Identification and characterisation of novel associations in the *CASP8/ALS2CR12* region on chromosome 2 with breast cancer risk

Authors: Wei-Yu Lin[1,2], Nicola J Camp[3], Maya Ghoussaini[4], Jonathan Beesley[5], Kyriaki Michailidou[4], John L. Hopper[6], Carmel Apicella[6], Melissa C. Southey[7], Jennifer Stone[8], Marjanka K Schmidt[9], Annegien Broeks[9], Laura J Van't Veer[9], Emiel J Th Rutgers[9] Kenneth Muir[10,11], Artitaya Lophatananon[10], Sarah Stewart-Brown[10], Pornthep Siriwanarangsan[12], Peter A. Fasching[13,14], Lothar Haeberle[13], Arif B. Ekici[13,15], Matthias W. Beckmann[13], Julian Peto[16], Isabel dos-Santos-Silva[16], Olivia Fletcher[17], Nichola Johnson[17], Manjeet K. Bolla[4], Qin Wang[4], Joe Dennis[4], Elinor J. Sawyer[18], Timothy Cheng[19], Ian Tomlinson[19], Michael J. Kerin[20], Nicola Miller[20], Frederik Marmé[21,22], Harald M. Surowy[21,23], Barbara Burwinkel[21,23], Pascal Guénel[24,25], Thérèse Truong[24,25], Florence Menegaux[24,25], Claire Mulot[26], Stig E. Bojesen[27-29], Børge G. Nordestgaard[27-29], Sune F Nielsen[27,28], Henrik Flyger[30], Javier Benitez[31,32], M. Pilar Zamora[33], Jose Ignacio Arias Perez[34], Primitiva Menéndez[35], Anna González-Neira[36], Guillermo Pita[36], M. Rosario Alonso[36], Nuria Álvarez[36], Daniel Herrero[36], Hoda Anton-Culver[37], Hermann Brenner[38,39], Aida Karina Dieffenbach[38,39], Volker Arndt[38], Christa Stegmaier[40], Alfons Meindl[41], Peter Lichtner[42], Rita K. Schmutzler[43], Bertram Müller-Myhsok[44], Hiltrud Brauch[45-47], Thomas Brüning[48], Yon-Dschun Ko[49], The GENICA Network[47,50], Daniel C. Tessier[51], Daniel Vincent[51], Francois Bacot[51], Heli Nevanlinna[52,53], Kristiina Aittomäki[53,54], Carl Blomqvist[53,55], Sofia Khan[52,53], Keitaro Matsuo[56], Hidemi Ito[57], Hiroji Iwata[58], Akiyo Horio[58], Natalia V. Bogdanova[59,60], Natalia N. Antonenkova[61], Thilo Dörk[59], Annika Lindblom[62], Sara Margolin[63], Arto Mannermaa[64,65], Vesa Kataja[66,67], Veli-Matti Kosma[64,65], Jaana M. Hartikainen[64,65], kConFab Investigators[68], Australian Ovarian Cancer Study Group[68,69], Anna H. Wu[70], Chiu-chen Tseng[70], David Van Den Berg[70], Daniel O. Stram[70], Patrick Neven[71], Els Wauters[72,73], Hans Wildiers[74], Diether Lambrechts[72,75], Jenny Chang-Claude[76], Anja Rudolph[76], Petra Seibold[76], Dieter Flesch-Janys[77], Paolo Radice[78],

Paolo Peterlongo[79], Siranoush Manoukian[80], Bernardo Bonanni[81], Fergus J. Couch[82], Xianshu Wang[82], Celine Vachon[83], Kristen Purrington[84], Graham G. Giles[6,85], Roger Milne[6,85], Catriona McLean[86], Christopher A. Haiman[70], Brian E. Henderson[70], Fredrick Schumacher[70], Loic Le Marchand[87], Jacques Simard[88], Mark S. Goldberg[89,90], France Labrèche[91], Martine Dumont[88], Soo Hwang Teo[92,93], Cheng Har Yip[93], Norhashimah Hassan[92,93], Eranga Nishanthie Vithana[94], Vessela Kristensen[95,96], Wei Zheng[97], Sandra Deming-Halverson[97], Martha Shrubsole[97], Jirong Long[97], Robert Winqvist[98], Katri Pylkäs[98], Arja Jukkola-Vuorinen[99], Saila Kauppila[100], Irene L. Andrulis[101,102], Julia A. Knight[103,104], Gord Glendon[101], Sandrine Tchatchou[105], Peter Devilee[106], Robert A.E.M. Tollenaar[107], Caroline Seynaeve[108], Christi J. Van Asperen[109], Montserrat García-Closas[110,111], Jonine Figueroa[112], Jolanta Lissowska[113], Louise Brinton[112], Kamila Czene[114], Hatef Darabi[114], Mikael Eriksson[114], Judith S. Brand[114], Maartje J. Hooning[115], Antoinette Hollestelle[115], Ans M.W. van den Ouweland[116], Agnes Jager[115], Jingmei Li[117], Jianjun Liu[117], Keith Humphreys[114], Xiao-Ou Shu[97], Wei Lu[118], Yu-Tang Gao[119], Hui Cai[97], Simon S. Cross[120], Malcolm W. R. Reed[1], William Blot[97,121], Lisa B. Signorello[97,121], Qiuyin Cai[97], Paul D.P. Pharoah[4,122], Barbara Perkins[122], Mitul Shah[122], Fiona M. Blows[122], Daehee Kang[123,124], Keun-Young Yoo[125], Dong-Young Noh[126], Mikael Hartman[127-129], Hui Miao[127,129], Kee Seng Chia[127,129], Thomas Choudary Putti[130], Ute Hamann[131], Craig Luccarini[122], Caroline Baynes[122], Shahana Ahmed[122], Mel Maranian[122], Catherine S. Healey[122], Anna Jakubowska[132], Jan Lubinski[132], Katarzyna Jaworska-Bieniek[132], Katarzyna Durda[132], Suleeporn Sangrajrang[133], Valerie Gaborieau[134], Paul Brennan[134], James McKay[134], Susan Slager[83], Amanda E. Toland[135], Drakoulis Yannoukakos[136], Chen-Yang Shen[137,138], Chia-Ni Hsiung[137], Pei-Ei Wu[139], Shian-ling Ding[140], Alan Ashworth[111], Michael Jones[110], Nick Orr[141], Anthony J Swerdlow[110,142], Helen Tsimiklis[143], Enes Makalic[144], Daniel F. Schmidt[144], Quang M. Bui[144], Stephen J. Chanock[112], David J. Hunter[145], Rebecca Hein[146,147], Norbert Dahmen[148], Lars Beckmann[149], Kirsimari Aaltonen[52,53], Taru A. Muranen[52,53], Tuomas Heikkinen[52,53], Astrid Irwanto[117], Nazneen Rahman[110], Clare

Turnbull[110], The Breast and Ovarian Cancer Susceptibility Study[150], Quinten Waisfisz[151], Hanne E. J. Meijers-Heijboer[151], Muriel A. Adank[151], Rob B van der Luijt[152], Per Hall[114], Georgia Chenevix-Trench[5], Alison Dunning[122], Douglas F. Easton[4,122], and Angela Cox[1]*

[1]Department of Oncology, University of Sheffield, Sheffield S10 2RX, UK, [2]Department of Neurosurgery, Chang Gung Memorial Hospital, Taoyuan County, 333, Taiwan, [3]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City Utah 84108-1266 USA, [4]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK, [5]Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia, [6]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria 3010 Australia, [7]Department of Pathology, The University of Melbourne, Melbourne, Victoria 3010, Australia, [8]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Melbourne School of Population Health, The University of Melbourne, Melbourne, Victoria 3010, Australia, [9]Netherlands Cancer Institute, Antoni van Leeuwenhoek hospital, 1066 CX Amsterdam, the Netherlands, [10]Division of Health Sciences, Warwick Medical School, Warwick University, Coventry CV4 7AL, UK, [11]Institute of Population Health, University of Manchester, Manchester M13 9QQ, UK, [12]Ministry of Public Health, Nonthaburi 11000, Thailand, [13]University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, D-91054 Erlangen, Germany, [14]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles 90095, CA, USA, [15]Institute of Human Genetics, University Hospital Erlangen, Friedrich Alexander University Erlangen-Nuremberg, D-91054 Erlangen, Germany, [16]Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, [17]Breakthrough Breast

Cancer Research Centre, The Institute of Cancer Research, London SW3 6JB, UK, [18]Division of Cancer Studies, Kings College London, Guy's Hospital, London SE1 9RT,UK, [19]Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford OX3 7BN, UK, [20]Clinical Science Institute, University Hospital Galway, Galway, Ireland, [21]Department of Obstetrics and Gynecology, University of Heidelberg, 69117 Heidelberg, Germany, [22]National Center for Tumor Diseases, University of Heidelberg, 69117 Heidelberg, Germany, [23]Molecular Epidemiology Group, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany, [24]Inserm (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, 94807 Villejuif, France, [25]University Paris-Sud, UMRS 1018, 94807 Villejuif, France, [26]Université Paris Sorbonne Cité, UMR-S775 Inserm, Paris 75015, France, [27]Copenhagen General Population Study, Herlev Hospital, 2730 Herlev, Copenhagen, Denmark, [28]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, 2100 Copenhagen Denmark, [29]Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark, [30]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, 2100 Copenhagen Denmark, [31]Human Genetics Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), E-28029 Madrid, Spain, [32]Centro de Investigación en Red de Enfermedades Raras (CIBERER) Valencia 28029, Spain, [33]Servicio de Oncología Médica, Hospital Universitario La Paz, 28046 Madrid, Spain, [34]Servicio de Cirugía General y Especialidades, Hospital Monte Naranco, 33012 Oviedo, Spain, [35]Servicio de Anatomía Patológica, Hospital Monte Naranco, 33012 Oviedo, Spain, [36]Human Genotyping-CEGEN Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), E-28029 Madrid, Spain, [37]Department of Epidemiology, University of California Irvine, Irvine, CA 92697, USA, [38]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany, [39]German Cancer Consortium

(DKTK), Heidelberg, Germany, [40]Saarland Cancer Registry, 66024 Saarbrücken, Germany, [41]Division of Gynaecology and Obstetrics, Technische Universität München, D-80333 Munich, Germany, [42]Insitute of Human Genetics, Technische Universität München, D-80333 Munich, Germany, [43]Division of Molecular Gyneco-Oncology, Department of Gynaecology and Obstetrics, University Cologne, 50923 Cologne, Germany, [44]Max Planck Institute of Psychiatry, 80804 Munich, Germany, [45]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, 70376 Stuttgart, Germany, [46]University of Tübingen, 72074 Tübingen, Germany, [47]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany, [48]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-University Bochum (IPA), D-44789 Bochum, Germany, [49]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, 53113 Bonn, Germany, [50]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, 70376 Stuttgart, Germany, University of Tübingen, 72074 Tubingen, Germany, Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), 69120 Heidelberg, Germany, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-University Bochum (IPA), D-44789 Bochum, Germany, Institute for Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany, Institute of Pathology, Medical Faculty of the University of Bonn, 53127 Bonn, Germany, Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, 53113 Bonn, Germany, [51]Centre d'innovation Génome Québec et Université McGill, Montréal, QC H3A 0G4 Canada, [52]Department of Obstetrics and Gynecology, University of Helsinki, 00029 Helsinki, Finland, [53]Helsinki University Central Hospital, 00029 Helsinki, Finland, [54]. Department of Clinical Genetics, University of Helsinki, 00029 Helsinki, Finland, [55]Department of Oncology, University of Helsinki, 00029 Helsinki, Finland, [56]Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences,

Fukuoka, 812-8582 Japan, [57]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, 464-8681 Japan, [58]Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya, 464-8681 Japan, [59]Department of Obstetrics and Gynaecology, Hannover Medical School, 30625 Hannover, Germany, [60]Department of Radiation Oncology, Hannover Medical School, 30625 Hannover, Germany, [61]N.N. Alexandrov Research Institute of Oncology and Medical Radiology, 223040 Minsk, Belarus, [62]Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 77 Stockholm, Sweden, [63]Department of Oncology - Pathology, Karolinska Institutet, 171 77 Stockholm, Sweden, [64]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, 70210 Kuopio, Finland, [65]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, 70029 Kuopio, Finland, [66]School of Medicine, Institute of Clinical Medicine, Oncology, Cancer Center of Eastern Finland, University of Eastern Finland, 70210 Kuopio, Finland, [67]Cancer Center, Kuopio University Hospital, 70029 Kuopio, Finland, [68]Peter MacCallum Cancer Center, Melbourne, VIC 3002, Australia, [69]QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia, [70]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA, [71]University Hospital Gashuisberg, 3000 Leuven, Belgium, [72]Vesalius Research Center (VRC), VIB, 3000 Leuven, Belgium, [73]Department of Oncology, University of Leuven, 3000 Leuven, Belgium, [74]Department of General Medical Oncology, University Hospitals Leuven, 3000 Leuven, Belgium, [75]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, 3000 Leuven, Belgium, [76]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany[77]Department of Cancer Epidemiology/Clinical Cancer Registry and Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, D - 20246 Hamburg, Germany, [78]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei

Tumori (INT), 20133 Milan, Italy, [79]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, 20139 Milan, Italy, [80]Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), 20133 Milan, Italy, [81]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia (IEO), 20141 Milan, Italy, [82]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA, [83]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA, [84]Karmanos Cancer Center, Detroit, MI 48201, USA, [85]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria 3004 Australia, [86]Anatomical Pathology, The Alfred Hospital, Melbourne, Victoria 3004 Australia, [87]Epidemiology Program, University of Hawaii Cancer, Honolulu, HI 96813, USA, [88]Centre Hospitalier Universitaire de Québec Research Centre and Laval University, Quebec City, Quebec, G1V 0A6 Canada, [89]Department of Medicine, McGill University, Montreal, Quebec H3A 0G4 Canada, [90]Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, Quebec, H3A 1A1 Canada, [91]Départements de Santé environnementale et santé au travail et de Médecine sociale et préventive, École de santé publique, Université de Montréal, Montreal, Quebec H3T 1J4, Canada, [92]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Ss 12 Subang Jaya, Selangor, Malaysia, [93]Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya Medical Centre, 50603 Kuala Lumpur, Malaysia, [94]Singapore Eye Research Institute, National University of Singapore, Singapore 119077, [95]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, 0372 Oslo, Norway, [96]Faculty of Medicine (Faculty Division Ahus), University of Oslo (UiO), 0316 Oslo, Norway, [97]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37232, USA, [98]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, Northern Finland Laboratory Centre

NordLab, 90570 Oulu, Finland, [99]Department of Oncology, Oulu University Hospital, University of Oulu, 90570 Oulu, Finland, [100]Department of Pathology, Oulu University Hospital, University of Oulu, 90570 Oulu, Finland, [101]Ontario Cancer Genetics Network, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada, [102]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 2J7, Canada, [103]Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada, [104]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario M5S 2J7, Canada, [105]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada, [106]Department of Human Genetics & Department of Pathology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands, [107]Department of Surgical Oncology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands, [108]Family Cancer Clinic, Department of Medical Oncology, Erasmus MC-Daniel den Hoed Cancer Center, 3075 EA Rotterdam, The Netherlands, [109]Department of Clinical Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands, [110]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK, [111]Breakthrough Breast Cancer Research Centre, Division of Breast Cancer Research, The Institute of Cancer Research, London SW3 6JB, UK, [112]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20850, USA, [113]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center & Institute of Oncology, 02-781 Warsaw, Poland, [114]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden, [115]Family Cancer Clinic, Department of Medical Oncology, Erasmus MC Cancer Institute, 3075 EA Rotterdam, The Netherlands, [116]Department of Clinical Genetics, Erasmus University Medical Center, 3075 EA Rotterdam, The Netherlands, [117]Human Genetics Division, Genome Institute of Singapore, Singapore 138672, Singapore, [118]Shanghai Center for Disease Control

and Prevention, Shanghai 200336, China, [119]Department of Epidemiology, Shanghai Cancer Institute, Shanghai 220025, China, [120]Department of Neuroscience, University of Sheffield, Sheffield S10 2JF, UK, [121]International Epidemiology Institute, Rockville, MD 20850, USA, [122]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN UK, [123]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul 110-799, Korea, [124]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 151-742, Korea, [125]Seoul National University College of Medicine, Seoul 110-799, Korea, [126]Department of Surgery, Seoul National University College of Medicine, Seoul 110-799, Korea, [127]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, [128]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, [129]National University Health System, Singapore 119228, [130]Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, [131]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany, [132]Department of Genetics and Pathology, Pomeranian Medical University, 70-204 Szczecin, Poland, [133]National Cancer Institute, Bangkok 10400, Thailand, [134]International Agency for Research on Cancer, 69372 Lyon, France, [135]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA, [136]Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos", Aghia Paraskevi Attikis, 153 10 Athens, Greece, [137]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, [138]College of Public Health, China Medical University, Taichung 40402, Taiwan, [139]Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, [140]Department of Nursing, Kang-Ning Junior College of Medical Care and Management, 11529 Taipei, Taiwan, [141]Breakthrough Breast Cancer Research Centre and Division of Breast Cancer Research, The Institute of Cancer Research, London SW3 6JB, UK,

[142]Division of Breast Cancer Research , Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK, [143]Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria 3010 Australia, [144]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, Melbourne School of Population Health, The University of Melbourne, Melbourne, Victoria 3010 Australia, [145]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA, [146]Division of Cancer Epidemiology, Deutsches Krebsforschungszentrum, 69120 Heidelberg, Germany, [147]PMV Research Group at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, 50923 Cologne, Germany, [148]Department of Psychiatry, University of Mainz, 55122 Mainz, Germany, [149]Institute for Quality and Efficiency in Health Care (IQWiG), 50670 Cologne, Germany, [150]Division of Cancer Genetics, Institute of Cancer Research, Sutton SM2 5NG, UK, [151]Department of Clinical Genetics, VU University Medical Center, section Oncogenetics, 1081 HZ Amsterdam, The Netherlands, [152]Department of Medical Genetics, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.

The authors wish it to be known that, in their opinion, the first 2 and 4 final authors should be regarded as having contributed equally to this work.

[*]To whom correspondence should be addressed, at: Sheffield University Medical School, Beech Hill Road, Sheffield, S10 2RX, UK.

Tel: +44 114 2712373

FAX: +44 114 2711602

email: a.cox@sheffield.ac.uk

Abstract

Previous studies have suggested that polymorphisms in *CASP8* on chromosome 2 are associated with breast cancer risk. To clarify the role of *CASP8* in breast cancer susceptibility we carried out dense genotyping of this region in the Breast Cancer Association Consortium (BCAC). Single nucleotide polymorphisms (SNPs) spanning a 1Mb region around *CASP8* were genotyped in 46,450 breast cancer cases and 42,600 controls of European origin from 41 studies participating in the BCAC as part of a custom genotyping array experiment (iCOGS). Missing genotypes and SNPs were imputed and, after quality exclusions, 501 typed and 1,232 imputed SNPs were included in logistic regression models adjusting for study and ancestry principal components. The SNPs retained in the final model were investigated further in data from nine genome-wide association studies (GWAS) comprising in total 10,052 case and 12,575 control subjects. The most significant association signal observed in European subjects was for the imputed intronic SNP rs1830298 in *ALS2CR12* (telomeric to *CASP8*), with per allele odds ratio and 95% confidence interval [OR (95% CI)] for the minor allele of 1.05 (1.03-1.07), $p=1x10^{-5}$. Three additional independent signals from intronic SNPs were identified, in *CASP8* (rs36043647), *ALS2CR11* (rs59278883) and *CFLAR* (rs7558475). The association with rs1830298 was replicated in the imputed results from the combined GWAS ($p=3 x10^{-6}$), yielding a combined OR (95% CI) of 1.06 (1.04-1.08), $p=1x10^{-9}$. Analyses of gene expression associations in peripheral blood and normal breast tissue indicate that *CASP8* might be the target gene, suggesting a mechanism involving apoptosis.

Introduction

Breast cancer is a complex disease with high, moderate and low penetrance germ-line variants involved in its aetiology (1). In recent years around 80 low penetrance breast cancer alleles have been identified, with modest odds ratios, ranging from 1.05-1.4, and together accounting for around 15% of familial breast cancer risk (2, 3). It is likely that there are many more loci

with even smaller effect sizes that remain to be identified, accounting for a further 14-15% of familial risk (2). One of the first low penetrance breast cancer variant associations to be convincingly replicated by large case-control studies was the single nucleotide polymorphism (SNP) rs1045485 encoding the mis-sense alteration D302H in the caspase 8 apoptosis-related cysteine peptidase (CASP8) gene at chromosome region 2q33 (4, 5). This association was first identified by a candidate gene study and replicated in 2007 by the Breast Cancer Association Consortium (BCAC), in a study of more than 17,000 cases and 16,000 controls (4, 5). The minor C allele, common in Europeans and rare in Asians, was found to be associated with a 10% reduction in risk of breast cancer (5). However, further fine-mapping studies have shown that other variants in the region are associated with an increased risk of breast cancer, and in the recent large-scale genotyping study carried out by the BCAC as part of the COGS (Collaborative Oncology Gene- environment Study), rs1045485 showed only weak evidence of association with breast cancer risk (2, 6, 7). In addition, in 2010 a UK genome-wide association study (GWAS) of 3,659 cases and 4,897 controls found suggestive evidence of association [OR (95% CI) 1.14 (1.06-1.22); $p=1.5x10^{-4}$] with an independent variant in the region; rs10931936, a *CASP8* intronic SNP, that is only weakly correlated with rs1045485 ($r^2=0.083$) (8).

In order to clarify the breast cancer risk association(s) at this locus, we have analysed 501 SNPs across a 1Mb region surrounding *CASP8,* for 89,050 women, as part of a custom-designed Illumina genotyping chip – the iCOGS array. We present here the results of this fine-mapping analysis, together with a meta-analysis across iCOGS and the combined data from nine breast cancer GWAS, followed by an examination of associations between the key SNPs and RNA expression levels.

Results

Breast cancer risk associations in the CASP8 region on chromosome 2

A summary of the breast cancer risk associations of 1,733 typed and imputed SNPs across a 1Mb region surrounding *CASP8*, based on the iCOGS European data, is shown in Figure 1. The most significant associations were for SNPs in the *CASP8* and *ALS2CR12* (amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 12) genes (Figure 1 and Table S2). The strongest signals came from imputed SNP rs1830298 in *ALS2CR12*, with MAF (minor allele frequency) of 0.29 and an estimated OR (95% CI) per copy of the minor allele of 1.05 (1.03-1.07), $p=1.1x10^{-5}$, and the genotyped SNP rs10197246, (MAF=0.28), with odds ratio (95% CI) 1.05 (1.02-1.07), $p=2.5x10^{-5}$. These two SNPs are highly correlated and likely reflect the same signal ($r^2=0.9$).

Two previously reported susceptibility SNPs, *CASP8* D302H (rs1045485) and rs10931936, were weakly replicated in iCOGS European data (Table S2), with minor allele OR in the same direction; however, the iCOGS OR estimates were much weaker than those from the original reports (5, 8). The minor C allele of rs1045485 (MAF=0.11) yielded an OR (95% CI) of 0.97 (0.94-1.0), p=0.03, in contrast to 0.88 (0.84-0.92) reported in Cox et al (5). Similarly, the rs10931936 minor allele (MAF=0.28) was associated with a 4% increased breast cancer risk [OR (95% CI)=1.04 (1.02-1.06), $p=1.9x10^{-4}$], compared to the 12% increase presented in Turnbull *et al.* (8). The latter SNP is strongly correlated with the iCOGS best hit rs1830298 ($r^2=0.96$), but there is very little correlation between rs1045485 and rs1830298 ($r^2=0.055$).

Identification of possible independent signals in iCOGS European data

The SNPs in the main association peak have similar ORs for breast cancer, are strongly correlated with one another ($r^2>0.66$) and confined to an 82 Kb region spanning the *CASP8* and *ALS2CR12* genes, and are therefore likely to reflect a single association signal, but this

337 does not preclude the possibility of other signals in the region. To test this hypothesis, we
338 carried out a regression analysis testing the association of individual SNPs adjusted for the
339 top hit rs1830298, in the iCOGS European data set (Table S3). Interestingly, whilst this
340 resulted in the loss of the signal from the main peak in *CASP8/ALS2CR12*, residual
341 associations remained (for example 43 SNPs with $p \leq 1x10^{-3}$), suggesting that there may be
342 further signals present in the region, albeit weaker (Table S3 and Figure S1). To investigate
343 this further, we carried out penalised logistic regression analysis of all 1,733 SNPs to identify
344 the best subset of SNPs that explain the association, using HyperLasso (9). This identified 59
345 models containing combinations of 27 SNPs (Table S2), but many of these models were
346 equivalent after taking into account linkage disequilibrium between SNPs. To obtain the most
347 parsimonious model, we carried out stepwise forward logistic regression on the 27 SNPs,
348 which resulted in a model containing four SNPs; rs1830298 (*ALS2CR12*; $p_{conditional}=9.3x10^{-3}$,
349 MAF=0.29), rs36043647 (*CASP8*; $p_{conditional}=1.9x10^{-4}$, MAF=0.06), rs59278883 *(ALS2CR11*;
350 $p_{conditional}=6.1x10^{-4}$, MAF=0.07) and rs7558475 (*CFLAR;* CASP8 and FADD-like apoptosis
351 regulator; $p_{conditional}=9.2x10^{-4}$, MAF=0.07). We refer to these four SNPs, marking four
352 independent sets of correlated highly associated variants (iCHAVs), as index SNPs.
353
354 ***Meta-analysis of iCOGS and combined nine GWAS data***
355 We first examined the results for the four index SNPs, together with the previous hits
356 rs1045485 and rs10931936, in the combined nine GWAS meta-analysis, and then carried out
357 a further meta-analysis combining the iCOGS European data with the combined nine GWAS
358 for these SNPs (total sample size 56,502 cases and 55,175 controls; Tables S4 and S5). We
359 found that the top index SNP, rs1830298, replicated in the combined GWAS data alone
360 ($p=2.7x10^{-6}$), and reached genome-wide significance ($p=1.1x10^{-9}$) in the meta-analysis
361 containing both the iCOGS and combined GWAS data (Table S5 and Figure 2). The
362 genotyped proxy rs10197246 also reached genome-wide significance ($p=1.7x10^{-8}$). When we

examined the other three index SNPs in the combined GWAS data, we found a replicated association ($p=1.8x10^{-3}$) for rs59278883, a null result for rs36043647 (p=0.58) and borderline evidence for rs7558475 (p=0.05) (Table S5 and Figure 2). However these three index SNPs all showed some evidence of association in the meta analysis of iCOGS and combined GWAS (Table S5 and Figure 2), providing some support for the existence of four signals in the region. Consistent with its strong correlation with rs1830298, a similar but slightly weaker signal was found for rs10931936 in the combined analysis ($p=1.0x10^{-7}$). Weak evidence for association was observed for *CASP8 D302H* rs1045485 ($p=1.1x10^{-3}$).

Analysis of index SNPs in different ethnic groups

We next explored these four associations in the available Asian and African-American populations genotyped as part of COGS (Figure 3 and Table S6). Figure 3 shows the study-specific OR for rs1830298 by the three ethnic groups. The rs1830298 OR were homogeneous among European studies ($p_{het}=0.54$, $I^2=0$) and African-American studies ($p_{het}=0.40$, $I^2=0$), but were more heterogeneous amongst the nine Asian studies ($p_{het}=0.025$, $I^2=54$), although the combined effect size in Asians was similar to that seen in Europeans [OR (95% CI)=1.04 (0.95-1.13); p=0.44], and slightly stronger in African-Americans [OR (95% CI)=1.12 (0.96-1.30); p=0.16]. Although estimates in both Asian and African-American populations were not statistically significant, the OR were consistent with the European data, and the pooled OR (95% CI) was 1.05 (1.03-1.07); $p=4.1x10^{-6}$ for all populations combined. The MAF of *CASP8* rs36043647 was much lower in Asians, in whom the association was in the opposite direction to that in Europeans and African-Americans, with an OR (95% CI) of 1.69 (1.13-2.51), p=0.009, for the minor allele (Table S6). We did not observe any association of rs59278883 and rs7558475 in Asian and African-American populations (Table S6).

Subtype and survival analysis in iCOGS

To investigate whether these SNP associations vary with clinical subtypes of breast cancer, we explored potential subtype-specific associations by comparing different subtypes to all controls in the iCOGS European data. The OR estimates by tumour estrogen receptor (ER) status, triple negative status and invasiveness of breast cancer were all similar and close to the OR of 1.05 seen in overall breast cancer for rs1830298 (Figure 4). Similarly no significant differences in OR were seen when cases were stratified by family history, tumour grade, tumour stage, tumour size and lymph node status (Figure S2). A broadly similar picture was seen for the other index SNPs (Figures S2 and S3).

SNP effects were also evaluated for overall survival and breast-cancer-specific survival. There were 4,191 deaths among 39,140 breast cancer patients with known vital status in the European dataset. Of these deaths, 1,979 died from breast cancer. We did not observe any associations between the index SNPs or previous hit SNPs with either overall or breast cancer-specific survival, and all hazard ratios (HR) were close to unity (data not shown).

In silico functional and eQTL annotations

We examined available *in silico* functional and expression quantitative trait loci (eQTL) data for the four iCHAVs. Of interest in iCHAV1, rs3769823 is a missense alteration encoding K14R in the 4th exon of *CASP8*, which encodes the N-terminus of protein isoform 9. In addition, this SNP and rs3769821 are both located in a region of deoxyribonuclease I (DNase I) hypersensitivity and histone H3K27 acetylation in breast cell lines (Figure 5). The minor alleles of both of these SNPs, together with four others in iCHAV1 for which data were available, were associated with a reduction in *CASP8* mRNA levels in peripheral blood samples in the eQTL meta-analysis of Westra et al ($p \leq 9.4 \times 10^{-5}$; Table S7, Figure 5) (10). The cancer genome atlas (TCGA) dataset only had data available for two SNPs from iCHAV1,

and both were associated with a reduction in CASP8 mRNA in normal breast tissue ($p \leq 1x10^{-3}$; Table S7, Figure 5). No strong eQTL associations were seen for other genes in the region in either the Westra *et al.* or the TCGA data. Taken together, these data suggest that one or more variants in iCHAV1 may affect levels of *CASP8* gene expression. As shown in Figure 5, iCHAVs 3 and 4 overlap enhancer sites identified in Hnisz et al; a *CASP8* enhancer in MCF7 cells and a *CFLAR* enhancer in human mammary epithelial cells (HMEC), respectively (11). However, there was limited eQTL data available for these iCHAVs, with no evidence of any significant eQTLs (Table S7).

Discussion

In our analysis of the genomic region surrounding *CASP8* for association with breast cancer, the strongest signal came from an imputed SNP, rs1830298, in the *ASL2CR12* gene (iCHAV1). A strongly correlated genotyped SNP, (rs10197246; $r^2=0.9$, 23.5 Kb telomeric in the same gene), yielded a similar association signal ($p=1.1x10^{-5}$ and $2.5x10^{-5}$ respectively). In each case, the rare allele (MAF=0.28) was associated with an increase in the risk of breast cancer of 5% [OR(95% CI) 1.05 (1.03, 1.07) and 1.05 (1.02, 1.07) respectively]. The odds ratios for both SNPs are consistent in Europeans, Asians and African-Americans, (although not statistically significant in the smaller non-European cohorts), and were replicated in the combined GWAS data, achieving a genome-wide level of significance when the iCOGS and GWAS data were combined ($p=1.1x10^{-9}$ and $p=1.7x10^{-8}$ respectively). This association is consistent between ER positive and negative disease, and between invasive and *in situ* cancers (Figure 4). The previously published result for rs10931936 in the UK GWAS is consistent with its correlation with rs1830298 (8).

Several of the SNPs in iCHAV1were associated with *CASP8* eQTLs. The minor alleles of SNPs in this group, associated with increased risk of breast cancer, are associated with

reduced *CASP8* mRNA levels in both peripheral blood lymphocytes and normal breast tissue (Table S7, Figure 5). These data suggest that *CASP8* may be the target gene of iCHAV1, and are consistent with a hypothesis in which the effect of the risk alleles is via reduced levels of apoptosis, thus promoting tumour initiation. However, further functional studies are required to demonstrate a direct interaction between iCHAV1 and the *CASP8* promoter and to investigate the allele-specific functional effects of these SNPs in different tissue types.

Our results also suggest three other independent signals in the region; the most significant SNPs for these three signals are in *CASP8* (iCHAV2), *ALS2CR11* (iCHAV3), and in the anti-apoptotic gene *CFLAR* (iCHAV4); see Figure 2 and Table S5. The signals for iCHAVs 3 and 4 were replicated in the combined GWAS, but since they did not achieve genome-wide levels of significance even in the very large data sets analysed here, they are harder to interpret. However, it is interesting that both these iCHAVs overlap enhancer regions (Figure 5).

As previously noted, we find only very weak support for an association of rs1045485/D302H in the iCOGS data (p=0.03; (2)), although the odds ratio in the combined GWAS data was more consistent with the original report [OR (95% CI)=0.90(0.85, 0.96), p=0.0007] (5). At present the reasons for the discrepancy with the original report are not clear. D302H is only weakly correlated with any of the four index SNPs identified here (max r^2=0.06 with rs1830298). However, it is correlated with rs28845859 (r^2=0.67); the latter SNP is associated with reduced breast cancer risk in the iCOGS data (OR 0.95, p=1.9x10^{-4}; Table S2) and combined GWAS (p=4.0x10^{-5}). We found no significant differences between sub-types, although the associated effect for D302H was stronger (and borderline significant) for triple negative disease, despite the smaller sample size (Figure S3). Further investigation with a larger sample of triple negative cases may help clarify this point.

The association for the top *CASP8* index SNP, rs1830298, represents one of the smaller effect sizes identified to date for breast cancer. However it is worth noting that the *CASP8* region has recently been reported to be associated with other cancers at genome-wide levels of significance, including melanoma and chronic lymphocytic leukaemia (CLL) (12, 13). The alleles associated with increased risk in melanoma are correlated with rs1830298, but the signal in CLL appears to be due to uncorrelated SNPs in the region. This difference may reflect the different cell type of origin and it will be interesting to determine the relative importance and function of alleles of the *CASP8* gene family in immune cell lineages, compared to that in epithelial cancers.

Materials and Methods

Study samples

The iCOGS and nine breast cancer GWAS data sets have been described in detail previously (2). Briefly, the COGS includes a total of 103,991 women from 50 studies participating in the BCAC whose DNA samples were genotyped with the iCOGS array. These were 89,050 Europeans (46,450 cases; 42,600 controls), 12,893 Asians (6,269 cases; 6,624 controls), and 2,048 African-Americans (1,116 cases and 932 controls). The numbers of subjects by study are detailed in Table S1. Approximately 93% of cases had invasive breast cancer (Table S1). The combined nine breast cancer GWAS data set comprised 10,052 cases and 12,575 controls of European ancestry from USA, UK, Australia, Germany, Finland, Sweden and the Netherlands (2).

Ethics statement

Each study was approved by the relevant local/institutional Research Ethics Committee, and all subjects gave written informed consent to take part.

SNP selection for fine-scale mapping on the iCOGS array The region for analysis on chromosome 2 was defined such that it contained all SNPs correlated ($r^2 \geq 0.1$) with the SNPs previously reported to be associated with breast cancer, namely, *CASP8 D302H* (rs1045485) and rs10931936 (5, 8). This identified a 1Mb region from 201,566,128 to 202,566,128 (hg19). In March 2010 when the iCOGS array was designed, 2,191 SNPs had been catalogued in this region by the 1000 genomes and HapMap3 projects. Of these, 1,723 SNPs had a MAF ≥2%, and of these 1,723, there were 988 SNPs with Illumina assay design scores of ≥ 0.8. We selected a total of 280 SNPs correlated at $r^2 \geq 0.1$ with rs1045485 or rs10931936, plus 288 tagSNPs which tagged the remaining 708 SNPs at $r^2 \geq 0.9$. Another 45 SNPs in the region, nominated by other consortia members, were included as part of the genotyping array that comprised 211,155 SNPs in total (2).

Genotyping & quality control

Genotyping, allele calling, quality control and principal components analysis for COGS are described in detail in Michailidou *et al.* (2). Genotyping was carried out at four centres using the Illumina Infinium iCOGS array, including 2% duplicates from each participating study. Final genotype calls were made using Illumina's proprietary GenCall algorithm. SNPs were excluded from analysis if the overall call rate was <95%, duplicate concordance rate was <98%, or if deviation from Hardy-Weinberg equilibrium in controls was significant at $p<1x10^{-7}$ (2). Subjects were excluded from analysis for the following reasons: genotypically non female; overall call rate < 95%; low or high heterozygosity ($P<1\times10^{-6}$); discordant replicates or cryptic duplicates. Genotype data and ancestry principal components (seven

514 principal components for the European and two each for the Asian and African-American
515 populations) were thus available for 103,991 individuals.
516
517 ***Statistical analysis***
518 The iCOGS *CASP8* region genotype data were split into four groups for efficiency of
519 imputation of missing genotypes and untyped SNPs. These comprised 36,793 European
520 ancestry subjects from North American and UK studies in group 1, with 26,129 and 26,128 of
521 the remaining European subjects in groups 2 and 3 respectively, and 14,941 Asians and
522 African-Americans in group 4. Imputations were carried out separately by group based on the
523 1000 genomes phase I reference panel with singleton variants excluded, using IMPUTE2
524 version 2.3 (14, 15). SNPs were included in the subsequent analyses if the mean information
525 score of the European groups was ≥ 0.9, and untyped imputed SNPs were only included if
526 their MAF was ≥3%; these criteria resulted in inclusion of 501 typed and 1,232 imputed SNPs
527 in the final analysis. The imputation accuracy for rs1830298 was verified in whole genome
528 sequence data from 197 individuals; the correlation between the observed and imputed
529 genotypes was 0.974. The imputation step increases the number of common SNPs captured at
530 r^2>0.9 from 76% (1198/1583) to 84% (1333/1583).
531
532 The main analyses were based on the data for individuals of European ancestry. For each SNP,
533 allelic dosage of the minor allele was estimated, and included in a logistic regression model,
534 to estimate OR and corresponding 95% CI. Covariates for each study plus the seven ancestry
535 principal components were included in the model (2). These analyses were implemented in R.
536 P-values from the Wald test are reported in the text (uncorrected for multiple testing). FDR
537 values in Table S2 were calculated according to the Benjamini & Hochberg method, as
538 implemented in the R p.adjust function (16). Penalised logistic regression models (based on
539 the normal exponential gamma probability density) were implemented in HyperLasso (9),

including all 501 typed and 1,232 imputed SNPs, to identify the best subsets of SNPs to account for the observed association data. Based on the sample size and a type I error of 0.001, a lambda of 0.05 and penalty of 491 were specified in HyperLasso, according to equation 7 in Hoggart *et al.* (9). Candidate SNPs were then compiled from the resulting HyperLasso models and included in a stepwise forward logistic regression procedure with penalty k=10 in the step function in R to identify the most parsimonious model, as described previously (17). The SNPs retained in the final model are referred to as index SNPs.

Index SNPs were further examined by means of meta-analysis of iCOGS European, Asian and African-American data, and also with individual SNP results from the combined nine breast cancer GWAS (2). Due to an overlap of 1,955 samples that exist in both the iCOGS and the combined GWAS data, we removed these samples from the iCOGS data before carrying out the meta-analysis. The meta-analysis was carried out using the MetaFor package in R, with inverse-variance weights and the DerSimonian-Laird estimator for the random effects model (18). We used the threshold of $p=5x10^{-8}$ to define genome-wide significance (2).

The index SNPs were also examined for associations with breast cancer specific and overall survival in Cox's proportional hazard models, including age at diagnosis, study and seven principal components as covariates, and accounting for the left-censoring time between study entry and diagnosis. Further adjustment was carried out for stage, grade, tumour size, and lymph node involvement for SNPs with nominally significant associations with survival ($p<0.05$). These analyses were implemented in R.

In silico functional and eQTL annotations

We defined independent sets of correlated highly associated variants (iCHAVs) with likelihood (determined from the individual-SNP logistic regression analysis) relative to an

index SNP of greater than 1/100 and degree of correlation with the index SNP of greater than 0.65. The ENCODE integrated regulation data for each SNP were retrieved from the UCSC Genome Browser by use of ANNOVAR (19). Predicted enhancers and target genes were retrieved from Hnisz *et al*. (11). Expression QTL data were obtained by interrogation of the GTEx Portal, the online results of the peripheral blood eQTL meta-analysis based on 5,311 samples from 7 studies by Westra and colleagues (10), and from breast cancer cases in the TCGA Project. For the latter, RNAseq data in the form of fragments per Kb of transcript per million mapped reads (FPKM) were available for uninvolved breast tissue from 97 TCGA breast cancer cases. Peripheral blood DNA SNP genotypes for these individuals were extracted from the TCGA Level 2 Affymetrix 6.0 array birdseed files. Mean FPKM were compared between individuals homozygous for the common allele and those carrying one or two copies of the rare allele by use of an unpaired, unequal variance T-test in Stata.

Acknowledgements

Part of this work was supported by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). 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), 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. The ABCFS, NC-BCFR and OFBCR work was supported by the United States National Cancer Institute, National Institutes of Health (NIH) under RFA-CA-06-503 and through cooperative agreements with members of the Breast Cancer Family Registry (BCFR) and Principal Investigators, including Cancer Care Ontario (U01 CA69467), Northern California Cancer Center (U01 CA69417), University of Melbourne (U01 CA69638). Samples from the NC-BCFR were processed and distributed by the Coriell Institute for Medical Research. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the US 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 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 is funded by the Breast Cancer Research Trust, UK. 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 Breakthrough Breast Cancer and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BBCS GWAS received funding from The Institut National de Cancer. 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 study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). The 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 was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital. 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 was supported by the California Breast Cancer Act of 1993; National Institutes of Health (grants R01 CA77398 and the Lon V Smith Foundation [LVS39420].); the California Breast Cancer Research Fund (contract 97-10500). Collection of cancer incidence data used in this study 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. 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 GC-HBOC was supported by Deutsche Krebshilfe (107 352). 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, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. 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 GWS population allele and genotype frequencies were obtained from the data source funded by the Nordic Center of Excellence in Disease Genetics based on samples regionally selected from Finland, Sweden and Denmark. 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 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 and the Gustav V Jubilee foundation. The KBCP was financially supported by the special Government Funding (EVO)

of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the Academy of Finland and by the strategic funding of the University of Eastern Finland. kConFab is supported by grants from the National Breast Cancer Foundation, the NHMRC, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia and the Cancer Foundation of Western Australia. The kConFab Clinical Follow Up Study was funded by the NHMRC [145684, 288704, 454508]. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], the Cancer Council of Tasmania and Cancer Foundation of Western Australia and the NHMRC [199600]. G.C.T. and P.W. are supported by the NHMRC. 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. 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 was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I], the Hamburg Cancer Society, the German Cancer Research Center and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402, 01KH0408, 01KH0409]. 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 grant CA128978, an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], the Breast Cancer Research Foundation, 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 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 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. 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 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 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 was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland Centre of Excellence grant 251314, the Sigrid Juselius Foundation, the University of Oulu, and the Oulu University Hospital special Govermental EVO Research Funds. The OFBCR work 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. 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 was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The pKARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). 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 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 SBCS was supported by Yorkshire Cancer Research awards S295, S299, S305PA, and by the Sheffield Experimental Cancer Medicine Centre Network. NJC was supported by NCI grant R01 CA163353 and The Avon Foundation (02-2009-080). 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 programme grants from Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The SEBCS was supported by the Korea Health 21 R&D Project [AO30001], Ministry of Health and Welfare, Republic of Korea. SGBCC is funded by the National Medical Research Council start-up Grant and Centre Grant (NMRC/CG/NCIS /2010). 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. 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: MCBCS (National Institutes of Health Grants CA122340 and a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), a generous gift from the David F. and Margaret T. Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation. This research has been partly financed by 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) - Research Funding Program of the General Secretariat for Research & Technology: ARISTEIA. Investing in knowledge society through the European Social Fund; and the Stefanie Spielman Breast Fund and the Ohio State University Comprehensive Cancer Center. The TWBCS is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences, Academia Sinica, Taiwan. The UKBGS is funded by Breakthrough Breast Cancer and the Institute of Cancer Research (ICR). ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The Nurses' Health Studies (CGEMS) are supported by NIH grants CA 65725, CA87969, CA49449, CA67262,

762 CA50385 and 5UO1CA098233. The UK2 GWAS was funded by Wellcome Trust and Cancer
763 Research UK. It included samples collected through the FBCS study, which is funded by
764 Cancer Research UK [C8620/A8372]. It included control data obtained through the WTCCC
765 which was funded by the Wellcome Trust. The DFBBCS GWAS was funded by The
766 Netherlands Organisation for Scientific Research (NWO) as part of a ZonMw/VIDI grant
767 number 91756341. Control GWA genotype data from the Rotterdam Study were funded by
768 NWO Groot Investments (project nr. 175.010.2005.011).
769
770 We thank all the individuals who took part in these studies and all the researchers, clinicians,
771 technicians and administrative staff who have enabled this work to be carried out. This study
772 would not have been possible without the contributions of the following: Andrew Berchuck
773 (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch
774 (PRACTICAL), Antonis Antoniou, Lesley McGuffog, Ken Offit (CIMBA), Andrew Lee, and
775 Ed Dicks, and the staff of the Centre for Genetic Epidemiology Laboratory, the staff of the
776 CNIO genotyping unit, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill
777 University and Génome Québec Innovation Centre, the staff of the Copenhagen DNA
778 laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey
779 Meyer and the staff of Mayo Clinic Genotyping Core Facility. We also thank Maggie
780 Angelakos, Judi Maskiell, Gillian Dite (ABCFS), and extend our thanks to the many women
781 and their families that generously participated in the Australian Breast Cancer Family Study
782 and consented to us accessing their pathology material. JLH is a National Health and Medical
783 Research Council Australia Fellow. MCS is a National Health and Medical Research Council
784 Senior Research Fellow. JLH and MCS are both group leaders of the Victoria Breast Cancer
785 Research Consortium. We thank Sten Cornelissen, Richard van Hien, Linde Braaf, Frans
786 Hogervorst, Senno Verhoef, Ellen van der Schoot, Femke Atsma (ABCS). The ACP study
787 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-General of Disease Control who have supported the study throughout. We thank Eileen Williams, Elaine Ryder-Mills, Kara Sargus (BBCS), Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones (BIGGS), Peter Bugert, and Medical Faculty Mannheim (BSUCH). We thank the staff and participants of the Copenhagen General Population Study, and for excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, and Dorthe Kjeldgård Hansen. The Danish Breast Cancer Group (DBCG) is acknowledged for the tumor information. We thank Guillermo Pita, Charo Alonso, Daniel Herrero, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO), Hartwig Ziegler, Sonja Wolf, and Volker Hermann (ESTHER). We thank Heide Hellebrand, Stefanie Engert (GC-HBOC). GC-HBOC would like to thank the following persons for providing additional informations and samples: Prof. Dr. Norbert Arnold, Dr. Sabine Preissler-Adams, Dr. Monika Mareeva-Varon, Dr. Dieter Niederacher, Prof. Dr. Brigitte Schlegelberger, Dr. Clemens Mül. 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 [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] and 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]; Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany

[Volker Harth]. HEBCS thanks Kirsimari Aaltonen, Tuomas Heikkinen, and Dr. Karl von Smitten and RN Irja Erkkilä for their help with the HEBCS data and samples. We thank Peter Hillemanns, Hans Christiansen and Johann H. Karstens (HMBCS), Eija Myöhänen, Helena Kemiläinen (KBCP). kConFab thanks Heather Thorne, Eveline Niedermayr, the AOCS Management Group (D Bowtell, G Chenevix-Trench, A deFazio, D Gertig, A Green, P Webb), the ACS Management Group (A Green, P Parsons, N Hayward, P Webb, D Whiteman). LAABC thanks all the study participants and the entire data collection team, especially Annie Fung and June Yashiki. We thank Gilian Peuteman, Dominiek Smeets, Thomas Van Brussel and Kathleen Corthouts (LMBC). MARIE would like to thank Alina Vrieling, Katharina Buck, Muhabbet Celik, Ursula Eilber and Sabine Behrens for their valuable contributions to data management and assessment, and Lars Beckmann, Thomas Illig, Kirsten Mittelstraß for their valuable contributions to analysis and generation of the GWAS data. MBCSG thank Bernard Peissel, Giulietta Scuvera and Daniela Zaffaroni of the Fondazione IRCCS Istituto Nazionale dei Tumori (INT); Monica Barile and Irene Feroce of the Istituto Europeo di Oncologia (IEO) and the personnel of the Cogentech Cancer Genetic Test Laboratory. We would like to 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 (MTLGEBCS). J.S. is Chairholder of the Canada Research Chair in Oncogenetics. We 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). We thank Mervi Grip, Meeri Otsukka, Kari Mononen (OBCS), and Teresa Selander, Nayana Weerasooriya (OFBCR). ORIGO thanks 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. We thank Mark Sherman, Stephen Chanock, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner (PBCS). SASBAC and pKARMA thank The Swedish Medical Research Counsel. We thank Petra Bos, Jannet Blom, Ellen Crepin, Anja Nieuwlaat, Annette Heemskerk, the Erasmus MC Family Cancer Clinic (RBCS), and Sue Higham, Helen Cramp, Ian Brock, Sabapathy Balasubramanian, Jenny Barrett, Mark Iles, John Taylor and Dan Connley (SBCS). We thank the SEARCH and EPIC teams, and Breakthrough Breast Cancer and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study. We thank Muriel Adank for selecting the samples and Margreet Ausems, Christi van Asperen, Senno Verhoef, and Rogier van Oldenburg for providing samples from their Clinical Genetic centers (DFBBCS). The SGBCC would like to thank the participants and research coordinator Kimberley Chua. The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI (dbGAP phs000178.v8.p7, accessed 15th November 2013). Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov. We also thank Westra et al. for the provision of eQTL data at http://genenetwork.nl/bloodeqtlbrowser/. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health (commonfund.nih.gov/GTEx). Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to the The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F

(HHSN261200800001E). Statistical Methods development grants were made to the University of Geneva (MH090941), the University of Chicago (MH090951 & MH090937), the University of North Carolina - Chapel Hill (MH090936) and to Harvard University (MH090948). The datasets used for the analyses described in this manuscript were obtained from the GTEx Portal on 01/16/2014.

Conflict of Interest Statement. None declared

References

1 Stratton, M.R. and Rahman, N. (2008) The emerging landscape of breast cancer susceptibility. *Nat. Genet.*, **40**, 17-22.

2 Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K. *et al.* (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, **45**, 353-361.

3 Welter D., MacArthur J., Morales J., Burdett T., Hall P., Junkins H., Klemm A., Flicek P., Manolio T., Hindorff L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001-D1006.

4 MacPherson, G., Healey, C.S., Teare, M.D., Balasubramanian, S.P., Reed, M.W.R., Pharoah, P.D.P., Ponder, B.A.J., Meath, M., Bhattacharyya, N.P. and Cox, A. (2004) Association of a common variant of the CASP8 gene with reduced risk of breast cancer. *J. Natl. Cancer Inst.*, **96**, 1866-1869.

5 Cox, A., Dunning, A.M., Garcia-Closas, M., Balasubramanian, S., Reed, M.W.R., Pooley, K.A., Scollen, S., Baynes, C., Ponder, B.A.J., Chanock, S. *et al.* (2007) A common coding variant in CASP8 is associated with breast cancer risk. *Nat. Genet.*, **39**, 352-358.

6 Shephard, N.D., Abo, R., Rigas, S.H., Frank, B., Lin, W.Y., Brock, I.W., Shippen, A., Balasubramanian, S.P., Reed, M.W.R., Bartram, C.R. *et al.* (2009) A Breast Cancer Risk Haplotype in the Caspase-8 Gene. *Cancer Res.*, **69**, 2724-2728.

7 Camp, N.J., Parry, M., Knight, S., Abo, R., Elliott, G., Rigas, S.H., Balasubramanian, S.P., Reed, M.W., McBurney, H., Latif, A. *et al.* (2012) Fine-mapping CASP8 risk variants in breast cancer. *Cancer Epidemiol. Biomarkers Prev.*, **21**, 176-181.

8 Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C.S. *et al.* (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.*, **42**, 504-U547.

9 Hoggart, C.J., Whittaker, J.C., De Iorio, M. and Balding, D.J. (2008) Simultaneous Analysis of All SNPs in Genome-Wide and Re-Sequencing Association Studies. *PLoS Genet.*, **4**, e1000130.

10 Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E. *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238-1243.

11 Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andr√©, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-Enhancers in the Control of Cell Identity and Disease. *Cell*, **155**, 934-947.

12 Barrett, J.H., Iles, M.M., Harland, M., Taylor, J.C., Aitken, J.F., Andresen, P.A., Akslen, L.A., Armstrong, B.K., Avril, M.F., Azizi, E. *et al.* (2011) Genome-wide association study identifies three new melanoma susceptibility loci. *Nat. Genet.*, **43**, 1108-1113.

13 Berndt, S.I., Skibola, C.F., Joseph, V., Camp, N.J., Nieters, A., Wang, Z., Cozen, W., Monnereau, A., Wang, S.S., Kelly, R.S. *et al.* (2013) Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat. Genet.*, **45**, 868-876.

14 Howie, B.N., Donnelly, P. and Marchini, J. (2009) A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet.*, **5**, e1000529.

15 Howie, B., Marchini, J. and Stephens, M. (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda, Md.)*, **1**, 457-470.

16 Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B-Methodol.*, **57**, 289-300.

17 French, J.D., Ghoussaini, M., Edwards, S.L., Meyer, K.B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M.J., O'Reilly, M., Hillman, K.M. *et al.* (2013) Functional Variants at the 11q13 Risk Locus for Breast Cancer Regulate Cyclin D1 Expression through Long-Range Enhancers. *Am. J. Hum. Genet.*, **92**, 489-503.

18 Viechtbauer, W. (2010) Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.*, **36**, 1-48.

19 Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.

20 Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336-2337.

Legends to Figures

Figure 1

Breast cancer associations within the 1Mb region surrounding *CASP8*.

The upper panel plots SNPs based on their chromosomal coordinates on the x axis and their p values on the $-\log_{10}$ scale on the y axis. Circle and diamond symbols represent typed and imputed SNPs respectively. The colours indicate the pairwise r^2 with index SNP for iCHAV1, rs1830298 (highlighted in purple); r^2 are calculated based on the European panel in the 1000 genomes project. The ranges of iCHAVs 1-4 are indicated with coloured shading. Genes within the region are indicated in the lower panel, with arrows indicating transcript direction, dense blocks for exons and lines for introns. The plot was generated using LocusZoom (20).

Figure 2

Associations of the 4 index SNPs corresponding to iCHAVs 1-4, and the two previous associations, in iCOGS European subjects and GWAS data.

Squares denote the per-allele OR for the minor allele based on iCOGS and nine GWAS data, with the size of the square proportional to the sample size. Diamonds represent the pooled estimates of ORs under the fixed effect model after exclusion of the 1955 samples from the iCOGS data that were also in the combined GWAS data. Index SNPs correspond to iCHAVs as follows: rs1830298; iCHAV1, rs36043647; iCHAV2, rs59278883; iCHAV3, rs7558475 iCHAV4.

Figure 3

Study-specific OR for the minor allele of rs1830298 in iCOGS European, Asian and African-American subjects.

Squares denote the individual study per-allele OR and diamonds indicate the combined effects, with the size of the symbol indicating sample size. Fixed effect models (FE model) were used

to combine the study ORs if p for the Cochran's Q test (p_{het}) was greater than 0.05, otherwise random effect models (RE model) were used. Pooled OR across the 3 populations is shown, with p_{het} and I-squared for heterogeneity in parenthesis.

Figure 4

Associations between rs1830298 and clinical subtypes of breast cancer in iCOGS European subjects.

Squares denote the individual study per-allele OR with the size of the symbol indicating sample size. Cases in each subtype group were compared to all controls.

Figure 5

Summary of the *CASP8/ALS2CR12* locus.

The locations of iCHAVs and lead SNPs are shown relative to genes. eQTL SNPs are displayed as red marks. ENCODE DNaseI hypersenitive sites derived from various mammary cell types are depicted as gray marks. H3K27ac histone modification ChIP-seq data is shown as well as predicted enhancers and target genes from Hnisz et al (11).

Abbreviations

BCAC: Breast Cancer Association Consortium

CI: confidence interval

CLL: chronic lymphocytic leukaemia

COGS: Collaborative Oncology Gene- environment Study

DNase: deoxyribonuclease

eQTL: expression quantitative trait locus/loci

ER: estrogen receptor

FPKM: fragments per Kb of transcript per million mapped reads

GWAS: genome-wide association studies

HMEC: human mammary epithelium cells

HR: hazard ratio

iCHAV independent sets of correlated highly associated variants

iCOGS: Collaborative Oncology Gene- environment Study genotyping array

Kb: kilobase

MAF: minor allele frequency

Mb: megabase

OR: odds ratio

SNP: Single nucleotide polymorphism

TCGA: The Cancer Genome Atlas