* Protein: protein coding genes; lncRNA: long non-coding RNAs; miRNA: microRNA; transcript: processed transcript; IG_gene: immunoglobulin genes.

P value: nominal *p* value from association analysis; R^2: prediction performance (R^2) derived using GTEx data.

Supplementary Table 3. Associations of predicted expression of identified genes with breast cancer risk in each of the three assessed datasets (OncoArray, iCOGS, and GWAS sets)

Gene name	OncoArray z-score	OncoArray *p*-value	iCOGS z-score	iCOGS *p*-value	GWAS z-score	GWAS p-value	Cochran's Q	I^2
Table 1								
ZSWIM5	2.98	0.003	4.32	1.57×10^{-5}	1.39	0.17	0.32	0
LRRC3B	-7.48	7.19×10^{-14}	-4.89	1.02×10^{-6}	-3.61	3.11×10^{-4}	2.09	0.04
SPATA18	-3.09	0.002	-2.59	0.01	-2.33	0.02	0.21	0
UBD	-1.55	0.12	-4.07	4.67×10^{-5}	-3.46	5.48×10^{-4}	1.54	0
KLHDC10	2.15	0.03	4.39	1.16×10^{-5}	2.87	0.004	0.92	0
MIR31HG	-4.35	1.35×10^{-5}	-2.90	0.004	-0.53	0.60	0.98	0
RIC8A	-3.28	0.001	-3.12	0.002	-2.71	0.007	0.17	0
B3GNT1	-2.70	0.007	-5.00	5.83×10^{-7}	-2.38	0.02	0.82	0
RP11-867G23.10	2.78	0.005	3.13	0.002	2.18	0.03	0.04	0
RP11-218M22.1	3.84	1.22×10^{-4}	3.33	8.82×10^{-4}	0.86	0.39	0.35	0
GALNT16	-4.45	8.74×10^{-6}	-6.17	6.82×10^{-10}	-3.67	2.40×10^{-4}	0.38	0
PLEKHD1	5.21	1.85×10^{-7}	3.96	7.43×10^{-5}	3.96	7.36×10^{-5}	0.90	0
MAN2C1	-4.08	4.47×10^{-5}	-3.49	4.88×10^{-4}	-0.86	0.39	0.43	0
CTD-2323K18.1	-3.69	2.23×10^{-4}	-2.62	0.009	-1.27	0.21	0.42	0
Table 2								
RP11-439A17.7	-4.35	1.37×10^{-5}	-3.39	6.90×10^{-4}	-0.32	0.75	0.88	0
NUDT17	-3.53	4.19×10^{-4}	-4.99	5.91×10^{-7}	-1.98	0.047	0.30	0
ANKRD34A	-4.27	1.97×10^{-5}	-2.54	0.01	-1.13	0.26	0.94	0
ALK	3.84	1.23×10^{-4}	3.23	0.001	-0.08	0.94	0.84	0
PRSS46	-4.33	1.51×10^{-5}	-3.51	4.41×10^{-4}	-1.78	0.08	0.31	0
RP11-114I8.4	-4.20	2.66×10^{-5}	-3.15	0.002	-2.74	0.006	0.48	0
RP11-53O19.1	8.29	1.17×10^{-16}	5.75	8.85×10^{-9}	3.23	0.001	2.16	0.07
UBLCP1	4.72	2.34×10^{-6}	3.12	0.002	1.98	0.047	0.80	0
RP11-32D16.1	-3.75	1.75×10^{-4}	-3.66	2.51×10^{-4}	-1.53	0.13	0.09	0
BTN3A2	3.16	0.002	2.74	0.006	2.06	0.04	0.12	0
RP11-73O6.3	-5.34	9.31×10^{-8}	-2.24	0.03	-4.32	1.53×10^{-5}	3.39	0.41

AP006621.6	3.58	3.40×10^{-4}	3.92	8.75×10^{-5}	1.98	0.048	0.01	0
RPLP2	3.43	5.93×10^{-4}	2.77	0.006	1.57	0.12	0.19	0
CTD-3051D23.1	-2.60	0.009	-3.36	7.85×10^{-4}	-3.34	8.30×10^{-4}	0.45	0
RP11-467J12.4	5.75	8.73×10^{-9}	5.41	6.28×10^{-8}	1.93	0.054	0.38	0
CTD-3032H12.1	2.93	0.003	2.95	0.003	2.60	0.009	0.15	0
LRRC37A	-4.13	3.56×10^{-5}	-3.08	0.002	-3.07	0.002	0.58	0
KANSL1-AS1	-3.83	1.28×10^{-4}	-3.17	0.002	-2.61	0.009	0.27	0
CRHR1	-3.58	3.39×10^{-4}	-2.81	0.005	-2.93	0.003	0.45	0
LINC00671	-4.40	1.11×10^{-5}	-4.15	3.32×10^{-5}	-0.25	0.80	0.82	0
LRRC37A2	-3.93	8.47×10^{-5}	-3.18	0.001	-2.96	0.003	0.39	0
HAPLN4	-5.49	4.01×10^{-8}	-5.10	3.46×10^{-7}	-0.31	0.75	1.28	0
RP11-15A1.7	3.65	2.59×10^{-4}	4.26	2.00×10^{-5}	0.78	0.44	0.38	0
Table 3								
KLHDC7A	-4.69	2.77×10^{-6}	-3.53	4.11×10^{-4}	-0.62	0.53	0.91	0
ALS2CR12	4.98	6.25×10^{-7}	2.80	0.005	4.24	2.21×10^{-5}	2.09	0.04
CASP8	-5.97	2.42×10^{-9}	-3.63	2.78×10^{-4}	-4.66	3.20×10^{-6}	2.30	0.13
ATG10	-3.00	0.003	-5.83	5.60×10^{-9}	-2.65	0.008	1.27	0
ATP6AP1L	-2.40	0.02	-4.24	2.20×10^{-5}	-1.87	0.06	0.51	0
L3MBTL3	-5.42	5.89×10^{-8}	-2.38	0.02	-4.13	3.65×10^{-5}	3.13	0.36
RMND1	3.14	0.002	2.76	0.006	2.41	0.02	0.18	0
SNX32	2.41	0.02	3.80	1.45×10^{-4}	1.78	0.08	0.27	0
RCCD1	-5.58	2.36×10^{-8}	-4.08	4.56×10^{-5}	-2.21	0.03	0.81	0
STXBP4	4.77	1.85×10^{-6}	4.01	6.05×10^{-5}	2.46	0.01	0.28	0
ZNF404	4.76	1.96×10^{-6}	5.28	1.28×10^{-7}	2.28	0.02	0.06	0
Table 4								
UQCRH	-3.13	0.002	-2.14	0.03	-1.19	0.23	0.32	0
LMO4	-2.42	0.02	-2.53	0.01	-1.42	0.16	0.002	0
DNAJC27-AS1	3.41	6.47×10^{-4}	1.37	0.17	1.77	0.08	1.12	0
KLHL5	2.34	0.02	1.96	0.05	1.59	0.11	0.09	0
AC008391.1	-2.84	0.004	-3.00	0.003	-0.36	0.72	0.27	0
HCG14	-2.65	0.008	-2.54	0.01	0.02	0.99	0.37	0
TRNAI2	-2.26	0.02	-2.46	0.01	-1.67	0.09	0.02	0

MTHFD1L	2.26	0.02	2.81	0.005	1.58	0.11	0.02	0
PVT1	2.12	0.03	2.73	0.006	1.82	0.07	0.06	0
RP11-123K19.1	-3.80	1.42×10^{-4}	-1.49	0.14	-1.44	0.15	1.37	0
RP11-57H14.3	3.54	3.98×10^{-4}	1.50	0.13	0.09	0.93	1.31	0
RP11-500G22.2	3.09	0.002	3.15	0.002	1.01	0.31	0.10	0
PTDSS2	-1.69	0.09	-2.98	0.003	-1.33	0.18	0.25	0
AP006621.5	2.80	0.005	3.13	0.002	1.26	0.21	0.03	0
PIDD1	1.61	0.11	3.70	2.16×10^{-4}	2.42	0.02	0.82	0
MRPL23-AS1	-2.29	0.02	-2.04	0.04	-2.89	0.004	0.48	0
PACS1	-1.40	0.16	-3.53	4.19×10^{-4}	-1.09	0.27	0.81	0
RP11-860B13.1	2.86	0.004	2.15	0.03	0.66	0.51	0.26	0
KLF5	-2.16	0.03	-2.38	0.017	-3.16	0.002	0.54	0
CTD-2566J3.1	-2.53	0.01	-2.65	0.008	-1.19	0.24	0.02	0
C14orf79	3.60	3.17×10^{-4}	1.89	0.06	2.03	0.04	0.86	0
FES	3.48	4.95×10^{-4}	1.82	0.07	2.34	0.02	0.90	0
BBS2	2.65	0.008	3.08	0.002	0.53	0.59	0.21	0
CRNDE	2.82	0.005	0.50	0.61	2.76	0.006	1.90	0
RP11-482M8.1	2.54	0.01	1.82	0.07	1.37	0.17	0.18	0
GOSR1	2.87	0.004	1.61	0.11	2.22	0.03	0.62	0
ATP6V0A1	2.23	0.03	2.74	0.006	0.94	0.34	0.06	0
RP11-400F19.8	-4.18	2.91×10^{-5}	0.36	0.72	-3.47	5.28×10^{-4}	5.84	0.66
RP11-105N13.4	-2.92	0.004	-2.64	0.008	-2.32	0.02	0.15	0
CBX8	1.82	0.07	3.61	3.04×10^{-4}	2.44	0.01	0.60	0
CTD-2538G9.5	1.61	0.11	3.17	0.002	1.48	0.14	0.39	0
HOMER3	-1.67	0.09	-2.92	0.004	-2.45	0.01	0.41	0
CTD-3216D2.5	1.40	0.16	3.10	0.002	3.18	0.001	0.98	0
TRIOBP	3.77	1.63×10^{-4}	0.55	0.58	0.92	0.36	2.46	0.19
RP5-1039K5.13	2.43	0.02	1.68	0.09	2.67	0.008	0.56	0
CBY1	2.13	0.03	2.60	0.009	2.29	0.02	0.14	0
APOBEC3A	-3.44	5.87×10^{-4}	-1.37	0.17	-2.60	0.009	1.40	0
RP1-85F18.6	1.68	0.09	2.94	0.003	1.48	0.14	0.24	0

Supplementary Table 4. Full list of all index SNPs within the same genomic loci/region of the identified associated genes in Tables 1-4 and their distances with the associated genes

Gene	Index SNP(s)[#]	Distance to the index SNP (kb)
Table 1		
ZSWIM5	rs1707302	829
LRRC3B	rs12493607 rs653465 rs4973768	3931 591 705
SPATA18	rs6815814	14,101
UBD	rs9257408	597
KLHDC10	rs4593472	892
MIR31HG	rs1011970	502
RIC8A	rs6597981 rs3817198	588 1694
B3GNT1	rs3903072	530
RP11-867G23.10	rs3903072	594
RP11-218M22.1	rs12422552	13,641
GALNT16	rs999737	691
PLEKHD1	rs999737	917
MAN2C1	rs2290203	15,851
CTD-2323K18.1	rs2290203	15,619
Table 2		
RP11-439A17.7	rs11249433	442
NUDT17	rs12405132	56
ANKRD34A	rs12405132	169
ALK	rs4577244	295
PRSS46	rs6796502	89
RP11-114I8.4	rs9833888	356
RP11-53O19.1	rs10941679 rs4415084	39 82

UBLCP1	rs1432679	446
RP11-32D16.1	rs1432679	283
BTN3A2	rs71557345	229
RP11-73O6.3	rs6569648	105
AP006621.6	rs6597981 rs909116 rs3817198	21 1160 1127
RPLP2	rs6597981 rs909116 rs3817198	7 1129 1096
CTD-3051D23.1	rs10623258	97
RP11-467J12.4	rs12922061 rs17817449 rs11075995 rs3112612 rs3803662 rs28539243	434-1595
CTD-3032H12.1	rs12922061 rs17817449 rs11075995 rs3112612 rs3803662 rs28539243	290-2385
LINC00671	rs72826962	190
LRRC37A	rs2532263	118
KANSL1-AS1	rs2532263	18
CRHR1	rs2532263	339
LRRC37A2	rs2532263	336
HAPLN4	rs8170 rs2363956 rs4808801 rs2965183	1977 1972 795 172
RP11-15A1.7	rs3760982	215

Table 3		
KLHDC7A	rs2992756	0.085
ALS2CR12	rs3769821 rs13393577 rs1830298	30 11075 inside the gene
CASP8	rs3769821 rs13393577	inside the gene 11144
ATG10	rs7707921	inside the gene
ATP6AP1L	rs7707921	37
L3MBTL3	rs6569648	208
RMND1	rs9383951 rs9485372 rs3757322 rs9397437 rs851984 rs9918437 rs2747652	169-2117
SNX32	rs3903072 rs75915166 rs78540526	18 3755 3707
RCCD1	rs2290203	6
STXBP4	rs6504950 rs2787486	inside the gene inside the gene
ZNF404	rs3760982	90
Table 4		
UQCRH	rs1707302	168
LMO4	rs17426269 rs12118297	342 15
DNAJC27-AS1	rs6725517 rs200648189	65 455
KLHL5	rs6815814	230
AC008391.1	rs16886113 rs16886181	242 276

	rs16886397	381
	rs2229882	416
	rs7726354	504
	rs62355902	301
HCG14	rs9257408	61
TRNAI2	rs71557345	307
	rs3757318	491
	rs2046210	525
MTHFD1L	rs9383938	564
	rs11780156	81
	rs13281615	451
PVT1	rs1562430	419
RP11-123K19.1	rs10760444	20
RP11-57H14.3	rs7904519	108
	rs2981582	336
	rs11199914	594
	rs35054928	347
RP11-500G22.2	rs45631563	339
PTDSS2	rs6597981	312
AP006621.5	rs6597981	19
PIDD1	rs6597981	inside the gene
MRPL23-AS1	rs3817198	95
PACS1	rs3903072	255
	rs10771399	221
RP11-860B13.1	rs7297051	241
KLF5	rs6562760	306
	rs2588809	64
CTD-2566J3.1	rs999737	438
C14orf79	rs10623258	240
FES	rs2290203	73
BBS2	rs2432539	80
CRNDE	rs28539243	271
RP11-482M8.1	rs4496150	441

GOSR1	rs146699004	376
ATP6V0A1	rs72826962	162
RP11-400F19.8	rs72826962	122
RP11-105N13.4	rs2532263	359
CBX8	rs745570	6
CTD-2538G9.5	rs8170 rs2363956 rs67397200	432 437 444
HOMER3	rs4808801 rs2965183	469 494
CTD-3216D2.5	rs2284378	281
TRIOBP	rs738321	396
RP5-1039K5.13	rs738321	99
CBY1	rs738321 chr22:39359355	484 289
APOBEC3A	rs738321 chr22:39359355	780 0.2
RP1-85F18.6	rs73161324 rs6001930	460 689

[#] Index SNPs identified in previous GWAS or fine-mapping studies

Supplementary Table 5. In-depth individual level association analyses of predicted expression of 41 identified genes with breast cancer risk in iCOGS and OncoArray datasets identified similar results to those obtained using summary statistics

Gene name	iCOGS dataset individual level analysis (n=84,740)		iCOGS dataset summary statistics analysis (n=89,677)		OncoArray dataset individual level analysis (n=112,133)		OncoArray dataset summary statistics analysis (n=106,776)	
	z-score[a]	*p*-value[a]	z-score[a]	*p*-value[a]	z-score[b]	*p*-value[b]	z-score[b]	*p*-value[b]
Table 1								
ZSWIM5	3.86	1.12×10^{-4}	4.32	1.57×10^{-5}	3.50	4.73×10^{-4}	2.98	0.003
LRRC3B	-4.76	1.95×10^{-6}	-4.89	1.02×10^{-6}	-7.44	1.04×10^{-13}	-7.48	7.19×10^{-14}
SPATA18	-2.02	0.04	-2.59	0.01	-2.89	3.90×10^{-3}	-3.09	0.002
KLHDC10	3.53	4.12×10^{-4}	4.39	1.16×10^{-5}	2.39	0.02	2.15	0.03
MIR31HG	-2.87	4.07×10^{-3}	-2.90	0.004	-4.99	6.11×10^{-7}	-4.35	1.35×10^{-5}
RIC8A	-3.11	1.86×10^{-3}	-3.12	0.002	-4.15	3.26×10^{-5}	-3.28	0.001
B3GNT1	-3.68	2.35×10^{-4}	-5.00	5.83×10^{-7}	-3.18	1.49×10^{-3}	-2.70	0.007
RP11-218M22.1	2.82	4.82×10^{-3}	3.33	8.82×10^{-4}	3.58	3.47×10^{-4}	3.84	1.22×10^{-4}
GALNT16	-5.07	3.93×10^{-7}	-6.17	6.82×10^{-10}	-4.70	2.62×10^{-6}	-4.45	8.74×10^{-6}
PLEKHD1	2.92	3.50×10^{-3}	3.96	7.43×10^{-5}	5.73	1.01×10^{-8}	5.21	1.85×10^{-7}
MAN2C1	-3.24	1.19×10^{-3}	-3.49	4.88×10^{-4}	-3.69	2.24×10^{-4}	-4.08	4.47×10^{-5}
CTD-2323K18.1	-2.91	3.56×10^{-3}	-2.62	0.009	-3.63	2.88×10^{-4}	-3.69	2.23×10^{-4}
Table 2								
RP11-439A17.7	-3.37	7.61×10^{-4}	-3.39	6.90×10^{-4}	-3.51	4.50×10^{-4}	-4.35	1.37×10^{-5}
ALK	3.27	1.06×10^{-3}	3.23	0.001	4.51	6.62×10^{-6}	3.84	1.23×10^{-4}
PRSS46	-3.22	1.26×10^{-3}	-3.51	4.41×10^{-4}	-5.00	5.80×10^{-7}	-4.33	1.51×10^{-5}
RP11-114I8.4	-3.22	1.28×10^{-3}	-3.15	0.002	-3.77	1.65×10^{-4}	-4.20	2.66×10^{-5}
UBLCP1	2.17	0.03	3.12	0.002	5.10	3.44×10^{-7}	4.72	2.34×10^{-6}
RP11-32D16.1	-2.68	7.31×10^{-3}	-3.66	2.51×10^{-4}	-4.63	3.63×10^{-6}	-3.75	1.75×10^{-4}
BTN3A2	1.51	0.13	2.74	0.006	3.65	2.65×10^{-4}	3.16	0.002
RP11-73O6.3	-1.62	0.11	-2.24	0.03	-5.72	1.08×10^{-8}	-5.34	9.31×10^{-8}
AP006621.6	4.29	1.82×10^{-5}	3.92	8.75×10^{-5}	3.45	5.58×10^{-4}	3.58	3.40×10^{-4}

RPLP2	2.93	3.44×10^{-3}	2.77	0.006	3.39	6.92×10^{-4}	3.43	5.93×10^{-4}
CTD-3051D23.1	-2.83	4.62×10^{-3}	-3.36	7.85×10^{-4}	-2.64	8.39×10^{-3}	-2.60	0.009
RP11-467J12.4	4.78	1.71×10^{-6}	5.41	6.28×10^{-8}	5.63	1.83×10^{-8}	5.75	8.73×10^{-9}
CTD-3032H12.1	3.79	1.50×10^{-4}	2.95	0.003	3.33	8.60×10^{-4}	2.93	0.003
LRRC37A	-3.07	2.11×10^{-3}	-3.08	0.002	-3.75	1.77×10^{-4}	-4.13	3.56×10^{-5}
KANSL1-AS1	-3.12	1.83×10^{-3}	-3.17	0.002	-3.53	4.10×10^{-4}	-3.83	1.28×10^{-4}
CRHR1	-2.67	7.59×10^{-3}	-2.81	0.005	-3.35	7.94×10^{-4}	-3.58	3.39×10^{-4}
HAPLN4	-4.73	2.26×10^{-6}	-5.10	3.46×10^{-7}	-5.87	4.44×10^{-9}	-5.49	4.01×10^{-8}
RP11-15A1.7	3.57	3.54×10^{-4}	4.26	2.00×10^{-5}	4.71	2.45×10^{-6}	3.65	2.59×10^{-4}
Table 3								
KLHDC7A	-2.87	4.06×10^{-3}	-3.53	4.11×10^{-4}	-4.51	6.54×10^{-6}	-4.69	2.77×10^{-6}
ALS2CR12	2.47	0.01	2.80	0.005	5.09	3.53×10^{-7}	4.98	6.25×10^{-7}
CASP8	-3.72	2.03×10^{-4}	-3.63	2.78×10^{-4}	-5.85	4.98×10^{-9}	-5.97	2.42×10^{-9}
ATG10	-4.55	5.28×10^{-6}	-5.83	5.60×10^{-9}	-4.04	5.44×10^{-5}	-3.00	0.003
ATP6AP1L	-3.33	8.80×10^{-4}	-4.24	2.20×10^{-5}	-3.72	2.02×10^{-4}	-2.40	0.02
L3MBTL3	-1.77	0.08	-2.38	0.02	-5.77	8.06×10^{-9}	-5.42	5.89×10^{-8}
RMND1	2.44	0.01	2.76	0.006	3.64	2.68×10^{-4}	3.14	0.002
SNX32	3.56	3.70×10^{-4}	3.80	1.45×10^{-4}	2.99	2.76×10^{-3}	2.41	0.02
RCCD1	-3.49	4.92×10^{-4}	-4.08	4.56×10^{-5}	-5.76	8.26×10^{-9}	-5.58	2.36×10^{-8}
STXBP4	3.53	4.22×10^{-4}	4.01	6.05×10^{-5}	5.26	1.42×10^{-7}	4.77	1.85×10^{-6}
ZNF404	4.76	1.91×10^{-6}	5.28	1.28×10^{-7}	5.97	2.44×10^{-9}	4.76	1.96×10^{-6}

[a] adjusted for study, the first eight principal components, and a principal component derived specifically for the study LMBC (set to zero for all other studies).
[b] adjusted for country and the first ten principal components.

Supplementary Table 6. INQUISIT scores of the identified genes showing a significant association with breast cancer risk in the TWAS ($p \leq 5.82 \times 10^{-6}$)

Gene	Distal	Promoter	GTEx eQTL
From Table 1			
ZSWIM5	none	none	
LRRC3B	none	none	
SPATA18	none	none	
UBD	0.05	none	
KLHDC10	none	none	
MIR31HG	none	none	
RIC8A	none	none	
B3GNT1	none	none	
RP11-867G23.10	none	none	
RP11-218M22.1	none	none	
GALNT16	none	none	
PLEKHD1	none	none	
MAN2C1	none	none	
CTD-2323K18.1	none	none	
From Table 2			
RP11-439A17.7	none	none	yes
NUDT17	3	none	
ANKRD34A	1	none	
ALK	none	none	
PRSS46	none	none	
RP11-114I8.4	none	none	
RP11-53O19.1	none	none	
UBLCP1	none	none	
RP11-32D16.1	none	none	
BTN3A2	none	none	yes
RP11-73O6.3	none	none	
AP006621.6	none	none	yes

RPLP2	1	none	
CTD-3051D23.1	none	none	
RP11-467J12.4	none	none	
CTD-3032H12.1	none	none	
LINC00671	none	none	
LRRC37A2	1	none	
LRRC37A	1	none	
KANSL1-AS1	3	none	
CRHR1	1	none	
HAPLN4	1	none	
RP11-15A1.7	None	none	
From Table 3			
KLHDC7A	none	3	
ALS2CR12	1	none	
CASP8	3	none	
ATG10	3	4	
ATP6AP1L	0.1	none	
L3MBTL3	2	2	
RMND1	4	none	
SNX32	2	none	
RCCD1	5	none	
STXBP4	1	none	
ZNF404	2	none	

Supplementary Table 7. Canonical pathways, diseases and bio functions, and networks associated with identified breast cancer associated genes, and highly co-expressed protein-coding genes of the identified novel susceptibility long non-coding RNAs

Gene(s)	**Top canonical pathways**	**Related diseases and disorders**	**Molecular and Cellular Functions**	**Top networks**	List of highly co-expressed genes for each long non-coding RNA
Protein-coding genes with Bonferroni corrected significant associations	Granzyme B Signaling (p=0.024); Inflammasome pathway (p=0.030); Tumoricidal Function of Hepatic Natural Killer Cells (p=0.036); Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells (p=0.046)	Cancer; Developmental Disorder; Hematological Disease; Hereditary Disorder; Immunological Disease	Cell Death and Survival; Cell-To-Cell Signaling and Interaction; Cellular Compromise; Cell Cycle; Cellular Morphology	Cell Death and Survival, Cellular Compromise, Nervous System Development and Function; Cancer, Dermatological Diseases and Conditions, Organismal Injury and Abnormalities; Cardiovascular System Development and Function, Cell Cycle, Cellular Development; Cellular Assembly and Organization, DNA Replication, Recombination, and Repair, Cell Cycle; Developmental Disorder, Hereditary Disorder, Ophthalmic Disease	NA
MIR31HG	BER pathway (p=7.56 × 10^{-3}); Dermatan Sulfate Biosynthesis (Late Stages) (p=0.026); Chondroitin Sulfate Biosynthesis (Late Stages) (p=0.028); Ephrin A Signaling (p=0.030); Heparan Sulfate	Cardiovascular Disease; Connective Tissue Disorders; Dermatological Diseases and Conditions; Developmental Disorder;	Cell-To-Cell Signaling and Interaction; Cellular Assembly and Organization; Cellular Movement; Gene Expression;	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease; Cell Cycle, Connective Tissue Disorders, Dermatological Diseases and Conditions; Cardiovascular Disease, Cellular Development, Organismal Injury and Abnormalities; Cancer,	*STMN4,ROCK1,APOL2,PRSS35,RPP38 ,RPUSD3,HS3ST6,LRR1,DIRC1, KLHL38,POLE,TREX2,CACNA1H, AC078883.4,RP5-826L7.1,MYLKP1,TSSK1A,MTHFD1P1 ,RP11-527F13.1,RP11-32B5.1, PRKCQ-AS1,RP11-834C11.3, CTD-2127H9.1,RP11-454K7.1, CTD-2561B21.10*

	Biosynthesis (Late Stages) (p=0.030)	Hereditary Disorder	Molecular Transport	Gastrointestinal Disease, Organismal Injury and Abnormalities; Hereditary Disorder, Neurological Disease, Organismal Injury and Abnormalities	
RP11-218M22.1	Netrin Signaling (p=0.024); ATM Signaling (p=0.037); Role of BRCA1 in DNA Damage Response (p=0.048)	Cancer; Dermatological Diseases and Conditions; Developmental Disorder; Hereditary Disorder; Neurological Disease	Cell Cycle; DNA Replication, Recombination, and Repair; Cell Death and Survival; Cell Morphology; Cellular Assembly and Organization	Cell Cycle, Cellular Development, Cellular Growth and Proliferation; Cancer, Cell Death and Survival, Organismal Injury and Abnormalities; Developmental Disorder, Hereditary Disorder, Organismal Injury and Abnormalities	*FFAR2,PKIB,TP53BP2,LSM14B,NSA2, SYAP1,ZNF738,MAGEF1,FOXI2,DCC, NCR1,XRCC2,BLM,RP11-94I2.1,LINC00160,AC092664.1, RP11-83M16.2,CTD-2325P2.4, CTD-3099C6.7*
CTD-2323K18.1	D-glucuronate Degradation I (p=3.31 × 10^{-3}); Methylglyoxal Degradation III (p=0.012); Mevalonate Pathway I (p=0.013);; Superpathway of Geranylgeranyldiphosphate Biosynthesis I (via Mevalonate) (p=0.018);; Tryptophan Degradation X (Mammalian, via Tryptamine) (p=0.020);	Cancer; Cardiovascular Disease; Dermatological Diseases and Conditions; Endocrine System Disorders; Hereditary Disorder	DNA Replication, Recombination, and Repair; Post-Translational Modification; Carbohydrate Metabolism; Cell Morphology; Cellular Assembly and Organization	Cellular Assembly and Organization, Hereditary Disorder, Organismal Injury and Abnormalities; Cellular Development, Cellular Growth and Proliferation, Cell Death and Survival; Cell Morphology, Cellular Function and Maintenance, Hematological System Development and Function; Cell Cycle, Cell Morphology, Organ Morphology	*VCAN,UBE2T,SUV39H1,MVK,AKR1A1 ,ZC3H13,MCM8,CASD1,CBLN2,DTL, DGKQ,RPL7A,CCDC74B,CTRC, RHEBL1,SNUPN,PKIA,KIF24,BMPR2, MUC19,LINC00612, RP11-157J24.1,LINC00035,RP11-460N20.4,FTH1P1,HNRNPA1P27, FAM203A,RP5-903G2.2, RP11-532F12.5,RP11-340F14.5, RP11-120M18.2,RP11-168F9.2*
RP11-439A17.7	Tetrahydrobiopterin Biosynthesis I (p=2.21 × 10^{-3}); Tetrahydrobiopterin	Developmental Disorder; Hereditary Disorder;	Cell Signaling; DNA Replication, Recombination, and Repair;	Cell Morphology, Gastrointestinal Disease, Organismal Injury and Abnormalities; Cellular	*TBC1D23,SPR,GNA13,TRMT5, BAIAP2L2,GATA5,GUCY2D,NIPA2, CEP170,ADAMTS16,GTF2H2C,CEND 1,IFITM1,SRRM2-AS1,*

	Biosynthesis II (p=2.21 × 10^{-3}); Relaxin Signaling (p=4.13 × 10^{-3}); Synaptic Long Term Depression (p=4.38 × 10^{-3}); Endothelin-1 Signaling (p=6.60 × 10^{-3})	Metabolic Disease; Neurological Disease; Ophthalmic Disease	Nucleic Acid Metabolism; Small Molecule Biochemistry; Cell Morphology	Development, Reproductive System Development and Function, Cell Cycle; Organ Morphology, Reproductive System Development and Function, Connective Tissue Disorders	*AC091167.3,RP11-30P6.1, RP5-956O18.3, RP11-137H2.4,COL6A4P1, RP5-836N10.1,VN1R20P, RP11-381E24.1,TET2-AS1, RP11-361I14.2,RP11-732A19.9, CTD-2329K10.1,LA16c-385E7.1,RN7SL15P,SH3GL1P2, MIR3942, RP11-820I16.1, HMGB2P1, RP11-479O17.10*
RP11-114I8.4	ErbB2-ErbB3 Signaling (p=0.043); ErbB4 Signaling (p=0.045)	Dermatological Diseases and Conditions; Developmental Disorder; Hereditary Disorder; Metabolic Disease; Organismal Injury and Abnormalities	Cellular Function and Maintenance; Molecular Transport; Cell Morphology; Gene Expression; Protein Trafficking	Cell Morphology, Cellular Compromise, Cellular Function and Maintenance; Lipid Metabolism, Small Molecule Biochemistry, Dermatological Diseases and Conditions; Hereditary Disorder, Nephrosis, Ophthalmic Disease; Molecular Transport, Cellular Assembly and Organization, Cell Morphology; Cellular Assembly and Organization, Cellular Function and Maintenance, Cell Signaling	*ARHGAP31,PEX3,DPP8,CPNE3, KDELR3,A4GNT,RIBC2,NCBP1, SLC25A26,ANKAR,CERS3,CCDC28B, PRORSD1P,NRG4,FAM26F,ZNF789, NFKBIL1,Y_RNA,RP1-13D10.2,LINC00205,AC002117.1, RP13-216E22.4,MED4-AS1, ZRANB2-AS1,AD000090.2, RP11-206L10.9,RP11-353N4.2, RP11-499P20.2, TMEM161B-AS1,AC008592.4, RP11-15A1.2,CTD-2639E6.9*
RP11-53O19.1	Inosine-5'-phosphate Biosynthesis II (p=5.44 × 10^{-3}); Retinoate Biosynthesis II (p=7.25 × 10^{-3}); Purine Nucleotides De Novo Biosynthesis II (p=0.020); Cleavage and	Dermatological Diseases and Conditions; Developmental Disorder; Hereditary Disorder; Neurological	Cell Cycle; Cell Morphology; Cellular Assembly and Organization; Cellular Function and Maintenance;	Cell Cycle, Cell-To-Cell Signaling and Interaction, Cellular Growth and Proliferation; Cell Death and Survival, Neurological Disease, Organismal Injury and Abnormalities; Cancer, Cell Death and Survival, Cell-	*MCM10,RRP15,EPN2,MTMR3,CPSF6, RBBP5,FBXO30,PAICS,CAPN11, RNF144B,HAUS2,KIF21A,FAM222A, CTDSPL,HNRNPU,TOMM70A,RIBC1, RRP1B,RBP7,FUBP1,S100Z,C17orf66, NUDT4,DSG4,MED16,OR10A6,GCNT4 ,TMEM139,ZNF320,C11orf72,CXorf38, ZNF566,ZNF197,TNFRSF18,MAGI2,*

	Polyadenylation of Pre-mRNA (p=0.022); Epithelial Adherens Junction Signaling (p=0.028)	Disease; Ophthalmic Disease	Nucleic Acid Metabolism	To-Cell Signaling and Interaction; Cardiovascular Disease, Cell Death and Survival, Cell Morphology; Embryonic Development, Organismal Development, Tissue Morphology	*VWC2,GLRA4,AC104472.1,UBE2Q2P2 ,AC073621.2,AC073850.6, FGF12-AS2,AC073342.12, RP11-557H15.4,RP11-640M9.1,RN7SL331P,AP000322.53, RP11-51J9.4,RP11-1055B8.4, RP11-135L13.4,RP11-64C12.8*
RP11-32D16.1	AMPK Signaling (p=1.96 × 10^{-4}); Tyrosine Degradation I (p=2.20 × 10^{-4}); Phenylalanine Degradation IV (Mammalian, via Side Chain) (p=1.95 × 10^{-3}); LPS/IL-1 Mediated Inhibition of RXR Function (p=3.11 × 10^{-3}); Valine Degradation I (p=3.23 × 10^{-3})	Metabolic Disease; Endocrine System Disorders; Gastrointestinal Disease; Hepatic System Disease; Organismal Injury and Abnormalities	Lipid Metabolism; Small Molecule Biochemistry; Energy Production; Molecular Transport; Carbohydrate Metabolism	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry; Cell Signaling, Nucleic Acid Metabolism, Small Molecule Biochemistry; Cell Cycle, Gene Expression, Organ Morphology; Carbohydrate Metabolism, Molecular Transport, Small Molecule Biochemistry; Skeletal and Muscular Disorders, Cell Morphology, Organ Development	*TNMD,CX3CL1,DAPK2,HEBP2, TRAF3IP2,GYG2,LIPE,SEPHS1,FMO2, APMAP,SLC6A2,FAH,ETFB,SH3D19, CPT1A,TMED5,ITGB1BP1,KCNIP2, CAT,WIPF3,ACVR1C,PCK1,ARHGEF6 ,SLC7A10,INPP5K,KL,TCN1,ADAM19, SDS,CD36,PPP1R1A,TTLL4,SLC19A3, RTN1,ETFA,ADAMTS18,SELENBP1, CDKN2B,TM7SF2,FLI1,PDE3B, PLOD2,RWDD2B,NAA11,RPUSD3, C2CD2,HPD,PFKFB1,FBXO27, CCDC51,CTSB,SUN1,NIPSNAP3B, AQP7,ZNF219,GPT2,PLIN1,ACAA2, GPD1,PLIN4,ANGPTL4,GLYCTK, GSG1L,MRAP,FABP4,PFKFB3,MUC7, AQPEP,DCTN2,FOXG1,CIDEA, PLA2G16,BOK,RPLP2,PNPLA2, COX14,ADIPOQ,ABAT,TUSC5, NAT8L,CIDEC,DHRS4L2,MAOA, MAN2A2,VKORC1L1,KCNRG,NUDT16 ,GJC2,LINC00222,PCDHA7, AC022007.5,MIR135A1,ZNF259P1, KCNIP2-AS1,AC022596.6, ADIPOQ-AS1,RP11-445L13__B.3,RP1-28O10.1,AC008738.1,VN1R108P, RP11-573D15.1,RP5-1172A22.1,*

					RP1-293L6.1,LINC00263,TRHDE-AS1,AC159540.2,VWFP1,GLYCTK-AS1,CEBPA,RP11-768F21.1, CTD-2589H19.4,RP13-884E18.2,RP11-1101K5.1,PAICSP4,AP006621.8, RP11-317P15.5,RP11-663N22.1
RP11-73O6.3	Pentose Phosphate Pathway (Oxidative Branch) (p=5.88 × 10^{-3}); Selenocysteine Biosynthesis II (Archaea and Eukaryotes) (p=8.81 × 10^{-3}); GDP-mannose Biosynthesis (p=8.81 × 10^{-3}); p53 Signaling (p=9.13 × 10^{-3}); Tryptophan Degradation to 2-amino-3-carboxymuconate Semialdehyde (p=0.010)	Cardiovascular Disease; Connective Tissue Disorders; Developmental Disorder; Hematological Disease; Hereditary Disorder	Cell Death and Survival; Carbohydrate Metabolism; Cell Cycle; Cell Morphology; Cell-To-Cell Signaling and Interaction	Cellular Development, Cellular Growth and Proliferation, Reproductive System Development and Function; Cell-mediated Immune Response, Cellular Development, Cellular Function and Maintenance	*HIVEP2,LAMA3,SEPHS1,ARHGAP28, DHX32,GGA1,EIF3D,PSMD7,GPI, URGCP,XPNPEP1,PPP1R9B,SMURF2 ,DDX25,NR5A1,TP53BP2,PIK3R1,SYB U,CDYL,NCAM2,DHRS4,G6PD,ZER1, SHANK1,HAAO,SH3TC2,PACS1, L3MBTL3,LINC00162, RP11-536C5.7,RP11-213G2.2,AC002401.1,MTX1P1, RP11-247I13.11,LINC00461, RP11-281P23.2,RP11-10A14.4,RP11-578O24.2,FTLP14,AC005702.1, RP11-138E2.1,RP11-216P16.2*
AP006621.6	Primary Immunodeficiency Signaling (p=1.40 × 10^{-3}); Acetate Conversion to Acetyl-CoA (p=5.04 × 10^{-3}); T Cell Receptor Signaling (p=6.49 × 10^{-3}); G12/13 Signaling (p=9.50 × 10^{-3}); Tec Kinase Signaling (p=0.016)	Cancer; Cardiovascular Disease; Connective Tissue Disorders; Dermatological Diseases and Conditions; Developmental Disorder	Cellular Function and Maintenance; Cell Death and Survival; Cell Morphology; Cell-To-Cell Signaling and Interaction; Cellular Development	Humoral Immune Response, Protein Synthesis, Hematological System Development and Function; Cellular Compromise, Cell Cycle, Cellular Assembly and Organization; Cell Cycle, Hereditary Disorder, Neurological Disease; Embryonic Development, Organismal Development, Tissue Development	*BTK,HERPUD1,PRDM1,KCNAB2, ZFAND6,DERL3,SRRD,ACSS3,LAX1, ZBP1,RGS13,TEC,DUOXA2,RPL11, UGCG,LPCAT1,GFI1B,RNF187, SHCBP1,AP006621.1,PIDD,NUGGC, IGHA1,FIS1,IGLC6,RPL32P1, RP11-162O12.2,LINC00568,GS1-124K5.11,LINC00582,AC007285.7, AC005162.5, RP11-181G12.4, LINC01010,RPS20P21, H2AFJ, RP11-510M2.2, RP11-1084E5.1,AP006621.5,AC091171.1, RP11-554D14.4,RP11-42I10.1,*

					RP11-61A14.3,RP11-325K4.3, RP11-174G6.5,RP11-849F2.8, RP4-713A8.1,STAG3L5P-PVRIG2P-PILRB
CTD-3051D23.1	Granulocyte Adhesion and Diapedesis (p=0.039); Agranulocyte Adhesion and Diapedesis (p=0.043); IL-22 Signaling (p=0.045); Role of JAK family kinases in IL-6-type Cytokine Signaling (p=0.046); B Cell Development (p=0.050)	Cancer; Cardiovascular Disease; Developmental Disorder; Endocrine System Disorders; Hematological Disease	Cellular Development; Cell Morphology; Cellular Growth and Proliferation; Lipid Metabolism; Molecular Transport	Cell Cycle, Cell Death and Survival, Cellular Compromise; Cellular Development, Cellular Growth and Proliferation, Hematological System Development and Function; Cellular Assembly and Organization, Cellular Function and Maintenance, Tissue Morphology; Connective Tissue Disorders, Organismal Injury and Abnormalities, Reproductive System Development and Function; Infectious Diseases, Cancer, Organismal Injury and Abnormalities	*RC3H2,TDRD3,GPATCH2L,IL5RA, EZR,TFIP11,MIB1,CD79A,EPB41L5, SLC1A4,STAT5A,CASZ1,VSTM2L, SPIRE1,KLF4,NAA30,TBX19,RNF145, AZGP1,SCIMP,PPP1R32,VANGL2, LEO1,LMAN2,SERPINA6,NAT1, NDUFA11,SOX11,TSSK6,PCED1B, CRELD2,LCN12,CCDC157,NKAPL, TMEM240,IGLV2-23, IGLV3-19,IGLC3,CCL27,HLA-F-AS1,RP11-216N14.5,AC068587.2, RP11-557J10.4,CTD-2240H23.2, KB-1460A1.5,LA16c-431H6.6, RP11-283I3.6,RP11-597M12.1, RP11-694I15.7,RP4-734G22.3*
RP11-467J12.4	Glycerol-3-phosphate Shuttle (p=2.32 × 10^{-3}); Glycerol Degradation I (p=5.78 × 10^{-3})	Cancer; Organismal Injury and Abnormalities; Reproductive System Disease; Cardiovascular Disease; Developmental Disorder	Cell-To-Cell Signaling and Interaction; Cellular Assembly and Organization; Cellular Growth and Proliferation; Cell Morphology; Cellular Development	Cardiovascular System Development and Function, Cellular Development, Cellular Function and Maintenance; Cell Morphology, Connective Tissue Development and Function, Tissue Morphology	*PREX2,CD82,MARCH2,INTS10,TPD52 L1,MORN1,HSDL2, CHST8,FGD3,IL17B,BTBD3,PRSS23, ANKFN1,ZKSCAN2,SUSD3,ALDH16A1 ,HDAC11,GPD1,EMR1,CTDNEP1, SRGAP3,SFTA2,AC074391.1, AL021068.1,RP11-553A21.3, RP4-669P10.16,LL0XNC01-237H1.3,DPY19L2P4,AC007750.5, KRT8P36,RP11-368I23.2, RP11-16P20.3,FMR1-AS1*

CTD-3032H12.1	ERK/MAPK Signaling (p=2.00 × 10^{-4}); FLT3 Signaling in Hematopoietic Progenitor Cells (p=1.14 × 10^{-3}); Acute Myeloid Leukemia Signaling (p=1.50 × 10^{-3}); -Adrenergic Signaling (p=1.92 × 10^{-3}); Corticotropin Releasing Hormone Signaling (p=3.69 × 10^{-3})	Inflammatory Response; Cancer; Organismal Injury and Abnormalities; Auditory Disease; Cardiovascular Disease	Cell-To-Cell Signaling and Interaction; Cell Death and Survival; Cell Cycle; Cell Morphology; Cellular Function and Maintenance	Embryonic Development, Organ Development, Organismal Development; Cell Morphology, Cell Death and Survival, Cellular Development; Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry; Post-Translational Modification, Cell Morphology, Cellular Function and Maintenance; Dermatological Diseases and Conditions, Organismal Injury and Abnormalities, Hair and Skin Development and Function	*MBTD1,AGA,CHI3L2,FNDC3B,GRHL2 ,CRLS1,ESR1,DHRS7,HTATSF1, DNAJB6,GATA3,ATE1,STC2,MOGS, PDCL3,SLC1A4,NR5A2,IPO13,STAT5A ,TRPV5,EPS15L1,YWHAH,RAF1, HRASLS2,GLCE,C1RL,GRTP1,IGLON 5,UCK2,TNFRSF21,RNF44,RASSF3, C17orf103,SH3RF2,PLXDC1,ADCY9, C1orf50,MOCS2,FBP1,TAF1D,DEGS2, PLA2G4F,C3orf33,IRX5,IRX3,SETD3, C1orf64,PIP5K1C,EIF4EBP1,HUS1B, ZKSCAN7,CALM1,CAPN8,SPINK9, SNORA40,LINC00277,RP11-223A3.1,ATP8A2P2,AK3P3, RP11-5P18.5,LINC00941,KIFC1, RP11-69L16.4,RP1-93H18.1,RP11-64D22.2,RGAG1,CTB-33O18.1,RP11-676M6.1,POLG2,AC040173.1,RP11-114F10.2,LINC00567,AC009133.20, RP11-106E15.1,CTD-2206N4.4, CTD-3214H19.6,CTD-3099C6.11, CTD-2376I4.2,RP11-383I23.2*
KANSL1-AS1	Endoplasmic Reticulum Stress Pathway (p=0.024); Tumoricidal Function of Hepatic Natural Killer Cells (p=0.027); Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells (p=0.035); TWEAK Signaling (p=0.038)	Developmental Disorder; Hereditary Disorder; Neurological Disease; Organismal Injury and Abnormalities; Psychological Disorders	Cell Morphology; Cellular Function and Maintenance; Lipid Metabolism; Molecular Transport; Small Molecule Biochemistry	Gene Expression, Cell Cycle, Lipid Metabolism; Cell Cycle, Reproductive System Development and Function, Embryonic Development; Developmental Disorder, Hereditary Disorder, Ophthalmic Disease; Cell Cycle, Endocrine System Development and Function, Lipid Metabolism	*BTN3A1,MRPL43,USP14,KANSL1, EMC6,NLGN4X,TMEM219,CDC42SE2, ZNF230,NEU3,CASP7,TRAPPC2L, ACYP2,LRRC37A,CECR6,FAM227A, NUDT17,EPHB4,CRHR1-IT1,ARHGAP19,LRRC37A4P,MICD, RP4-782L23.1,FAM215B,AC007386.4, LRRC37A2,DECR2,CCDC153,BANF1P 1,AC004449.6,CTD-2555A7.3, RP11-259G18.2,RP11-259G18.3, RP11-1055B8.8,DND1P1,*

					RP11-798G7.8,RP11-622C24.1
LINC00671	Dolichyl-diphosphooligosaccharide Biosynthesis (p=0.016); Hereditary Breast Cancer Signaling (p=0.017); Antiproliferative Role of TOB in T Cell Signaling (p=0.037); Inhibition of Angiogenesis by TSP1 (p=0.046)	Inflammatory Response; Cancer; Cardiovascular Disease; Developmental Disorder; Gastrointestinal Disease	DNA Replication, Recombination, and Repair; Cell-To-Cell Signaling and Interaction; Cellular Function and Maintenance; Cell Cycle; Cellular Development	Cell Death and Survival, Cancer, Organismal Injury and Abnormalities; Cell Morphology, Developmental Disorder, Digestive System Development and Function; Cellular Movement, Nervous System Development and Function, Embryonic Development; Cell Morphology, Cellular Function and Maintenance, Cellular Movement; Organ Morphology, Organismal Development, Organismal Injury and Abnormalities	*RHBDF1,NCAPD2,UBE2D1,PALB2, PHF14,EHD1,AMOTL2,CHST10, TXLNG2P,KDELC1,TUFT1,TRIM50, CABYR,N6AMT1,SIGLEC11,C2orf44, TGFBR2,MAP9,ARSK,WEE1,TEF, DPAGT1,C11orf68,HIC1,LINC00670, CEP97,TNFAIP2,KTN1-AS1,ZNF567,Y_RNA,LINC00449, LINC00963,SYNJ2BP,RPL39P5, LRRC37A11P,U3,SMIM13,RP11-61N20.3,RP11-222A11.1,RN7SL165P,RN7SL244P, RP11-463J10.3, RP11-407G23.4,AOC4P, RP11-2I17.4,LINC00565, RP11-703I16.1,MIR24-2,CLEC4GP1*
RP11-15A1.7	Induction of Apoptosis by HIV1 (p=1.09 × 10^{-4}); Docosahexaenoic Acid (DHA) Signaling (p=1.72 × 10^{-3}); Molecular Mechanisms of Cancer (p=2.33 × 10^{-3}); CD27 Signaling in Lymphocytes (p=2.93 × 10^{-3}); Small Cell Lung Cancer Signaling (p=5.60 × 10^{-3})	Infectious Diseases; Cancer; Cardiovascular Disease; Dermatological Diseases and Conditions; Developmental Disorder	Cellular Compromise; Cellular Assembly and Organization; Cell Morphology; Cell Death and Survival; Cell-To-Cell Signaling and Interaction	Cell Morphology, Cellular Assembly and Organization, Behavior; Cell Morphology, Cellular Function and Maintenance, Cellular Compromise; Cell Signaling, Nucleic Acid Metabolism, Molecular Transport; Cell Death and Survival, Cellular Development, Cellular Growth and Proliferation	*SLC25A13,BID,WAPAL,PBX4,RASD1, KLF9,ADCY7,ZNF211,TMEM115, CDKN2D,TULP4,WTAP,ZNF7, ANXA8L1,PAN3,ZNF761,UVSSA, INTS1,FAM102A,BCL2L1,C19orf18, PER1,CDC42EP4,HKR1,GPR1, ANKRD19P,ZNF34,DNM3,ZFP2, ZNF155,FAM71F2,AKR1C7P,RAB1C, RP11-538P18.2,ZBED5-AS1, RP11-301H24.3,RP11-420A23.1,RP11-521B24.5,CHMP4BP1,NT5CP1,RP11-109N23.6,SMIM6,HCCAT3, RP11-727F15.13,RP11-130L8.2, RP11-274B21.9,RP5-1024N4.4, RP11-434H6.7*

NA: not available

Supplementary Table 8. Cell line and media information.

Cell Line	Media constituents
MCF10A	DMEM/F12 + 5% Horse Serum + 20ng/mL EGF + 0.5µg/mL Hydrocortisone + 100ng/mL Cholera Toxin + 10 µg/mL Insulin from bovine pancreas + 1% Penicillin-Streptomycin
Bre80-Tert	DMEM/F12 + 5% Horse Serum + 20ng/mL EGF + 0.5µg/mL Hydrocortisone + 100ng/mL Cholera Toxin + 10 µg/mL Insulin from bovine pancreas + 1% Penicillin-Streptomycin
184A1	MEGM + BPE 52ug/mL + HC 500ng/mL + EGF 10ng/ml + I 5ug/ml + transferrin 5ug/mL + cholera toxin 1ng/mL
ZR751	RPMI-1640 + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin + 10 µg/mL Insulin from bovine pancreas
MCF7	RPMI-1640 + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
KPL1	DMEM + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
T47D	RPMI-1640 + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
SKBR3	DMEM + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
BT474	RPMI-1640 + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
MDA-MB-453	DMEM/F12 + 20% Fetal Bovine Serum + 1% Penicillin-Streptomycin + 10 µg/mL Insulin from bovine pancreas
MDA-MB-231	DMEM + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
MDA-MB-436	DMEM + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin

BT549	RPMI-1640 + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
MDA-MB-157	DMEM + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
HCC1937	RPMI-1640 + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
HS578T	DMEM + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin
SUM159PT	RPMI-1640 + 10% Fetal Bovine Serum + 1%Penicillin-Streptomycin
MDA-MB-468	DMEM + 10% Fetal Bovine Serum + 1% Penicillin-Streptomycin

Supplementary Table 9. siRNA sequences.

Name	Sense sequence (5'-3')	Antisense sequence (5'-3')
RMND1-1	CCACGGAUAUGUUGAAGUATT	UACUUCAACAUAUCCGUGGGA
RMND1-2	CAAACCAAAUCUGUUGGGUUCUAAA	UUUAGAACCCAACAGAUUUGGUUUG
KLHDC10-1	CAACCUAUAUGUGUUUGGAGGUUAU	AUAACCUCCAAACACAUAUAGGUUG
KLHDC10-2	GAGAUAUCUGGAAGUUGAAUCUGCA	UGCAGAUUCAACUUCCAGAUAUCUC
ZSWIM5-1	GGGAAAGUGAAAGACUACUCUUUAA	UUAAAGAGUAGUCUUUCACUUUCCC
ZSWIM5-2	CCUCAUUGGCCAUGAGCCAUCUUAA	UUAAGAUGGCUCAUGGCCAAUGAGG
UBLCP1-1	GCACCUAAAUCGUGAUAAATT	UUUAUCACGAUUUAGGUGCGC
UBLCP1-2	CAGGAGUAUUCAGUGACCACACUUU	AAAGUGUGGUCACUGAAUACUCCUG
PLEKHD1-1	UCAAAGAGAGCUUUCUGCUUUACUA	UAGUAAAGCAGAAAGCUCUCUUUGA
PLEKHD1-2	AAGAUGCCUUAAGGGUGUAGAACA	UGUUCUACACCCUUAAGGCAUCUUG
ALS2CR12-1	AACUCCACAGGGAGUUCCAAGCUAA	UUAGCUUGGAACUCCCUGUGGAGUU
ALS2CR12-2	CAGCAAGGCAAGAAGAGACUAAUAA	UUAUUAGUCUCUUCUUGCCUUGCUG
STXBP4-1	(CCUGGAGGAGACUGUUAUA)dTdT	(UAUAACAGUCUCCUCCAGG)dAdA
STXBP4-2	(GGACCUCAAGCCUCAACAU)dTdT	(AUGUUGAGGCUUGAGGUCC)dAdT
ZNF404-1	UGCGUACCAUCAGGAGACAUGGAAA	UUUCCAUGUCUCCUGAUGGUACGCA
ZNF404-2	GGGAAACGUUUAGAUUAUAUCGACA	UGUCGAUAUAAUCUAAACGUUUCCC
PIDD-1	GACUGUUCCUGACCUCAGAtt	UCUGAGGUCAGGAACAGUCtg
PIDD-2	AGGGCAGAAUCUGCUUUGUCUUCUA	UAGAAGACAAAGCAGAUUCUGCCCU
NRBF2-1	UGUGAAAUGCGCUGCGUAUUU	AUACGCAGCGCAUUUCACAUU
NRBF2-2	CCGGAGGAGGAAGUGGUGAGGUUGU	ACAACCUCACCACUUCCUCCUCCGG
NRBF2-3	AGGAAGUGGUGAGGUUGUUGCUCCU	AGGAGCAACAACCUCACCACUUCCU
ABHD8-1	GAGCAAUCUUCAAGCGCUAUGCCAA	UUGGCAUAGCGCUUGAAGAUUGCUC
ABHD8-2	CAUUCCUACGGUGUCUCUUUCUGCA	UGCAGAAAGAGACACCGUAGGAAUG
RP11-218M22-R1-1	UGAGCGCAGGAACCAUGGUCUUCAU	AUGAAGACCAUGGUUCCUGCGCUCA
RP11-218M22-R1-2	CGCAGGAACCAUGGUCUUCAUUGCU	AGCAAUGAAGACCAUGGUUCCUGCG
RP11-218M22-R2-1	CCAGUGGGUUUGGAUAUAAUCCUGA	UCAGGAUUAUAUCCAAACCCACUGG
RP11-218M22-R2-2	CAGACUGCGAGACAAUCUCUCUUUA	UAAAGAGAGAUUGUCUCGCAGUCUG
AP006621.6-1	GGGUACCUUCACCUGGGCGUCAGAA	UUCUGACGCCCAGGUGAAGGUACCC

AP006621.6-2	UCACCUGGGCGUCAGAAGCACUUGA	UCAAGUGCUUCUGACGCCCAGGUGA
RP11-467J12.4-1	CACCAUAUCAUGGUUCCCACUAGCA	UGCUAGUGGGAACCAUGAUAUGGUG
RP11-467J12.4-2	UAUGAGAGUUCCAGUUGCUCCACAA	UUGUGGAGCAACUGGAACUCUCAUA
RP11-15A1.7-1	CACCCUCCUCAUACUUCCGUAGUUU	AAACUACGGAAGUAUGAGGAGGGUG
RP11-15A1.7-2	GGAAUCCACCUAAGUGUCUAUCAAU	AUUGAUAGACACUUAGGUGGAUUCC
CTD-3032H12.1-1	CAAGCUCCCGAGGCGAUCUGCUGUU	AACAGCAGAUCGCCUCGGGAGCUUG
CTD-3032H12.1-2	AGGCCCAAGUCGCAGUUCUCGUGAA	UUCACGAGAACUGCGACUUGGGCCU
B2M-1	CCAGCGUACUCCAAAGAUUTT	AAUCUUUGGAGUACGCUGGTT
B2M-2	GGTTTACTCACGTCATCCATT	TGGATGACGTGAGTAAACCTT
ARHGDIA-1	CCCGUCUAACCAUGAUGCCUUAACA	UGUUAAGGCAUCAUGGUUAGACGGG
ARHGDIA-2	CCUUAACAUGUGGAGUGUACCGUGG	CCACGGUACACUCCACAUGUUAAGG
ZAP70-1	UAACCUCCUCAUAGCUGACAUUGAA	UUCAAUGUCAGCUAUGAGGAGGUUA
ZAP70-2	CCGAAUGCAUCAACUUCCGCAAGUU	AACUUGCGGAAGUUGAUGCAUUCGG

Supplementary Table 10. Literature reported link between genes identified in our study that have not been reported from eQTL and/or following functional studies as target genes of risk variants (Tables 1-2) and breast cancer

Gene	Reported link with breast cancer	Study type	Consistency with the direction of effect identified in our study	PMID of literature
Table 1				
ZSWIM5	NA	NA	NA	NA
	inhibits bupivacaine-induced breast cancer cell invasion	*In vitro*	consistent	29085514
	reduced expression in breast cancer tissues compared with breast fibroma tissues; low gene expression associated with higher tissue grade	Human tissues		24839112
LRRC3B	methylated and/or deleted in ~32% breast carcinoma samples	Human tissues		22321817
SPATA18	downregulated ~ 5-folds in human ductal breast carcinomas compared with normal breast samples	Human tissues	consistent	21300779; 16473279
	inhibits growth in MCF-7 breast carcinoma cells	*In vitro*	consistent	12170760
UBD	increased expression in breast cancer tissues compared with surrounding tissues; expression correlated with triple-negative breast cancer (TNBC)	Human tissues	inconsistent	26185453
KLHDC10	NA	NA	NA	NA
	down-regulated in TNBC cell lines of basal subtype; heavily methylated in the TNBC cell lines	*In vitro*	consistent	22289355
MIR31HG	increased expression in breast cancer tissues compared with normal control; expression associated with advanced pathologic stage and tumor size; knockdown decreases breast cancer cell proliferation, induces apoptosis, inhibits migration/invasion and impedes tumorigenesis	*In vitro*, *in vivo*, and human tissues	inconsistent	24631686

RIC8A	undergoes a classical double-hit genetic inactivation in a breast cancer cell line; loss of expression in a subgroup of aggressive *TP53* mutant breast cancers	*In vitro* and *in vivo*	consistent	19432969
B3GNT1	NA	NA	NA	NA
RP11-867G23.10	NA	NA	NA	NA
RP11-218M22.1	NA	NA	NA	NA
GALNT16	NA	NA	NA	NA
PLEKHD1	NA	NA	NA	NA
MAN2C1	NA	NA	NA	NA
CTD-2323K18.1	NA	NA	NA	NA
Table 2				
RP11-439A17.7	NA	NA	NA	NA
NUDT17	NA	NA	NA	NA
ANKRD34A	NA	NA	NA	NA
ALK	overexpressed in 36% of breast cancer patients; gene amplification present in 13.3 % of cases; overexpression associated with aggressive behavior	Human tissues	consistent	26384210
	amplified in a large proportion of Inflammatory Breast Cancers (IBC), a highly aggressive subtype of breast cancer	*In vitro* and human tissues		22215853
	copy number gain observed in 47.2% of IBC patients; copy number gain associated with poorer recurrence free survival	Human tissues		25803816
PRSS46	NA	NA	NA	NA
RP11-114I8.4	NA	NA	NA	NA
RP11-53O19.1	NA	NA	NA	NA
UBLCP1	NA	NA	NA	NA
RP11-32D16.1	NA	NA	NA	NA
BTN3A2	higher expression associated with improved distant metastasis-free survival in HR-/HER2+ breast cancer	Human tissues	NA	28409241

RP11-73O6.3	NA	NA	NA	NA
AP006621.6	NA	NA	NA	NA
RPLP2	differentially expressed for breast cancer apoptosis (both up- and down-regulation)	*In vitro*	NA	22133146
CTD-3051D23.1	NA	NA	NA	NA
RP11-467J12.4	NA	NA	NA	NA
CTD-3032H12.1	NA	NA	NA	NA
LINC00671	NA	NA	NA	NA
LRRC37A2	NA	NA	NA	NA
LRRC37A	NA	NA	NA	NA
KANSL1-AS1	NA	NA	NA	NA
CRHR1	encodes a receptor of corticotropin-releasing hormone (CRH), which suppresses TGFβ1-induced Epithelial-Mesenchymal Transition in breast cancer cells	*In vitro*	consistent	24412750 26138318
HAPLN4	NA	NA	NA	NA
RP11-15A1.7	NA	NA	NA	NA

NA: not available

Supplementary Table 11. Performance of prediction models and association results for breast cancer target genes reported previously in GWAS-identified loci

Chromosome regions	Target genes	Reference	Evidence from original paper for supporting this gene as the target gene	Performance of expression prediction model (R^2) in GTEx/TCGA	Association of predicted expression with breast cancer risk
1p33	*NSUN4*	[1]	eQTL analyses in GTEx, TCGA (tumor tissue) and METABRIC (tumor adjacent normal tissue), prediction by ChIA-PET in MCF7 cells	0.01/0.006	$p=1.95 \times 10^{-4}$ (z: negative)
1p36.22	*PEX14*	[2]	eQTL analyses in TCGA (tumor and adjacent normal tissue)	0.02/0	$p=0.002$ (z: positive)
2p23.2	*TRMT61B*	[3]	eQTL analyses in TCGA (tumor tissue) and Norwegian normal breast cohort (normal tissue)	0.23/0.33	$p=0.30$
2q33	*PPIL3*, *CASP8*	[3,4]	eQTL analyses in TCGA (tumor tissue); eQTL analyses in TCGA (tumor adjacent normal tissue) and Westra *et al.* (peripheral blood samples)	0.44/0.59, 0.22/0.30	$p=0.02$ (z: positive), $p=8.51 \times 10^{-16}$ (z: negative)
2q35	*IGFBP5*	[5]	eQTL analyses in the Norwegian Breast Cancer Study and METABRIC (tumor adjacent normal tissue) (marginal significant associations with levels of one of the tested probes, but not any others)	0.04/0.004	NA
4q24	*TET2*	[6]	eQTL analyses in TCGA (tumor tissue) and METABRIC (tumor adjacent normal tissue)	0.007/0.02	$p=0.08$
5p12	*FGF10, MRPS30*	[7]	eQTL analyses in GTEx (normal tissue) and Norwegian Breast Cancer Study (tumor and tumor adjacent normal tissue); eQTL analyses in GTEx (normal tissue), and Norwegian Breast Cancer Study and TCGA (both tumor and tumor adjacent normal tissue)	0.02/0, 0.006/0.16	$p=0.26$, $p=1.43 \times 10^{-25}$ (z: positive)
5p15.33	*TERT*	[8]	luciferase reporter assays	NA	NA

5q11.2	*MAP3K1*	[9-11]	Chromosome Conformation Capture and luciferase reporter assays etc, while, no detectable differences in expression were found across genotypes of the index SNPs	0.06/0	p=0.32
5q14	*ATP6AP1L*	[12]	eQTL analyses in TCGA (tumor tissue)	0.63/0.32	$p=6.32 \times 10^{-7}$ (z: negative)
6p24.3	*GCNT2*	[13]	eQTL analyses in TCGA (tumor tissue)	NA	NA
6q25	*ESR1, RMND1, CCDC170, AKAP12*	[14,15]	eQTL analyses in TCGA (tumor tissue) and METABRIC (tumor and tumor adjacent normal tissue); eQTL analyses in TCGA (tumor tissue); eQTL analyses in TCGA (tumor tissue) and GTEx (normal tissue); eQTL analyses in TCGA (tumor tissue)	NA, 0.13/0.02, 0.02/NA, NA	NA, $p=1.95 \times 10^{-6}$ (z: positive), p=0.002 (z: negative), NA
7q35	*OR2A7*	[10]	eQTL analyses in TCGA (tumor tissue)	0.23/0.12	p=0.34
8q24	*POU5F1B, PVT1*	[16]	eQTL analyses in TCGA (tumor tissue)	NA, 0.03/0.01	NA, $p=1.12 \times 10^{-4}$(z: positive)
9q31.2	*KLF4*	[11,17]	eQTL analyses in TCGA (tumor tissue)	0.02/0	p=0.007 (z: positive)
10q21.2	*NRBF2*	[18]	eQTL analyses in Normal breast I (normal tissue) and Breast carcinomas I (tumor tissue)	NA	NA
10q26.13	*FGFR2*	[19]	prediction by ChIA-PET in MCF7 cells, while no association in eQTL analyses in METABRIC (tumor tissue)	0.13/0.02	p=0.73
11p15.5	*TH*	[10]	eQTL analyses in TCGA (tumor tissue)	NA	NA
11q13.1	*AP5B1*	[10]	eQTL analyses in TCGA (tumor tissue)	NA	NA
11q13.3	*CCND1*	[20]	eQTL analyses in the Helsinki Breast Cancer Study (tumor tissue) suggests borderline association for one SNP rs554219 in a recessive model; while there was no linear trend, and no signal detected in analyses of 40 normal breast tissue samples or TCGA tumor samples	NA	NA
15q26.1	*RCCD1*	[21]	eQTL analyses in TCGA (tumor and adjacent normal tissue)	0.13/0.07	$p=3.33 \times 10^{-13}$ (z: negative)
16q12.1	*TOX3*	[10,11]	eQTL analyses in TCGA (tumor tissue)	$0.02/4.27 \times 10^{-5}$	p=0.09

16q13	*AMFR*	[1]	eQTL analyses in METABRIC (tumor adjacent normal tissue); prediction by ChIA-PET in MCF7 cells	NA	NA
16q23.2	*DYNLRB2*	[10]	eQTL analyses in TCGA (tumor tissue)	NA	NA
17q22	*STXBP4*	[22]	Index SNP associated with differential transcript expression in TCGA (tumor tissue)	0.03/0.01	$p=2.21 \times 10^{-11}$ (z: positive)
19p13	*LRRC25, ABHD8*	[10,23,24]	eQTL analyses in TCGA (tumor tissue); eQTL analyses in normal breast tissue	5.36×10^{-6}/0, NA	$p=0.65$, NA
19q13.31	*ZNF404, ZNF155*	[2,10]	eQTL analyses in TCGA (tumor tissue); eQTL analyses in TCGA (tumor tissue)	0.15/0.21, 0.13/0.19	$p=1.15 \times 10^{-13}$ (z: positive), $p=0.03$ (z: positive)
21q22.12	*KCNE1, RUNX1, RCAN1*	[25]	eQTL analyses in TCGA (tumor tissue); eQTL analyses in METABRIC (tumor tissue); eQTL analyses in METABRIC (tumor tissue)	0.08/0.06, 0.04/0, NA	$p=0.65$, $p=0.76$, NA

NA, not applicable

Supplementary Table 12. Primer sequences.

Name	Sequence 5'-> 3'
GUSB Fwd	GAAAATATGTGGTTGGAGAGCTCATT
GUSB Rev	CGAGTGAAGATCCCCTTTTTA
PUM1 Fwd	AATGCAGGCGCGAGAAAT
PUM1 Rev	TTGTGCAGCTGAGGAACTAATGA
RPLP0 Fwd	CCATTGAAATCCTGAGTGATGTG
RPLP0 Rev	CTTCGCTGGCTCCCACTTT
ZSWIM5_H_FWD1	AAGACGGTGGCGGAAAAGTG
ZSWIM5_H_REV1	GAAGGACCAGTAGACGATGCG
ZSWIM5_H_FWD2	AGTCGGCTTTCATCTGAGTGG
ZSWIM5_H_REV2	AGGAAGACGCAATTTGACTTGG
ZSWIM5_H_FWD3	CTATCTCCGAAACCCTTTTCCAG
ZSWIM5_H_REV3	TGTGGTGTGCCGTGATTAAATA
KLHDC10_H_FWD1	CTCAACCGCTTCGTGCAAC
KLHDC10_H_REV1	CCTAACTGGGTCCCATCGTATTT
KLHDC10_H_FWD2	TACGATGGGACCCAGTTAGGA
KLHDC10_H_REV2	TGTGGCCTCTCAAAAACCTGT
KLHDC10_H_FWD3	GCACGAAGTGGACATCGTTG
KLHDC10_H_REV3	CCTCCCGATTCATCATAATCTGG
UBLCP1_H_FWD1	GTGGACAGGAGTATTCAGTGACC
UBLCP1_H_REV1	CAAGTAACTTTTGGCGTTCTGG
UBLCP1_H_FWD2	CTCGCAGAGTGAAAGAGTACAAA
UBLCP1_H_REV2	GCACAAGACCTGTGGTCAAATA
PLEKHD1_H_FWD1	TCCCGGCGGTTTTTCATCATC
PLEKHD1_H_REV1	CCACTGGGTCTGCTCAAACT
PLEKHD1_H_FWD2	GGAAGAGACCGAAGAACTCTGC
PLEKHD1_H_REV2	TGCAAGGACTCCGTGAGGT
ALS2CR12_H_FWD1	ACTTGGGACCACGGAAGCTA
ALS2CR12_H_REV1	GGAGCTGGTACAAGAGGAGTTA

ALS2CR12_H_FWD2	ATGCACAAGCCCTTATCCTAGA
ALS2CR12_H_REV2	AGAGGCCAATCTCCCAGAACA
RMND1_H_FWD1	CAGTGCCGAAGAATCGGTCAT
RMND1_H_REV1	CGAGCAGCATTTAATGGAGACA
RMND1_H_FWD2	GCACACCTTCCAACCATGAAA
RMND1_H_REV2	TGGATGCTTTTAGTGGTCTCTTC
RMND1_H_FWD3	GAGACCACTAAAAGCATCCAGG
RMND1_H_REV3	GCAGTGCATTAGGTCCTCGT
STXBP4_H_FWD1	CCTTGGCCTGAAGGTACTAGG
STXBP4_H_REV1	AGCAGATTCTAACCTCAACTTGG
STXBP4_H_FWD2	GAATCTGCTTGGGAGATAGCATT
STXBP4_H_REV2	TGAGGCTTGAGGTCCATATTCT
STXBP4_H_FWD3	ATCCCTCTGTTCGCTTTAAGGC
STXBP4_H_REV3	TCAGGGCTTGGTGTTGTTCC
ZNF404_H_FWD1	AAGTAAATGCGTACCATCAGGAG
ZNF404_H_REV1	TCCCACTTTAGGTCTCTGTTGT
ZNF404_H_FWD2	GGCCTTTGTTCGCAGCTATCT
ZNF404_H_REV2	AGGCTTGAGCCCTTACCAAAA
ZNF404_H_FWD3	GGCCTTTTGTAGAGGCTCTCA
ZNF404_H_REV3	AAGGTCTCCAACACGACTGAA
PIDD1_H_FWD1	TCAGAGGATTCGGACGCAG
PIDD1_H_REV1	GTGAGTGCTCAGACGCAAGAA
PIDD1_H_FWD2	GAGCCTCGTCGAGTCTCCAT
PIDD1_H_REV2	GGCCCAGTACAACAGGTGC
PIDD1_H_FWD3	CTCACCCACCTGTACGCAC
PIDD1_H_REV3	CAGAGCGATGAGGTTCACAC
NRBF2_H_FWD1	CAGACGAGCAGACCGTTTATT
NRBF2_H_REV1	TGCTGGGCTTTCAATCTTTCTT
ABHD8_H_FWD1	GGGGTGACCGACGGTATCT
ABHD8_H_REV1	GGCTTGACCTCTACAAAGGTG
ABHD8_H_FWD2	TCGAGCCGACCTCCTACAC
ABHD8_H_REV2	TTTGCAGCTAGTGATGCGCTT

ABHD8_H_FWD3	CTGAGGACATGCGAGCAATCT
ABHD8_H_REV3	GAAAGAGACACCGTAGGAATGG
RP11-218M22.1_H_R1_FWD1	CGGGAAAAGATGGAGTGAAGGT
RP11-218M22.1_H_R1_REV1	GGCACTTCCGCTAATGCTG
RP11-218M22.1_H_R1_FWD2	TGAGCCGGGAAAAGATGGAGT
RP11-218M22.1_H_R1_REV2	GCACTTCCGCTAATGCTGAGG
RP11-218M22.1_H_R2_FWD1	CACTGAGAGAAGCAGGAGAATGT
RP11-218M22.1_H_R2_REV1	AAGAGAGATTGTCTCGCAGTC
RP11-218M22.1_H_R2_FWD2	ACTGAGAGAAGCAGGAGAATGT
RP11-218M22.1_H_R2_REV2	AAAGAGAGATTGTCTCGCAGTC
AP006621.6_H_FWD1	TCCTGAGGGCCGACTCTAC
AP006621.6_H_REV1	CGTCTTAGCGGCTGTCACTT
AP006621.6_H_FWD2	ACTGAGAGAAGCAGGAGAATGTT
AP006621.6_H_REV2	CACTAAAGAGAGATTGTCTCGCA
RP11-467J12.4_H_FWD1	GGGGTGGTGGGTGTCACTAA
RP11-467J12.4_H_REV1	ATTCACCTTCACCAGGGCAC
RP11-467J12.4_H_FWD2	TCACTAAAAGGAACCAGCCCC
RP11-467J12.4_H_REV2	CTCTGACTGATTCACCTTCACCA
RP11-15A1.7_H_FWD1	CAGAGTGTGTCTGGACTCCG
RP11-15A1.7_H_REV1	CCAGGCGCTCAGAGATATGG
RP11-15A1.7_H_FWD2	GCGACTCAGAGTGTGTCTGG
RP11-15A1.7_H_REV2	ATGGAATACGTTCCCGGTGG
CTD-3032H12.1_H_FWD1	CCTACACGAGGCCAGAGATCC
CTD-3032H12.1_H_REV1	CCTAACAGCAGATCGCCTCG
CTD-3032H12.1_H_FWD2	GCCCGTGGCCTACACGAG
CTD-3032H12.1_H_REV2	CGGGTCTTCCTTTGTGTCCAG
B2M-FWD-1	GAGGCTATCCAGCGTACTCCA
B2M-REV-1	CGGCAGGCATACTCATCTTTT
B2M-FWD-2	CTCACGTCATCCAGCAGAGA
B2M-REV-2	CGGCAGGCATACTCATCTTT
B2M-FWD-3	AGGCTATCCAGCGTACTCCA
B2M-REV-3	CGGCAGGCATACTCATCTTT

ARHGDIA-FWD-1	GGATGAGCACTCGGTCAACTA
ARHGDIA-REV-1	GGCCTCCTTGTACTTTCGCAG
ARHGDIA-FWD-2	GAGCCTGCGAAAGTACAAGG
ARHGDIA-REV-2	TCCTTCAGCACAAACGACTG
ARHGDIA-FWD-2	TGCCTCTGCCTTTTCTGTCT
ARHGDIA-REV-3	GCACTTGGTCCCTTGTTTGT
ZAP70-FWD-1	CGAGCGTGTATGAGAGCCC
ZAP70-REV-1	ATGAGGAGGTTATCGCGCTTC
ZAP70-FWD-2	ACGCCAAGATCAGCGACTTT
ZAP70-REV-2	GGGTGCGTACCACTTGAGC
ZAP70-FWD-3	CTGGAGCTATGGGGTCACCA
ZAP70-REV-3	CAGGCTGTAGTAACAGGCTCG

Supplementary Table 13. Predicting variants in gene expression prediction models for the identified associated genes after Bonferroni correction

RP11-439A17.7	rs17023394, rs12037207, rs838518, rs17023457, rs838522, rs699774, rs838532, rs838530, rs838528, rs3820032, rs3753264, rs3753263, rs3753262, rs2185556, rs10754396, rs12405488, rs1417610, rs1417609, rs12024495, rs2050892, rs10923824, rs4659178, rs10802098, rs10923836, rs2275609, rs7547046, rs3949342, rs947130, rs4659182, rs7553527, rs6692504, rs4659200, rs346670, rs2024838, rs10754414, rs12046880, rs10802122, rs4659221, rs10494228, rs347910, rs838990, rs404937, rs380155, rs598100, rs12025390, rs4659226, rs663807, rs616111, rs539304, rs539426, rs17258425, rs4391705, rs532208, rs17186233, rs753424, rs10923902, rs947269, rs3009197, rs3009196, rs2994815, rs2994816, rs2994817, rs3009182, rs3009184, rs3009186, rs2487573, rs835578, rs4659245, rs4659246, rs4659247, rs2843021, rs12075536, rs947273, rs699780, rs835574, rs835573, rs2793830, rs6688004, rs4659248, rs2453056, rs5025718, rs2493420, rs1493696, rs2493411, rs2453044, rs10903159, rs4844381, rs12145080, rs11249348, rs6600671, rs2319969, rs2319971, rs11249431, rs2222371
ZSWIM5	rs6690437, rs12754891, rs12091565, rs11210998, rs2120823, rs17386059, rs12732315, rs16832024, rs12744658, rs12749754, rs1889759, rs6688710, rs7525308, rs6429550, rs7517439, rs7517639, rs6692487, rs11579580, rs12733586, rs7528461, rs12139143, rs6703452, rs7531019, rs11211053, rs11577974, rs12735637, rs12125367, rs12755554, rs11211059, rs7519454, rs263997, rs263992, rs263991, rs263989, rs183809, rs11556200, rs264025, rs264022, rs12749130, rs12738542, rs7553658, rs6692713, rs4399199, rs12126314, rs12126318, rs2202152, rs11579411, rs10789463, rs6429566, rs12743512, rs937291, rs10789465, rs2275276, rs11580609, rs7903, rs11211129, rs644915, rs512026, rs518365, rs12141928, rs6696085, rs12406217, rs12141269, rs17102087, rs12137934, rs12146051, rs12142240
KLHDC7A	rs4920399, rs11203247, rs17435018, rs7517220, rs6665151, rs11261017, rs11261020, rs4920322, rs4920323, rs11261021, rs2992745, rs3000058, rs2816030, rs2230705, rs6683394
ALK	rs4665406, rs7576048, rs13029274, rs12995493, rs10190267, rs2940806, rs12052472, rs1992810, rs2276551, rs2276549, rs4666201, rs2293564, rs12997783, rs4666202, rs7561975, rs6731724, rs12993746, rs12997218, rs11897665, rs7564775, rs7562088, rs6753532, rs4414641, rs4665485, rs11127243, rs13010777, rs12476676, rs12465499, rs4233750, rs12478888, rs12478928, rs6547981, rs7603844, rs7560160, rs4502372, rs5018731, rs11690664, rs7576793, rs12714298, rs829602, rs2253121, rs1474194, rs17324662, rs7594598, rs7597567, rs13388219, rs13385578, rs1862960
CASP8	rs6728002, rs3754935, rs7603014, rs6735656, rs6754084, rs1861270, rs2293554, rs10931936, rs1035142, rs700635, rs6743068, rs13016963, rs9288316, rs7560328, rs2597900
ALS2CR12	rs1035142, rs13016963, rs9288316, rs7560328

PRSS46	rs11716028, rs6808473, rs7632165, rs11717357, rs9843503, rs4413345, rs12495098, rs10510751, rs3918357, rs743661, rs916092, rs1520484, rs9819159, rs9836993, rs9880885, rs9829227, rs9810013, rs3796367, rs3796369, rs3796370, rs9834713, rs9835025, rs11915788, rs9820361, rs9820372, rs9820785, rs9820845, rs9820861, rs9821418, rs9841203, rs9841229, rs9864097, rs9868357, rs7632176, rs7623501, rs7626129, rs7428736, rs7428787, rs7639979, rs12489663, rs12496832, rs6793235, rs6787229, rs6784957, rs376737976, rs1014229, rs11714840
LRRC3B	rs2052760, rs11711082, rs11719046, rs11715434, rs10510576, rs11719770, rs11719901, rs17018100, rs994169, rs973603, rs11712421, rs17018155, rs1158545, rs1158544, rs12633309, rs17018167, rs2036430, rs2036428, rs9820211, rs1602349, rs4435583, rs1907178, rs6808839, rs1488240, rs9845198, rs6794554, rs6551142, rs7646852, rs1488215, rs17018761, rs1386884, rs11717214, rs1915915, rs9841537, rs12495557, rs1522140, rs7618127, rs13080907, rs13098500, rs6765909, rs4586769, rs10510594, rs1603051, rs7611368, rs11707188, rs3892373
NUDT17	rs9286836, rs11587364, rs12402787, rs12732381, rs2040085, rs2040086, rs34695381
ANKRD34A	rs704985
RP11-114I8.4	rs12639465, rs6441308, rs1287283, rs7616988, rs9873709, rs10511177, rs9875640, rs6807176, rs6799379, rs6806178, rs9815439, rs1021341, rs9814359, rs6790535
SPATA18	rs7440594, rs10012938, rs10434448, rs17577020, rs4864836, rs6856794, rs11724730, rs225160, rs225163, rs13989, rs225165, rs225170, rs419792, rs17612170, rs11947242, rs4470701, rs7665551, rs730284, rs4865271, rs12510605, rs9683559, rs7692441, rs7690931, rs1501614, rs1841263, rs6858566, rs10517269, rs11939448, rs11938159, rs6838718, rs7693568, rs6835977, rs17644026, rs11133238, rs4864440, rs11133239, rs11929934, rs13434989, rs10012324, rs17082294, rs10517278, rs4864717, rs17646076
RP11-53O19.1	rs80316101, rs150134525, rs7720551, rs76768074, rs148946381, rs181072007, rs186001811, rs111765202, rs35601455, rs112494990, rs144785376, rs62366821, rs112679498, rs13179565, rs201180654, rs191324191
ATG10	rs4703825, rs12187089, rs11738172, rs432872, rs457049, rs456778, rs463247, rs457700, rs386424, rs462122, rs11740142, rs1384256, rs11741569, rs1485587, rs11740648, rs11741303, rs2860007, rs11748868, rs12515069, rs1428939, rs178957, rs2407064, rs10068160, rs3857369, rs10061458, rs17245188, rs9293290, rs4703537, rs891159, rs4703870, rs10066167, rs10065463, rs1543911, rs2059891, rs6895884, rs6888977, rs6884232, rs1019806, rs4703879, rs2407153, rs11747683, rs749402, rs749401, rs862240, rs146991557, rs226204, rs11743578, rs2407156, rs17247678, rs6861268, rs58757861, rs715888, rs1827391, rs168534, rs355285, rs4703914, rs6885480, rs10473857, rs4605761, rs1505073, rs40214, rs256795
ATP6AP1L	rs6881927, rs12517153, rs13174473, rs10042996, rs12188888, rs2972230, rs10514220, rs1561150, rs442417, rs461802, rs4703852, rs2406905, rs6892261, rs16899359, rs7727483, rs178957, rs3857369, rs10061458, rs9293290, rs4703537, rs10066167, rs10065463, rs3738, rs226202, rs226199, rs6880209, rs178931, rs226196, rs862240, rs146991557, rs862239, rs3756683, rs3734115, rs224844, rs224843, rs6872917, rs12187334, rs4703894, rs905221, rs2385882, rs12153244, rs10062095, rs2015904, rs2015911, rs4354012, rs4282294, rs16899813, rs924615, rs10514242,

	rs10041952, rs16899864, rs9293313, rs17205439, rs10462383, rs6863549, rs13165887, rs13166060, rs10474061, rs10474062, rs10051128, rs12658884, rs7713051, rs1363948, rs9283781, rs11744735, rs433820, rs37558, rs4703936
UBLCP1	rs31199, rs10054046, rs10070382
RP11-32D16.1	rs11466807, rs11744671, rs11466784, rs3097837, rs254664, rs1896606, rs13155377, rs13159354, rs7709209, rs2984629, rs17242576, rs11741271, rs2988321, rs1952657, rs12187534, rs10045014, rs12188478, rs10515764, rs17055744, rs2419654, rs1317414, rs864821, rs824869, rs824871, rs699083, rs10040448, rs13171583, rs11750665, rs11747709, rs11741746, rs13175305, rs11743605, rs7715522, rs11748882, rs31193, rs13155183, rs1345721, rs7736575, rs2009612, rs1469070, rs7730360, rs13172711, rs4130208, rs2287763, rs6881547, rs12187372
BTN3A2	rs9467504, rs13203202, rs2328879, rs6456693, rs10946795, rs6924948, rs9467701, rs6923139, rs9467704, rs6903015, rs9467707, rs6939978, rs9366653, rs9379851, rs9348709, rs9393705, rs9393706, rs9393707, rs9393708, rs9358932, rs9358934, rs9357006, rs9379855, rs9348712, rs9379856, rs9379857, rs9379858, rs9393710, rs9379859, rs9358935, rs12173854, rs9379864, rs12176317, rs12174602, rs12174631, rs13216828, rs9393713, rs9393714, rs9358937, rs2073529, rs2073531, rs9348716, rs9366655, rs1977, rs1978, rs1979, rs3799380, rs4518487, rs6456728, rs9348721, rs9467772, rs3799383, rs6920256, rs16901784, rs2494691, rs2451746, rs9366673, rs9393777, rs17687319, rs17687372, rs17687396, rs17739727, rs12213055, rs9468076, rs2076305, rs2064219
L3MBTL3	rs9321204, rs9388762, rs7754426, rs6569648, rs7740107
RP11-73O6.3	rs9388721, rs9372945, rs17755387, rs9402157, rs4499953, rs17811901, rs12191170, rs12202100, rs6569644, rs9388762, rs7754426, rs7740107, rs4364506, rs9492440, rs10499172, rs7759381, rs9492441, rs7769599, rs12190724, rs7764762, rs4548027, rs9492443, rs12198331, rs9492445, rs9492446, rs9492447
RMND1	rs11759741, rs11751703, rs17800315, rs17800327, rs11759502, rs11757075, rs2223451, rs7451945, rs742315, rs9371462, rs7769835, rs6557545, rs9384561, rs9371463, rs11752947, rs9383832, rs11155740, rs7742124, rs4870532, rs17080057, rs17080062, rs17080069, rs17080087, rs17080089, rs17080091, rs17080093, rs17080102, rs9372087, rs10872665, rs1494309, rs782665, rs803401, rs2295083, rs9397029, rs742829, rs6902496, rs12203650, rs7349940, rs12205664, rs1575219, rs2180927, rs9478167, rs730489, rs4869987, rs4869988, rs6923696, rs17362091, rs4869991, rs6939639, rs2179544, rs12210237, rs3800270, rs6904364, rs6557140, rs9322321, rs9397050, rs7752091, rs3736175, rs6557141, rs9383571, rs954238, rs2982558, rs2982556, rs1124674, rs9479117, rs1709180, rs3020404, rs2982683, rs9340978, rs3798577, rs9383963, rs7450824, rs2813549, rs2813550, rs2763033, rs2635469, rs1324452, rs2813504, rs12174080, rs2788381, rs1890100, rs3816850, rs4472361, rs7776230, rs725235, rs4870095, rs6913579, rs4870101, rs4870102, rs214955, rs214992
KLHDC10	rs4728160, rs721691, rs2896415, rs4731568, rs7800983, rs11764547, rs10246160, rs10225672, rs2727455, rs10480805, rs1046691, rs10246707, rs17162050, rs10249037, rs17162123, rs9656386, rs17162136, rs10273782, rs6467267, rs17162295, rs10500121, rs10253233, rs6962745, rs7803795, rs7801603, rs7805980, rs10279425, rs12531444, rs12534580, rs901799, rs1488009, rs1574704, rs6943386, rs7793239, rs2129902, rs6467309, rs10257888, rs10263075,

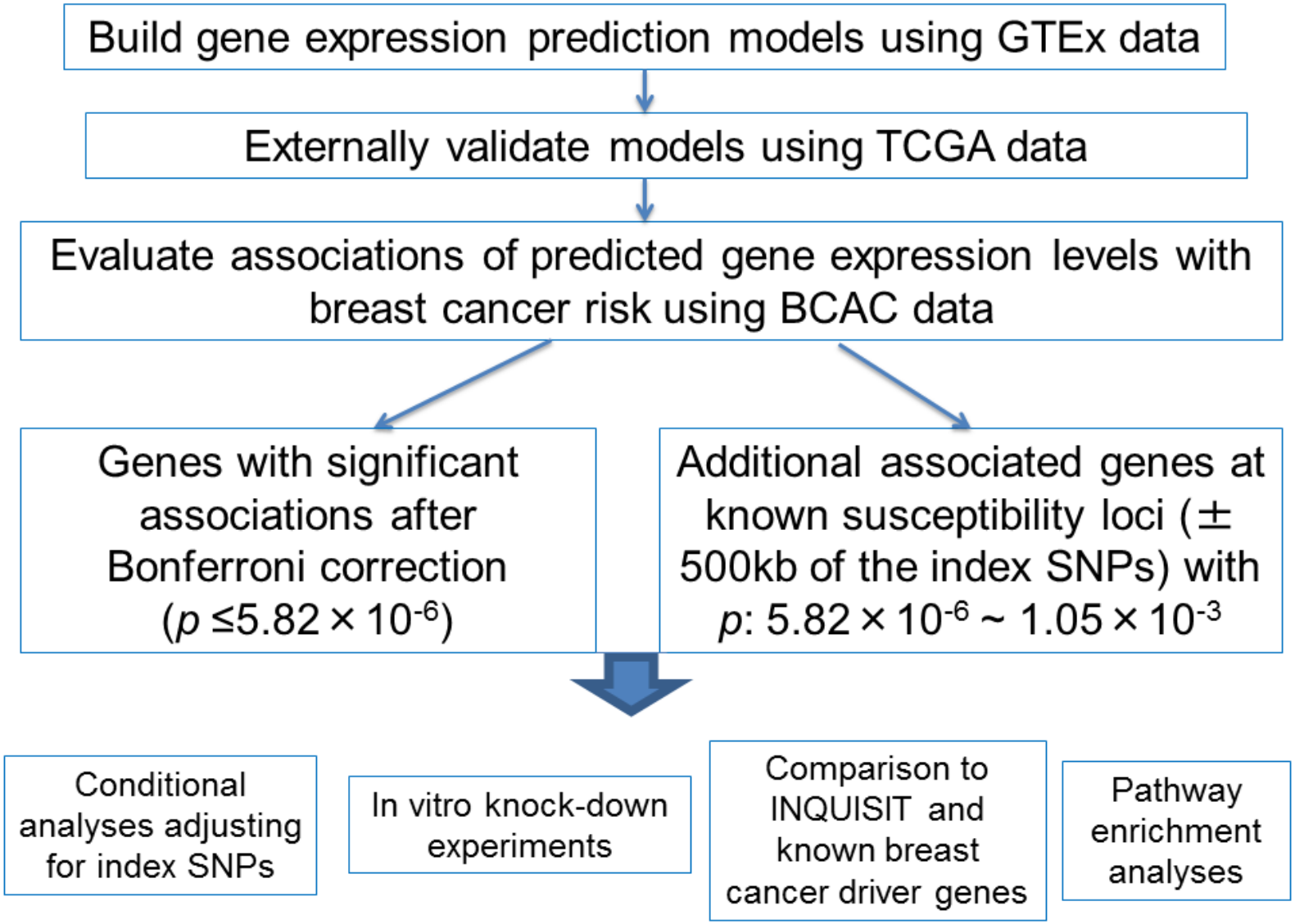
	rs1035596, rs10281580, rs10272075, rs10272206, rs10248834, rs10248294, rs10279517, rs6969737, rs6945822, rs17165066, rs290794, rs290805, rs290804, rs17165262, rs17688449
MIR31HG	rs10965219
AP006621.6	rs6597947, rs12270802, rs7928098, rs7927765, rs4077757, rs3793964, rs7395835, rs7394830, rs12801744, rs11037265, rs11038276, rs16927520, rs2292958, rs11246048, rs12801980, rs2242566, rs909098, rs7396812, rs7481525, rs7395918, rs7481685, rs11246175, rs1056812, rs7942569, rs7395822, rs4078520, rs10902208, rs11246300, rs7952095, rs6597984, rs11246311, rs7104929, rs4963153, rs10902221, rs6597981, rs11246314, rs11246316, rs4131364, rs11246319, rs11246327, rs11246340
RIC8A	rs3782116, rs7115703, rs11246062, rs7947900, rs7395319, rs7396812, rs10751657, rs7481525, rs12361394, rs7484182, rs7102822, rs17585, rs7113204, rs12577324, rs35579818
RPLP2	rs11245936, rs7103978, rs28514396, rs2943510, rs11029039, rs6578471, rs4752763, rs11604009, rs10400297, rs10838484, rs1038727, rs7110331, rs575488, rs2292958, rs2292963, rs1317356, rs11041082, rs7127542, rs6598055, rs7925234, rs11246052, rs9737419, rs10902120, rs11602841, rs7947900, rs7395116, rs11246068, rs4131942, rs7117996, rs10794314, rs10794315, rs11246108, rs11246130, rs11246131, rs10902165, rs12806187, rs4963166, rs7635, rs10902208, rs7952095, rs12277141, rs10902221, rs11246314, rs11246316, rs28633403
SNX32	rs17304039, rs11601767, rs11227332, rs1939212, rs583887, rs596002, rs601863, rs658524, rs645900, rs687672, rs658938, rs568617, rs694994, rs641018, rs656980, rs2231884, rs531784
B3GNT1	rs694243, rs512715, rs674485, rs1194758, rs686320, rs1787666, rs616599, rs4102217, rs4099470, rs11227226, rs2298615, rs491666, rs610497, rs684546, rs668210, rs539046, rs17307346, rs512421, rs559298, rs10219183, rs1791682, rs2242663, rs7103627, rs7119426, rs7120256, rs3862391, rs11227678
RP11-867G23.10	rs78407319, rs190536043, rs118019315, rs2270448, rs118151305
RP11-218M22.1	rs10849596, rs11064617, rs11614523, rs7967165, rs16932084, rs2286036, rs17223490, rs7973873, rs2189669, rs11609462, rs7313155, rs10848486, rs7966350, rs4765827, rs4766400, rs7135126, rs10505717, rs7967909, rs510714, rs1051104, rs542736, rs2075228, rs518685, rs2300127, rs11062163, rs11063111, rs215227, rs215231, rs11063281, rs11063286, rs4980927, rs2286781, rs1860612, rs756502, rs10849215, rs6489652, rs10849328, rs2607918, rs7955627, rs2075032, rs4980841, rs11615697, rs2061317, rs10744703, rs3858703, rs11064524, rs12314329, rs12301299
GALNT16	rs17105278, rs7150454, rs12889206, rs8003738, rs4902567, rs1810623, rs6573834, rs1570106, rs17105586, rs916962, rs2247048, rs2525521, rs1476586, rs1859302, rs2525523, rs2525524, rs2525525, rs2525526, rs2525527, rs2842331, rs7153476, rs8007194, rs2257111, rs2257116, rs2257127, rs4899246, rs10137893, rs4902611, rs181464, rs17835996, rs1275195, rs1950712, rs7143336, rs1890941, rs2185492, rs7155178, rs7151003, rs8008770, rs12889279, rs1958184, rs10873226, rs12884186, rs11158830, rs1009256, rs2182972, rs9323533, rs41350744, rs1469253, rs7141710, rs4902805, rs12590132, rs1015585, rs4902806

PLEKHD1	rs9323513, rs11158749, rs10134446, rs2189517, rs7140266, rs2525530
CTD-3051D23.1	rs8095, rs1053419, rs11850704, rs4906423, rs4555088, rs3809461, rs3809457, rs7144812, rs4454893, rs7142224, rs10136937, rs4340263, rs10135130, rs3809454, rs2841214, rs2582559, rs4983589, rs3825761, rs7151594, rs880616, rs10139596, rs1882848, rs10140111, rs11625865, rs4983590, rs11160839
MAN2C1	rs2075589, rs4886649, rs1809714, rs1984586, rs1984587, rs7164976, rs3866545, rs7183520, rs7166852, rs7164429, rs8028182, rs12708519, rs8029112, rs28610581, rs8030802, rs6495182, rs8023268, rs8023815, rs11636031, rs11636199, rs12708520, rs7163907, rs28693593, rs4886716, rs4075522, rs7184046, rs13380103
CTD-2323K18.1	rs17336243, rs2304900, rs12908919, rs1822324, rs12911696, rs12899456, rs12909554, rs8038911, rs12905302, rs1809714, rs1984586, rs1984587, rs7164976, rs3866545, rs7183520, rs7166852, rs7164429, rs9673084, rs5745935, rs8027749, rs11635996, rs3210683, rs1128585
RCCD1	rs11855570, rs8030486, rs2227935, rs7167216, rs734252, rs2677744, rs1266489, rs1266483, rs4773, rs1550636, rs2290202, rs2301825, rs3826033, rs4392040, rs7402585, rs12915069, rs9744944, rs8028382, rs4583214, rs4244910, rs4932591, rs4306482
RP11-467J12.4	rs17257857, rs12597737, rs7200881, rs17201162, rs12598784, rs1344490, rs3910446, rs9889099, rs1362380, rs9936470, rs16950876, rs8048212, rs17268400, rs3095536, rs3095537, rs1076081, rs1861315, rs16951015, rs1074734, rs16951035, rs1345312, rs16951056, rs12597728, rs4477699, rs4480800, rs1362558, rs11075488, rs9921890, rs1861527, rs3095599, rs3095600, rs194392, rs194394, rs12930211, rs7191789, rs1362553, rs8048309, rs1362554, rs1345389, rs2075236, rs3095660, rs1420546, rs3095661, rs40841, rs1362560, rs3095616, rs1420548, rs8051542, rs4784220, rs12922061, rs11647542, rs11866049, rs16951465, rs11867085, rs3104823, rs3112587, rs16951525, rs12925035, rs4784253, rs12919531, rs4238756, rs4783785, rs1420257, rs6499105, rs7205069, rs17298178, rs17370363, rs11639509, rs7500472, rs12919591, rs12922267, rs2387879, rs551415417, rs9925367, rs9936502, rs10153135, rs16951919, rs4783804, rs9925003, rs7198530, rs12933494, rs12919486, rs12930884, rs8058720, rs3760010, rs4456500, rs8051064, rs8047647, rs1833205, rs1833207, rs12051480, rs8045574, rs2388117, rs9924562, rs2160290, rs3743772, rs7186754, rs3095633, rs17802269, rs16952362, rs1477199, rs11861870, rs16952517, rs6499642, rs6499643, rs16945088, rs11075994, rs7195539, rs6499653, rs12149433, rs9926180, rs2111113, rs7500562, rs13335343, rs1362570, rs16952634, rs9937234, rs10852525, rs12921721, rs1344503, rs12232391, rs8053966, rs17821714, rs11864972, rs7185301, rs7185479, rs17224310, rs17224394, rs11643535, rs7194907, rs8053888, rs7185783, rs16952730, rs9933107, rs8056502, rs8061928, rs16952756, rs13335453, rs9922370, rs1971037, rs1125337, rs1125338, rs11076015
CTD-3032H12.1	rs6499657, rs17823199, rs17196003, rs748815, rs13333140, rs16953503, rs12926529, rs9926409, rs1420289, rs2160294, rs933517, rs1420290, rs1420292, rs8059628, rs1186818, rs7205346, rs16953806, rs12325292, rs7198507, rs4435250, rs4783866, rs2589010, rs8053467
LINC00671	rs72826975

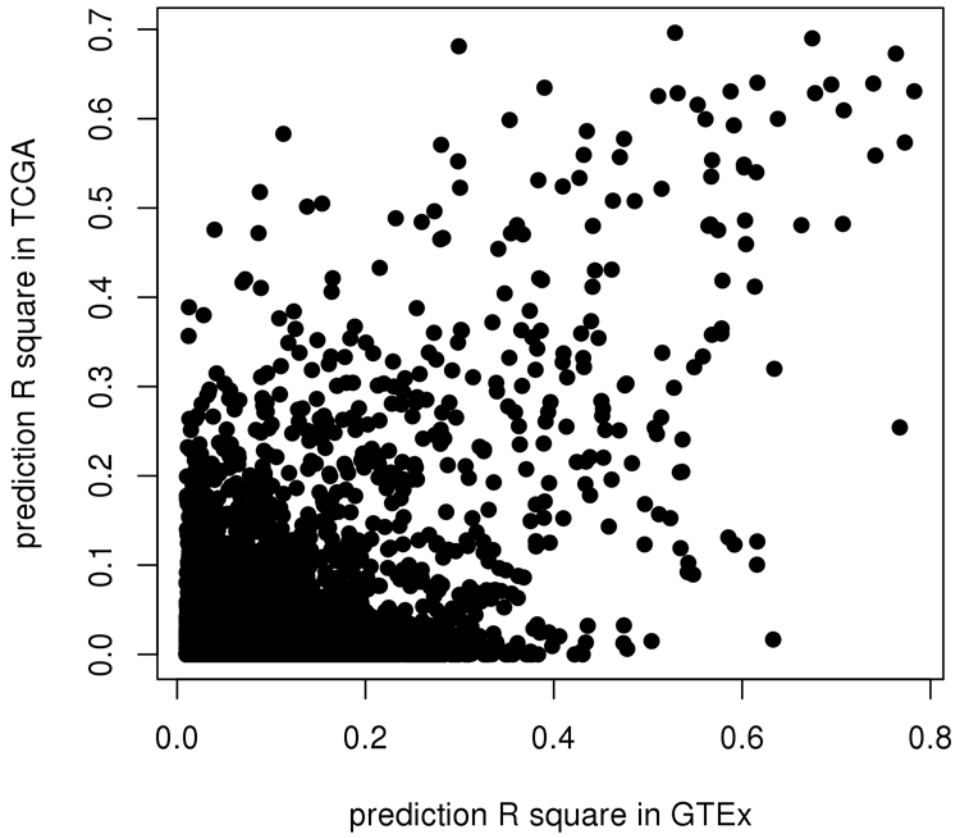
LRRC37A	rs12947718, rs12942666, rs7222389, rs34018943, rs11012, rs9730, rs17631303, rs2077606, rs3946526, rs17631676, rs1880750, rs1358071, rs17690703, rs16940758, rs4510068, rs7225002, rs2696531, rs538628, rs169201, rs199439, rs199457, rs199456, rs199451, rs199448, rs199445, rs199443, rs199535, rs199534, rs9896243, rs199533, rs199498, rs199497
KANSL1-AS1	rs1133458, rs7213493, rs9911183, rs11079502, rs2072090, rs16939943, rs7222751, rs7214661, rs7222377, rs1059504, rs1230106, rs439558, rs17687796, rs1358071, rs1989480, rs8082105, rs12953076, rs12938031, rs4076452, rs9892359, rs171441, rs242939, rs171443, rs17690703, rs753235, rs2435205, rs3785883, rs2471738, rs2532345, rs2696531, rs148126555, rs538628, rs169201, rs199439, rs199457, rs199456, rs199451, rs199448, rs199445, rs199443, rs199535, rs199534, rs9896243, rs199533, rs35732828, rs2074404, rs199498, rs199497, rs3851781, rs7214920, rs7213046, rs1004485, rs3851788, rs2317996, rs2317995, rs3851797, rs2316757, rs3851800, rs6504640, rs8078196, rs8074704, rs9905068, rs7210126, rs2316759, rs7221913, rs11653838, rs7216950, rs4968304, rs4074249, rs11079750, rs12950699, rs11871364
CRHR1	rs12942842, rs17544947, rs6503404, rs10853009, rs1071682, rs1000354, rs12947718, rs12942666, rs7222389, rs34018943, rs11012, rs9730, rs17631303, rs2077606, rs3946526, rs17631676, rs9890016, rs2435200, rs4630591, rs183211, rs199436, rs199438, rs142167, rs7224296, rs199453, rs199452, rs199449, rs199444, rs199442, rs199524, rs199520
STXBP4	rs2164160, rs1434739, rs974588, rs11653274, rs11079129, rs17709877, rs17675596, rs7210359, rs12952253, rs7209926, rs7208123, rs9916547, rs8065361, rs9891704, rs9900816, rs9907961, rs17745183, rs2787497, rs2787481, rs244317, rs7218719, rs7223718, rs7213282, rs3931318, rs11079178, rs7218226, rs8069305, rs8073227, rs4794602, rs17211444, rs11653426, rs906580, rs12940838, rs8066578, rs11656691, rs12938239, rs12936661, rs8076964, rs7208778, rs13380851, rs11656314, rs11656373, rs6504993, rs6504994, rs12951953, rs6504995, rs11658412, rs10515108, rs10515109, rs7218936, rs376416681, rs7223017, rs7225097, rs8068391, rs7220718, rs2055661, rs8069548, rs16956323, rs9903179, rs9902928, rs9903177
HAPLN4	rs12976333, rs12985909, rs1363119, rs10415765, rs10423237, rs10415319, rs10401774, rs8102478, rs3746183, rs3746181, rs1363120, rs888663, rs888669, rs35742476, rs10420384, rs2303692, rs7246788, rs2891676, rs4808803, rs11672385, rs1971093, rs8111582, rs8111397, rs10426768, rs4808807, rs2013069, rs11670392, rs7249760, rs10409346, rs10409408, rs4808135, rs1469412, rs6512269, rs4580317, rs4808812, rs8106068, rs4808814, rs10422974, rs10415320, rs7249453, rs34992005, rs17214879, rs247765, rs247781, rs248969, rs248950, rs7247625, rs7248612, rs2116889, rs12980778, rs2080261, rs7250793, rs258583
ZNF404	rs12610476, rs3786956, rs4803662, rs423765, rs376069, rs417699, rs407731, rs425221, rs388685, rs388706, rs399098, rs378109, rs424729, rs423320, rs423752, rs375066, rs384329, rs17656688, rs413093, rs403137, rs379785, rs454559, rs367741, rs373168, rs411803, rs10422017, rs12972550, rs16989260, rs12459705, rs17714676, rs2965109, rs7254776

RP11-15A1.7	rs364691, rs349032
UBD	rs16894184, rs16894189, rs6909302, rs6941946, rs7768299, rs2394512, rs7741520, rs9461498, rs9468471, rs9380105, rs9468473, rs7749435, rs6919044, rs9501291, rs6924824, rs6456886, rs6456889, rs2064365, rs11758255, rs6917293, rs9295790, rs6456908, rs9380110, rs3749977, rs6908651, rs9380111, rs4713203, rs9283885, rs238883, rs1419640, rs12110437
LRRC37A2	rs241036, rs241035, rs241033, rs241031, rs241030, rs241027, rs241026, rs241023, rs241022, rs241021, rs241020, rs17760577, rs17760631, rs17687462, rs17760733, rs17687504, rs17687534, rs17687571, rs17687625, rs17687667, rs17687740, rs757502, rs757501, rs757500, rs735423, rs4486953, rs17688032, rs17688056, rs17688068, rs17688090, rs17761387, rs17688205, rs17688296, rs17688391, rs17688410, rs17688434, rs17688452, rs10491144, rs10491143, rs17688534, rs17761838, rs12150141, rs12150610, rs12150547, rs17688682, rs12150454, rs1526129, rs17762308, rs968028, rs968027, rs17762361, rs17689104, rs17689116, rs17689218, rs17762535, rs1568949, rs1105571, rs1105569, rs17563433, rs17649019, rs17563501, rs10514879, rs1358071, rs4401083, rs1880752, rs4617909, rs2902662, rs2864087, rs4471726, rs17563599, rs17649138, rs4390635, rs17649162, rs17563683, rs17563718, rs1526125, rs1526126, rs17563787, rs17563800, rs17563827, rs17563861, rs17563889, rs17563923, rs12150683, rs17334797, rs12150658, rs12150048, rs12150455, rs12150604, rs17334944, rs17426064, rs17426106, rs17426174, rs17426195, rs4264434, rs6503447, rs11079723, rs11079724, rs17690703, rs17769490, rs17769552, rs885639, rs2873269, rs4627402, rs2696531, rs17665188, rs538628, rs169201, rs199439, rs199457, rs199456, rs199451, rs199448, rs199445, rs199443, rs199535, rs199534, rs9896243, rs199533, rs199528, rs199498

Supplementary Figure 1. **Study design flow chart**

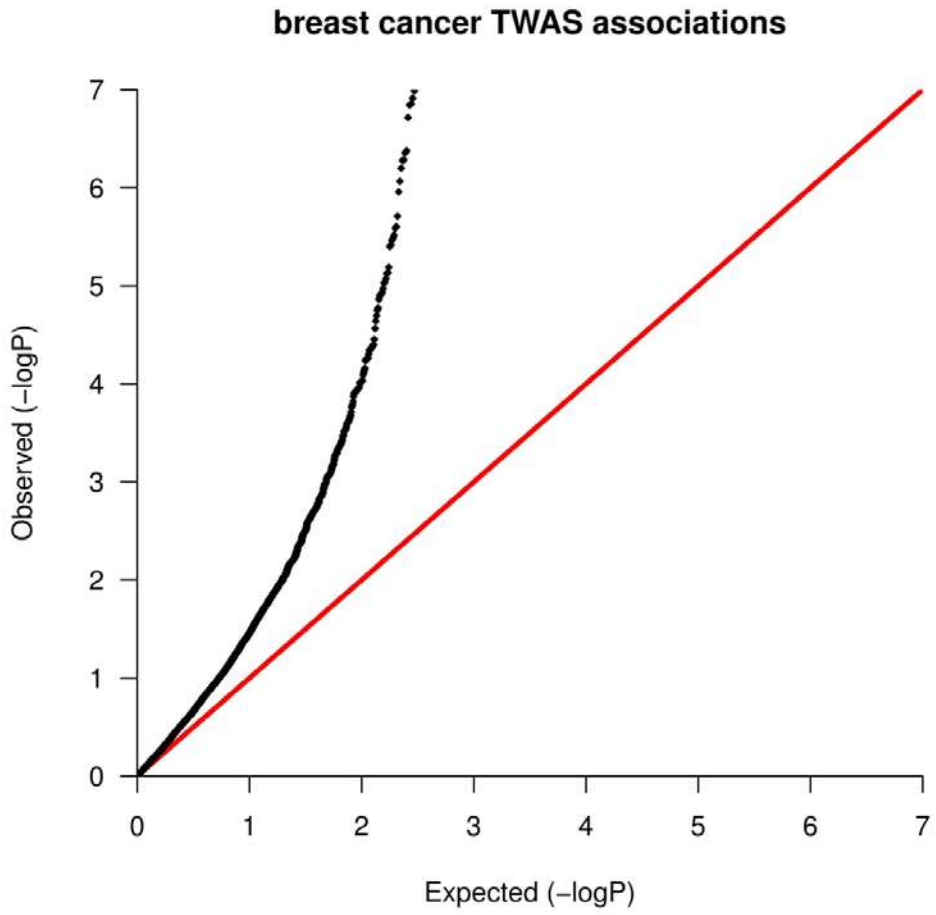

Build gene expression prediction models using GTEx data
Externally validate models using TCGA data
Evaluate associations of predicted gene expression levels with breast cancer risk using BCAC data
Genes with significant associations after Bonferroni correction ($p \leq 5.82 \times 10^{-6}$)
Additional associated genes at known susceptibility loci (± 500kb of the index SNPs) with p: $5.82 \times 10^{-6} \sim 1.05 \times 10^{-3}$
Conditional analyses adjusting for index SNPs
In vitro knock-down experiments
Comparison to INQUISIT and known breast cancer driver genes
Pathway enrichment analyses

Supplementary Figure 2. **Performance of expression prediction models in GTEx and TCGA datasets for genes with at least 10% correlation in GTEx data.** The x axis represents the prediction performance (R^2) in GTEx dataset. The y axis represents the prediction performance in TCGA dataset. Each dot represents the expression prediction model for one gene. There is a trend that genes with a high internal prediction performance in GTEx data also have a high external prediction performance in TCGA data (correlation coefficient: 0.55).

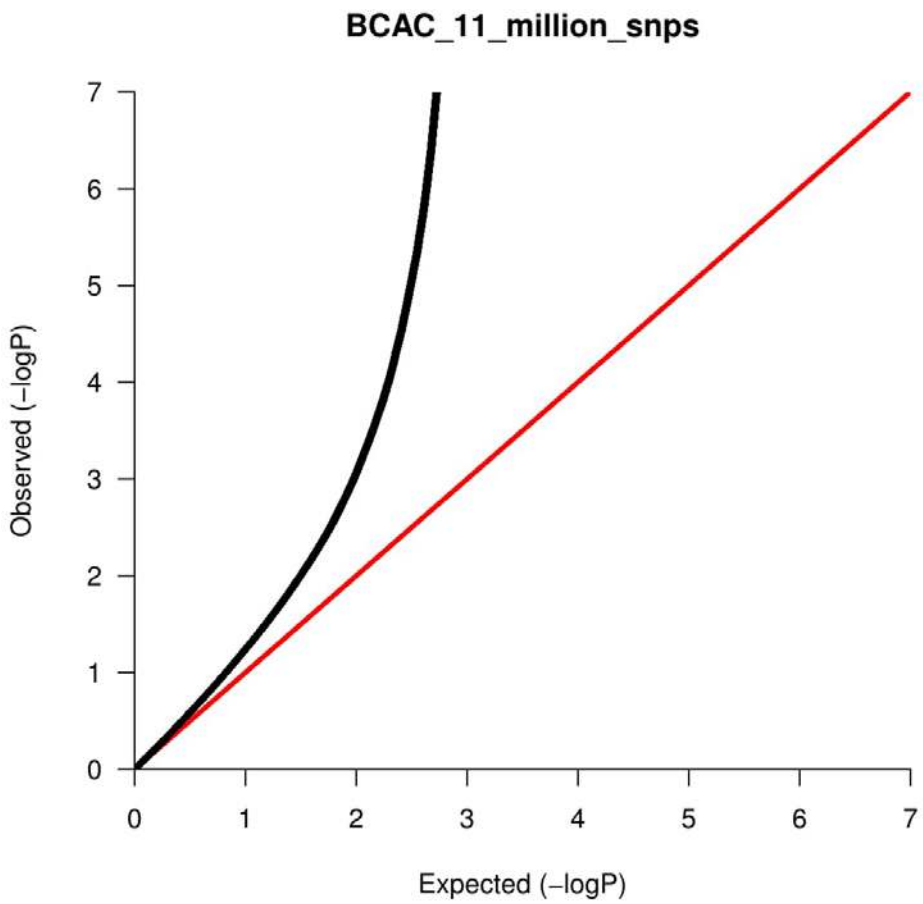


Supplementary Figure 3. **Quantile-quantile plots.** (A) Quantile-quantile plot of P values in –log scale of associations between genetically predicted expression levels of 8,597 genes and breast cancer risk; (B) Quantile-quantile plot of P values in –log scale of associations between the over 250,000 SNPs predicting expression levels of the 8,597 genes and breast cancer risk in BCAC; (C) Quantile-quantile plot of P values in –log scale of associations between all 11.8 million SNPs and breast cancer risk in BCAC

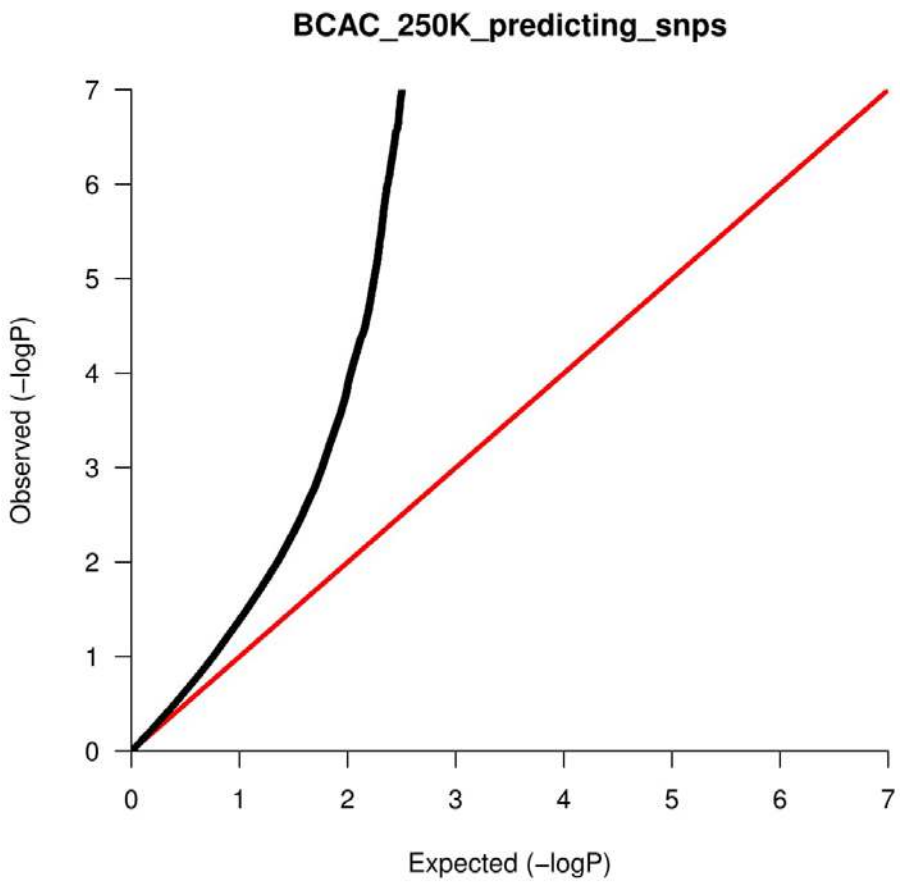
(A)



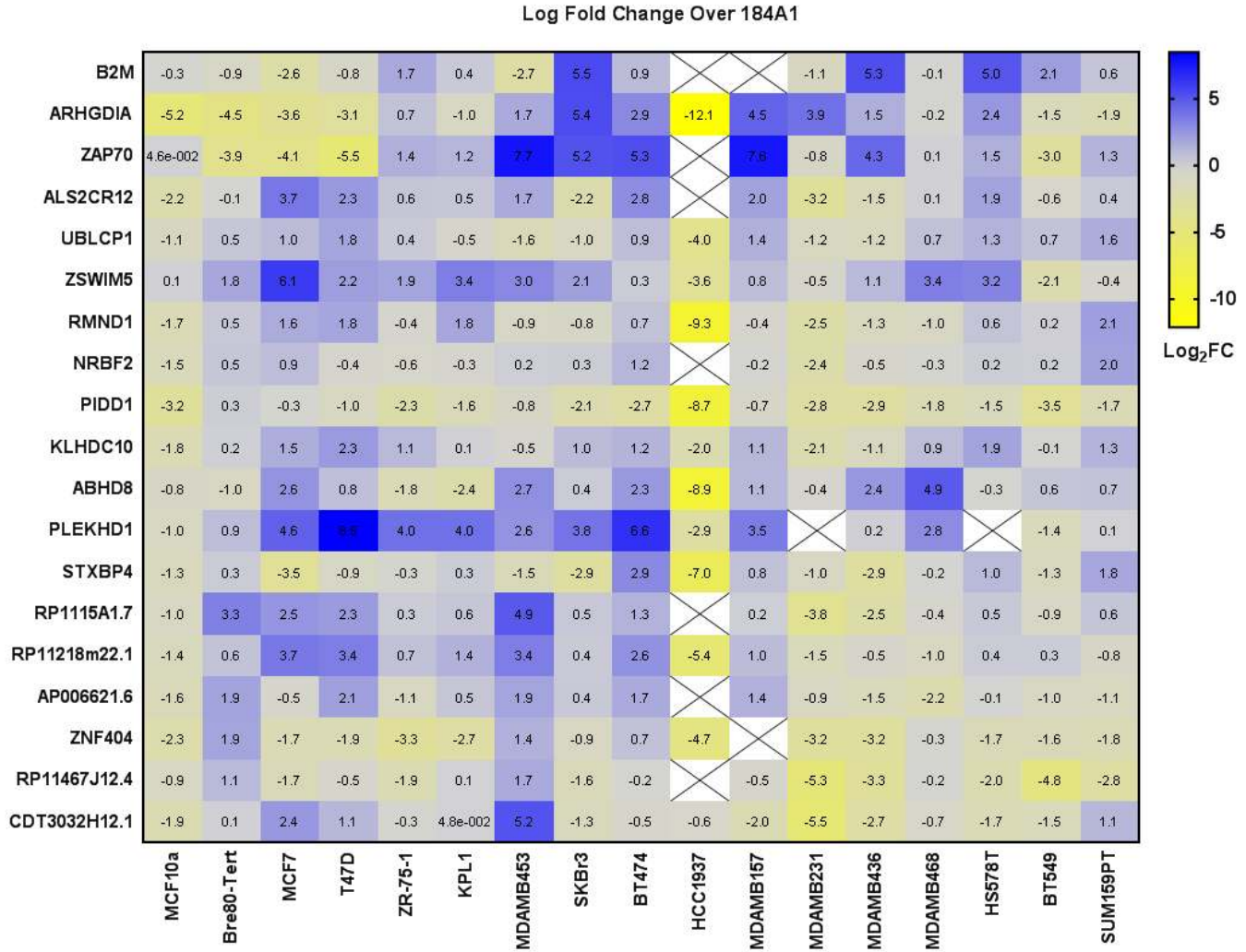
(B)



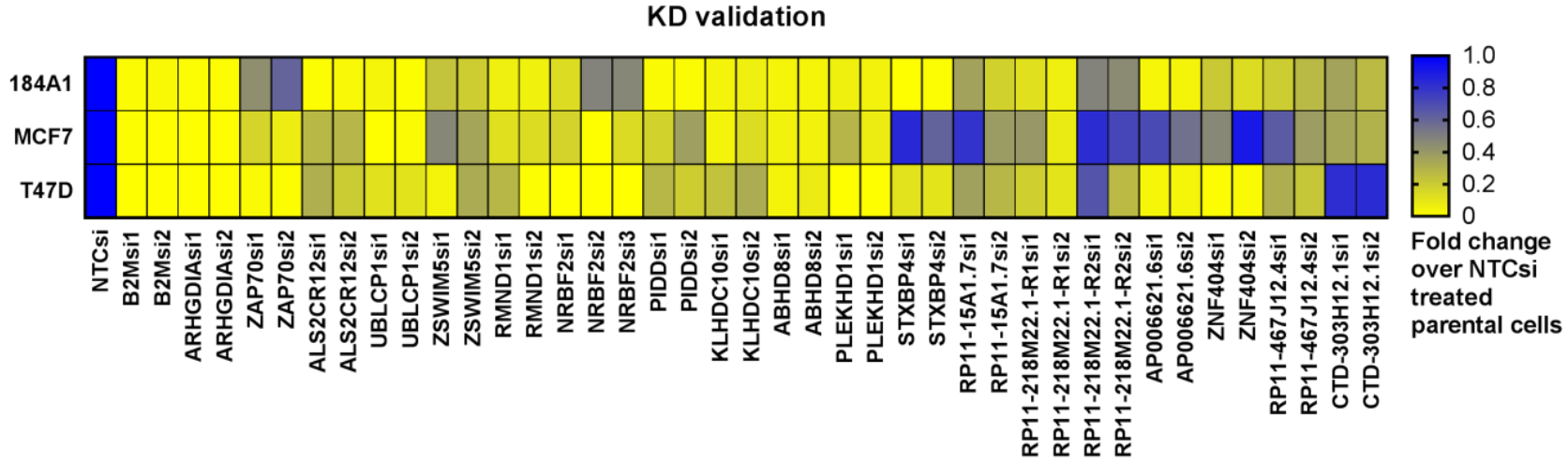
(C)


BCAC_250K_predicting_snps
Observed (−logP)
Expected (−logP)
0
1
2
3
4
5
6
7

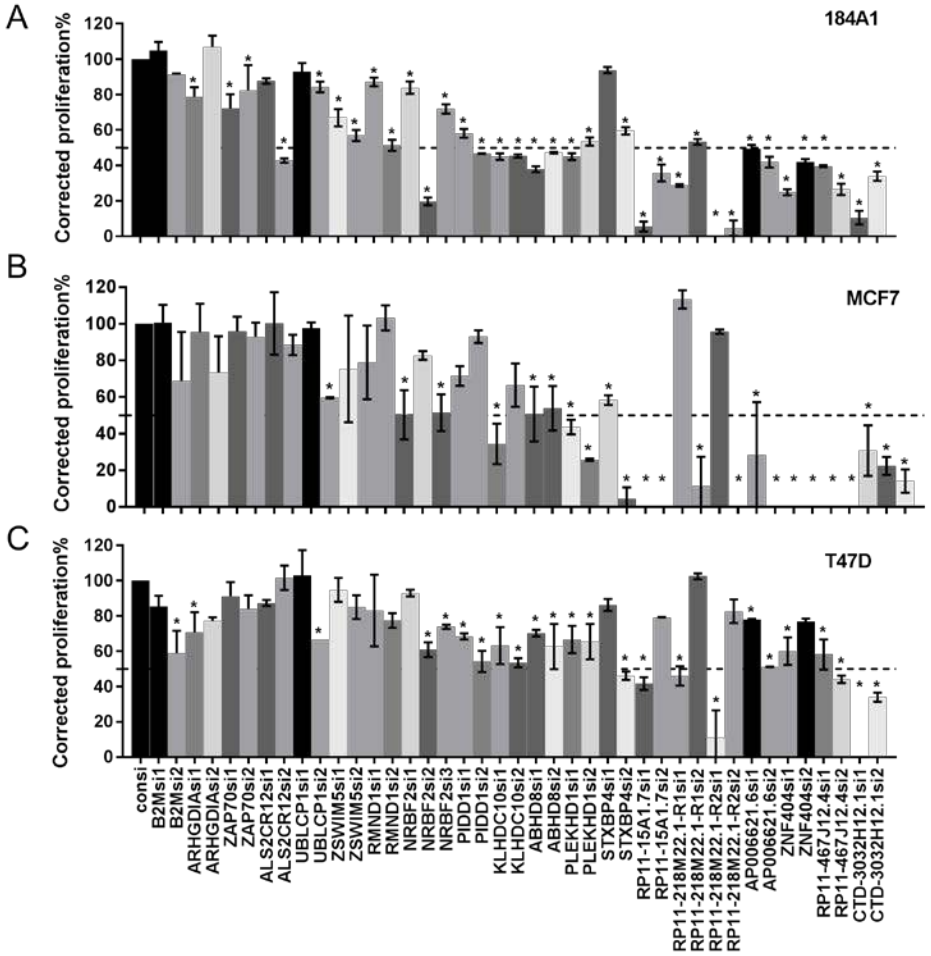
Supplementary Figure 4. **Heat map of log fold change (FC) of selected genes normalized to expression levels in 184A1 breast cells.** Two or three primer sets were designed for each gene (y-axis) and mRNA levels quantified by qPCR in indicated cells lines (x-axis), including 184A1. The FC of genes normalized to that in 184A1 = mRNA level in indicated cells / mRNA level in 184A1. The log_2FC over 184A1 is depicted as a heat map. An X represents "not detectable" with all primer sets.



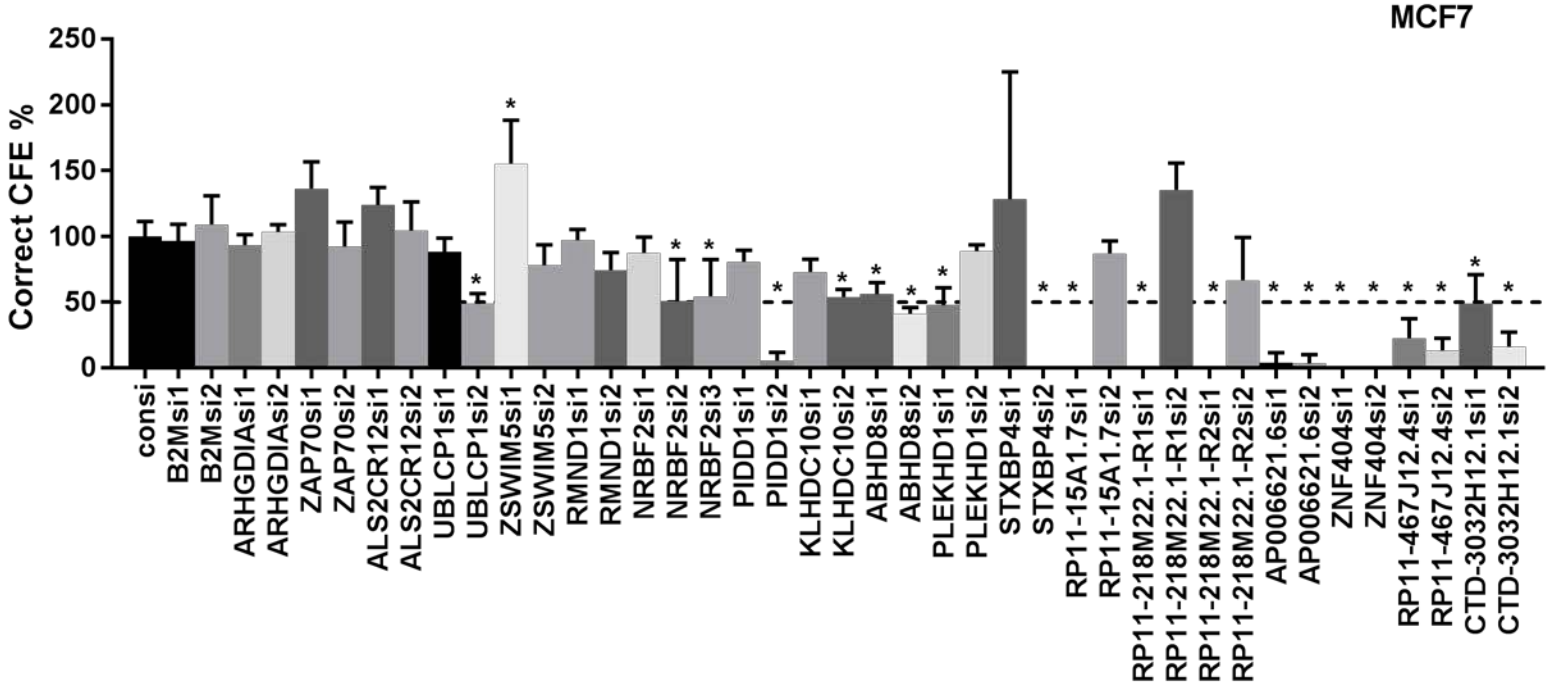
Supplementary Figure 5. Validation of knockdown. 184A1, MCF7 and T47D cells, transfected with the indicated siRNAs, were harvested after 36 hours for qPCR analysis to assess knockdown efficiency. The fold changes over NTCsi-transfected parental cells were plotted.



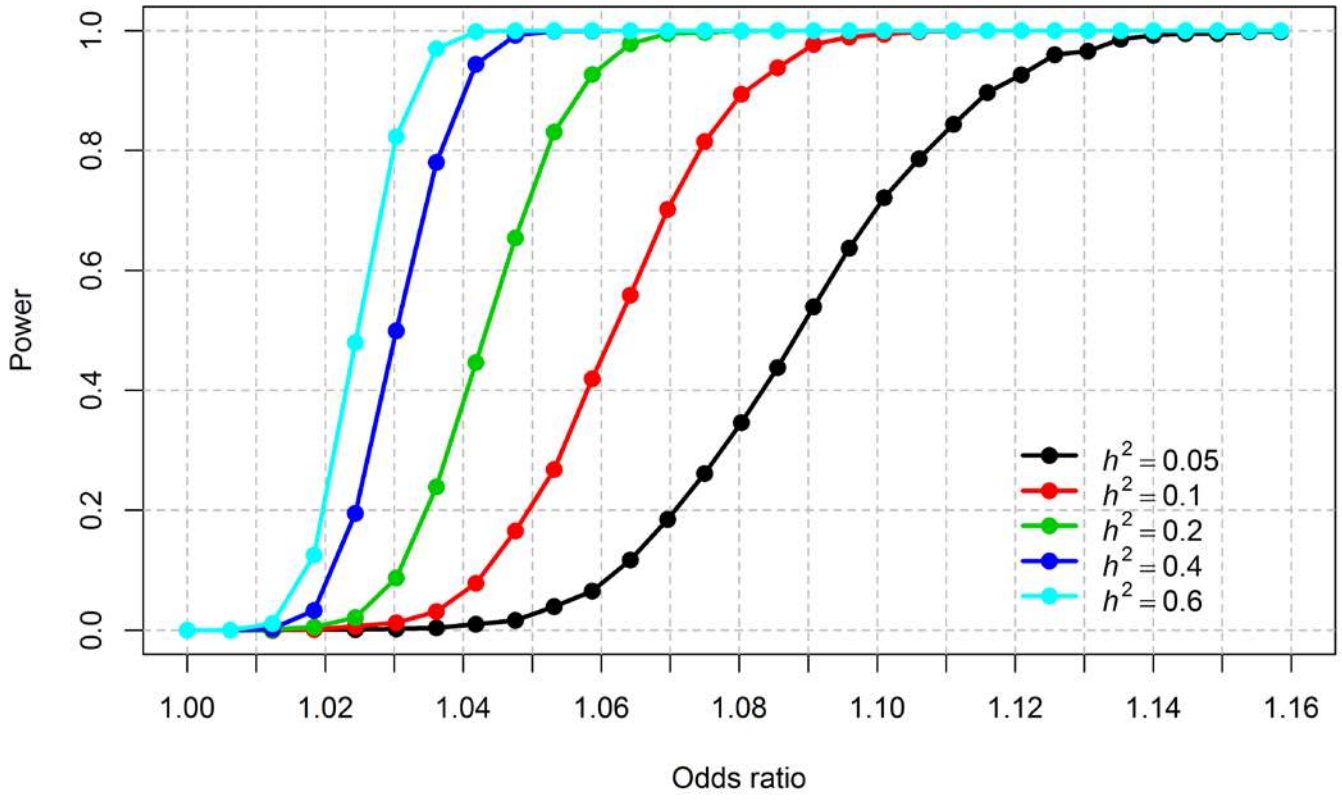
Supplementary Figure 6. **Proliferation in breast cells using two independent siRNAs** (related to Figure 2A). **(A)** 184A1, **(B)** MCF7 or **(C)** T47D cells were transfected with indicated siRNAs over seven days and phase-contrast images collected using an IncuCyte ZOOM. Each cell proliferation time-course was normalized to the baseline confluency and analyzed in GraphPad Prism. Corrected proliferation % = 100 +/- (relative proliferation in indicated siRNA - proliferation in control siRNA (consi))/knockdown efficiency. Error bars, SD (*N*=2). *P*-values were determined by one-way ANOVA followed by Dunnett's multiple comparisons test: **P*-value < 0.05.



Supplementary Figure 7. **Colony formation efficiency in MCF7 cells using two independent siRNAs** (related to Fig 2B). MCF7 cells were transfected with indicated siRNAs, then reseeded after 16 hours for colony formation (CF) assays. At day 14, colonies were fixed with methanol, stained with crystal violet, scanned and batch analyzed by ImageJ. Corrected CF efficiency (CFE) % = 100 +/- (relative CFE in indicated siRNA - CFE in control siRNA (consi))/knockdown efficiency. Error bars, SD (*N*=2). *P*-values were determined by one-way ANOVA followed by Dunnett's multiple comparisons test: **P*-value < 0.05.



Supplementary Figure 8. **Power calculation of the TWAS analysis.** The simulation analysis is based on 122,977 cases and 105,974 controls. The gene expression was generated from the empirical distribution of predicted gene expression levels in the BCAC. Statistical power was calculated at $P<5.82\times10^{-6}$ (the significance level used in main TWAS analyses) according to cis-heritability (h^2) which we aim to capture using gene expression prediction models (R^2). The figure shows results per one standard deviation increase (or decrease) in the gene expression based on 1000 replicates.



Supplementary Note

Acknowledgements for BCAC studies

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme grant numbers 634935 and 633784 for BRIDGES (Breast cancer RIsk after Diagnostic GEne Sequencing) and B-CAST (Breast CAncer STratification) respectively, and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. Genotyping of the **OncoArray** was funded by Genome Canada Grant GPH-129344, NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563. Funding for the **iCOGS** infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. This study would not have been possible without the contributions of the following: Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Antonis Antoniou, Lesley McGuffog and Ken Offit (CIMBA), Andrew Lee, Ed Dicks and the staff of the Centre for Genetic Epidemiology Laboratory, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Francois Bacot, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, and the staff of the Copenhagen DNA laboratory, and Sharon A. Windebank, Christopher A. Hilker and the staff of Mayo Clinic Genotyping Core Facility. The **DRIVE** Consortium was funded by U19 CA148065. The **PERSPECTIVE** project was supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344), the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec, and the Quebec Breast Cancer Foundation. The Australian Breast Cancer Family Study (**ABCFS**) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). 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 Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. We would like to thank Maggie Angelakos, Judi Maskiell, and Gillian Dite. The **ABCS** study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]. We thank Blood bank Sanquin, The Netherlands. The Australian Breast Cancer Tissue Bank (**ABCTB**) is generously supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. ABCTB Investigators include Christine Clarke,Rosemary Balleine,

Robert Baxter,Stephen Braye, Jane Carpenter, Jane Dahlstrom, John Forbes, Soon Lee, Debbie Marsh, Adrienne Morey, Nirmala Pathmanathan, Rodney Scott, Allan Spigelman, Nicholas Wilcken, Desmond Yip. Samples are made available to researchers on a non-exclusive basis. The work of the **BBCC** was partly funded by ELAN-Fond of the University Hospital of Erlangen. The **BBCS** is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). We would like to thank Eileen Williams, Elaine Ryder-Mills, and Kara Sargus. The **BCEES** was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). We thank Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, and Rachel Peake. **BCFR-NY** was supported by an award from NIH grants R01 CA159868 and UM1 CA164920 from the USA National Cancer Institute. **BCFR-PA** was supported by grant UM1 CA164920 from the National Cancer Institute. **BCFR-UT** was supported by grant UM1 CA164920 from the National Cancer Institute. 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 Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. **BCINIS** would not have been possible without the contributions of Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. For **BIGGS**, ES is supported by NIHR Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, United Kingdom. IT is supported by the Oxford Biomedical Research Centre. We thank Niall McInerney, Gabrielle Colleran, Andrew Rowan, and Angela Jones. **BOCS** is supported by funds from Cancer Research UK (C8620/A8372) and the Institute of Cancer Research (UK). BOCS acknowledges NHS funding to the Royal Marsden / Institute of Cancer Research NIHR Specialist Cancer Biomedical Research Centre. The BREast Oncology GAlician Network (**BREOGAN**) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica (IBI) Orense-Pontevedra-Vigo. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. This study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREOGAN), José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS, Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica (IBI) Orense-Pontevedra-Vigo,

Vigo-SERGAS, Vigo, Spain. The **BSUCH** study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). We thank Peter Bugert, Medical Faculty Mannheim. Tha **CAMA** study was funded by Consejo Nacional de Ciencia y Tecnología (CONACyT) (SALUD-2002-C01-7462). Sample collection and processing was funded in part by grants from the National Cancer Institute (NCI R01CA120120 and K24CA169004). We would like to recognize CONACyT for the financial support provided for this work and all physicians responsible for the project in the different participating hospitals: Dr. Germán Castelazo (IMSS, Ciudad de México, DF), Dr. Sinhué Barroso Bravo (IMSS, Ciudad de México, DF), Dr. Fernando Mainero Ratchelous (IMSS, Ciudad de México, DF), Dr. Joaquín Zarco Méndez (ISSSTE, Ciudad de México, DF), Dr. Edelmiro Pérez Rodríguez (Hospital Universitario, Monterrey, Nuevo León), Dr. Jesús Pablo Esparza Cano (IMSS, Monterrey, Nuevo León), Dr. Heriberto Fabela (IMSS, Monterrey, Nuevo León), Dr. Fausto Hernández Morales (ISSSTE, Veracruz, Veracruz), Dr. Pedro Coronel Brizio (CECAN SS, Xalapa, Veracruz) and Dr. Vicente A. Saldaña Quiroz (IMSS, Veracruz, Veracruz). **CBCS** was funded by the Canadian Institutes of Health Research. We thank study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. **CCGP** is supported by funding from the University of Crete. We thank Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou, Georgia Sevastaki, and Konstantinos Pompodakis. The **CECILE** study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The **CGPS** was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. We thank staff and participants of the Copenhagen General Population Study. We also would like to thank the excellent technical assistance from Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, and Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. The **CNIO-BCS** was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). We thank Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO). **COLBCCC** is supported by the German Cancer Research Center (DKFZ), Heidelberg, Germany. Diana Torres was in part supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation. We thank all patients, the physicians Justo G. Olaya, Mauricio Tawil, Lilian Torregrosa, Elias Quintero, Sebastian Quintero, Claudia Ramírez, José J. Caicedo, and Jose F. Robledo, the researchers Ignacio Briceno, Fabian Gil, Angela Umana, Angela Beltran and Viviana Ariza, and the technician Michael Gilbert for their contributions and commitment to this study. The American Cancer Society funds the creation, maintenance, and updating of the **CPS-II** cohort. Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. The **CTS** was initially supported by the California Breast Cancer Act of 1993 and the California Breast Cancer Research Fund (contract 97-10500) and is currently funded through the National Institutes of Health (R01 CA77398, UM1 CA164917, and U01 CA199277).

Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. HAC receives support from the Lon V Smith Foundation (LVS39420). The CTS Steering Committee includes Leslie Bernstein, Susan Neuhausen, James Lacey, Sophia Wang, Huiyan Ma, and Jessica Clague DeHart at the Beckman Research Institute of City of Hope, Dennis Deapen, Rich Pinder, and Eunjung Lee at the University of Southern California, Pam Horn-Ross, Peggy Reynolds, Christina Clarke Dur and David Nelson at the Cancer Prevention Institute of California, Hoda Anton-Culver, Argyrios Ziogas, and Hannah Park at the University of California Irvine, and Fred Schumacher at Case Western University. The University of Westminster curates the **DietCompLyf** database funded by Against Breast Cancer Registered Charity No. 1121258 and the NCRN. We thank the patients, nurses and clinical staff involved in the study. The DietCompLyf study was funded by the charity Against Breast Cancer (Registered Charity Number 1121258) and the NCRN. The coordination of **EPIC** is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). The **ESTHER** study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). We thank Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, and Katja Butterbach. The **GC-HBOC** (German Consortium of Hereditary Breast and Ovarian Cancer) is supported by the German Cancer Aid (grant no 110837, coordinator: Rita K. Schmutzler, Cologne). This work was also funded by the European Regional Development Fund and Free State of Saxony, Germany (LIFE - Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). We thank Stefanie Engert, Heide Hellebrand, Sandra Kröber and LIFE - Leipzig Research Centre for Civilization Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber). The **GENICA** was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of

Tübingen, Germany [HB, Wing-Yee Lo, Christina Justenhoven], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) [HB], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [Yon-Dschun Ko, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [TB, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. The **GEPARSIXTO** study was conducted by the German Breast Group GmbH. The **GESBC** was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). We thank Ursula Eilber. **GLACIER** was supported by Breast Cancer Now, CRUK and Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. We thank Kelly Kohut, Patricia Gorman, and Maria Troy. The **HABCS** study was supported by by the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. We thank Michael Bremer. For **HCSC**, AR is suported by Joan Rodes grant (JR14/00017), Instituto de Salud Carlos III. The **HEBCS** was financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (266528), the Finnish Cancer Society, The Nordic Cancer Union and the Sigrid Juselius Foundation. We thank Sofia Khan, Johanna Kiiski, Kristiina Aittomäki, Rainer Fagerholm, Kirsimari Aaltonen, Karl von Smitten, and Irja Erkkilä. The **HMBCS** was supported by a grant from the Friends of Hannover Medical School and by the Rudolf Bartling Foundation. We thank Peter Hillemanns, Hans Christiansen and Johann H. Karstens. The **HUBCS** was supported by a grant from the German Federal Ministry of Research and Education (RUS08/017), RFBR 14-04-97088. We thank Shamil Gantsev. **ICICLE** was supported by Breast Cancer Now, CRUK and Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. We thank Kelly Kohut, Michele Caneppele, and Maria Troy. Financial support for **KARBAC** was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and Bert von Kantzows foundation. The **KARMA** study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. We thank The Swedish Medical Research Counsel. 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, and by the strategic funding of the University of Eastern Finland. We thank Eija Myöhänen and Helena Kemiläinen. **kConFab** is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, 199600). G.C.T. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. We wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer

Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. **LMBC** is supported by the 'Stichting tegen Kanker'. Diether Lambrechts is supported by the FWO. We thank Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. The **MABCS** study is funded by the Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov" and supported by the German Academic Exchange Program, DAAD. We thank Milena Jakimovska (RCGEB "Georgi D. Efremov), Katerina Kubelka, Mitko Karadjozov (Adzibadem-Sistina" Hospital), Andrej Arsovski and Liljana Stojanovska (Re-Medika" Hospital) for their contributions and commitment to this study. The **MARIE** study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. We thank Petra Seibold, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. **MBCSG** is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects "5x1000"). We thank MBCSG (Milan Breast Cancer Study Group), Siranoush Manoukian, Bernard Peissel, Jacopo Azzollini, Daniela Zaffaroni and Lidia Pezzani of the Fondazione IRCCS Istituto Nazionale dei Tumori (INT); Bernardo Bonanni, Monica Barile and Irene Feroce of the Istituto Europeo di Oncologia (IEO) and the personnel of the Cogentech Cancer Genetic Test Laboratory. The **MCBCS** was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. **MCCS** cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. The **MEC** was support by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The **MISS** study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson. The **MMHS** study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. **MSKCC** is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. We thank Marina Corines and Lauren Jacobs. The work of **MTLGEBCS** was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program – grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade – grant # PSR-SIIRI-701. We would like to thank Martine Tranchant (CHU de Québec Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management and skillful technical assistance. J.S. is Chairholder of the Canada Research Chair in Oncogenetics. The **NBCS** has been supported by the Research Council of Norway grant

193387/V50 (to A-L Børresen-Dale and V.N. Kristensen) and grant 193387/H10 (to A-L Børresen-Dale and V.N. Kristensen), South Eastern Norway Health Authority (grant 39346 to A-L Børresen-Dale and 27208 to V.N.Kristensen) and the Norwegian Cancer Society (to A-L Børresen-Dale and 419616 - 71248 - PR-2006-0282 to V.N. Kristensen). It has received funding from the K.G. Jebsen Centre for Breast Cancer Research (2012-2015). The **NBHS** was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. We thank study partcipants and research staff for their contributions and commitment to this study. The Northern California Breast Cancer Family Registry (**NC-BCFR**) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). 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 Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The **NHS** was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. We would like to thank the participants and staff of the NHS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. The **NHS2** was supported by NIH grants UM1 CA176726 and U19 CA148065. We would like to thank the participants and staff of the NHS2 for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. The **OBCS** was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant number 250083, 122715 and Center of Excellence grant number 251314), the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation and the special Governmental EVO funds for Oulu University Hospital-based research activities. We thank Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri Otsukka, Leena Keskitalo and Kari Mononen for their contributions to this study. The Ontario Familial Breast Cancer Registry (**OFBCR**) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). 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 Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. We thank Teresa Selander and Nayana Weerasooriya. The **ORIGO** study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). We thank E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. The LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCDOC) with the help of Dr. J. Molenaar. The **PBCS** was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. We thank Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, and Michael Stagner. Genotyping for **PLCO** was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from

the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. The **POSH** study is funded by Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, C1275/A15956 and Breast Cancer Campaign 2010PR62, 2013PR044. The ethical approval for the POSH study is MREC /00/6/69, UKCRN ID: 1137. We thank staff in the Experimental Cancer Medicine Centre (ECMC) supported Faculty of Medicine Tissue Bank and the Faculty of Medicine DNA Banking resource. **PREFACE** study thanks Sonja Oeser and Silke Landrith. The **RBCS** was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). We thank Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huijskens, Anja Kromwijk-Nieuwlaat, Annette Heemskerk, the Erasmus MC Family Cancer Clinic. The **SASBAC** study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. We thank The Swedish Medical Research Counsel. The **SBCS** was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now. We thank Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed. The **SCCS** is supported by a grant from the National Institutes of Health (R01 CA092447). Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry. **SEARCH** is funded by a programme grant from Cancer Research UK [C490/A10124] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. We thank The SEARCH and EPIC teams. The Sister Study (**SISTER**) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The Two Sister Study (**2SISTER**) was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES102245), and, also by a grant from Susan G. Komen for the Cure, grant FAS0703856. **SKKDKFZS** is supported by the DKFZ. We thank all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to this study. The **SMC** is funded by the Swedish Cancer Foundation. **SUCCESSB** and **SUCCESSC** thank The SUCCESS Study teams in Munich, Duessldorf, Erlangen and Ulm. The **SZBCS** was supported by Grant PBZ_KBN_122/P05/2004. We thank Ewa Putresza. The **TNBCC** was supported by: a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), a grant from the Breast Cancer Research Foundation, a generous gift from the David F. and Margaret T. Grohne Family Foundation. The **UCIBCS** component of this research was supported by the NIH [CA58860, CA92044] and the Lon V Smith Foundation [LVS39420]. We thank Irene Masunaka. The **UKBGS** is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR

acknowledges NHS funding to the NIHR Biomedical Research Centre. We thank Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. The **UKOPS** study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The **US3SS** study was supported by Massachusetts (K.M.E., R01CA47305), Wisconsin (P.A.N., R01 CA47147) and New Hampshire (L.T.-E., R01CA69664) centers, and Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The **USRT** Study was funded by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services. The **WAABCS** study was supported by grants from the National Cancer Institute of the National Institutes of Health (R01 CA89085 and P50 CA125183 and the D43 TW009112 grant), Susan G. Komen (SAC110026), the Dr. Ralph and Marian Falk Medical Research Trust, and the Avon Foundation for Women. The **WHI** program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01. The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible. A full listing of WHI investigators can be found at:
http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf

Consortia Membership

NBCS Collaborators

Prof. Em. Anne-Lise Børresen-Dale, PhD (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway), Dr. Kristine K. Sahlberg, PhD (Department of Research, Vestre Viken Hospital, Drammen, Norway and Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway), Dr. Lars Ottestad, MD (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway), Prof. Em. Rolf Kåresen, MD (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway and Department of Breast- and Endocrine Surgery, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital Ullevål, Oslo, Norway), Dr. Anita Langerød, PhD (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway), Dr. Ellen Schlichting, MD (Section for breast- and endocrine surgery, Department of oncology, Oslo university hospital), Dr. Marit Muri Holmen, MD (Department of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway), Prof. Toril Sauer, MD (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway and Department of Pathology at Akershus University hospital, Lørenskog, Norway), Dr. Vilde Haakensen, MD (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway), Dr. Olav Engebråten, MD (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway, Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway and Department of Oncology, Division of Surgery and Cancer and Transplantation Medicine, Oslo University Hospital Radiumhospitalet, Oslo, Norway), Prof. Bjørn Naume, MD (Department of Oncology, Oslo University Hospital, Oslo, Norway and Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway), Dr. Alexander Fosså, MD (National Advisory Unit on Late Effects after Cancer Treatment, Department of Oncology, Oslo University Hospital, Oslo, Norway and Department of Oncology, Oslo University Hospital Ullevål, Oslo, Norway), Dr. Cecile E. Kiserud, MD (National Advisory Unit on Late Effects after Cancer Treatment, Department of Oncology, Oslo University Hospital, Oslo, Norway and Department of Oncology, Oslo University Hospital Ullevål, Oslo, Norway), Dr. Kristin V. Reinertsen, MD (National Advisory Unit on Late Effects after Cancer Treatment, Department of Oncology, Oslo University Hospital, Oslo, Norway and Department of Oncology, Oslo University Hospital Ullevål, Oslo, Norway), Assoc. Prof. Åslaug Helland, MD (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway), Dr. Margit Riis, MD (Department of Breast- and Endocrine Surgery, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital Ullevål, Oslo, Norway), Prof. Jürgen Geisler, MD (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway and Department of Oncology, Akershus University Hospital, Oslo, Norway), Prof. Per Eystein Lønning, MD (Section of Oncology, Institute of Medicine, University of Bergen and Department of Oncology, Haukeland University Hospital, Bergen, Norway), Grethe I. Grenaker Alnæs, M.Sc. (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway) and OSBREAC (Oslo Breast Cancer Research Consortium, Oslo University Hospital, Oslo, Norway)

Kathleen Cuningham Foundation Consortium for research into Familial Breast cancer (kConFab)
Morteza Aghmesheh, David Amor, Lesley Andrews, Yoland Antill, Shane Armitage, Leanne Arnold, Rosemary Balleine, Agnes Bankier, Patti Bastick, Jonathan Beesley, John Beilby, Barbara Bennett, Ian Bennett, Geoffrey Berry, Anneke Blackburn, Michael Bogwitz , Meagan Brennan, Melissa Brown, Michael Buckley, Matthew Burgess , Jo Burke, Phyllis Butow, Keith Byron, David Callen, Ian Campbell, Deepa Chauhan, Manisha Chauhan, Georgia Chenevix-Trench, Alice Christian, Christine Clarke, Alison Colley, Dick Cotton, Ashley Crook, James Cui, Bronwyn Culling, Margaret Cummings, Sarah-Jane Dawson, Anna deFazio, Martin Delatycki, Rebecca Dickson, Joanne Dixon, Alexander Dobrovic, Tracy Dudding, Ted Edkins, Stacey Edwards, Maurice Eisenbruch, Gelareh Farshid, Susan Fawcett, Andrew Fellows, Georgina Fenton, Michael Field, Frank Firgaira, James Flanagan, Jean Fleming, Peter Fong, John Forbes, Stephen Fox, Juliet French, Michael Friedlander, Clara Gaff, Mac Gardner, Mike Gattas, Peter George, Graham Giles, Grantley Gill, Jack Goldblatt, Sian Greening, Scott Grist, Eric Haan, Kate Hardie, Marion Harris, Stewart Hart, Nick Hayward, Sue Healey, Louise Heiniger, John Hopper, Evelyn Humphrey, Clare Hunt, Paul James, Mark Jenkins, Alison Jones, Rick Kefford, Alexa Kidd, Belinda Kiely, Judy Kirk, Jessica Koehler, James Kollias, Serguei Kovalenko, Sunil Lakhani, Amanda Leaming, Jennifer Leary, Jacqueline Lim, Geoff Lindeman, Lara Lipton, Liz Lobb, Graham Mann, Deborah Marsh, Sue Anne McLachlan, Bettina Meiser, Cliff Meldrum, Roger Milne, Gillian Mitchell, Beth Newman, Eveline Niedermayr, Sophie Nightingale, Shona O'Connell, Imelda O'Loughlin, Richard Osborne, Nick Pachter, Briony Patterson, Lester Peters, Kelly Phillips, Melanie Price, Lynne Purser, Tony Reeve, Jeanne Reeve, Robert Richards, Edwina Rickard, Bridget Robinson, Barney Rudzki, Mona Saleh, Elizabeth Salisbury, Joe Sambrook, Christobel Saunders, Jodi Saunus, Robyn Sayer, Elizabeth Scott, Rodney Scott, Clare Scott, Ram Seshadri, Adrienne Sexton, Raghwa Sharma, Andrew Shelling, Peter Simpson, Melissa Southey, Amanda Spurdle, Graeme Suthers, Pamela Sykes, Margaret Tassell, Donna Taylor, Jessica Taylor, Benjamin Thierry, Susan Thomas, Ella Thompson, Heather Thorne, Sharron Townshend, Alison Trainer, Lan Tran, Kathy Tucker, Janet Tyler Jane Visvader, Logan Walker, Ian Walpole, Robin Ward, Paul Waring, Bev Warner, Graham Warren, Rachael Williams, Judy Wilson, Ingrid Winship, Kathy Wu, Mary Ann Young

Australian Ovarian Cancer Study (AOCS Management Group):
D Bowtell (Peter MacCallum Cancer Centre), G Chenevix-Trench, A Green, P Webb (QIMRBerghofer), A deFazio (Westmead lnstitute for Cancer Research, WMI), D Gertig (Victorian Cervical Cytology Registry)

References

1. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
2. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-61, 361e1-2 (2013).
3. Couch, F.J. *et al.* Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. *Nat Commun* **7**, 11375 (2016).
4. Lin, W.Y. *et al.* Identification and characterization of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. *Hum Mol Genet* **24**, 285-98 (2015).
5. Ghoussaini, M. *et al.* Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun* **4**, 4999 (2014).
6. Guo, X. *et al.* Fine-scale mapping of the 4q24 locus identifies two independent loci associated with breast cancer risk. *Cancer Epidemiol Biomarkers Prev* **24**, 1680-91 (2015).
7. Ghoussaini, M. *et al.* Evidence that the 5p12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGF10 and MRPS30 Regulation. *Am J Hum Genet* **99**, 903-911 (2016).
8. Bojesen, S.E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* **45**, 371-84, 384e1-2 (2013).
9. Glubb, D.M. *et al.* Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. *Am J Hum Genet* **96**, 5-20 (2015).
10. Li, Q. *et al.* Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. *Hum Mol Genet* **23**, 5294-302 (2014).
11. Li, Q. *et al.* Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* **152**, 633-41 (2013).
12. Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* **47**, 373-80 (2015).
13. Gaudet, M.M. *et al.* Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. *PLoS Genet* **9**, e1003173 (2013).
14. Sun, Y. *et al.* Evaluation of potential regulatory function of breast cancer risk locus at 6q25.1. *Carcinogenesis* **37**, 163-8 (2016).
15. Dunning, A.M. *et al.* Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nat Genet* **48**, 374-86 (2016).
16. Shi, J. *et al.* Fine-scale mapping of 8q24 locus identifies multiple independent risk variants for breast cancer. *Int J Cancer* **139**, 1303-17 (2016).

17. Orr, N. *et al.* Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. *Hum Mol Genet* **24**, 2966-84 (2015).
18. Darabi, H. *et al.* Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate NRBF2 Expression. *Am J Hum Genet* **97**, 22-34 (2015).
19. Meyer, K.B. *et al.* Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet* **93**, 1046-60 (2013).
20. French, J.D. *et al.* 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 (2013).
21. Cai, Q. *et al.* Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet* **46**, 886-90 (2014).
22. Caswell, J.L. *et al.* Multiple breast cancer risk variants are associated with differential transcript isoform expression in tumors. *Hum Mol Genet* **24**, 7421-31 (2015).
23. Lawrenson, K.e.a. Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. *Nat Commun* (2016).
24. Lawrenson, K. *et al.* Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. *Nat Commun* **7**, 12675 (2016).
25. Han, M.R. *et al.* Genome-wide association study in East Asians identifies two novel breast cancer susceptibility loci. *Hum Mol Genet* (2016).