The American Journal of Human Genetics, Volume *96*

Supplemental Data

Fine-Scale Mapping of the 5q11.2 Breast Cancer Locus Reveals at Least Three Independent Risk Variants Regulating *MAP3K1*

Dylan M. Glubb, Mel J. Maranian, Kyriaki Michailidou, Karen A. Pooley, Kerstin B. Meyer, Siddhartha Kar, Saskia Carlebur, Martin O'Reilly, Joshua A. Betts, Kristine M. Hillman, Susanne Kaufmann, Jonathan Beesley, Sander Canisius, John L. Hopper, Melissa C. Southey, Helen Tsimiklis, Carmel Apicella, Marjanka K. Schmidt, Annegien Broeks, Frans B. Hogervorst, C. Ellen van der Schoot, Kenneth Muir, Artitaya Lophatananon, Sarah Stewart-Brown, Pornthep Siriwanarangsan, Peter A. Fasching, Matthias Ruebner, Arif B. Ekici, Matthias W. Beckmann, Julian Peto, Isabel dos-Santos-Silva, Olivia Fletcher, Nichola Johnson, Paul D.P. Pharoah, Manjeet K. Bolla, Qin Wang, Joe Dennis, Elinor J. Sawyer, Ian Tomlinson, Michael J. Kerin, Nicola Miller, Barbara Burwinkel, Frederik Marme, Rongxi Yang, Harald Surowy, Pascal Guénel, Thérèse Truong, Florence Menegaux, Marie Sanchez, Stig E. Bojesen, Børge G. Nordestgaard, Sune F. Nielsen, Henrik Flyger, Anna González-Neira, Javier Benitez, M. Pilar Zamora, Jose Ignacio Arias Perez, Hoda Anton-Culver, Susan L. Neuhausen, Hermann Brenner, Aida Karina Dieffenbach, Volker Arndt, Christa Stegmaier, Alfons Meindl, Rita K. Schmutzler, Hiltrud Brauch, Yon-Dschun Ko, Thomas Brüning, The GENICA Network, Heli Nevanlinna, Taru A. Muranen, Kristiina Aittomäki, Carl Blomqvist, Keitaro Matsuo, Hidemi Ito, Hiroji Iwata, Hideo Tanaka, Thilo Dörk, Natalia V. Bogdanova, Sonja Helbig, Annika Lindblom, Sara Margolin, Arto Mannermaa, Vesa Kataja, Veli-Matti Kosma, Jaana M. Hartikainen, kConFab Investigators, Anna H. Wu, Chiu-chen Tseng, David Van Den Berg, Daniel O. Stram, Diether Lambrechts, Hui Zhao, Caroline Weltens, Erik van Limbergen, Jenny Chang-Claude, Dieter Flesch-Janys, Anja Rudolph, Petra Seibold, Paolo Radice, Paolo Peterlongo, Monica Barile, Fabio Capra, Fergus J. Couch, Janet E. Olson, Emily Hallberg, Celine Vachon, Graham G. Giles, Roger L. Milne, Catriona McLean, Christopher A. Haiman, Brian E. Henderson, Fredrick Schumacher, Loic Le Marchand, Jacques Simard, Mark S. Goldberg, France Labrèche, Martine Dumont, Soo Hwang Teo, Cheng Har Yip, Mee-Hoong See, Belinda Cornes, Ching-Yu Cheng, M. Kamran Ikram, Vessela Kristensen, NBCS, Wei Zheng, Sandra L. Halverson, Martha

Shrubsole, Jirong Long, Robert Winqvist, Katri Pylkäs, Arja Jukkola-Vuorinen, Saila Kauppila, Irene L. Andrulis, Julia A. Knight, Gord Glendon, Sandrine Tchatchou, Peter Devilee, Robert A.E.M. Tollenaar, Caroline Seynaeve, Christi J. Van Asperen, Montserrat García-Closas, Jonine Figueroa, Stephen J. Chanock, Jolanta Lissowska, Kamila Czene, Daniel Klevebring, Hatef Darabi, Mikael Eriksson, Maartje J. Hooning, Antoinette Hollestelle, John W.M. Martens, J. Margriet Collée, Per Hall, Jingmei Li, Keith Humphreys, Xiao-Ou Shu, Wei Lu, Yu-Tang Gao, Hui Cai, Angela Cox, Simon S. Cross, Malcolm W.R. Reed, William Blot, Lisa B. Signorello, Qiuyin Cai, Mitul Shah, Maya Ghoussaini, Daehee Kang, Ji-Yeob Choi, Sue K. Park, Dong-Young Noh, Mikael Hartman, Hui Miao, Wei Yen Lim, Anthony Tang, Ute Hamann, Diana Torres, Anna Jakubowska, Jan Lubinski, Katarzyna Jaworska, Katarzyna Durda, Suleeporn Sangrajrang, Valerie Gaborieau, Paul Brennan, James McKay, Curtis Olswold, Susan Slager, Amanda E. Toland, Drakoulis Yannoukakos, Chen-Yang Shen, Pei-Ei Wu, Jyh-Cherng Yu, Ming-Feng Hou, Anthony Swerdlow, Alan Ashworth, Nick Orr, Michael Jones, Guillermo Pita, M. Rosario Alonso, Nuria Álvarez, Daniel Herrero, Daniel C. Tessier, Daniel Vincent, Francois Bacot, Craig Luccarini, Caroline Baynes, Shahana Ahmed, Catherine S. Healey, Melissa A. Brown, Bruce A.J. Ponder, Georgia Chenevix-Trench, Deborah J. Thompson, Stacey L. Edwards, Douglas F. Easton, Alison M. Dunning, and Juliet D. French

Supplemental Acknowledgments

Manuscript writing group: DMG, MJM, KM, KAP, KBM, AMD, JDF. Locus SNP selection: MG, ED, AMD. iCOGS genotyping, calling and QC: MJM, SEB, SFN, AG-N, MRA, DH, JB, DCT, DV, FB, FR, SA, CL, CB, DC, JC, JD, CSH, JS, AMD, GC-T, DFE. Imputation: KM, DFE. Statistical analyses and programming: KM, KAP, DB, ED, JT, SK, DFE. Functional analysis and bioinformatics: DMG, SLE, KBM, SC, MO, MJM JAB, KMH, SK, JB and JDF. COGS coordination: PH, DFE, JB, AMD, BCAC coordination: DFE, GC-T, PDP. BCAC data management: MKB, QW. All other authors provided participant samples and phenotype information, and read and approved the manuscript. BCAC thanks 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. BCAC is funded by Cancer Research UK [C1287/A10118, C1287/A12014] and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). Meetings of the BCAC have been funded by the European Union COST programme (BM0606). 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, the staff of the CNIO genotyping unit, Sylvie LaBoissière, Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. Genotyping of the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710 and C8197/A16565), 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 of Quebec – grant # PSR-SIIRI-701. This study makes use of data generated by the Molecular Taxonomy of Breast Cancer International Consortium. Funding for the project was provided by Cancer Research UK and the British Columbia Cancer Agency Branch. The QIMR Berghofer group was supported by project grants from the National Health and Medical Research Council of Australia (1021731, 1058415). The work by KBM, MO, SC and BJAP is supported by Cancer Research UK and the Breast Cancer Research Foundation. DFE is a Principal Research Fellow of CR-UK. GCT is an NHMRC Senior Principal Research Fellow. SLE and JDF are supported by Fellowships from the National Breast Cancer Foundation (NBCF) Australia. The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

BCAC investigating groups:

The Australian Breast Cancer Family Study (ABCFS) would like to thank Maggie Angelakos, Judi Maskiell, Gillian Dite. 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) Australia Fellow and a Victorian Breast Cancer Research Consortium Group Leader. M.C.S. is a NHMRC Senior Research Fellow and a Victorian Breast Cancer Research Consortium Group Leader. The ABCS study would like to thank Sten Cornelissen, Richard van Hien, Linde Braaf, Senno Verhoef, Laura van 't Veer, Emiel Rutgers. The **ABCS** study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]; BBMRI-NL, which is a Research Infrastructure financed by the Dutch government (NWO 184.021.007); and the Dutch National Genomics Initiative. The ACP study wishes to thank the participants in the Thai Breast Cancer study. Special Thanks also go to the Thai Ministry of Public Health (MOPH), doctors and nurses who helped with the data collection process. Finally, the study would like to thank Dr Prat Boonyawongviroj, the former Permanent Secretary of MOPH and Dr Pornthep Siriwanarungsan, the Department Director-Generalof Disease Control who have supported the study throughout. The ACP study is funded by the Breast Cancer Research Trust, UK. The BBCC study would like to thank Matthias Rübner, Alexander Hein, Michael Schneider. The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BBCS would like to thank Eileen Williams, Elaine Ryder-Mills, Kara Sargus. The **BBCS** is funded by Cancer Research UK and Breakthrough Breast Cancer and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BIGGS would like to thank Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones. 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. The BSUCH would like to thank Peter Bugert, Medical Faculty Mannheim. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). The CECILE study would like to thank Claire Mulot and Pierre Laurent-Puig (Biobank Saints-Pères Paris), Pierre Kerbrat (Centre Eugène Marquis, Rennes), Patrick Arveux (Centre Georges François Leclerc, DijonThe **CECILE** study was funded by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Ligue contre le Cancer Grand Ouest, Agence Nationale de Sécurité Sanitaire (ANSES), Agence Nationale de la Recherche (ANR). The CGPS study would like to thank Staff and participants of the Copenhagen General Population Study. It would also like to thank Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank and Dorthe Kjeldgård Hansen for the excellent technical assistance. The Danish Breast Cancer Group (DBCG) is acknowledged for the tumor information. The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital. The CNIO-BCS would like to thank Guillermo Pita, Charo Alonso, Daniel Herrero, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO). The CNIO-BCS was supported by the Genome Spain Foundation, 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 PI081120). The Human Genotyping-CEGEN Unit (CNIO) is supported by the Instituto de Salud Carlos III. The CTS study would like to thank the CTS Steering Committee: Leslie Bernstein, Susan Neuhausen, James Lacey, Sophia Wang, Huiyan Ma, Yani Lu, Jessica Clague DeHart at the Beckman Research Institute of City of Hope, Dennis Deapen, Rich Pinder, Eunjung Lee, and Fred Schumacher at the University of Southern California, Pam Horn-Ross, Peggy Reynolds, Christina Clarke Dur and David Nelson at the Cancer Prevention Institute of California, and Hoda Anton-Culver, Argyrios Ziogas, and Hannah Park at the University of California Irvine.. 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). 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 ESTHER study would like to thank Hartwig Ziegler, Sonja Wolf, Volker Hermann. 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). The GENICA Network would like to thank Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany; [HB, Wing-Yee Lo, Christina Justenhoven], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [YDK, 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 **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 HEBCS would like to thank Kirsimari Aaltonen, Karl von Smitten, Sofia Khan, Tuomas Heikkinen, Irja Erkkilä. 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. The **HERPACC** was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports, Culture and Technology of Japan, by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from Ministry Health, Labour and Welfare of Japan, by Health and Labour Sciences Research Grants for Research on Applying Health Technology from Ministry Health, Labour and Welfare of Japan and by National Cancer Center Research and Development Fund. The HMBCS would like to thank Peter Hillemanns, Hans Christiansen and Johann H. Karstens. . The **HMBCS** was supported by a grant from the Friends of Hannover Medical School and by the Rudolf Bartling Foundation. 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 and Bert von Kantzows foundation. The KBCP would like to thank Eija Myöhänen, Helena Kemiläinen. 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. The kConFab/AOCS study would like 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**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. **LAABC** is supported by grants (1RB-0287, 3PB-0102, 5PB-0018, 10PB-0098) from the California Breast Cancer Research Program. Incident breast cancer cases were collected by the USC Cancer Surveillance Program (CSP) which is supported under subcontract by the California Department of Health. The CSP is also part of the National Cancer Institute's Division of Cancer Prevention and Control Surveillance, Epidemiology, and End Results Program, under contract number N01CN25403. The LMBC would like to thank Gilian Peuteman, Dominiek Smeets, Thomas Van Brussel and Kathleen Corthouts. LMBC is supported by the 'Stichting tegen Kanker' (232-2008 and 196-2010). Diether Lambrechts is supported by the FWO and the KULPFV/10/016-SymBioSysII. The MARIE study would like to Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. The **MARIE** study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419], the Hamburg Cancer Society, the German Cancer Research Center and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The MBCSG would like to thank Siranoush Manoukian, Bernard Peissel, Giulietta Scuvera and Daniela Zaffaroni of the Fondazione IRCCS Istituto Nazionale dei Tumori (INT); Bernardo Bonanni and Irene Feroce of the Istituto Europeo di Oncologia (IEO) and Loris Bernard and the personnel of the Cogentech Cancer Genetic Test Laboratory. **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”). The **MCBCS** was supported by the NIH grants CA128978, 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 and the Ting Tsung and Wei Fong Chao Foundation. **MCCS** cohort recruitment in the 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. The MEC was support by NIH grants CA63464, CA54281, CA098758 and CA132839. The MTLGEBCS 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 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. The MYBRCA study would like to thank Phuah Sze Yee, Peter Kang, Kang In Nee, Kavitta Sivanandan, Shivaani Mariapun, Yoon Sook-Yee, Daphne Lee, Teh Yew Ching and Nur Aishah Mohd Taib for DNA Extraction and patient recruitment. **MYBRCA** is funded by research grants from the Malaysian Ministry of Science, Technology and Innovation (MOSTI), Malaysian Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Initiatives Foundation (CARIF). Additional controls were recruited by the Singapore Eye Research Institute, which was supported by a grant from the Biomedical Research Council (BMRC08/1/35/19/550), Singapore and the National medical Research Council, Singapore (NMRC/CG/SERI/2010). The NBCS was supported by grants from the Norwegian Research council, 155218/V40, 175240/S10 to ALBD, FUGE-NFR 181600/V11 to VNK and a Swizz Bridge Award to ALBD. The NBHS_TN study would like to thank participants and research staff for their contributions and commitment to this study. 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. The OBCS study would like to thank Meeri Otsukka, Kari Mononen, Mervi Grip. 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. The OFBCR study would like to thank Teresa Selander, Nayana Weerasooriya. 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. The ORIGO study would like to 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 ORIGO study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS study would like to thank Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The pKARMA study would like to thank The Swedish Medical Research Counsel. The pKARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. The RBCS study would like to thank Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huijskens, Annette Heemskerk, the Erasmus MC Family Cancer Clinic. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The SASBAC would like to thank The Swedish Medical Research Counsel. 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. The SBCGS would like to thank participants and research staff for their contributions and commitment to this study. The SBCGS was supported primarily by NIH grants R01CA64277, R01CA148667, and R37CA70867. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The scientific development and funding of this project were, in part, supported by the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network U19 CA148065. The SBCS would like to thank Sue Higham, Helen Cramp, Sabapathy Balasubramanian, Ian Borck and Dan Connley. The **SBCS** was supported by Yorkshire Cancer Research S295, S299, S305PA and Sheffield Exeprimental Cancer Medicine Centre. 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.The SEARCH study would like to thank The SEARCH and EPIC teams. **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. **SEBCS** was supported by the BRL (Basic Research Laboratory) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2012-0000347). **SGBCC** is funded by the National Medical Research Council start-up Grant and National University Cancer Institute Singapore (NCIS) Centre Grant. The study would like to thank the participants and research coordinator Kimberley Chua. Additional controls were recruited by the Singapore Consortium of Cohort Studies-Multi-ethnic cohort (SCCS-MEC), which was funded by the Biomedical Research Council, grant number: 05/1/21/19/425. The SKKDKFZS study thanks all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to this study. SKKDKFZS is supported by the DKFZ. The SZBCS was supported by Grant PBZ_KBN_122/P05/2004. The TBCS was funded by The National Cancer Institute Thailand. 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 Stefanie Spielman Breast Cancer fund and the OSU Comprehensive Cancer Center, the Hellenic Cooperative Oncology Group research grant (HR R_BG/04) and the Greek General Secretary for Research and Technology (GSRT) Program, Research Excellence II, the European Union (European Social Fund – ESF), and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - ARISTEIA. The **TWBCS** is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences, Academia Sinica, Taiwan. The UKBGS study would like to thank Breakthrough Breast Cancer 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 UKBGS is funded by Breakthrough Breast Cancer and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.

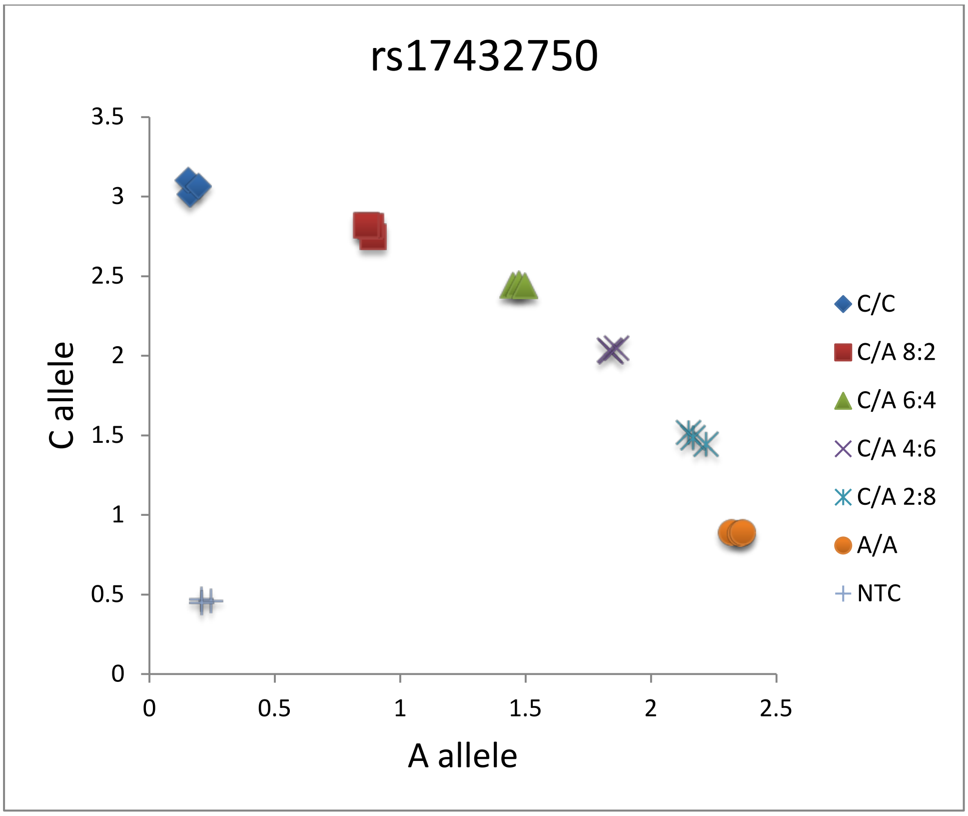
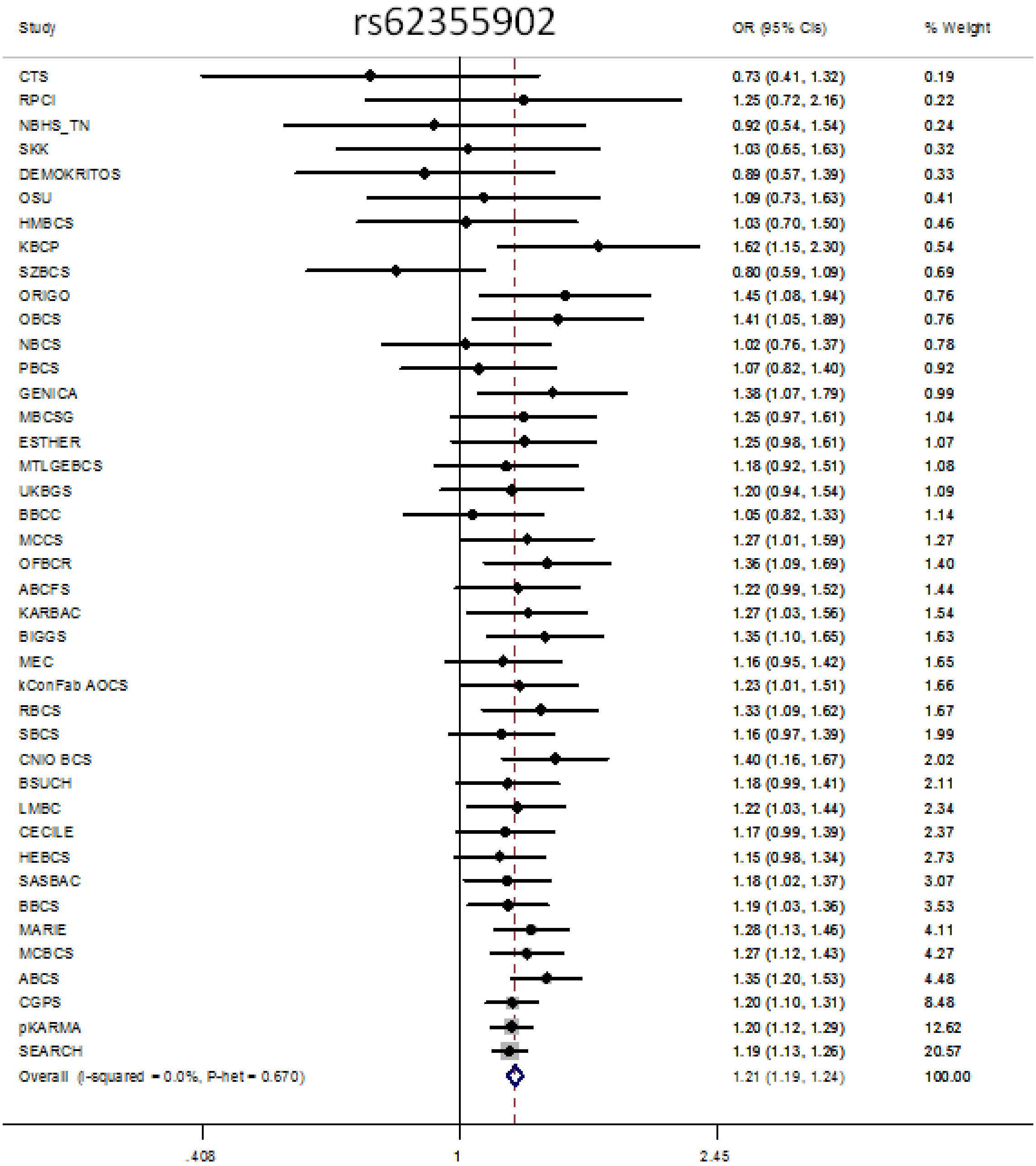
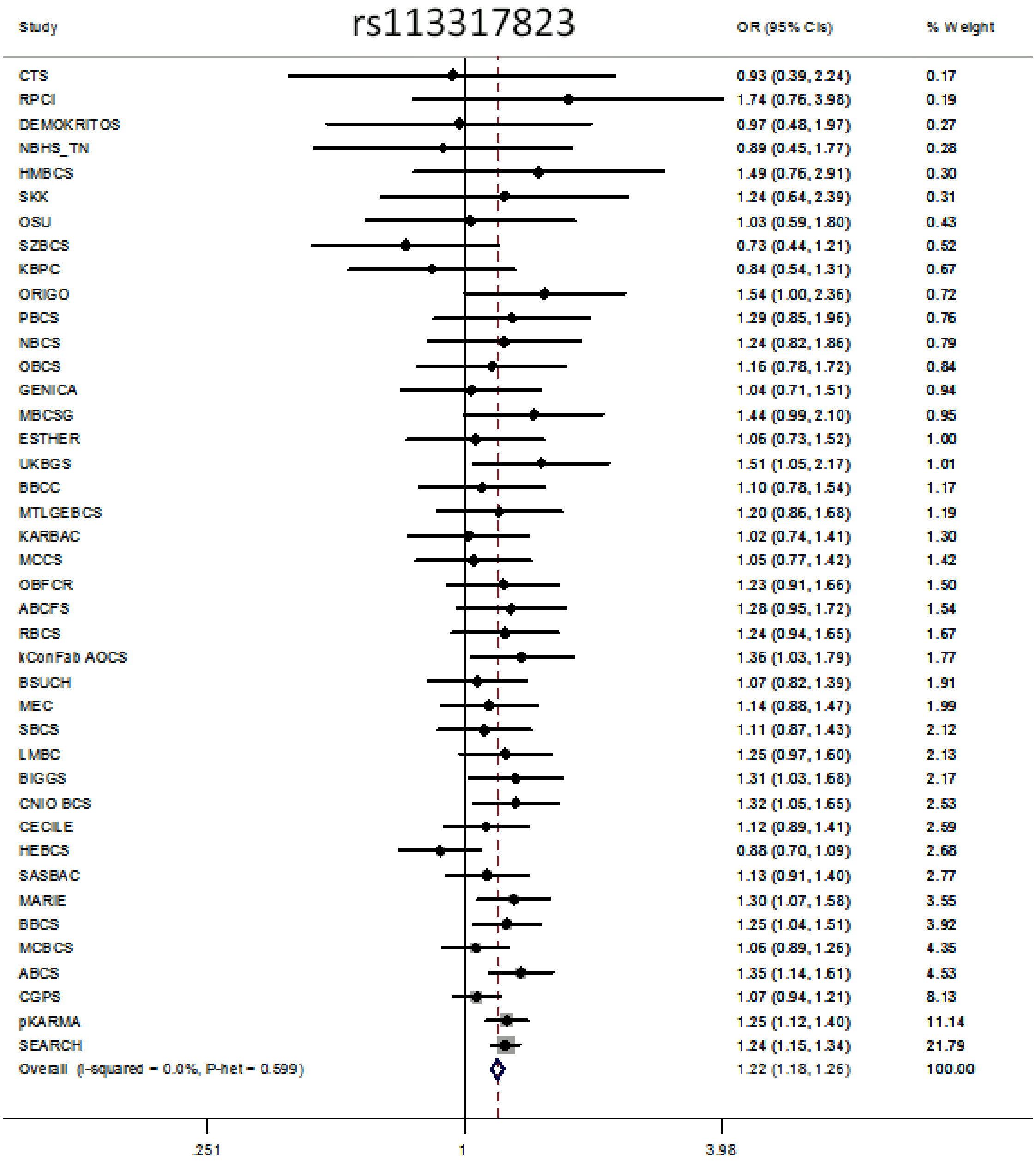


Figure S1. Allelic Discrimination plot of known ratios of genomic DNA from breast cell lines carrying the C/C (ZR751) or A/A (MDA-MB-468) at rs17432750.



Study	OR (95% CIs)	% Weight
CTS	0.73 (0.41, 1.32)	0.19
RPCI	1.25 (0.72, 2.16)	0.22
NBHS_TN	0.92 (0.54, 1.54)	0.24
SKK	1.03 (0.65, 1.63)	0.32
DEMOKRITOS	0.89 (0.57, 1.39)	0.33
OSU	1.09 (0.73, 1.63)	0.41
HMBCS	1.03 (0.70, 1.50)	0.46
KBCP	1.62 (1.15, 2.30)	0.54
SZBCS	0.80 (0.59, 1.09)	0.69
ORIGO	1.45 (1.08, 1.94)	0.76
OBCS	1.41 (1.05, 1.89)	0.76
NBCS	1.02 (0.76, 1.37)	0.78
PBCS	1.07 (0.82, 1.40)	0.92
GENICA	1.38 (1.07, 1.79)	0.99
MBCSG	1.25 (0.97, 1.61)	1.04
ESTHER	1.25 (0.98, 1.61)	1.07
MTLGEBCS	1.18 (0.92, 1.51)	1.08
UKBGS	1.20 (0.94, 1.54)	1.09
BBCC	1.05 (0.82, 1.33)	1.14
MCCS	1.27 (1.01, 1.59)	1.27
OFBCR	1.36 (1.09, 1.69)	1.40
ABCFS	1.22 (0.99, 1.52)	1.44
KARBAC	1.27 (1.03, 1.56)	1.54
BIGGS	1.35 (1.10, 1.65)	1.63
MEC	1.16 (0.95, 1.42)	1.65
kConFab AOCS	1.23 (1.01, 1.51)	1.66
RBCS	1.33 (1.09, 1.62)	1.67
SBCS	1.16 (0.97, 1.39)	1.99
CNIO BCS	1.40 (1.16, 1.67)	2.02
BSUCH	1.18 (0.99, 1.41)	2.11
LMBC	1.22 (1.03, 1.44)	2.34
CECILE	1.17 (0.99, 1.39)	2.37
HEBCS	1.15 (0.98, 1.34)	2.73
SASBAC	1.18 (1.02, 1.37)	3.07
BBCS	1.19 (1.03, 1.36)	3.53
MARIE	1.28 (1.13, 1.46)	4.11
MCBCS	1.27 (1.12, 1.43)	4.27
ABCS	1.35 (1.20, 1.53)	4.48
CGPS	1.20 (1.10, 1.31)	8.48
pKARMA	1.20 (1.12, 1.29)	12.62
SEARCH	1.19 (1.13, 1.26)	20.57
Overall (I-squared = 0.0%, P-het = 0.670)	1.21 (1.19, 1.24)	100.00



Study	OR (95% CIs)	% Weight
CTS	0.93 (0.39, 2.24)	0.17
RPCI	1.74 (0.76, 3.98)	0.19
DEMOKRITOS	0.97 (0.48, 1.97)	0.27
NBHS_TN	0.89 (0.45, 1.77)	0.28
HMBCS	1.49 (0.76, 2.91)	0.30
SKK	1.24 (0.64, 2.39)	0.31
OSU	1.03 (0.59, 1.80)	0.43
SZBCS	0.73 (0.44, 1.21)	0.52
KBPC	0.84 (0.54, 1.31)	0.67
ORIGO	1.54 (1.00, 2.36)	0.72
PBCS	1.29 (0.85, 1.96)	0.76
NBCS	1.24 (0.82, 1.86)	0.79
OBCS	1.16 (0.78, 1.72)	0.84
GENICA	1.04 (0.71, 1.51)	0.94
MBCSG	1.44 (0.99, 2.10)	0.95
ESTHER	1.06 (0.73, 1.52)	1.00
UKBGS	1.51 (1.05, 2.17)	1.01
BBCC	1.10 (0.78, 1.54)	1.17
MTLGEBCS	1.20 (0.86, 1.68)	1.19
KARBAC	1.02 (0.74, 1.41)	1.30
MCCS	1.05 (0.77, 1.42)	1.42
OBFCR	1.23 (0.91, 1.66)	1.50
ABCFS	1.28 (0.95, 1.72)	1.54
RBCS	1.24 (0.94, 1.65)	1.67
kConFab AOCS	1.36 (1.03, 1.79)	1.77
BSUCH	1.07 (0.82, 1.39)	1.91
MEC	1.14 (0.88, 1.47)	1.99
SBCS	1.11 (0.87, 1.43)	2.12
LMBC	1.25 (0.97, 1.60)	2.13
BIGGS	1.31 (1.03, 1.68)	2.17
CNIO BCS	1.32 (1.05, 1.65)	2.53
CECILE	1.12 (0.89, 1.41)	2.59
HEBCS	0.88 (0.70, 1.09)	2.68
SASBAC	1.13 (0.91, 1.40)	2.77
MARIE	1.30 (1.07, 1.58)	3.55
BBCS	1.25 (1.04, 1.51)	3.92
MCBCS	1.06 (0.89, 1.26)	4.35
ABCS	1.35 (1.14, 1.61)	4.53
CGPS	1.07 (0.94, 1.21)	8.13
pKARMA	1.25 (1.12, 1.40)	11.14
SEARCH	1.24 (1.15, 1.34)	21.79
Overall (I-squared = 0.0%, P-het = 0.599)	1.22 (1.18, 1.26)	100.00

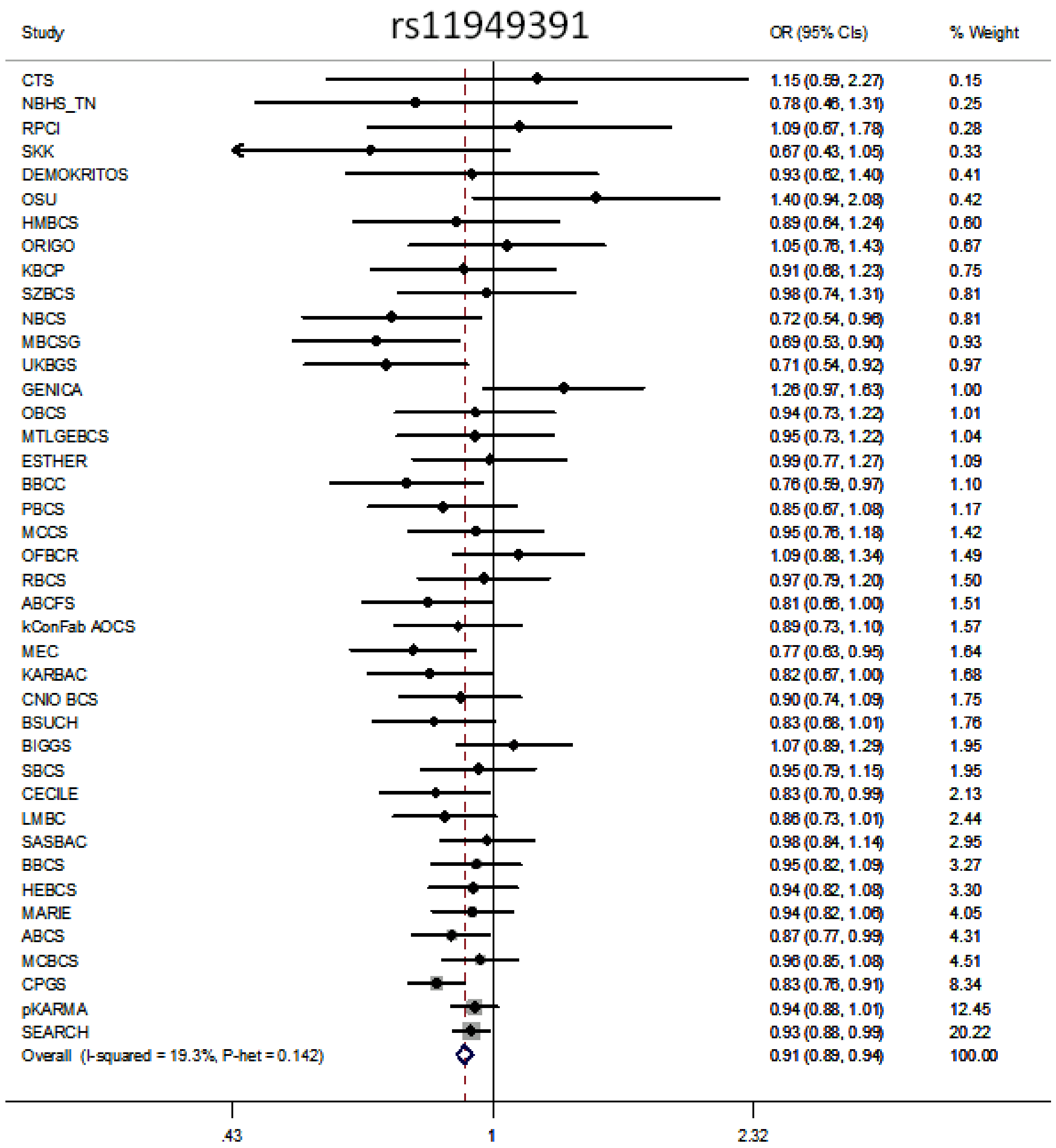


Figure S2. Forest plots for the three independently-associated SNPs marking iCHAVs 1-3.

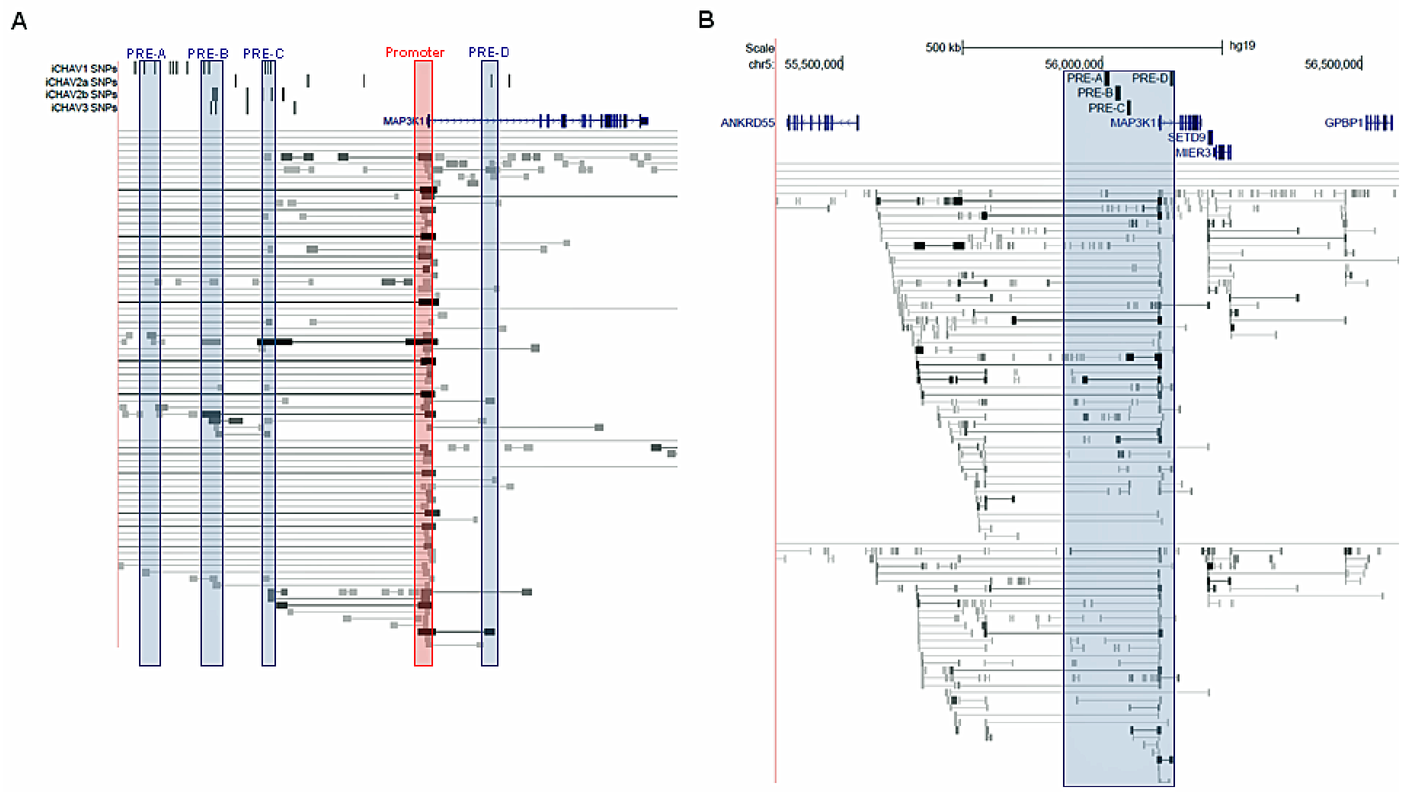


Figure S3. RNA polymerase II ChIA-PET at the 5q11.2 breast cancer risk locus in MCF7 cells shows chromatin interactions between discrete genomic regions and the *MAP3K1* promoter. ENCODE data (GEO sample accession GSM970209) was accessed using the UCSC genome browser. iCHAV variants and PREs are shown. Panel (A) shows the zoomed in region containing the PREs and *MAP3K1*. Panel (B) shows the 5q11.2 risk locus and flanking genes.

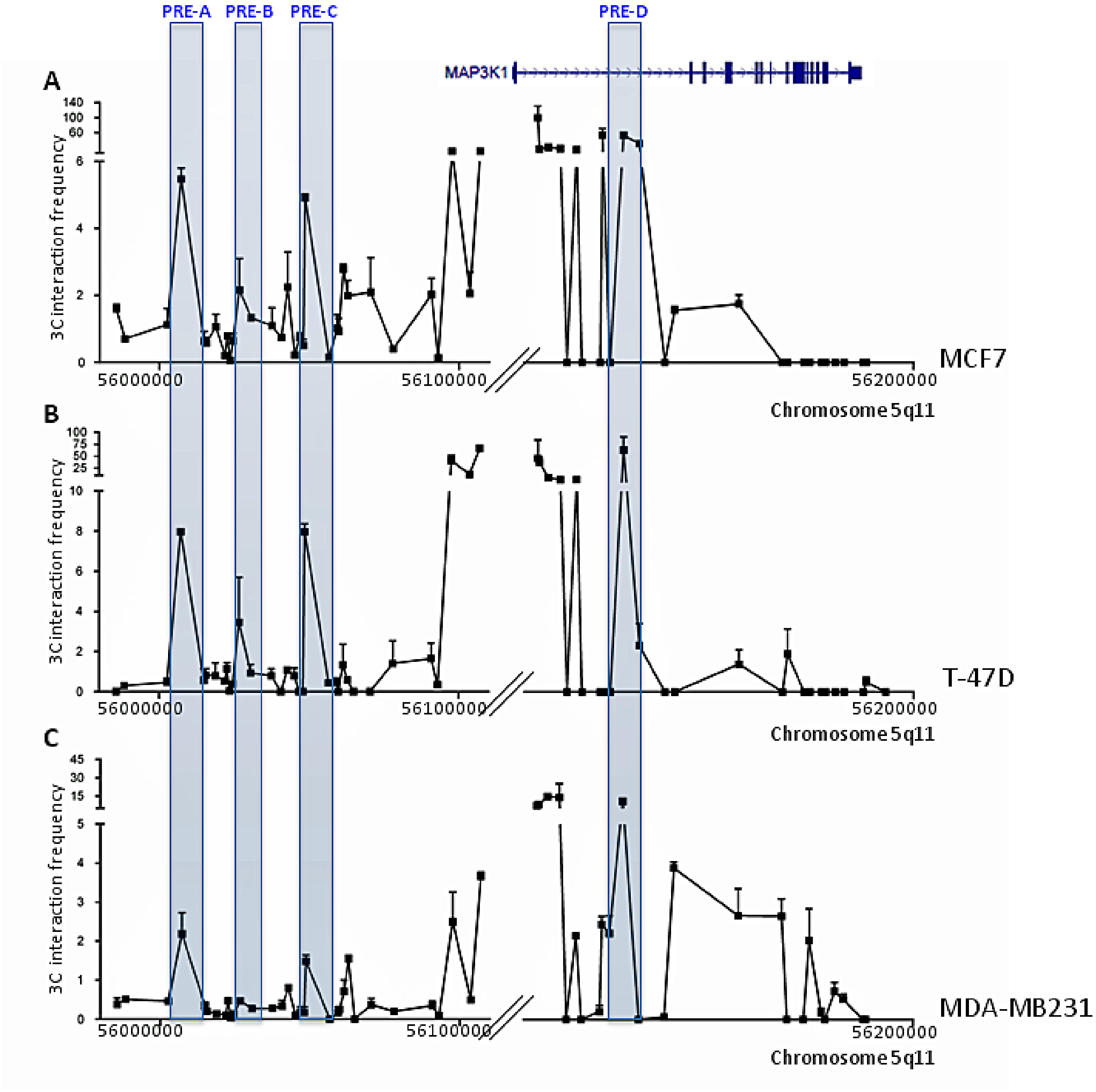


Figure S4. Chromatin interactions with the *MAP3K1* promoter in breast cancer cell lines at the 5q11.2 risk locus. 3C libraries were generated with EcoRI. Chromatin interaction frequencies were plotted at the chromosomal position of the corresponding EcoRI fragments for MCF7 **(A)**, T-47D **(B)** and MDA-MB231 **(C)** libraries, respectively. PREs are highlighted at their chromosomal locations. A representative graph of at least two biological replicates is shown and error bars represent SD.

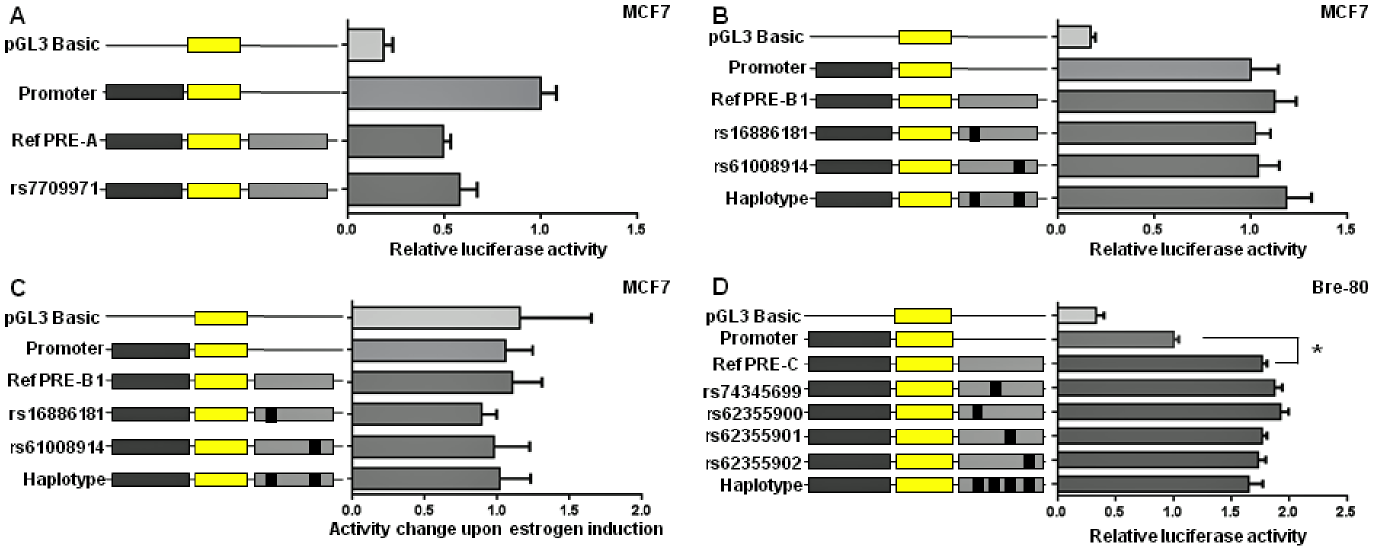


Figure S5. Luciferase reporter assays of PRE constructs containing iCHAV1 SNPs. MCF7 cells were transiently transfected with PRE-A (A) or PRE-B1 (B) and Bre-80 cells with PRE-C (D) constructs under basal conditions and assayed for luciferase activity after 24 h. Panel (C) shows results from luciferase assays after estrogen induction of MCF7 cells transfected with PRE-B1 constructs. For each reporter construct, the luciferase activity of estrogen treated cells was normalized to the activity of the corresponding vehicle treated cells. Error bars denote SEM from three experiments performed in triplicate. *P*-values were determined by repeated-measures ANOVA followed by Dunnett's multiple comparisons test (**P*<0.05).

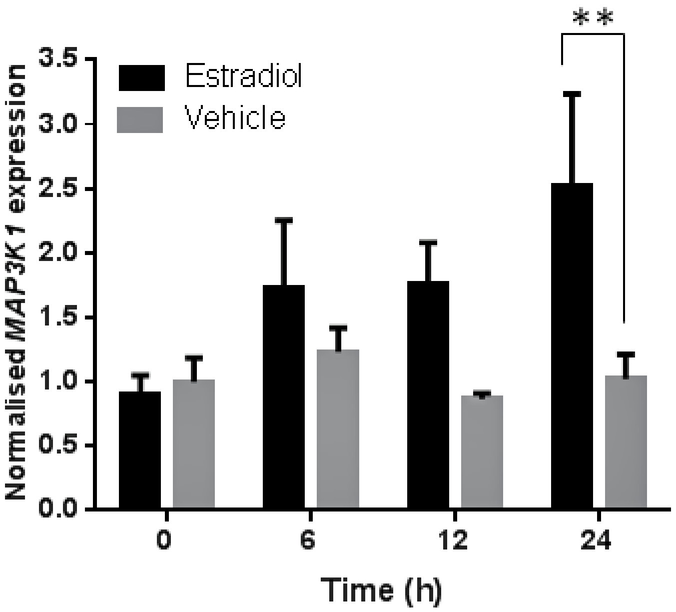


Figure S6. ***MAP3K1*** **expression is induced in MCF7 cells after estrogen treatment.** MCF7 cells were treated with either 100 nM estradiol or vehicle for 24 h. *MAP3K1* expression was measured at baseline and after 6, 12 and 24 h by qPCR and normalized using *GUS* expression as an internal control. Error bars denote SEM from three experiments performed in triplicate. Differences in expression between vehicle and estradiol treated cells were tested using a two-way repeated measures ANOVA followed by Sidak's multiple comparisons test (***P*<0.01).

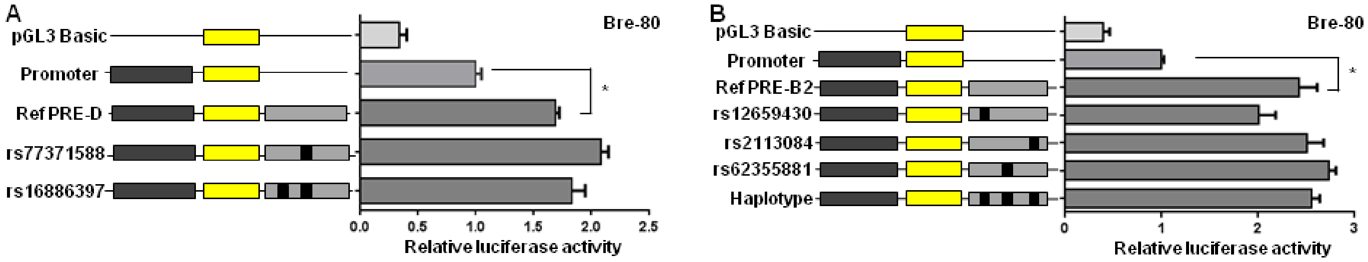


Figure S7. Luciferase reporter assays of PRE-D constructs containing iCHAV2a SNPs and PRE-B2 constructs containing iCHAV2b variants in Bre-80 cells. Bre-80 cells were transfected with PRE-D and PRE-B2 constructs. Cells were assayed for luciferase activity after 24 h. Error bars denote SEM from three experiments performed with triplicates. Statistical significance was determined by repeated-measures ANOVA followed by Dunnett's multiple comparisons test (*P<0.05).

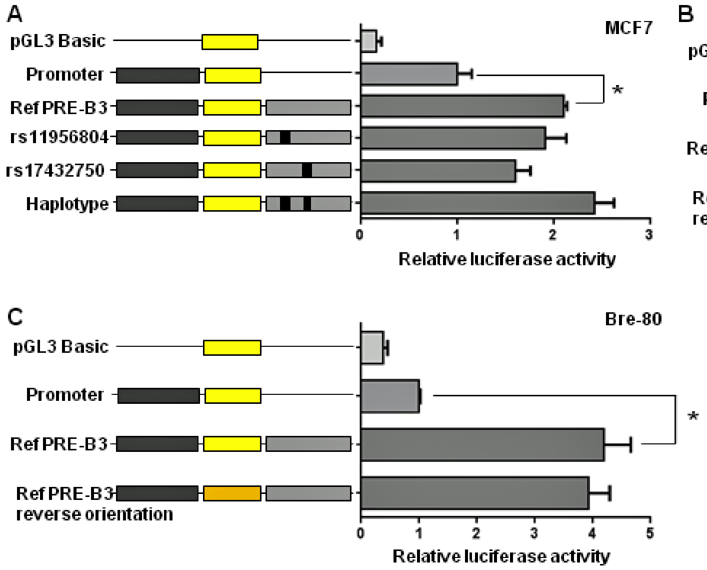


Figure S8. Luciferase reporter assays of PRE-B3 constructs containing iCHAV3 SNPs in MCF7 and Bre-80 cells. MCF7 cells were transiently transfected the PRE-B3 reference and variant luciferase constructs (A) or the PRE-B3 reference construct in the reverse orientation (B). Bre-80 cells were also transiently transfected with the reference PRE-B3 construct in the reverse orientation (C). Cells were assayed for luciferase activity after 24 h. Error bars denote SEM from three experiments performed with triplicates. *P*-values were determined by ANOVA followed by Dunnett's multiple comparisons test (**P*<0.05).

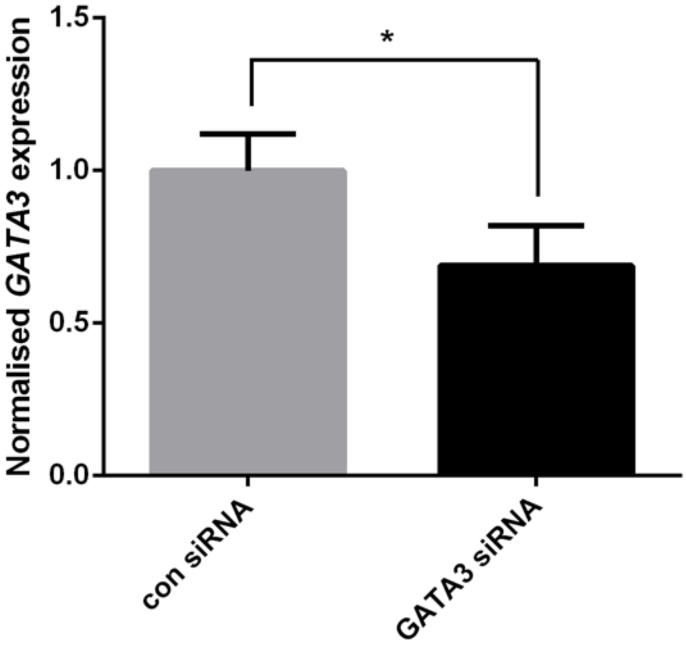


Figure S9. TaqMan qPCR assay confirming knockdown of *GATA3* in Bre80 cells. *GATA3* expression from cells transfected with non-targeting siRNA and *GATA3* is denoted by con siRNA and GATA3 siRNA, respectively. Error bars denote SEM from three experiments performed with triplicates. Statistical significance was determined by a paired t-test (*P<0.05).

Table S1. 3C qPCR primers

3C primers	EcoRI site (GRCh37 chr 5 coordinates)	Sequence
Fragment 1	55,988,536	aaccttttcctgttctgcttaaggtgggtgg
Fragment 2	56,001,687	aagacagccaaagactctccaattccttctgc
Fragment 3	56,002,653	cactggctcagtccctcattgattcagtcc
Fragment 4	56,015,010	caaagaaacgagctctaagaggtgggcagc
Fragment 5	56,015,005	gaggcagatttccctggaaccttcttttcc
Fragment 6	56,015,605	cctgagtctcattccctcttctgcaagagacc
Fragment 7	56,021,916	aagagtctcgtctgtcagacaaaatgtcacgg
Fragment 8	56,022,688	aagaatcactggggtgttattgctggaaagg
Fragment 9	56,023,585	ctcatggcctatgtacctccaaaggctcc
Fragment 10	56,024,547	catgtaactgctgggtttcattccagttcagg
Fragment 11	56,026,749	ggttctatgtgaaggtgccctccaaaacc
Fragment 12	56,030,623	aatccagcacctgctatgaacacctcatcc
Fragment 13	56,037,547	cctaagtcttctccagagcattttgaccctgc
Fragment 14	56,040,674	atcaggaggactacaatctaggggtatgctgagg
Fragment 15	56,042,804	gcaggacttctttgtcagcatctttctctttcc
Fragment 16	56,045,182	tgggactgtgggagacacacagttttgc
Fragment 17	56,046,719	agcaatcagtaatgggaaaagatgccatgc
Fragment 18	56,047,998	ctcttctgccttgttcctgaccttgtgtcc
Fragment 19	56,048,601	tctaggtgtgtctgcaagggcatttctgg
Fragment 20	56,056,381	gccagtgtgtttgattccaaacctctgagc
Fragment 21	56,056,646	tggaacatgctttcactaaagccttctaggaagc
Fragment 22	56,059,823	gcagtgaatgaggttgcacacagatctcc
Fragment 23	56,061,410	gcgtttgaactatccagagttaacacgcatcc
Fragment 24	56,062,862	aaggtgctgtcttatgaaagaggaaaccgagg
Fragment 25	56,064,941	ggaggaaagagtaggagacggctgttcagc
Fragment 26	56,070,521	agagtaagtaatgacctgcccacggcagg
Fragment 27	56,078,074	gacgaggtagggataattagtgggaaccttaggc
Fragment 28	56,090,803	tcatagtgggtatggaatggtattgcattgtgc
Fragment 29	56,093,049	gtgaacacaagccgagattttcagaaggtagc
Fragment 30	56,097,598	ctatcctaccaatttgccagctaagtgatttggg
Fragment 31	56,103,646	cagtttccctgatcccctattcccgc
Fragment 32	56,107,073	ttacaggaggatcttgggaaaatgtatgaaccc
Fragment 33	56,108,495	tctagccccagcactcctgcaagtattcc
Fragment 34	56,109,036	gcatgcgtgaaatcattgcaaggtttacg
Bait (promoter fragment)	56,112,866	gcttaactgaaaagggtgttctccctctgc
Fragment 37	56,113,523	gcactgtaccatggcactactgagaccaaagg
Fragment 38	56,115,580	tttgtactccagtcatcttcttgggttgaggg
Fragment 39	56,118,301	cggccagaccattgcttgatgtttaatagc
Fragment 40	56,119,851	aatgagagaaggcctaggaccaccctgg
Fragment 41	56,121,912	cgtcatttcagcatcgtgctcagttttagtcc
Fragment 42	56,123,317	gatttacatcagagctgggacaagacagtctgg
Fragment 43	56,127,536	ttcgagtttaatcagaccaactatcaggcaggc
Fragment 44	56,128,078	gaccgttccccattactgtcctaacactttgg
Fragment 45	56,129,808	aactaggatgattgaggctggatcccagacc
Fragment 46	56,132,919	cataagcagttcacataaaagacttatgggccagc

Fragment 47	56,136,559	gaagcaacccaagaggaagaaagggaactagc
Fragment 48	56,142,493	cttcacctcctgctttttctcctcacctgc
Fragment 49	56,144,772	gtaaatggagttattgtgcatgaccatgtgg
Fragment 50	56,159,585	ggacatttgagctgctcctgtttggtgg
Fragment 51	56,169,636	gaagctgacaaaaacgctcctagggtgagg
Fragment 52	56,170,833	gatgggtcatgaaagggcaggtgagg
Fragment 53	56,174,572	tgtaggctcctgggtattcattttgttctgtgg
Fragment 54	56,175,998	cctctgcaagtcataatgacttagtttgcacagg
Fragment 55	56,178,759	gtctcaggatgccctccccatagttcc
Fragment 56	56,179,669	ggctcaagatgtgggaactggaactttaatgg
Fragment 57	56,181,895	cggtggctcatttgctgagtaaatatggagc
Fragment 58	56,183,388	agattttggagctgcagccaggttgg
Fragment 59	56,188,350	ccatacatcacagctccttccatactgttgacc
Fragment 60	56,189,045	tactaagaatcagaatttgtccaacacggaagg
Fragment 61	56,193,516	gtctatgaaaaaattcatccactggatgggagg
Fragment 62	56,211,583	ttttctggattctgacatgaagactgtcaccc

Table S2. Reporter gene construct cloning and overlapping mutagenesis primers

Primer	Sequence
MAP3K1 promoter forward	acgcgtccacctctctgcagtaacatag
MAP3K1 promoter reverse	aagcttgcttcctctcggcaatctcg
rs74762363 forward	ccagtttcaaccactgaacaaacagtcaaaaatcag
rs74762363 reverse	ctgatttttgactgtttgttcagtggttgaaactgg
PRE-A forward	accggtgctcaactagccaaagcactcc
PRE-A reverse	cctgcagggggcaagtctaaagcagtgtgg
rs7709971 forward	cagttcagccactgtggaaagcagtgagatctggag
rs7709971 reverse	ctccagatctcactgctttccacagtggctgaactg
PRE-B3 forward	accggtggaagatgggacaagccttgatgc
PRE-B3 reverse	cctgcaggggaagaacaacctgtctcaatgatgg
rs17432750 forward	caacctggattctttcactaatcacacaagtcagg
rs17432750 reverse	cctgacttgtgtgattagtgaaagaatccaggttg
rs11956804 forward	gctgagcagtttatctttgtcatatttagtaggatgaatg
rs11956804 reverse	cattcatcctactaaatatgacaaagataaactgctcagc
PRE2c forward	accggtaaaagatagctttcaaagg
PRE2c reverse	gcgtcgaccatagttacttcaaagg
rs12659430 forward	catgcatttggatgtgtcctataaaag
rs12659430 reverse	cttttataggacacatccaaatgcatg
rs62355881 forward	aaagtcacggatgcttctggtagc
rs62355881 reverse	gctaccagaagcatccgtgacttt
rs2113084 forward	taaatttgtggcatgcaaatattaaaac
rs2113084 forward	gttttaatatttgcatgccacaaattta
PRE-C forward	accggtccaagtttcatgcatggctctgtgg
PRE-C reverse	cctgcagggtggcctctttccagtacagtgg
rs74345699 forward	ggctcacgcctgtaatctcagcactttgg
rs74345699 reverse	ccaaagtgctgagattacaggcgtgagcc
rs62355900 forward	tgggggagcatccgaggtggatgaag
rs62355900 reverse	cttcatccacctcggatgctcccca
rs62355901 forward	ggagatcaagaccatcccggctagcacg
rs62355901 reverse	cgtgctagccgggatggtcttgatctcc
rs62355902 forward	tgggcaacagagcgagactccatctcaaaaaaca
rs62355902 reverse	tgtttttgagatggagtctcgctctgttgccca
PRE-D forward	accggtggccattatagcagtgctctttgc
PRE-D reverse	cctgcaggggctgatgcctagtagtcaattaagc
rs77371588 forward	ctgggcagtgagcccgtcttccagtg
rs77371588 reverse	cactggaagacgggctcactgcccag

Underlined sequences correspond to restriction enzyme recognition sites and bases highlighted in red denote SNP loci.

Table S3. Primer pairs used in ChIP-RT-PCR and sequence confirmation

Primer	Sequence	Ref or comment
rs17432750-for1	GGCCATCTGTTTTACCAACC	RT-PCR
rs17432750-rev1	ATTTGCACATGCCTTTCTCC	RT-PCR
rs17432750-for2	CAATGCAAATCTTCCTTGCTT	RT-PCR & sequencing
rs17432750-rev2	TGGGAAGGAGTCGTTGAGTT	RT-PCR & sequencing
ER-α-ENH-for	TGTAGGCTAGTTTTGTTTAACGATTTTT	Cancer Research 2007; 67:6477-6483
ER-α-ENH-rev	GGTGATGGGAGAATTGCTTAGAA	as above
CCND1-for	TGCCACACACCAGTGACTTT	Genes Dev. 2006; 20:2513-2526.
CCND1-rev	ACAGCCAGAAGCTCCAAAAA	as above

Table S6. Associations of the iCHAV representative SNPs with breast cancer risk in Asian and African American studies

		Asian Studies						African American Studies					
SNP	**Position**	**MAF**	**Imp r^2**	***P*-trend**	**OR**	**LCI**	**UCI**	**MAF**	**Imp r^2**	***P*-trend**	**OR**	**LCI**	**UCI**
iCHAV1 rs62355902	56053723	0.38	0.94	3.30E-02	1.06	1.00	1.12	0.09	0.98	2.30E-01	1.14	0.92	1.36
iCHAV2a rs113317823	56087883	0.13	0.88	1.40E-05	1.19	1.11	1.27	0.06	0.78	7.80E-01	1.04	0.77	1.31
iCHAV2b rs62355899	56050465	0.26	1.00	7.46E-01	0.99	0.93	1.05	0.04	1.00	9.60E-01	0.99	0.72	1.36
iCHAV2c rs7721581	56087883	0.56	1.00	3.80E-01	0.98	0.93	1.03	0.06	0.78	7.78E-01	1.04	0.78	1.38
iCHAV3 rs11949391	56045081	0.05	1.00	2.60E-02	0.87	0.75	0.99	0.10	1.00	1.20E-01	0.84	0.62	1.06