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

Wei-Yu Lin[1,2,†], Nicola J. Camp[3,†], Maya Ghoussaini[4], Jonathan Beesley[6], Kyriaki Michailidou[4], John L. Hopper[8], Carmel Apicella[8], Melissa C. Southey[9], Jennifer Stone[10], Marjanka K. Schmidt[13], Annegien Broeks[13], Laura J. Van't Veer[13], Emiel J. Th Rutgers[13], Kenneth Muir[14,15], Artitaya Lophatananon[14], Sarah Stewart-Brown[14], Pornthep Siriwanarangsan[16], Peter A. Fasching[17,19], Lothar Haeberle[17], Arif B. Ekici[18], Matthias W. Beckmann[17], Julian Peto[20], Isabel Dos-Santos-Silva[20], Olivia Fletcher[21], Nichola Johnson[21], Manjeet K. Bolla[4], Qin Wang[4], Joe Dennis[4], Elinor J. Sawyer[24], Timothy Cheng[25], Ian Tomlinson[25], Michael J. Kerin[26], Nicola Miller[26], Frederik Marmé[27,28], Harald M. Surowy[27,29], Barbara Burwinkel[27,29], Pascal Guénel[35,36], Thérèse Truong[35,36], Florence Menegaux[35,36], Claire Mulot[37], Stig E. Bojesen[38,39,41], Børge G. Nordestgaard[38,39,41], Sune F. Nielsen[38,39], Henrik Flyger[40], Javier Benitez[42,44], M. Pilar Zamora[45], Jose Ignacio Arias Perez[46], Primitiva Menéndez[47], Anna González-Neira[43], Guillermo Pita[43], M. Rosario Alonso[43], Nuria Álvarez[43], Daniel Herrero[43], Hoda Anton-Culver[48], Hermann Brenner[30,49], Aida Karina Dieffenbach[30,49], Volker Arndt[30], Christa Stegmaier[50], Alfons Meindl[51], Peter Lichtner[52], Rita K. Schmutzler[53], Bertram Müller-Myhsok[54], Hiltrud Brauch[55,56,31], Thomas Brüning[57], Yon-Dschun Ko[58], The GENICA Network[31,33,55,56,57,58,59,60], Daniel C. Tessier[61], Daniel Vincent[61], Francois Bacot[61], Heli Nevanlinna[62,65], Kristiina Aittomäki[63,65], Carl Blomqvist[64,65], Sofia Khan[62,65], Keitaro Matsuo[66], Hidemi Ito[67], Hiroji Iwata[68], Akiyo Horio[68], Natalia V. Bogdanova[69,70], Natalia N. Antonenkova[71], Thilo Dörk[69], Annika Lindblom[72], Sara Margolin[73], Arto Mannermaa[75,77], Vesa Kataja[76,78], Veli-Matti Kosma[75,77], Jaana M. Hartikainen[75,77], kConFab Investigators[79], Australian Ovarian Cancer Study Group[7,79], Anna H. Wu[80], Chiu-Chen Tseng[80], David Van Den Berg[80], Daniel O. Stram[80], Patrick Neven[81], Els Wauters[82,83], Hans Wildiers[85], Diether Lambrechts[82,84], Jenny Chang-Claude[32], Anja Rudolph[32], Petra Seibold[32], Dieter Flesch-Janys[86], Paolo Radice[87], Paolo Peterlongo[89], Siranoush Manoukian[88], Bernardo Bonanni[90], Fergus J. Couch[91], Xianshu Wang[91], Celine Vachon[92], Kristen Purrington[93], Graham G. Giles[8,94], Roger L. Milne[8,94], Catriona Mclean[95], Christopher A. Haiman[80], Brian E. Henderson[80], Fredrick Schumacher[80], Loic Le Marchand[96], Jacques Simard[97], Mark S. Goldberg[98,99], France Labrèche[100], Martine Dumont[97], Soo Hwang Teo[101,102], Cheng Har Yip[102], Norhashimah Hassan[101,102], Eranga Nishanthie Vithana[103], Vessela Kristensen[104,105], Wei Zheng[106], Sandra Deming-Halverson[106], Martha J. Shrubsole[106], Jirong Long[106], Robert Winqvist[107], Katri Pylkäs[107], Arja Jukkola-Vuorinen[108], Saila Kauppila[109], Irene L. Andrulis[110,113], Julia A. Knight[111,114], Gord Glendon[110], Sandrine Tchatchou[112], Peter Devilee[115], Robert A.E.M. Tollenaar[116],

[†]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.

Caroline Seynaeve[118], **Christi J. Van Asperen**[117], **Montserrat García-Closas**[22,119], **Jonine Figueroa**[122], **Jolanta Lissowska**[123], **Louise Brinton**[122], **Kamila Czene**[74], **Hatef Darabi**[74], **Mikael Eriksson**[74], **Judith S. Brand**[74], **Maartje J. Hooning**[118], **Antoinette Hollestelle**[118], **Ans M.W. Van Den Ouweland**[124], **Agnes Jager**[118], **Jingmei Li**[126], **Jianjun Liu**[125], **Keith Humphreys**[74], **Xiao-Ou Shu**[106], **Wei Lu**[126], **Yu-Tang Gao**[127], **Hui Cai**[106], **Simon S. Cross**[128], **Malcolm W. R. Reed**[1], **William Blot**[106,129], **Lisa B. Signorello**[106,129], **Qiuyin Cai**[106], **Paul D.P. Pharoah**[4,5], **Barbara Perkins**[5], **Mitul Shah**[5], **Fiona M. Blows**[5], **Daehee Kang**[130,133], **Keun-Young Yoo**[132], **Dong-Young Noh**[131], **Mikael Hartman**[134,135,137], **Hui Miao**[134,137], **Kee Seng Chia**[134,137], **Thomas Choudary Putti**[136], **Ute Hamann**[33], **Craig Luccarini**[5], **Caroline Baynes**[5], **Shahana Ahmed**[5], **Mel Maranian**[5], **Catherine S. Healey**[5], **Anna Jakubowska**[138], **Jan Lubinski**[138], **Katarzyna Jaworska-Bieniek**[138], **Katarzyna Durda**[138], **Suleeporn Sangrajrang**[139], **Valerie Gaborieau**[140], **Paul Brennan**[140], **James Mckay**[140], **Susan Slager**[92], **Amanda E. Toland**[141], **Drakoulis Yannoukakos**[142], **Chen-Yang Shen**[143,145], **Chia-Ni Hsiung**[143], **Pei-Ei Wu**[144], **Shian-ling Ding**[146], **Alan Ashworth**[22], **Michael Jones**[119], **Nick Orr**[23], **Anthony J Swerdlow**[119,120], **Helen Tsimiklis**[11], **Enes Makalic**[12], **Daniel F. Schmidt**[12], **Quang M. Bui**[12], **Stephen J. Chanock**[122], **David J. Hunter**[147], **Rebecca Hein**[34,148], **Norbert Dahmen**[149], **Lars Beckmann**[150], **Kirsimari Aaltonen**[62,65], **Taru A. Muranen**[62,65], **Tuomas Heikkinen**[62,65], **Astrid Irwanto**[125], **Nazneen Rahman**[119], **Clare A. Turnbull**[119], **The Breast and Ovarian Cancer Susceptibility (BOCS) Study**[119], **Quinten Waisfisz**[151], **Hanne E. J. Meijers-Heijboer**[151], **Muriel A. Adank**[151], **Rob B. Van Der Luijt**[152], **Per Hall**[74], **Georgia Chenevix-Trench**[6,†], **Alison Dunning**[5,†], **Douglas F. Easton**[4,5,†] and **Angela Cox**[1,*,†]

[1]Department of Oncology, University of Sheffield Medical School, 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, UT 84108-1266, USA, [4]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, [5]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK, [6]Department of Genetics, [7]QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia, [8]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, [9]Department of Pathology, [10]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Melbourne School of Population Health, [11]Genetic Epidemiology Laboratory, Department of Pathology, [12]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, Melbourne School of Population Health, The University of Melbourne, Melbourne, VIC 3010, Australia, [13]Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam 1066 CX, the Netherlands, [14]Division of Health Sciences, Warwick Medical School, Warwick University, Coventry CV4 7AL, UK, [15]Institute of Population Health, University of Manchester, Manchester M13 9QQ, UK, [16]Ministry of Public Health, Nonthaburi 11000, Thailand, [17]University Breast Center Franconia, Department of Gynecology and Obstetrics, [18]Institute of Human Genetics, University Hospital Erlangen, Friedrich Alexander University Erlangen-Nuremberg, Erlangen D-91054, Germany, [19]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California, Los Angeles, CA 90095, USA, [20]Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, [21]Breakthrough Breast Cancer Research Centre, [22]Breakthrough Breast Cancer Research Centre, Division of Breast Cancer Research, [23]Breakthrough Breast Cancer Research Centre and Division of Breast Cancer Research, The Institute of Cancer Research, London SW3 6JB, UK, [24]Division of Cancer Studies, Kings College London, Guy's Hospital, London SE1 9RT, UK, [25]Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford OX3 7BN, UK, [26]School of Medicine, National University of Ireland, Galway, Ireland, [27]Department of Obstetrics and

*To whom correspondence should be addressed at: Sheffield University Medical School, Beech Hill Road, Sheffield S10 2RX, UK. Tel: +44 1142712373; Fax: +44 1142711602; Email: a.cox@sheffield.ac.uk

Gynecology, [28]National Center for Tumor Diseases, University of Heidelberg, Heidelberg 69117, Germany, [29]Molecular Epidemiology Group, [30]Division of Clinical Epidemiology and Aging Research, [31]German Cancer Consortium (DKTK), [32]Division of Cancer Epidemiology, [33]Molecular Genetics of Breast Cancer, [34]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, [35]Inserm (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, Villejuif 94807, France, [36]University Paris-Sud, UMRS 1018, Villejuif 94807, France, [37]Université Paris Sorbonne Cité, UMR-S775 Inserm, Paris 75015, France, [38]Copenhagen General Population Study, Herlev Hospital, 2730 Herlev, Copenhagen, Denmark, [39]Department of Clinical Biochemistry, [40]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Copenhagen 2100, Denmark, [41]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark, [42]Human Genetics Group, Human Cancer Genetics Program, [43]Human Genotyping-CEGEN Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid E-28029, Spain, [44]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia 28029, Spain, [45]Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid 28046, Spain, [46]Servicio de Cirugía General y Especialidades, [47]Servicio de Anatomía Patológica, Hospital Monte Naranco, Oviedo 33012, Spain, [48]Department of Epidemiology, University of California Irvine, Irvine, CA 92697, USA, [49]German Cancer Consortium (DKTK), Heidelberg, Germany, [50]Saarland Cancer Registry, Saarbrücken 66024, Germany, [51]Division of Gynaecology and Obstetrics, [52]Insitute of Human Genetics, Technische Universität München, Munich D-80333, Germany, [53]Center for Hereditary Breast and Ovarian Cancer, Center for Integrated Oncology (CIO) and Center for Molecular Medicine Cologne (CMMC), Medical Faculty, University of Cologne and University Hospital Cologne, 50932 Cologne, Germany, [54]Max Planck Institute of Psychiatry, Munich 80804, Germany, [55]Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart 70376, Germany, [56]University of Tübingen, Tübingen 72074, Germany, [57]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-University Bochum (IPA), Bochum D-44789, Germany, [58]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn 53113, Germany, [59]Institute for Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany, [60]Institute of Pathology, Medical Faculty of the University of Bonn, Bonn 53127, Germany, [61]Centre D'innovation Génome Québec et Université McGill, Montréal, QC, Canada H3A 0G1, [62]Department of Obstetrics and Gynecology, [63]Department of Clinical Genetics, [64]Department of Oncology, University of Helsinki, [65]Helsinki University Central Hospital, Helsinki 00029, Finland, [66]Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka 812-8582, Japan, [67]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya 464-8681, Japan, [68]Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya 464-8681, Japan, [69]Department of Obstetrics and Gynaecology, [70]Department of Radiation Oncology, Hannover Medical School, Hannover 30625, Germany, [71]N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk 223040, Belarus, [72]Department of Molecular Medicine and Surgery, [73]Department of Oncology–Pathology, [74]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden, [75]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine and Cancer Center of Eastern Finland, [76]School of Medicine, Institute of Clinical Medicine, Oncology, University of Eastern Finland, Kuopio 70210, Finland, [77]Imaging Center, Department of Clinical Pathology, [78]Cancer Center, Kuopio University Hospital, Kuopio 70029, Finland, [79]Peter MacCallum Cancer Center, Melbourne, VIC 3002, Australia, [80]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA, [81]University Hospital Gashuisberg, Leuven 3000, Belgium, [82]Vesalius Research Center (VRC), VIB, Leuven 3000, Belgium, [83]Department of Oncology, [84]Laboratory for Translational Genetics, Department of Oncology, [85]Department of General Medical Oncology, University Hospitals Leuven, Leuven 3000, Belgium, [86]Department of Cancer Epidemiology/Clinical Cancer Registry and Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg D-20246, Germany, [87]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, [88]Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan 20133, Italy, [89]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan 20139, Italy, [90]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia (IEO), Milan 20141, Italy, [91]Department of Laboratory Medicine and Pathology, [92]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA, [93]Karmanos Cancer Center, Detroit, MI 48201, USA, [94]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC 3004, Australia, [95]Anatomical Pathology, The Alfred Hospital, Melbourne, VIC 3004, Australia, [96]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA, [97]Centre Hospitalier Universitaire de Québec Research Centre and Laval University, Quebec City, QC, Canada G1V 4G2,

[98]Department of Medicine, McGill University, Montreal, QC, Canada H3A 0G4, [99]Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, QC, Canada H3A 1A1, [100]Départements de Santé Environnementale et Santé au Travail et de Médecine Sociale et Préventive, École de Santé Publique, Université de Montréal, Montreal, QC, Canada H3T 1J4, [101]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Selangor, Malaysia, [102]Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya Medical Centre, Kuala Lumpur 50603, Malaysia, [103]Singapore Eye Research Institute, National University of Singapore, Singapore 119077, Singapore, [104]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Oslo 0372, Norway, [105]Faculty of Medicine (Faculty Division Ahus), University of Oslo (UiO), Oslo 0316, Norway, [106]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37232, USA, [107]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, Northern Finland Laboratory Centre NordLab, [108]Department of Oncology, [109]Department of Pathology, Oulu University Hospital, University of Oulu, Oulu 90570, Finland, [110]Ontario Cancer Genetics Network, [111]Prosserman Centre for Health Research, [112]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada M5G 1X5, [113]Department of Molecular Genetics, [114]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada M5S 2J7, [115]Department of Human Genetics & Department of Pathology, [116]Department of Surgical Oncology, [117]Department of Clinical Genetics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands, [118]Family Cancer Clinic, Department of Medical Oncology, Erasmus MC Cancer Institute Rotterdam 3075 EA, The Netherlands, [119]Division of Genetics and Epidemiology, [120]Division of Breast Cancer Research, [121]Division of Cancer Genetics, The Institute of Cancer Research, London SW7 3RP, UK, [122]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20850, USA, [123]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center & Institute of Oncology, Warsaw 02-781, Poland, [124]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam 3075 EA, The Netherlands, [125]Human Genetics Division, Genome Institute of Singapore, Singapore 138672, Singapore, [126]Shanghai Center for Disease Control and Prevention, Shanghai 200336, China, [127]Department of Epidemiology, Shanghai Cancer Institute, Shanghai 220025, China, [128]Department of Neuroscience, University of Sheffield, Royal Hallamshire Hospital, Sheffield S10 2JF, UK, [129]International Epidemiology Institute, Rockville, MD 20850, USA, [130]Department of Preventive Medicine, [131]Department of Surgery, [132]Seoul National University College of Medicine, Seoul 110-799, Korea, [133]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 151-742, Korea, [134]Saw Swee Hock School of Public Health, [135]Department of Surgery, Yong Loo Lin School of Medicine, [136]Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore, [137]National University Health System, Singapore 119228, Singapore, [138]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin 70-204, Poland, [139]National Cancer Institute, Bangkok 10400, Thailand, [140]International Agency for Research on Cancer, Lyon 69372, France, [141]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA, [142]Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research 'Demokritos', Aghia Paraskevi Attikis, Athens 153 10, Greece, [143]Institute of Biomedical Sciences, [144]Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, [145]College of Public Health, China Medical University, Taichung 40402, Taiwan, [146]Department of Nursing, Kang-Ning Junior College of Medical Care and Management, Taipei 11529, Taiwan, [147]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA, [148]PMV Research Group at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, Cologne 50923, Germany, [149]Department of Psychiatry, University of Mainz, Mainz 55122, Germany, [150]Institute for Quality and Efficiency in Health Care (IQWiG), Cologne 50670, Germany, [151]Department of Clinical Genetics, Section Oncogenetics, VU University Medical Center, Amsterdam 1081 HZ, The Netherlands and [152]Department of Medical Genetics, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands

---

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 1 Mb 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 1232 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% confidence interval, CI)] for the minor allele of 1.05 (1.03–1.07), $P = 1 \times 10^{-5}$. 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 ($P = 3 \times 10^{-6}$), yielding a combined OR (95% CI) of 1.06 (1.04–1.08), $P = 1 \times 10^{-9}$. 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 etiology (1). In recent years, ~80 low penetrance breast cancer alleles have been identified, with modest odds ratios, ranging from 1.05 to 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 missense 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 >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 3659 cases and 4897 controls found suggestive evidence of association [OR (95% confidence interval, CI) 1.14 (1.06–1.22); $P = 1.5 \times 10^{-4}$] 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 analyzed 501 SNPs across a 1 Mb 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 1733 typed and imputed SNPs across a 1 Mb 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 (Fig. 1; Supplementary Material, Table S2). The strongest signals came from imputed SNP rs1830298 in *ALS2CR12*, with minor allele frequency (MAF) of 0.29 and an estimated OR (95% CI) per copy of the minor allele of 1.05 (1.03–1.07), $P = 1.1 \times 10^{-5}$, and the genotyped SNP rs10197246 (MAF = 0.28), with odds ratio (95% CI) 1.05 (1.02–1.07), $P = 2.5 \times 10^{-5}$. 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 (Supplementary Material, 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), $P = 1.9 \times 10^{-4}$], compared with 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 does not preclude the possibility of other signals in the region. To test this hypothesis, we carried out a regression analysis testing the association of individual
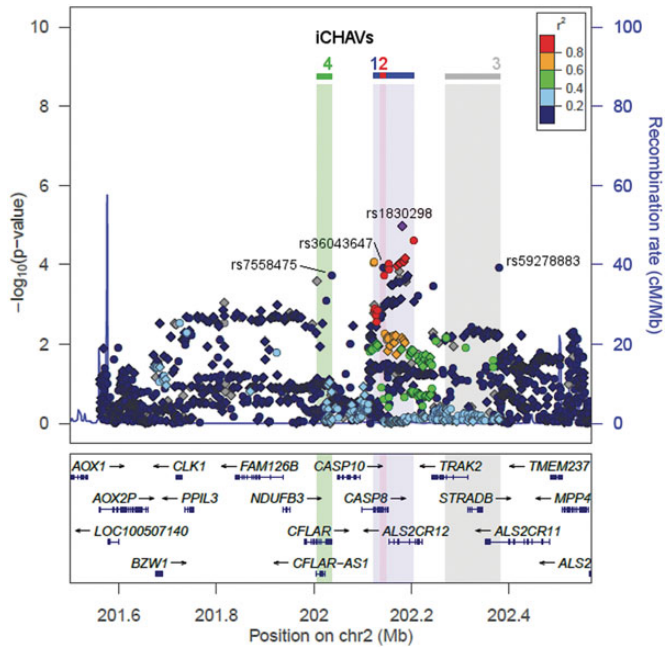
Figure 1. Breast cancer associations within the 1 Mb region surrounding *CASP8*. The upper panel plots SNPs based on their chromosomal coordinates on the *x*-axis and their *P*-values on the $-\log_{10}$ scale on the *y*-axis. Circle and diamond symbols represent typed and imputed SNPs, respectively. The colors indicate the pairwise r^2 with index SNP for iCHAV1, rs1830298 (highlighted in purple); r^2 is calculated based on the European panel in the 1000 genomes project. The ranges of iCHAVs 1–4 are indicated with colored 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 (9).

SNPs adjusted for the top hit rs1830298, in the iCOGS European dataset (Supplementary Material, Table S3). Interestingly, while this resulted in the loss of the signal from the main peak in *CASP8/ALS2CR12*, residual associations remained (e.g. 43 SNPs with $P \leq 1 \times 10^{-3}$), suggesting that there may be further signals present in the region, albeit weaker (Supplementary Material, Table S3 and Fig. S1). To investigate this further, we carried out penalized logistic regression analysis of all 1733 SNPs to identify the best subset of SNPs that explain the association, using HyperLasso (10). This identified 59 models containing combinations of 27 SNPs (Supplementary Material, Table S2), but many of these models were equivalent after taking into account linkage disequilibrium between SNPs. To obtain the most parsimonious model, we carried out stepwise forward logistic regression on the 27 SNPs, which resulted in a model containing four SNPs; rs1830298 (*ALS2CR12*; $p_{conditional} = 9.3 \times 10^{-3}$, MAF = 0.29), rs36043647 (*CASP8*; $p_{conditional} = 1.9 \times 10^{-4}$, MAF = 0.06), rs59278883 (*ALS2CR11*; $p_{conditional} = 6.1 \times 10^{-4}$, MAF = 0.07) and rs7558475 (*CFLAR*; CASP8- and FADD-like apoptosis regulator; $p_{conditional} = 9.2 \times 10^{-4}$, MAF = 0.07). We refer to these four SNPs, marking four independent sets of correlated highly associated variants (iCHAVs), as index SNPs.

Meta-analysis of iCOGS and combined nine GWAS data

We first examined the results for the four index SNPs, together with the previous hits rs1045485 and rs10931936, in the combined nine GWAS meta-analysis, and then carried out a further
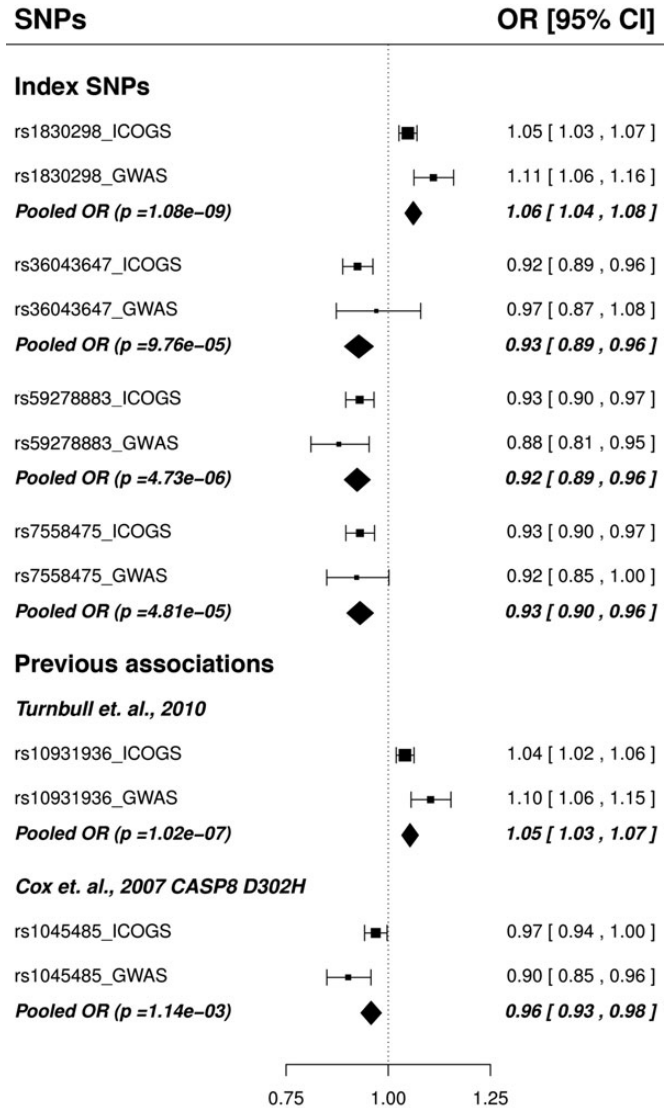
Figure 2. Associations of the four index SNPs corresponding to iCHAVs 1–4, and the two previous associations, in iCOGS European subjects and GWAS data. Squares denote the per-allele OR for the minor allele based on iCOGS and nine GWAS data, with the size of the square proportional to the sample size. 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.

meta-analysis combining the iCOGS European data with the combined nine GWAS for these SNPs (total sample size 56 502 cases and 55 175 controls; Supplementary Material, Tables S4 and S5). We found that the top index SNP, rs1830298, replicated in the combined GWAS data alone ($P = 2.7 \times 10^{-6}$), and reached genome-wide significance ($P = 1.1 \times 10^{-9}$) in the meta-analysis containing both the iCOGS and combined GWAS data (Supplementary Material, Table S5; Fig. 2). The genotyped proxy rs10197246 also reached genome-wide significance ($P = 1.7 \times 10^{-8}$). When we examined the other three index SNPs in the combined GWAS data, we found a replicated association ($P = 1.8 \times 10^{-3}$) for rs59278883, a
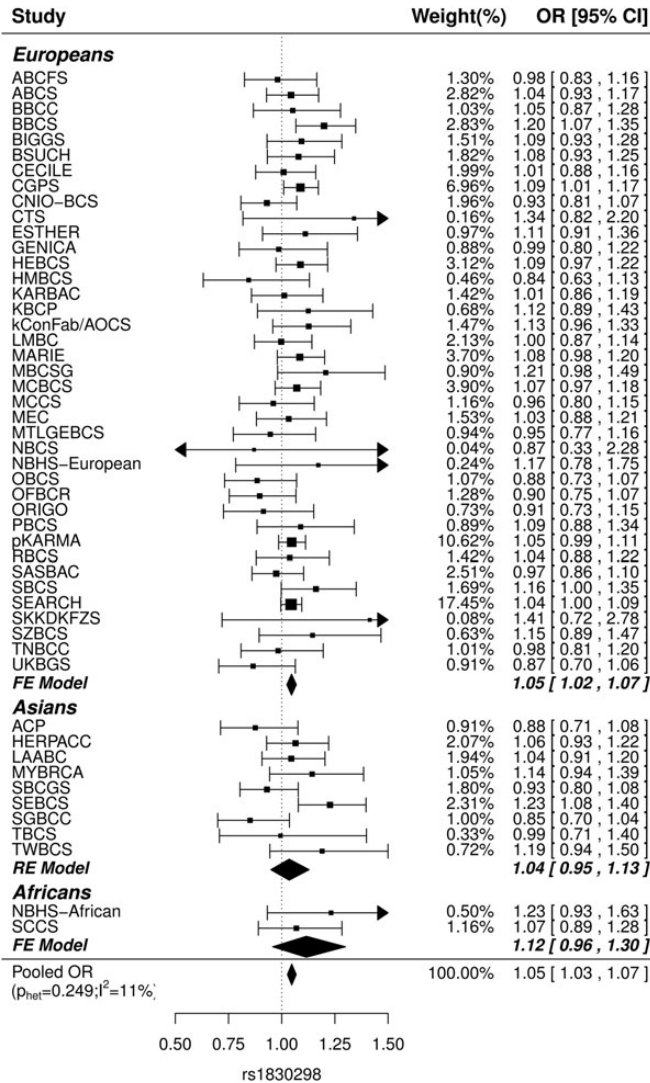
Figure 3. Study-specific OR for the minor allele of rs1830298 in iCOGS European, Asian and African-American subjects. Squares denote the individual study per-allele OR and diamonds indicate the combined effects, with the size of the symbol indicating sample size. Fixed effect models (FE model) were used to combine the study ORs if p for the Cochran's Q test (p_{het}) was >0.05, otherwise random effect models (RE model) were used. Pooled OR across the three populations is shown, with p_{het} and I^2 for heterogeneity in parenthesis.
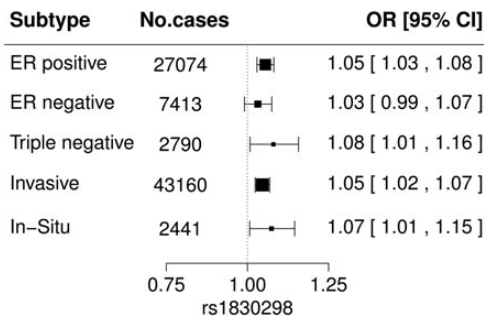
Figure 4. Associations between rs1830298 and clinical subtypes of breast cancer in iCOGS European subjects. Squares denote the individual study per-allele OR with the size of the symbol indicating sample size. Cases in each subtype group were compared with all controls.

null result for rs36043647 ($P = 0.58$) and borderline evidence for rs7558475 ($P = 0.05$) (Supplementary Material, Table S5; Fig. 2). However, these three index SNPs all showed some evidence of association in the meta-analysis of iCOGS and combined GWAS (Supplementary Material, Table S5 and Fig. 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 ($P = 1.0 \times 10^{-7}$). Weak evidence for association was observed for *CASP8 D302H* rs1045485 ($P = 1.1 \times 10^{-3}$).

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 (Fig. 3; Supplementary Material, 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 among 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 ORs were consistent with the European data, and the pooled OR (95% CI) was 1.05 (1.03–1.07); $P = 4.1 \times 10^{-6}$ 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 (Supplementary Material, Table S6). We did not observe any association of rs59278883 and rs7558475 in Asian and African-American populations (Supplementary Material, 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 tumor 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 (Fig. 4). Similarly, no significant differences in OR were seen when cases were stratified by family history, tumor grade, tumor stage, tumor size and lymph node status (Supplementary Material, Fig. S2). A broadly similar picture was seen for the other index SNPs (Supplementary Material, Figs S2 and S3).

SNP effects were also evaluated for overall survival and breast cancer-specific survival. There were 4191 deaths among 39 140 breast cancer patients with known vital status in the European dataset. Of these deaths, 1979 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).
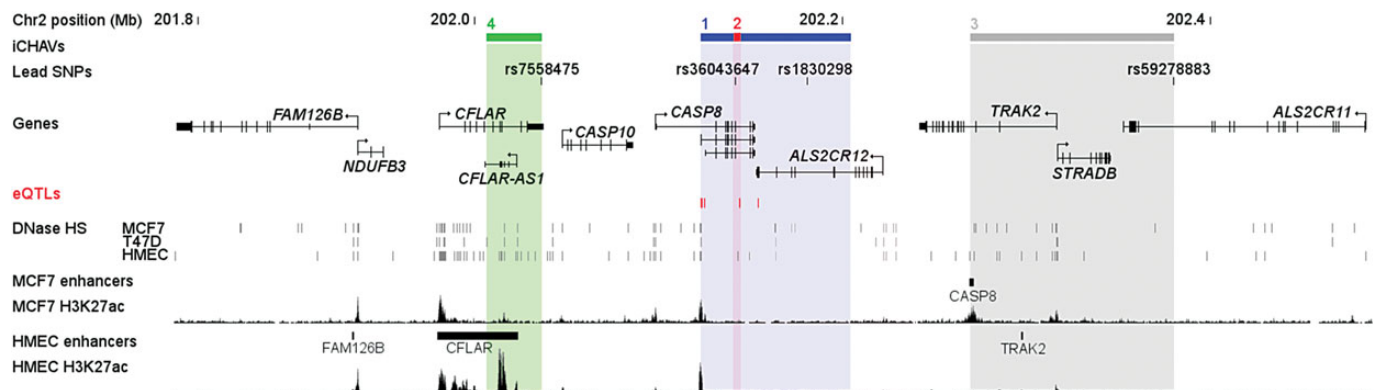
Figure 5. Summary of the *CASP8/ALS2CR12* locus. The locations of iCHAVs and lead SNPs are shown relative to genes. eQTL SNPs are displayed as red marks. ENCODE DNaseI hypersensitive sites derived from various mammary cell types are depicted as gray marks. H3K27ac histone modification ChIP-seq data are shown as well as predicted enhancers and target genes from Hnisz *et al.* (12).

In silico functional and expression quantitative trait loci 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 hypersensitivity and histone H3K27 acetylation in breast cell lines (Fig. 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.* ($P \leq 9.4 \times 10^{-5}$; Supplementary Material, Table S7; Fig. 5) (11). 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 ($P \leq 1 \times 10^{-3}$; Supplementary Material, Table S7; Fig. 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, respectively (12). However, there was limited eQTL data available for these iCHAVs, with no evidence of any significant eQTLs (Supplementary Material, 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 ($P = 1.1 \times 10^{-5}$ and 2.5×10^{-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 ($P = 1.1 \times 10^{-9}$ and $P = 1.7 \times 10^{-8}$, respectively). This association is consistent between ER-positive and -negative disease, and between invasive and *in situ* cancers (Fig. 4). The previously published result for rs10931936 in the UK GWAS is consistent with its correlation with rs1830298 (8).

Several of the SNPs in iCHAV1 were 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 (Supplementary Material, Table S7; Fig. 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 tumor 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 the anti-apoptotic gene *CFLAR* (iCHAV4); see Figure 2; Supplementary Material, 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 datasets analyzed here, they are harder to interpret. However, it is interesting that both these iCHAVs overlap enhancer regions (Fig. 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, $P = 1.9 \times 10^{-4}$; Supplementary Material, Table S2) and

combined GWAS ($P = 4.0 \times 10^{-5}$). We found no significant differences between subtypes, although the associated effect for D302H was stronger (and borderline significant) for triple negative disease, despite the smaller sample size (Supplementary Material, Fig. 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 leukemia (CLL) (13,14). 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 with that in epithelial cancers.

MATERIALS AND METHODS

Study samples

The iCOGS and nine breast cancer GWAS datasets 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 (6269 cases; 6624 controls) and 2048 African Americans (1116 cases and 932 controls). The numbers of subjects by study are detailed in Supplementary Material, Table S1. Approximately 93% of cases had invasive breast cancer (Supplementary Material, Table S1). The combined nine breast cancer GWAS dataset comprised 10 052 cases and 12 575 controls of European ancestry from United States, 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 1 Mb region from 201 566 128 to 202 566 128 (hg19). In March 2010 when the iCOGS array was designed, 2191 SNPs had been catalogued in this region by the 1000 genomes and HapMap3 projects. Of these, 1723 SNPs had an MAF $\geq 2\%$, and of these 1723, 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 and 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 centers 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 $P < 1 \times 10^{-7}$ (2). Subjects were excluded from analysis for the following reasons: genotypically non-female; overall call rate <95%; low or high heterozygosity ($P < 1 \times 10^{-6}$); discordant replicates or cryptic duplicates. Genotype data and ancestry principal components (seven principal components for the European and two each for the Asian and African-American populations) were thus available for 103 991 individuals.

Statistical analysis

The iCOGS *CASP8* region genotype data were split into four groups for efficiency of imputation of missing genotypes and untyped SNPs. These comprised 36 793 European ancestry subjects from North American and UK studies in Group 1, with 26 129 and 26 128 of the remaining European subjects in Groups 2 and 3, respectively, and 14 941 Asians and African Americans in Group 4. Imputations were carried out separately by group based on the 1000 genomes phase I reference panel with singleton variants excluded, using IMPUTE2 version 2.3 (15,16). SNPs were included in the subsequent analyses if the mean information score of the European groups was ≥ 0.9, and untyped imputed SNPs were only included if their MAF was $\geq 3\%$; these criteria resulted in inclusion of 501 typed and 1232 imputed SNPs in the final analysis. The imputation accuracy for rs1830298 was verified in whole-genome sequence data from 197 individuals; the correlation between the observed and imputed genotypes was 0.974. The imputation step increases the number of common SNPs captured at $r^2 > 0.9$ from 76% (1198/1583) to 84% (1333/1583).

The main analyses were based on the data for individuals of European ancestry. For each SNP, allelic dosage of the minor allele was estimated, and included in a logistic regression model, to estimate OR and corresponding 95% CI. Covariates for each study plus the seven ancestry principal components were included in the model (2). These analyses were implemented in R. *P*-values from the Wald test are reported in the text (uncorrected for multiple testing). FDR values in Supplementary Material, Table S2 were calculated according to the Benjamini & Hochberg method, as implemented in the R p.adjust function (17). Penalized logistic regression models (based on the normal exponential gamma probability density) were implemented in HyperLasso (10), including all 501 typed and 1232 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 λ of 0.05 and penalty of 491 were specified in HyperLasso, according to equation 7 in Hoggart

et al. (10). 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 (18). 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 1955 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 (19). We used the threshold of $P = 5 \times 10^{-8}$ 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, tumor 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 iCHAVs with likelihood (determined from the individual-SNP logistic regression analysis) relative to an index SNP of $>1/100$ and degree of correlation with the index SNP of >0.65. The ENCODE integrated regulation data for each SNP were retrieved from the UCSC Genome Browser by use of ANNOVAR (20). Predicted enhancers and target genes were retrieved from Hnisz *et al.* (12). Expression QTL data were obtained by interrogation of the GTEx Portal, the online results of the peripheral blood eQTL meta-analysis based on 5311 samples from seven studies by Westra and colleagues (11), and from breast cancer cases in the TCGA Project. For the latter, RNAseq data in the form of fragments per kilobase 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.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

FUNDING

ACKNOWLEDGEMENTS

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. 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.
10. 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.
11. 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.
12. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, 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.
13. 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.
14. 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.
15. 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.
16. Howie, B., Marchini, J. and Stephens, M. (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda, MD)*, **1**, 457–470.
17. 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.
18. 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.
19. Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.*, **36**, 1–48.
20. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.