Published in final edited form as:
Nat Genet. 2013 April ; 45(4): 353–361e2. doi:10.1038/ng.2563.

Large-scale genotyping identifies 41 new loci associated with breast cancer risk

Kyriaki Michailidou[1,138], Per Hall[2,138], Anna Gonzalez-Neira[3], Maya Ghoussaini[4], Joe Dennis[1], Roger L Milne[5], Marjanka K Schmidt[6,7], Jenny Chang-Claude[8], Stig E Bojesen[9,10], Manjeet K Bolla[1], Qin Wang[1], Ed Dicks[4], Andrew Lee[1], Clare Turnbull[11], Nazneen Rahman[11], The Breast and Ovarian Cancer Susceptibility Collaboration[12], Olivia Fletcher[13], Julian Peto[14], Lorna Gibson[14], Isabel dos Santos Silva[14], Heli Nevanlinna[15], Taru A Muranen[15], Kristiina Aittomäki[16], Carl Blomqvist[17], Kamila Czene[2], Astrid Irwanto[18], Jianjun Liu[18], Quinten Waisfisz[19], Hanne Meijers-Heijboer[19], Muriel Adank[19], Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON)[12], Rob B van der Luijt[20], Rebecca Hein[8,21], Norbert Dahmen[22], Lars Beckman[23], Alfons Meindl[24], Rita K

Correspondence should be addressed to: D.F.E. (dfe20@medschl.cam.ac.uk) or P.H. (per.hall@ki.se).

[12] A list of members is provided in the Supplementary Note

[138] These authors contributed equally to this work.

Note: Supplementary information is available in the online version of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

AUTHOR CONTRIBUTIONS

K. Michailidou and D.F.E. performed the statistical analysis and drafted the manuscript. D.F.E. conceived and coordinated the synthesis of the iCOGS array and led BCAC. P.H. coordinated COGS. J. Benitez led the iCOGS genotyping working group. A.G.-N., G.P., M.R.A., J. Benitez, D.V., F.B., D.C.T., J. Simard, A.M.D. and C.L. coordinated genotyping of the iCOGS array. M.G.-C., P.D.P.P. and M.K.S. led the BCAC pathology and survival working group. J.C.-C. led the BCAC risk factor working group. A.M.D. and G.C.-T. led the iCOGS quality control working group. J.D., E.D., M. Ghoussaini and A. Lee provided bioinformatics support. M.K.B. and Q. Wang provided data management support for BCAC. S.C. and L.F.A.W. provided analysis of the TCGA expression data. C.T., N.R. and D.F.E. led the UK2 GWAS. O.F., J.P. and I.d.S.S. led the BBCS GWAS. H.N., T.A.M., K. Aittomäki and C.B. led the HEBCS GWAS. P.H., K.C., A.I. and J. Liu led the SASBAC GWAS. Q. Waisfisz, H.M.-H., M.A. and R.B.v.d.L. led the DFBBCS GWAS. J.C.-C., R.H., N.D. and L. Beckman led the MARIE GWAS. A. Meindl, R.K.S., B.M.-M. and P.L. led the GC-HBOC GWAS. J.L.H., M.C.S., E.M., D.F.S. and H.T. led the ABCFS GWAS. A.G.U. and A. Hofman led the genotyping in the Rotterdam study. D.J.H. and S.J.C. led the CGEMS GWAS. F.J.C. and S. Slager coordinated TNBCC. C.A.H., B.E.H., F.S. and L.L.M. coordinated MEC. P.D.P.P., D.F.E. and M. Shah coordinated SEARCH. R.L. coordinated EPIC-Norfolk. J. Brown coordinated SIBS. P.H., K.C., N.S., K.H. and J. Li coordinated SASBAC and pKARMA. S.E.B., B.G.N., S.F.N. and H.F. coordinated CGPS. F.J.C., X.W., C.V. and K.N.S. coordinated MCBCS. D.L., M.M., R.P. and M.-R.C. coordinated LMBC. J.C.-C., A.R., S.N. and D.F.-J. coordinated MARIE. N.J., L.G. and Z.A. coordinated BBCS. K. Aaltonen and T.H. coordinated HEBCS. M.K.S., A.B., L.J.V.t.V. and C.E.v.d.S. coordinated ABCS. P.G., T.T., P.L.-P. and F. Menegaux coordinated CECILE. F. Marme, A. Schneeweiss, C. Sohn and B. Burwinkel coordinated BSUCH. R.L.M., A.G.-N., M.P.Z., J.I.A.P. and J. Benitez coordinated CNIO-BCS. A.C., I.W.B., S.S.C. and M.W.R.R. coordinated SBCS. E.J.S., I.T., M.J.K. and N.M. coordinated BIGGS. I.L.A., J.A.K., G.G. and A.M.M. coordinated OFBCR. A. Lindblom and S. Margolin coordinated KARBAC. M.J.H., A. Hollestelle, A.M.W.v.d.O. and A. Jager coordinated RBCS. J.L.H., M.C.S., Q.M.B., J. Stone, G.S.D. and C.A. coordinated ABCFS. J.L.H., M.C.S., G.G.G., G.S. and L. Baglietto coordinated MCCS. P.A.F., L.H., A.B.E. and M.W.B. coordinated BBCC. H. Brenner, H. Müller, V.A. and C. Stegmaier coordinated ESTHER. A. Swerdlow, A.A., N.O., M.J. and M.G.-C. coordinated UKBGS. M.G.-C., J.F., J. Lissowska and L. Brinton coordinated PBCS. M.S.G., F.L., M.D. and J. Simard coordinated MTLGEBCS. R.W., K.P., A.J.-V. and M. Grip coordinated OBCS. H. Brauch, U.H. and T.B. coordinated GENICA. P.R., P.P., S. Manoukian and B. Bonanni coordinated MBCSG. P.D., R.A.E.M.T., C. Seynaeve and C.J.v.A. coordinated ORIGO. A. Jakubowska, J. Lubinski, K.J. and K.D. coordinated SZBCS. A. Mannermaa, V.K., V.-M.K. and J.M.H. coordinated KBCP. N.V.B., N.N.A. and T.D. coordinated HMBCS. V.N.K. coordinated NBCS. H.A.-C. coordinated UCIBCS. A.E.T. coordinated OSU. S.E. coordinated RPCI. F.F. coordinated DEMOKRITOS. D.K., K.-Y.Y. and D.-Y.N. coordinated SEBCS. K. Matsuo, H. Ito, H. Iwata and A. Sueta coordinated HERPACC. A.H.W., C.-C.T., D.V.D.B. and D.O.S. coordinated LAABC. W.Z., X.-O.S., W.L., Y.-T.G. and H.C. coordinated SGBCS. S.H.T., C.H.Y., S.Y.P. and B.K.C. coordinated MYBRCA. M.H., H. Miao, W.Y.L. and J.-H.S. coordinated SGBCC. K. Muir, A. Lophatananon, S.S.-B. and P.S. coordinated ACP. C.-Y.S., C.-N.H., P.-E.W. and S.-L.D. coordinated TWBCS. S. Sangrajrang, V.G., P.B. and J.M. coordinated TBCS. W.J.B., L.B.S., Q.C. and W.Z. coordinated SCCS. W.Z., S.D.-H., M. Shrubsole and J. Long coordinated NBHS. G.C.-T. coordinated the genotyping component of kConFab. All authors provided critical review of the manuscript.

Schmutzler[25,26], Bertram Müller-Myhsok[27], Peter Lichtner[28], John L Hopper[29], Melissa C Southey[30], Enes Makalic[29], Daniel F Schmidt[29], Andre G Uitterlinden[31], Albert Hofman[32], David J Hunter[33], Stephen J Chanock[34], Daniel Vincent[35], François Bacot[35], Daniel C Tessier[35], Sander Canisius[36], Lodewyk F A Wessels[36], Christopher A Haiman[37], Mitul Shah[4], Robert Luben[1], Judith Brown[1], Craig Luccarini[4], Nils Schoof[2], Keith Humphreys[2], Jingmei Li[18], Børge G Nordestgaard[9,10], Sune F Nielsen[9,10], Henrik Flyger[38], Fergus J Couch[39], Xianshu Wang[39], Celine Vachon[40], Kristen N Stevens[40], Diether Lambrechts[41,42], Matthieu Moisse[41,42], Robert Paridaens[43], Marie-Rose Christiaens[43], Anja Rudolph[8], Stefan Nickels[8], Dieter Flesch-Janys[8,44,45], Nichola Johnson[13], Zoe Aitken[14], Kirsimari Aaltonen[15,16,17], Tuomas Heikkinen[15], Annegien Broeks[6], Laura J Van't Veer[6], C Ellen van der Schoot[46], Pascal Guénel[47,48], Thérèse Truong[47,48], Pierre Laurent-Puig[49], Florence Menegaux[47,48], Frederik Marme[50,51], Andreas Schneeweiss[50,51], Christof Sohn[50], Barbara Burwinkel[50,52], M Pilar Zamora[53], Jose Ignacio Arias Perez[54], Guillermo Pita[3], M Rosario Alonso[3], Angela Cox[55], Ian W Brock[55], Simon S Cross[56], Malcolm W R Reed[55], Elinor J Sawyer[57], Ian Tomlinson[58,59], Michael J Kerin[60], Nicola Miller[60], Brian E Henderson[37], Fredrick Schumacher[37], Loic Le Marchand[61], Irene L Andrulis[62,63], Julia A Knight[64,65], Gord Glendon[62], Anna Marie Mulligan[66,67], kConFab Investigators[12], Australian Ovarian Cancer Study Group[12], Annika Lindblom[68], Sara Margolin[69], Maartje J Hooning[70], Antoinette Hollestelle[70], Ans M W van den Ouweland[71], Agnes Jager[70], Quang M Bui[29], Jennifer Stone[29], Gillian S Dite[29], Carmel Apicella[29], Helen Tsimiklis[30], Graham G Giles[29,72], Gianluca Severi[29,72], Laura Baglietto[29,72], Peter A Fasching[73,74], Lothar Haeberle[73], Arif B Ekici[75], Matthias W Beckmann[73], Hermann Brenner[76], Heiko Müller[76], Volker Arndt[76], Christa Stegmaier[77], Anthony Swerdlow[11], Alan Ashworth[13,78], Nick Orr[13,78], Michael Jones[11], Jonine Figueroa[34], Jolanta Lissowska[79], Louise Brinton[34], Mark S Goldberg[80,81], France Labrèche[82], Martine Dumont[83], Robert Winqvist[84], Katri Pylkäs[84], Arja Jukkola-Vuorinen[85], Mervi Grip[86], Hiltrud Brauch[87,88], Ute Hamann[89], Thomas Brüning[90], The GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network[12], Paolo Radice[91,92], Paolo Peterlongo[91,92], Siranoush Manoukian[93], Bernardo Bonanni[94], Peter Devilee[95,96], Rob A E M Tollenaar[97], Caroline Seynaeve[98], Christi J van Asperen[99], Anna Jakubowska[100], Jan Lubinski[100], Katarzyna Jaworska[100,101], Katarzyna Durda[100], Arto Mannermaa[102,103,104], Vesa Kataja[104,105,106], Veli-Matti Kosma[102,103,104], Jaana M Hartikainen[102,103,104], Natalia V Bogdanova[107,108], Natalia N Antonenkova[109], Thilo Dörk[107], Vessela N Kristensen[110,111], Hoda Anton-Culver[112], Susan Slager[40], Amanda E Toland[113], Stephen Edge[114], Florentia Fostira[115], Daehee Kang[116], Keun-Young Yoo[116], Dong-Young Noh[116], Keitaro Matsuo[117], Hidemi Ito[117], Hiroji Iwata[118], Aiko Sueta[117], Anna H Wu[37], Chiu-Chen Tseng[37], David Van Den Berg[37], Daniel O Stram[37], Xiao-Ou Shu[119], Wei Lu[120], Yu-Tang Gao[121], Hui Cai[119], Soo Hwang Teo[122,123], Cheng Har Yip[123], Sze Yee Phuah[122], Belinda K Cornes[124], Mikael Hartman[125,126], Hui Miao[125], Wei Yen Lim[125], Jen-Hwei Sng[126], Kenneth Muir[127], Artitaya Lophatananon[127], Sarah Stewart-Brown[127], Pornthep Siriwanarangsan[128], Chen-Yang Shen[129,130], Chia-Ni Hsiung[129], Pei-Ei Wu[131], Shian-Ling Ding[132], Suleeporn Sangrajrang[133], Valerie Gaborieau[134], Paul Brennan[134], James McKay[134], William J Blot[119,135], Lisa B Signorello[119,135], Qiuyin Cai[119], Wei Zheng[119], Sandra Deming-Halverson[119], Martha Shrubsole[119], Jirong Long[119], Jacques Simard[83], Montse Garcia-Closas[11,13,78], Paul D P Pharoah[1,4], Georgia Chenevix-Trench[136], Alison M Dunning[4], Javier Benitez[3,137], and **Douglas F Easton[1,4]**

[1]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [2]Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden [3]Human Genotyping Unit–Centro Nacional de Genotipado (CEGEN), Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain [4]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK [5]Genetic & Molecular Epidemiology Group, Human

Cancer Genetics Programme, CNIO, Madrid, Spain [6]Division of Molecular Pathology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands [7]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands [8]Division of Cancer Epidemiology, Deutsches Krebsforschungszentrum, Heidelberg, Germany [9]Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark [10]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark [11]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, UK [13]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK [14]Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK [15]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland [16]Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland [17]Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland [18]Human Genetics Division, Genome Institute of Singapore, Singapore [19]Section of Oncogenetics, Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands [20]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands [21]PMV (Primär Medizinische Versorgung) Research Group at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany [22]Department of Psychiatry, University of Mainz, Mainz, Germany [23]Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany [24]Division for Gynaecological Tumor Genetics, Clinic of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany [25]Centre of Familial Breast and Ovarian Cancer, University of Cologne, Cologne, Germany [26]Centre for Molecular Medicine (CMMC), University of Cologne, Cologne, Germany [27]Max Planck Institute of Psychiatry, Munich, Germany [28]Institute of Human Genetics, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany [29]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, Melbourne School of Population Health, The University of Melbourne, Melbourne, Victoria, Australia [30]Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia [31]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands [32]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands [33]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA [34]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA [35]McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada [36]Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands [37]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA [38]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark [39]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA [40]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA [41]Vesalius Research Center (VRC), VIB, Leuven, Belgium [42]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium [43]Department of Oncology, University Hospital Gasthuisberg, University of Leuven, Leuven, Belgium [44]Department of Cancer Epidemiology/Clinical Cancer Registry, University Clinic Hamburg-Eppendorf, Hamburg, Germany [45]Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany [46]Sanquin Research, Amsterdam, The Netherlands [47]INSERM (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, Villejuif, France [48]Unité Mixte de Recherche Scientifique (UMRS) 1018, University Paris–Sud, Villejuif, France [49]UMRS 775, INSERM, Université Paris Sorbonne Cité, Paris, France [50]Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany

[51]National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany [52]Molecular Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany [53]Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid, Spain [54]Servicio de Cirugía General y Especialidades, Hospital Monte Naranco, Oviedo, Spain [55]Cancer Research UK/Yorkshire Cancer Research Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield, UK [56]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK [57]Division of Cancer Studies, National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Guy's & St. Thomas' National Health Service (NHS) Foundation Trust in partnership with King's College London, London, UK [58]Welcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK [59]Oxford Biomedical Research Centre, University of Oxford, Oxford, UK [60]Clinical Science Institute, University Hospital Galway, Galway, Ireland [61]University of Hawaii Cancer Center, Honolulu, Hawaii, USA [62]Ontario Cancer Genetics Network, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada [63]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada [64]Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada [65]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada [66]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada [67]Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada [68]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden [69]Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden [70]Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands [71]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands [72]Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Victoria, Australia [73]University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen–Nuremberg, Erlangen, Germany [74]Division of Hematology and Oncology, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA [75]Institute of Human Genetics, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany [76]Division of Clinical Epidemiology and Aging Research, DKFZ, Heidelberg, Germany [77]Saarland Cancer Registry, Saarbrücken, Germany [78]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK [79]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland [80]Department of Medicine, McGill University, Montreal, Quebec, Canada [81]Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, Quebec, Canada [82]Département de Médecine Sociale et Préventive, Département de Santé Environnementale et Santé au Travail, Université de Montréal, Montreal, Quebec, Canada [83]Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Québec and Laval University, Quebec City, Quebec, Canada [84]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Genetics and Biocenter Oulu, University of Oulu, Oulu University Hospital, Oulu, Finland [85]Department of Oncology, Oulu University Hospital, University of Oulu, Oulu, Finland [86]Department of Surgery, Oulu University Hospital, University of Oulu, Oulu, Finland [87]Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany [88]University of Tübingen, Tübingen, Germany [89]Molecular Genetics of Breast Cancer, DKFZ, Heidelberg, Germany [90]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr–Universität Bochum (IPA), Bochum, Germany [91]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale Tumori (INT), Milan, Italy [92]Istituto FIRC di Oncologia Molecolare (IFOM), Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy [93]Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS INT, Milan, Italy [94]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia, Milan,

Italy [95]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands [96]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands [97]Department of Surgical Oncology, Leiden University Medical Center, Leiden, The Netherlands [98]Family Cancer Clinic, Department of Medical Oncology, Erasmus Medical Center–Daniel den Hoed Cancer Center, Rotterdam, The Netherlands [99]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands [100]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland [101]Postgraduate School of Molecular Medicine, Warsaw Medical University, Warsaw, Poland [102]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland [103]Biocenter Kuopio, Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio, Finland [104]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland [105]School of Medicine, Institute of Clinical Medicine and Oncology, University of Eastern Finland, Kuopio, Finland [106]Cancer Center, Kuopio University Hospital, Kuopio, Finland [107]Department of Obstetrics and Gynaecology, Hannover Medical School, Hannover, Germany [108]Department of Radiation Oncology, Hannover Medical School, Hannover, Germany [109]NN Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus [110]Institute for Clinical Epidemiology and Molecular Biology (EpiGen), Faculty of Medicine, University of Oslo, Oslo, Norway [111]Group of Cancer Genome Variation, Department of Genetics, Institute for Cancer Research, Rikshospitalet-Radiumhospitalet, Oslo, Norway [112]Department of Epidemiology, University of California–Irvine, Irvine, California, USA [113]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA [114]Roswell Park Cancer Institute, Buffalo, New York, USA [115]Molecular Diagnostics Laboratory, Institute of Radioisotopes and Radiodiagnostic Products (IRRP), National Centre for Scientific Research Demokritos, Aghia Paraskevi Attikis, Athens, Greece [116]Seoul National University College of Medicine, Seoul, Korea [117]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan [118]Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya, Japan [119]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA [120]Shanghai Center for Disease Control and Prevention, Shanghai, China [121]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China [122]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Malaysia [123]Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya Medical Centre, Kuala Lumpur, Malaysia [124]Singapore Eye Research Institute, National University of Singapore, Singapore [125]Saw Swee Hock School of Public Health, National University of Singapore, Singapore [126]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore [127]Warwick Medical School, University of Warwick, Coventry, UK [128]Ministry of Public Health, Bangkok, Thailand [129]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan [130]Colleague of Public Health, China Medical University, Taichong, Taiwan [131]Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan [132]Department of Nursing, Kang-Ning Junior College of Medical Care and Management, Taipei, Taiwan [133]National Cancer Institute, Bangkok, Thailand [134]International Agency for Research on Cancer, Lyon, France [135]International Epidemiology Institute, Rockville, Maryland, USA [136]Department of Genetics, Queensland Institute of Medical Research, Brisbane, Queensland, Australia [137]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain

Abstract

Breast cancer is the most common cancer among women. Common variants at 27 loci have been identified as associated with susceptibility to breast cancer, and these account for ~9% of the familial risk of the disease. We report here a meta-analysis of 9 genome-wide association studies, including 10,052 breast cancer cases and 12,575 controls of European ancestry, from which we selected 29,807 SNPs for further genotyping. These SNPs were genotyped in 45,290 cases and 41,880 controls of European ancestry from 41 studies in the Breast Cancer Association Consortium (BCAC). The SNPs were genotyped as part of a collaborative genotyping experiment involving four consortia (Collaborative Oncological Gene-environment Study, COGS) and used a custom Illumina iSelect genotyping array, iCOGS, comprising more than 200,000 SNPs. We identified SNPs at 41 new breast cancer susceptibility loci at genome-wide significance ($P < 5 \times 10^{-8}$). Further analyses suggest that more than 1,000 additional loci are involved in breast cancer susceptibility.

Breast cancer is the most commonly occurring malignancy among women, with an estimated 1 million new cases and over 400,000 deaths annually worldwide[1]. Familial aggregation and twin studies have shown the substantial contribution of inherited susceptibility to breast cancer[2,3]. Many genetic loci are known to contribute to this familial risk, including genes with high-penetrance mutations (notably *BRCA1* and *BRCA2*), moderate-risk alleles in genes such as *ATM*, *CHEK2* and *PALB2*, and common lower penetrance alleles, of which 27 have been identified so far, principally through genome-wide association studies (GWAS)[4–16]. In total, these loci explain approximately 30% of the familial risk of breast cancer[15]. Global analysis of GWAS data suggests that a substantial fraction of the residual aggregation can be explained by other common variants not yet identified, but the relative contributions of common and rare variants are still uncertain.

RESULTS

To identify additional susceptibility loci for breast cancer, we first conducted a meta-analysis of 9 breast cancer GWAS in populations of European ancestry, including 10,052 cases and 12,575 controls (Supplementary Table 1). From this analysis, we selected 35,084 SNPs on the basis of evidence of association with breast cancer, derived from a 1-degree-of-freedom trend test, a test weighted for family history, a 2-degrees-of-freedom test and subset analyses based on cases of breast cancer diagnosed before 40 years of age and before 50 years of age (Online Methods). In particular, we were able to select all SNPs or surrogate SNPs with 1-degree-of-freedom $P_{\text{trend}} < 0.008$. To evaluate these SNPs, we then designed a custom Illumina iSelect genotyping array (iCOGS) in collaboration with three other consortia studying, in addition to breast cancer risk, susceptibility to ovarian cancer, prostate cancer and breast and ovarian cancers in *BRCA1* and *BRCA2* mutation carriers (COGS)[17–20]. The array included, in addition to SNPs selected from GWAS, SNPs selected for fine mapping of known susceptibility loci, functional candidate SNPs and SNPs related to other traits (Online Methods and Supplementary Note). The iCOGS array comprised 211,155 SNPs. These arrays were used to genotype 114,255 DNA samples from 52 studies participating in BCAC (Supplementary Table 2). After quality control exclusions (Online Methods and Supplementary Table 3), data were obtained for 199,961 SNPs in 52,675 cases and 49,436 controls. The analyses presented here are based on data from subjects of European ancestry (45,290 cases and 41,880 controls from 41 studies) and focus on 29,807 SNPs that were selected on the basis of the GWAS analysis that were successfully genotyped and were not located in regions previously known to be associated with breast cancer.

The association between each SNP and breast cancer risk was tested using a 1-degree-of-freedom trend test adjusted for study and seven principal components (Online Methods). There was some evidence for inflation in the test statistics, detected using data from 22,897 uncorrelated SNPs on iCOGS not selected on the basis of breast cancer risk ($\lambda = 1.20$, $\lambda_{1000} = 1.005$; Supplementary Fig. 1a). There was, however, clear evidence of an excess of statistically significant associations among the SNPs selected from the GWAS analysis

(Table 1 and Supplementary Fig. 1b). Although some excess was also observed among the SNPs not selected from the breast cancer GWAS, the excess of statistically significant associations was much more marked among the GWAS SNPs at all levels of statistical significance. In addition, of 21,128 SNPs not selected for breast cancer association that were also present in the combined GWAS data set, 10,864 (51%) had effects in the same direction in the GWAS and iCOGS data, and, for these SNPs, inflation was 1.26 ($\lambda_{1000} = 1.007$) compared with 1.14 ($\lambda_{1000} = 1.0035$) for SNPs with effects in opposite directions in the two stages. A similar direction of effect was seen for these SNPs in the combined GWAS ($\lambda = 0.87$ for SNPs with effects in the same direction versus $\lambda = 0.79$ for SNPs with effects in the opposite direction, with inflation being <1 because SNPs showing evidence of association were excluded). Taken together, these results suggest that much of the inflation in the test statistics for SNPs not selected for breast cancer association is also due to the effect of true associations. Moreover, some of the excess of statistically significant associations seen in the SNPs not selected for breast cancer association was due to SNPs close to breast cancer–associated SNPs. For example, of the 45 SNPs with significant association at $P < 0.00001$, 21 were within 1 Mb of 1 of the newly identified breast cancer loci identified at our set genome-wide significance threshold. Taken together, these results strongly suggest that most of the excess of significant association for the GWAS-selected SNPs reflect true associations.

Of the 27 previously established breast cancer–associated loci, all but 4 showed clear evidence of association with overall breast cancer risk in the iCOGS stage ($P = 2.2 \times 10^{-5}$ – $P = 5.9 \times 10^{-125}$; Supplementary Table 4). Three loci showed weaker evidence for association: rs1045485, encoding an Asp302His variant in *CASP8*, whose association was previously identified in a candidate gene study ($P = 0.054$ in the iCOGS stage; $P = 0.0013$ in combined data from the GWAS and iCOGS stages)[21]; rs2380205 at 10p15, identified in a GWAS but suggested to be a possible false positive association in a previous BCAC analysis[22,23] (iCOGS $P = 0.075$; combined $P = 0.0021$); and rs8170 at 19p13.1, for which the association has been shown to be specific to estrogen receptor (ER)-negative breast cancer[24] ($P = 0.0027$ in iCOGS; combined $P = 0.0012$). One locus, rs2284378 at 20q11, recently shown to be associated with ER-negative breast cancer, was not selected for the iCOGS array[16].

Identification of new susceptibility loci

When the results from the GWAS and the iCOGS array were combined, 263 SNPs in 37 new regions had associations that reached $P < 5 \times 10^{-8}$ (Fig. 1, Table 2 and Supplementary Figs. 2 and 3). In four regions (5q11.2, 8q21.11, 10p12.31 and 18q11.2), this set of SNPs included SNPs within 1 Mb of each other that were uncorrelated, such that a second SNP was associated with disease after adjustment for the most significantly associated SNP (Supplementary Fig. 4 and Supplementary Table 5). There was little or no evidence for heterogeneity in the per-allele odds ratios (ORs) among studies for any SNP (per-SNP I^2 and P values are given in Supplementary Fig. 2 and Supplementary Table 6). Genotype-specific OR estimates were consistent with a log-additive (allele dose) model for most SNPs, with the exception of three SNPs (rs616488, rs204247 and rs720475) for which the heterozygotes had a similar OR as homozygotes for the high-risk allele and two SNPs (rs11242675 and rs6472903) that were more consistent with a recessive model (Supplementary Table 6). Consistent with the pattern seen for previously established loci, there was strong evidence for specificity of the association to tumor subtype. For 13 of the loci, the per-allele OR was higher for ER-positive disease than for ER-negative disease (case-only $P < 0.05$), in most instances with little or no evidence of an association with ER-negative disease (based on data from 7,465 ER-negative cases and 27,074 ER-positive cases; Supplementary Table 7a). The most notable differences were for SNP rs6828523 at 4q34.1 (ER-positive OR = 0.87

(95% confidence interval (CI) = 0.84–0.90); ER-negative OR = 1.01 (95% CI = 0.96–1.07); *P* for difference = 1.2 × 10^{-7}) and for rs7072776 at 10p12.31, where the estimated effects were in opposite directions (ER-positive OR = 1.09 (95% CI = 1.06–1.12); ER-negative OR = 0.94 (95% CI = 0.90–0.98); *P* for difference = 3.1 × 10^{-10}). No such difference was observed for the neighboring SNP rs11814448, which was associated with both ER-positive and ER-negative disease in the same direction. For one locus, SNP rs17817449 on chromosome 16, the association was stronger for ER-negative than for ER-positive disease (*P* for difference = 0.039). All SNPs showed comparable ORs for invasive and *in situ* disease (based on data from 2,335 ductal carcinoma *in situ*, DCIS, and 42,118 invasive cases), with the exceptions of rs12493607 and rs3903072, for which associations seemed to be restricted to invasive disease (Supplementary Table 7b). Two loci (rs2588809 at 14q24.1 (*P* = 0.001) and rs941764 at 14q32.12 (*P* = 0.007)) showed higher per-allele ORs for cases diagnosed at a young age (Supplementary Table 7c). Consistent with the predictions of a polygenic model of susceptibility[25], for 26 of the loci, the estimated OR was higher when restricted to cases with a positive family history for disease (significant at $P < 0.05$ for 5 loci), whereas for only 6 loci was the OR lower when restricted to cases with a positive family history (Supplementary Table 7d).

Four of the newly associated loci (rs16857609 at 2q35, rs10759243 at 9q31, rs11199914 at 10q26 and rs2588809 at 14q24) lie close to regions previously associated with breast cancer risk. In each locus, however, the lead SNP was not correlated with the most strongly associated known association, and the association of the new SNP remained similarly statistically significant after adjustment for the previously associated SNP (Supplementary Table 5). In the case of rs2588809, which lies in *RAD51B* (also known as *RAD51L1*), the association was markedly stronger for ER-positive disease (*P* = 0.011; Supplementary Table 7a), whereas the previously associated SNPs (rs999737 and rs10483813), which lie ~370 kb telomeric, are associated with similar ORs for both ER-positive and ER-negative disease[26].

Two associated loci lie within or close to known breast cancer susceptibility genes. rs11571833 is a polymorphic variant in *BRCA2* that introduces a premature stop codon (p.Lys3326*), previously reported to have no association with breast cancer risk[27]. The results from the current study, however, indicate that this variant is associated with a modestly higher risk of breast cancer. Further work will be required to determine whether this association is due to a higher risk variant or variants in linkage disequilibrium (LD). SNP rs132390 at 22q12 lies within an intron of *EMID1* but is ~500 kb upstream of *CHEK2*, raising the possibility that this association is mediated through the latter. *CHEK2* c. 1100delC, the major deleterious *CHEK2* variant in European populations[28], occurs more frequently in association with the risk allele at rs132390 ($r^2 = 0.06$); however, the association between r132390 and breast cancer risk persisted after adjustment for *CHEK2* c. 1100delC, although attenuated (unadjusted OR in iCOGS = 1.12, *P* = 5.9 × 10^{-6}; adjusted OR = 1.09, *P* = 0.04).

In addition to rs11571833, one further SNP is a coding variant: rs11552449 encodes a missense substitution p.His61Tyr in *DCLRE1B* (also known as *SNM1B*), an evolutionarily conserved gene involved in DNA stability and the repair of interstrand cross-links[29]. The remaining loci are either intronic (20) or intergenic (19). Two loci lie within genes previously proposed as candidate breast cancer susceptibility genes. SNP rs12493607 lies in intron 2 of *TGFBR2*. An analysis of genes in the transforming growth factor (TGF)-β signaling pathway in European populations found weak evidence of an association between rs4522809 and breast cancer risk (*P* = 0.02)[30]. This SNP is weakly correlated with rs12493607 ($r^2 = 0.25$) and also showed some evidence of association in our study, although weaker than that seen for rs12493607 (iCOGS *P* = 0.00096; combined analysis of GWAS and iCOGS *P* = 0.0029). A similar analysis of candidate SNPs in Asian populations identified SNP rs1078985 as a potential breast cancer susceptibility variant[31]. This variant, however, was uncorrelated with rs12493607 in Europeans and showed no evidence of association in our study ($P = 0.33$ in the iCOGS stage). SNP rs7904519 lies in intron 4 of *TCF7L2*. A previous candidate gene study found weak evidence for an association between a correlated SNP, rs12255372, associated with type 2 diabetes ($r^2 = 0.37$ with rs7904519), and familial breast cancer ($P = 0.04$)[32].

The identification of the genes and variants underlying these associations will require more detailed fine mapping and functional analysis. Nevertheless, it is possible to discern some patterns. We identified 53 genes within 50 kb of the lead SNPs in the newly associated regions, totaling 96 genes when including the previously known loci. Analysis using Ingenuity Systems Pathway Analysis (IPA) identified an excess of genes reported to be involved in tumorigenesis (34 genes; $P = 0.0005$), breast cancer (15 genes; $P = 2 \times 10^{-5}$) and tumor incidence in model systems (10 genes; $P = 2 \times 10^{-7}$). The most consistently over-represented functions were cell death ($P = 0.0028$), differentiation ($P = 2 \times 10^{-5}$) and expression ($P = 2 \times 10^{-8}$).

Three loci are located in the vicinity of susceptibility regions for other cancer types. SNP rs11780156 lies ~400 kb downstream of *MYC*. Previous GWAS have identified multiple loci upstream of *MYC* that are associated with different cancer types, including a locus for breast cancer. Functional studies have indicated that these associations might be mediated through transcriptional regulation of *MYC*. The newly associated locus is ~300 kb centromeric to a previously reported susceptibility locus for ovarian cancer, rs10088218, but is uncorrelated with it ($r^2 = 0.02$, based on data from European subjects in BCAC), raising the possibility that these loci might also be regulating *MYC*[33]. SNP rs9790517 at 4q24 lies ~20 kb away from SNP rs7679673, previously reported to be associated with prostate cancer[34], and is correlated with it ($r^2 = 0.53$). SNP rs9790517 lies in intron 11 of *TET2*, which encodes a methylcytosine dioxygenase involved in myelopoiesis. Mutations in *TET2* are frequent in hematological malignancies but have also been reported in 2 of 47 breast tumors in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. In addition, Pharoah *et al.*[18] have found an association between rs1243180 and ovarian cancer. This SNP is ~120 kb telomeric to rs7072776 and is partially correlated with it ($r^2 = 0.51$); both SNPs and the neighboring breast cancer–associated locus rs11814448 lie within the region 400 kb upstream of *DNAJC1*.

To further investigate the likely genes underlying the susceptibility variants, we examined associations between the lead SNPs and the RNA expression of neighboring genes in 473 primary breast tumors and 61 normal breast tissue samples in The Cancer Genome Atlas (TCGA) database. We found strong evidence for an association between rs616402 (a surrogate for rs616488; $r^2 = 0.66$) and expression of *PEX14* in both tumor ($P = 4.7 \times 10^{-12}$) and normal tissue ($P = 0.00018$; Supplementary Table 8), between rs3760983 (a surrogate for rs3760982; $r^2 = 1$) and expression of both *ZNF404* ($P = 1.2 \times 10^{-6}$ in tumors) and *ZNF283* ($P = 0.0089$) and between rs3903072 and expression of *CTSW* ($P = 4.9 \times 10^{-5}$). SNP rs3760982 was also found to be associated with the expression of *ZNF45* ($P = 0.0077$), *ZNF283* ($P = 0.05$) and *ZNF222* ($P = 0.01$) in lymphoblastoid cell lines from HapMap samples using the Genevar database[35] (Supplementary Table 8c). After adjustment for the SNP in the region most strongly associated with expression, SNP rs616488 and *PEX14* ($P = 0.0071$) as well as rs1217396 (a proxy for rs11552449) and *PTPN22* ($P = 0.0055$) and *DCLRE1B* ($P = 0.0067$) reached nominal significance at $P < 0.01$ (Supplementary Table 8a). Although none of these passed Bonferroni correction for multiple testing, the three associations found exceeded the number expected by chance with 46 associations tested. This supports some transcriptional effect from the risk-associated SNPs. PEX14 is involved in peroxisome organization and protein and transmembrane transport; mutations in *PEX14* have been associated with Zellweger syndrome[36]. The functions of ZNF45, ZNF222 and ZNF283 are unknown but may involve transcriptional regulation.

In addition to the genes described above, plausible candidate genes exist in several of the newly associated regions. *MUS81* at 11q13 has a key role in the maintenance of genomic stability and in DNA repair pathways[37,38], and the cofilin gene (*CFL1*) is required for tumor cell motility and invasion, particularly in mammary tumors[39,40]. Several other genes have been associated with tumor aggressiveness; these include *PTH1R* at 3p21, *FOXQ1* at 6p25, *ARHGEF5* at 7q35 and *MKL1* at 22q13. PTH1R is the receptor for PTHLH, encoded by a previously identified breast cancer susceptibility locus[15]. PTHLH is required for normal mammary gland function and has been shown to be involved in the metastasis of breast cancer cells to bone[41,42]. *FOXQ1* encodes a transcription factor with a key role in cell proliferation and migration and in breast cancer metastasis[43]. Alterations in its expression level induce mesenchymal-epithelial transition[44]. Dysfunctional *ARHGEF5* acts as an oncogene specific for human breast tissue, with a crucial role in tumorigenesis and metastasis in breast cancer[45]. *MKL1* is also involved in tumor cell invasion and metastasis, particularly in human breast carcinoma[46]. Two of the newly associated SNPs lie within the *TCF7L2* and *FTO* genes, previously associated with type 2 diabetes and/or obesity through GWAS[47–49]. *TCF7L2* acts as a proto-oncogene and is involved in the Wnt pathway and in tumor formation[50]. *PAX9* at 14q13.3 encodes a transcription factor that regulates cell proliferation, migration and resistance to apoptosis[51,52]. *SSBP4* is involved in DNA recombination and repair and has been suggested to have tumor suppressor activity[53,54]. The expression of *KREMEN1* at 22q12.1 is lower or absent in human tumors compared to normal tissue[55,56]. This gene encodes a negative regulator of the Wnt/β-catenin pathway, which has a key role in cell fate determination, stem cell regulation and cell differentiation and proliferation. It has been suggested that lack of KREMEN1 would activate the Wnt/β-catenin pathway, thereby enhancing susceptibility to tumorigenesis[55,56]. Finally, *NTN4* at 12q22 encodes a secreted growth factor that regulates tumor growth. High levels of NTN4 have been found in ER-positive but not ER-negative breast tumors[57]. NTN4 expression in tumors has also been suggested as a potential prognostic marker for breast cancer[57].

Overall contribution to breast cancer susceptibility

On the assumption that the risks conferred by common susceptibility loci combine multiplicatively (no interaction on a log-additive scale) and on the basis of the per-allele OR estimates from the iCOGS stage, we determined that the 41 newly associated loci explain approximately 5% of the familial risk of breast cancer. However, the overall excess of significant associations for SNPs selected from the breast cancer GWAS for genotyping in the iCOGS stage suggests that a much larger number of loci contribute to susceptibility, although they did not have associations reaching genome-wide levels of significance in the current study. To assess this hypothesis more formally, we identified a set of 10,668 SNPs selected from the GWAS that were uncorrelated ($r^2 < 0.1$ between any pair). Of these, the estimated OR was in the same direction as in the combined GWAS for 5,918 SNPs and in the opposite direction for 4,750 SNPs. Assuming that SNPs with effects in opposite directions are not associated with risk, an estimated 1,168 loci selected from the GWAS are associated with risk. However, this is an underestimate because weakly associated SNPs might have effects in opposite directions in the two stages. As an alternative approach, we fitted the distribution of z scores for the iCOGS stage, aligned to the direction of the effect in the GWAS, as a mixture of two normal distributions representing those SNPs that were or were not associated with disease (Fig. 2 and Online Methods)[58]. On the basis of the posterior probabilities from this analysis, an estimated 92% of loci ($n = 9,815$) were associated with breast cancer risk (95% CI = 85–100%), and these contributed approximately 18% of the familial risk of breast cancer. It should be noted, however, that the large majority of the loci had very small individual effects on risk: for example, the estimated OR was >1.05 for only 10 loci, and 920 loci had an estimated OR of >1.02. When taking into account effects from the previously known loci, these analyses suggest that ~28% of familial risk is explained by common variants selected for iCOGS, of which ~14% can be explained by the 67 established loci (with a further ~20% due to higher penetrance loci).

DISCUSSION

To our knowledge, this is the largest genetic association study in cancer so far. The power of this approach is demonstrated by the fact that we have found evidence, at genome-wide levels of significance, for more than 40 new susceptibility loci, more than doubling the number of susceptibility loci for breast cancer. The effect sizes of the newly identified loci are generally modest (the highest OR was 1.26). However, the very high levels of statistical significance, the lack of heterogeneity among studies, the generally higher effect sizes for familial cases and the fact that most of the excess of significant associations was concentrated among SNPs selected on the basis of an association in the combined breast cancer GWAS all indicate that these are robust associations. Although the majority of the data are from populations of Northern and Western European ancestry, there was little or no evidence of heterogeneity in the OR estimates between studies, indicating that the associations apply broadly to populations of European ancestry. With more than 60 established breast cancer susceptibility loci, it is becoming possible to discern some more general patterns among the loci. Although most of the underlying genes and variants remain to be identified, there is a clear excess of genes either known to be involved in tumorigenesis in model systems or involved in processes relevant to cancer, such as cell death and differentiation. However, for other loci, such as *PEX14*, there is no obvious link to cancer susceptibility. Nine of the new loci lie in chromosomal regions with no known genes, suggesting that these may provide further examples of long-range regulation similar to that seen in the 8q24 region[59]. We have identified three additional examples of loci in the vicinity of susceptibility loci for other cancers (*TET2*, 8q24 and *DNAJC1*). These associations might reflect the tissue-specific regulation of key genes, and understanding the functional mechanisms underlying these associations may be particularly informative.

On the basis of the current set of loci and assuming that all loci combine multiplicatively, the currently known loci now define a genetic profile for which 5% of the female population has a risk that is ~2.3-fold higher than the population average and for which 1% of the population has a risk that is ~3-fold higher. However, the large excess of significant associations among the SNPs selected from the GWAS suggests that many more susceptibility loci exist that have not met our threshold for genome-wide-significant association in this study and that these explain a similar fraction of the heritability as the currently known loci. The observation, made by comparing effect sizes in the iCOGS stage with those in the GWAS, that a very large number of loci, perhaps several thousand, contribute to polygenic susceptibility to breast cancer is consistent with results from GWAS in other complex disorders such as schizophrenia, using a different analytical approach[60]. Incorporating these loci into risk models should substantially improve disease prediction, even if not all loci can be identified individually. Moreover, fine-scale mapping of the identified regions may uncover more of the missing heritability, either through identifying a more strongly associated variant (as found for the *CCND1* locus; see French *et al.*[61]) or by identifying additional signals (exemplified for the *TERT* region in Bojesen *et al.*[62]). Genetic profiling using these common susceptibility loci in combination with rarer high-risk loci and other risk factors may provide a rational basis for targeted breast cancer prevention.

URLs

TCGA, http://cancergenome.nih.gov/; IPA, http://www.ingenuity.com/products/ipa; COSMIC, http://www.sanger.ac.uk/genetics/CGP/cosmic/; BCAC, http://ccge.medschl.cam.ac.uk/consortia/bcac/index.html; CIMBA, http://ccge.medschl.cam.ac.uk/consortia/cimba/index.html; OCAC, http://ccge.medschl.cam.ac.uk/consortia/ocac/index.html; PRACTICAL, http://ccge.medschl.cam.ac.uk/consortia/practical/index.html; COGS, http://www.cogseu.org/; iCOGS, http://ccge.medschl.cam.ac.uk/research/consortia/icogs/; Illumina GenCall, http://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf; SNAP, http://www.broadinstitute.org/mpg/snap/ldplot.php.

ONLINE METHODS

GWAS analysis

Primary genotype data were obtained for nine breast cancer GWAS in populations of European ancestry (Supplementary Table 1). Standard quality control was performed on all scans as follows. We excluded all individuals with low call rate (<95%) and extremely high or low heterozygosity ($P < 1 \times 10^{-5}$), as well as all individuals evaluated to be of non-European ancestry (>15% non-European component, as determined by multidimensional scaling using the HapMap version 2 CEU, JPT/CHB and YRI populations as a reference). We excluded SNPs with MAF < 1%; call rate < 95%; or call rate < 99% and MAF < 5% and all SNPs with genotype frequencies that departed from Hardy-Weinberg equilibrium at $P < 1 \times 10^{-6}$ in controls or $P < 1 \times 10^{-12}$ in cases. For highly significant SNPs, genotype intensity cluster plots were examined manually to judge reliability, either centrally or by contacting the original investigators.

Data were imputed for all scans for ~2.6 million SNPs with the HapMap version 2 CEU panel (Utah residents of Northern and Western European ancestry) as a reference, using the program MaCH v1.0. Imputation was conducted separately for each scan. Estimated per-allele ORs and standard errors were generated from the imputed genotypes using ProbABEL[63]. For two studies (UK2 and HEBCS), estimates were adjusted by the first three principal components, as this was found to materially reduce the inflation of test statistics. Residual inflation was then adjusted for by multiplying the variance by a genomic control adjustment factor, based on the ratio of the median χ^2 test statistic to its expected value[64]. BBCS and UK2 used the same control data (WTCCC2) but different genotyping platforms. Data were imputed separately for these studies. For the combined analysis, the control set was divided randomly between the two studies, in proportion to the size of the case series, to provide disjoint strata. Overall significance tests for each SNP were performed using a fixed-effects meta-analysis; data were only included for a given study if the imputation accuracy r^2 was >0.3.

SNP selection

Details of SNP selection for the iCOGS array are given in the Supplementary Note.

For the purpose of the BCAC analyses, we included SNPs on the basis of the analysis of the nine GWAS described above. We ranked SNPs on the basis of the results from five analyses: an overall 1-degree-of-freedom trend test; a 1-degree-of-freedom trend test giving a weight of 2 to those studies selecting cases for a positive family history (UK2, BBCS, DFBBCS and GC-HBOC); a 2-degrees-of-freedom genotype test; and 1-degree-of-freedom tests based on cases diagnosed before the ages of 40 years or 50 years compared with all controls. We also defined lists based on 1-degree-of-freedom trend tests restricted to data from each of the nine component studies. SNPs were also selected from analyses of cases with ER-negative disease, but these are not reported here.

iCOGS genotyping

Samples for the iCOGS stage were drawn from 52 studies participating in BCAC, including 41 from populations of predominantly European ancestry, 9 of Asian ancestry and 2 of African-American ancestry. The majority of studies were population-based or hospital-based case-control studies, but some studies selected samples by age or oversampled for cases with a family history of breast cancer (Supplementary Table 2). Studies were required to provide ~2% of samples in duplicate.

Genotyping was conducted using a custom Illumina Infinium array (iCOGS) in seven centers, of which four were used for BCAC. Genotypes were called using Illumina's proprietary GenCall algorithm. Initial calling used a cluster file generated from 270 samples from HapMap 2. To generate the final calls, we first selected a subset of 3,018 individuals, including samples from each of the genotyping centers, each of the participating consortia and each major ancestry group. Only plates with a consistently high call rate in the initial calling were used. We also included 380 samples of European, Asian or African ancestry genotyped as part of the HapMap Project and 1000 Genomes Project and 160 samples that were known positive controls for rare variants on the array. This subset was used to generate a cluster file that was then applied to call the genotypes for the remaining samples. We also investigated two other calling algorithms: Illumnus[65] and GenoSNP[66]. All three algorithms were >99% concordant in their calling for 91% of the SNPs on the array. However, manual inspection of a sample of the SNPs with discrepancies indicated that the calls from GenCall were almost invariably superior (generally, because Illumnus or GenoSNP attempted to call SNPs that clustered poorly). Therefore, only the genotypes called by GenCall have been used in the analyses reported here.

Quality control

We excluded individuals for any of the following reasons: genotypically not female XX (XY, XXY or XO); overall call rate < 95%; low or high heterozygosity ($P < 1 \times 10^{-6}$, determined separately for individuals of European, East Asian and African-American ancestry); genotypes discordant with those determined in previous BCAC genotyping such that the individual appeared to be different; genotypes for the duplicate sample that seemed to be from a different individual; and cryptic duplicates where the phenotypic data indicated that the individuals were different. We searched for cryptic duplicates, both within each study and between studies from the same country. For known and cryptic concordant duplicates, the sample with the lower call rate was excluded. We attempted to identify first-degree relative pairs using identity-by-state estimates based on ~37,000 uncorrelated SNPs. For apparent first-degree relative pairs, we removed the control from a case-control pair; otherwise, we excluded the individual with the lower call rate. For the main analyses presented here, we also excluded 1,880 individuals who were included in any of the GWAS to allow the GWAS and iCOGS stages to be combined.

Ancestry outliers were identified by multidimensional scaling, combining the iCOGS data with genotypes from the HapMap 2 populations, on the basis of a subset of 37,000 uncorrelated markers that passed quality control (including ~1,000 that were selected as ancestry-informative markers). Most studies were predominantly of a single ancestry (European or East Asian), and individuals with >15% minority ancestry, as determined on the basis of the first two principal components, were excluded. Two studies from Singapore (SGBCC) and Malaysia (MYBRCA) contained a substantial fraction of individuals of mixed European and Asian ancestry (likely of South Asian ancestry). For these studies, no exclusions for ancestry outliers were made, but principal-components analysis adequately corrected for inflation in these studies. Similarly, for the two African-American studies (NBHS and SCCS), no exclusions for ancestry outliers were made.

Principal-components analyses were carried out separately for the European, Asian and African-American subgroups, on the basis of a subset of 37,000 uncorrelated SNPs. For the analyses of European subjects, we included the first six principal components as covariates, together with a seventh component derived specifically for one study (LMBC) for which there was substantial inflation not accounted for by the components derived from the analysis of all studies (this component was set to zero for all other studies). The addition of further principal components did not reduce inflation further. We included two principal components each for the studies in Asian and African-American populations.

We excluded SNPs with call rates of <95%. We also excluded SNPs that deviated from Hardy-Weinberg equilibrium in controls at $P < 1 \times 10^{-7}$, on the basis of a stratified 1-degrre-of-freedom test in which the deviations were summed across strata[67]. We also excluded SNPs for which the genotypes were discrepant in more than 2% of duplicate samples across all COGS consortia. The final analyses were based on data from 199,961 SNPs.

Genotype intensity cluster plots were examined manually for SNPs in each new region in which a genome-wide significant association was obtained, and SNPs were eliminated if the clustering was judged to be poor.

Statistical analysis

For each SNP, we estimated a per-allele log(OR) and standard error by logistic regression, including study and principal components as covariates. Genotype-specific ORs were also computed. Overall significance levels were obtained by combining the estimates from the combined GWAS and iCOGS using a fixed-effects meta-analysis to derive a 1-degree-of-freedom test. Inflation of the test statistics (λ) was estimated by dividing the 45th percentile of the test statistic by 0.357 (the 45th percentile for a χ^2 distribution on 1 degree of freedom). For this purpose, we used a subset of 22,897 SNPs that were uncorrelated ($r^2 <$ 0.1), which were not selected by BCAC and were not within 1 of the 4 common fine-mapping regions. This subset was used to minimize the selection of SNPs associated with disease, on the assumption that such SNPs are likely to be representative of common SNPs in terms of population structure. The inflation statistic was converted to an equivalent inflation statistic for a study with 1,000 cases and 1,000 controls ($\lambda_{1,000}$) by adjusting by effective study size, namely

$$\lambda_{1,000}=1+\frac{500(\lambda-1)}{\sum_k\left(\frac{1}{n_k}+\frac{1}{m_k}\right)^{-1}}$$

where n_k and m_k are the number of cases and controls, respectively, for study k. Heterogeneity in the per-allele OR by ER status, age at diagnosis, family history and tumor invasiveness (DCIS versus invasive) were evaluated using a case-only analysis.

Expression analysis

Gene expression, copy number and genotype data were retrieved from the TCGA breast cancer study. Gene expression profiles were measured by TCGA using a custom Agilent 244K expression array. We downloaded the raw expression data and performed preprocessing using the limma R package. Copy number and germline genotype were both measured using the Affymetrix Genome-Wide Human SNP 6.0 array. We used the segmented copy number and called genotype data as provided by TCGA. Intersecting the different genomic data types, we collected 458 primary tumor samples with germline genotypes from blood and both gene expression and somatic copy number data from the tumor. In addition, for 61 samples, we had germline genotype and gene expression data from normal breast tissue from individuals in the TCGA breast cancer study. Expression quantitative trait locus (eQTL) analysis was performed on both sets separately. For *cis*-eQTL analysis, we considered all genes 50 kb upstream or downstream of the lead SNP. Fourteen of the risk-associated SNPs are represented directly on the Affymetrix SNP array. For an additional 23, we were able to select proxies on the basis of maximum LD with minimum r^2 of 0.5. In case of equal LD, we used proximity on the genome to break the tie. LD estimates were extracted from the HapMap data for the CEU population. eQTL analysis was performed by regressing the gene expression of selected candidate genes on the genotype followed by a significance test of the t statistic for the genotype covariate. For both the normal and tumor analyses, the linear regression was adjusted for potential batch effects by including indicator variables for the plate identifier component of the TCGA sample barcode. In addition, the first principal component of the complete gene expression matrix was added as a covariate to adjust for other global, typically non-genetic contributions to the gene expression signal. To prevent spurious associations due to confounding by nearby eQTLs, we corrected the model for the most strongly associated eQTL SNP in the region. For the tumor analysis only, we also added the copy number of the candidate gene as a covariate because apparent associations between germline genotype and tumor expression may be confounded or obscured by somatic copy number alterations.

To assess the potential effects of the new SNPs on nearby gene expression in lymphocytes, we identified all genes that lie within a 500-kb window surrounding each of the SNPs and used Genevar (Gene Expression Variation), a public database with gene expression data quantified in lymphocytes from individuals in the HapMap 2 populations[35,68].

Estimation of the number of associated loci

To estimate the total number of newly associated loci selected for the iCOGS array, we first used the set of 29,807 SNPs selected from the GWAS and not selected for fine mapping, to exclude previously known loci. We then defined a set of 10,668 SNPs that were uncorrelated ($r^2 < 0.1$ between any pair) and determined the number of loci for which the estimated effect size in the iCOGS stage was in the same direction as in the combined GWAS and the number of loci for which the effect was in the opposite direction. Similar results were obtained using cutoffs of $r^2 < 0.05$ and $r^2 < 0.2$. On the assumption that none of the loci with effects in opposite directions in the two stages were associated with disease, the number of loci associated with disease can be estimated as the difference between the number of loci with effects in the same direction and the number with effects in opposite directions. This, however, is an underestimate because loci with weak effects may have estimated effects in opposite directions in the two stages. To allow for this possibility, we fitted the distribution of z scores as a mixture of a standard normal distribution (representing SNPs with no effect) and a normal distribution with unknown mean and variance, using an expectation-maximization algorithm[58]. The total contribution to heritability was then computed from the posterior estimates. To allow for the potential effect of residual population stratification, we conducted an additional analysis in which the null distribution was assumed to have variance of 1.2, based on the estimated inflation from the non-BCAC SNPs, but the estimates were essentially identical.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. BCAC is funded by Cancer Research UK (C1287/A10118 and C1287/A12014) and by the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS). Meetings of BCAC have been funded by the European Union European Cooperation in Science and Technology (COST) programme (BM0606). Genotyping of the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer program and the Ministry of Economic Development, Innovation and Export Trade of Quebec (grant PSR-SIIRI-701). Combining the GWAS data was supported in part by the US National Institutes of Health (NIH) Cancer Post-Cancer GWAS initiative grant 1 U19 CA 148065-01 (DRIVE, part of the GAME-ON initiative). A full description of funding and acknowledgments is provided in the Supplementary Note.

References

1. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol. 2006; 24:2137–2150. [PubMed: 16682732]
2. Lichtenstein P, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000; 343:78–85. [PubMed: 10891514]
3. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. Nat Genet. 2000; 26:411–414. [PubMed: 11101836]
4. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447:1087–1093. [PubMed: 17529967]
5. Hunter DJ, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007; 39:870–874. [PubMed: 17529973]
6. Stacey SN, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor–positive breast cancer. Nat Genet. 2007; 39:865–869. [PubMed: 17529974]
7. Stacey SN, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor–positive breast cancer. Nat Genet. 2008; 40:703–706. [PubMed: 18438407]
8. Ahmed S, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet. 2009; 41:585–590. [PubMed: 19330027]
9. Zheng W, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009; 41:324–328. [PubMed: 19219042]
10. Thomas G, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). Nat Genet. 2009; 41:579–584. [PubMed: 19330030]
11. Turnbull C, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010; 42:504–507. [PubMed: 20453838]
12. Antoniou AC, et al. A locus on 19p13 modifies risk of breast cancer in *BRCA1* mutation carriers and is associated with hormone receptor–negative breast cancer in the general population. Nat Genet. 2010; 42:885–892. [PubMed: 20852631]
13. Fletcher O, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J Natl Cancer Inst. 2011; 103:425–435. [PubMed: 21263130]
14. Haiman CA, et al. A common variant at the *TERT-CLPTM1L* locus is associated with estrogen receptor–negative breast cancer. Nat Genet. 2011; 43:1210–1214. [PubMed: 22037553]
15. Ghoussaini M, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet. 2012; 44:312–318. [PubMed: 22267197]

16. Siddiq A, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Hum Mol Genet. 2012; 21:5373–5384. [PubMed: 22976474]

17. Eeles RA, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. Mar 27.2013 published online. 10.1038/ng.2560

18. Pharoah PDP, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet. Mar 27.2013 published online. 10.1038/ng.2564

19. Couch FJ, et al. Genome-wide association study in *BRCA1* mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet. 2013; 9:e1003212. [PubMed: 23544013]

20. Gaudet MM, et al. Identification of a *BRCA2*-specific modifier locus at 6p24 related to breast cancer risk. PLoS Genet. 2013; 9:e1003173. [PubMed: 23544012]

21. Cox A, et al. A common coding variant in *CASP8* is associated with breast cancer risk. Nat Genet. 2007; 39:352–358. [PubMed: 17293864]

22. Turnbull C, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010; 42:504–507. [PubMed: 20453838]

23. Lambrechts D, et al. 11q13 is a susceptibility locus for hormone receptor positive breast cancer. Hum Mutat. 2012; 33:1123–1132. [PubMed: 22461340]

24. Stevens KN, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. Cancer Res. 2012; 72:1795–1803. [PubMed: 22331459]

25. Antoniou AC, Easton DF. Polygenic inheritance of breast cancer: implications for design of association studies. Genet Epidemiol. 2003; 25:190–202. [PubMed: 14557987]

26. Figueroa JD, et al. Associations of common variants at 1p11.2 and 14q24.1 (*RAD51L1*) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. Hum Mol Genet. 2011; 20:4693–4706. [PubMed: 21852249]

27. Mazoyer S, et al. A polymorphic stop codon in *BRCA2*. Nat Genet. 1996; 14:253–254. [PubMed: 8896551]

28. Schutte M, et al. Variants in *CHEK2* other than 1100delC do not make a major contribution to breast cancer susceptibility. Am J Hum Genet. 2003; 72:1023–1028. [PubMed: 12610780]

29. Hemphill AW, et al. Mammalian SNM1 is required for genome stability. Mol Genet Metab. 2008; 94:38–45. [PubMed: 18180189]

30. Scollen S, et al. TGF-β signaling pathway and breast cancer susceptibility. Cancer Epidemiol Biomarkers Prev. 2011; 20:1112–1119. [PubMed: 21527583]

31. Ma X, et al. Pathway analyses identify *TGFBR2* as potential breast cancer susceptibility gene: results from a consortium study among Asians. Cancer Epidemiol Biomarkers Prev. 2012; 21:1176–1184. [PubMed: 22539603]

32. Burwinkel B, et al. Transcription factor 7–like 2 (*TCF7L2*) variant is associated with familial breast cancer risk: a case-control study. BMC Cancer. 2006; 6:268. [PubMed: 17109766]

33. Goode EL, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat Genet. 2010; 42:874–879. [PubMed: 20852632]

34. Eeles RA, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat Genet. 2009; 41:1116–1121. [PubMed: 19767753]

35. Yang TP, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics. 2010; 26:2474–2476. [PubMed: 20702402]

36. Shimozawa N, et al. Identification of a new complementation group of the peroxisome biogenesis disorders and *PEX14* as the mutated gene. Hum Mutat. 2004; 23:552–558. [PubMed: 15146459]

37. Murfuni I, et al. The WRN and MUS81 proteins limit cell death and genome instability following oncogene activation. Oncogene. 2013; 32:610–620. [PubMed: 22410776]

38. Pamidi A, et al. Functional interplay of p53 and Mus81 in DNA damage responses and cancer. Cancer Res. 2007; 67:8527–8535. [PubMed: 17875692]

39. Leong S, McKay MJ, Christopherson RI, Baxter RC. Biomarkers of breast cancer apoptosis induced by chemotherapy and TRAIL. J Proteome Res. 2012; 11:1240–1250. [PubMed: 22133146]

40. Wang W, et al. The activity status of cofilin is directly related to invasion, intravasation, and metastasis of mammary tumors. J Cell Biol. 2006; 173:395–404. [PubMed: 16651380]

41. Dunbar ME, Wysolmerski JJ, Broadus AE. Parathyroid hormone–related protein: from hypercalcemia of malignancy to developmental regulatory molecule. Am J Med Sci. 1996; 312:287–294. [PubMed: 8969618]

42. Dunbar ME, et al. Stromal cells are critical targets in the regulation of mammary ductal morphogenesis by parathyroid hormone–related protein. Dev Biol. 1998; 203:75–89. [PubMed: 9806774]

43. Qiao Y, et al. FOXQ1 regulates epithelial-mesenchymal transition in human cancers. Cancer Res. 2011; 71:3076–3086. [PubMed: 21346143]

44. Kaneda H, et al. FOXQ1 is overexpressed in colorectal cancer and enhances tumorigenicity and tumor growth. Cancer Res. 2010; 70:2053–2063. [PubMed: 20145154]

45. Debily MA, et al. Expression and molecular characterization of alternative transcripts of the *ARHGEF5*/*TIM* oncogene specific for human breast cancer. Hum Mol Genet. 2004; 13:323–334. [PubMed: 14662653]

46. Muehlich S, et al. The transcriptional coactivators megakaryoblastic leukemia 1/2 mediate the effects of loss of the tumor suppressor deleted in liver cancer 1. Oncogene. 2012; 31:3913–3923. [PubMed: 22139079]

47. Frayling TM, et al. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316:889–894. [PubMed: 17434869]

48. Grant SF, et al. Variant of transcription factor 7–like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. Nat Genet. 2006; 38:320–323. [PubMed: 16415884]

49. Sladek R, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007; 445:881–885. [PubMed: 17293876]

50. Jingushi K, et al. DIF-1 inhibits the Wnt/β-catenin signaling pathway by inhibiting TCF7L2 expression in colon cancer cell lines. Biochem Pharmacol. 2012; 83:47–56. [PubMed: 22005519]

51. Dantuma NP, Heinen C, Hoogstraten D. The ubiquitin receptor Rad23: at the crossroads of nucleotide excision repair and proteasomal degradation. DNA Repair (Amst). 2009; 8:449–460. [PubMed: 19223247]

52. Lee JC, et al. Pax9 mediated cell survival in oral squamous carcinoma cell enhanced by c-myb. Cell Biochem Funct. 2008; 26:892–899. [PubMed: 18979497]

53. Castro P, Liang H, Liang JC, Nagarajan L. A novel, evolutionarily conserved gene family with putative sequence-specific single-stranded DNA-binding activity. Genomics. 2002; 80:78–85. [PubMed: 12079286]

54. Sanchez-Cespedes M, et al. Chromosomal alterations in lung adenocarcinoma from smokers and nonsmokers. Cancer Res. 2001; 61:1309–1313. [PubMed: 11245426]

55. Nakamura T, et al. Molecular cloning and characterization of Kremen, a novel kringle-containing transmembrane protein. Biochim Biophys Acta. 2001; 1518:63–72. [PubMed: 11267660]

56. Nakamura T, Nakamura T, Matsumoto K. The functions and possible significance of Kremen as the gatekeeper of Wnt signalling in development and pathology. J Cell Mol Med. 2008; 12:391–408. [PubMed: 18088386]

57. Esseghir S, et al. Identification of NTN4, TRA1, and STC2 as prognostic markers in breast cancer in a screen for signal sequence encoding proteins. Clin Cancer Res. 2007; 13:3164–3173. [PubMed: 17545519]

58. Morris AP, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012; 44:981–990. [PubMed: 22885922]

59. Ahmadiyeh N, et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with *MYC*. Proc Natl Acad Sci USA. 2010; 107:9742–9746. [PubMed: 20453196]

60. Purcell SM, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–752. [PubMed: 19571811]

61. French JD. Functional variants at the 11q13 risk locus regulate cyclin D1 expression through long-range enhancers. Am J Hum Genet. Mar 27.2013 published online. 10.1016/j.ajhg.2013.01.002

62. Bojesen SE, et al. Multiple independent variants at the *TERT* locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet. Mar 27.2013 published online. 10.1038/ng.2566

63. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics. 2010; 11:134. [PubMed: 20233392]

64. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]

65. Teo YY, et al. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics. 2007; 23:2741–2746. [PubMed: 17846035]

66. Giannoulatou E, Yau C, Colella S, Ragoussis J, Holmes CC. GenoSNP: a variational Bayes within-sample SNP genotyping algorithm that does not require a reference population. Bioinformatics. 2008; 24:2209–2214. [PubMed: 18653518]

67. Haldane JBS. An exact test for randomness of mating. J Genet. 1954; 52:631–635.

68. Stranger BE, et al. Patterns of *cis* regulatory variation in diverse human populations. PLoS Genet. 2012; 8:e1002639. [PubMed: 22532805]

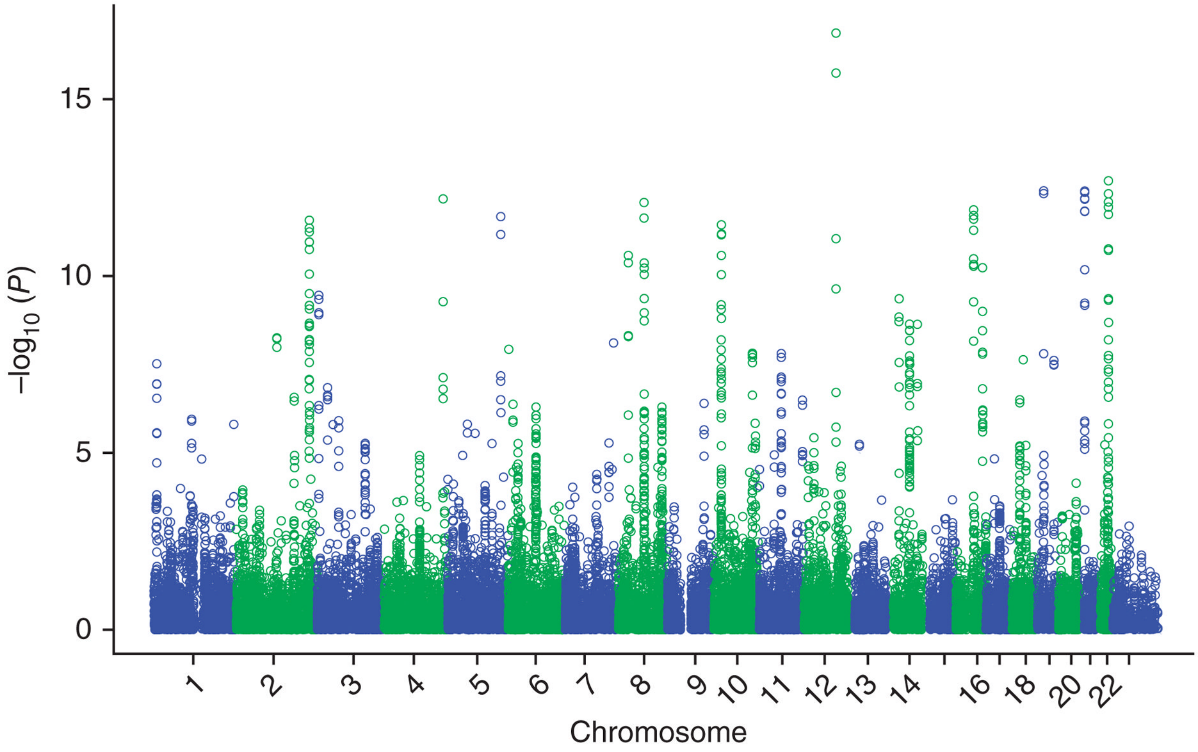

Figure 1.

One-degree-of-freedom trend-test statistics for 29,807 iCOGS SNPs selected from the combined GWAS, excluding those occurring in known susceptibility regions. The red horizontal line represents $P = 5 \times 10^{-8}$. The blue horizontal line represents $P = 1 \times 10^{-5}$.

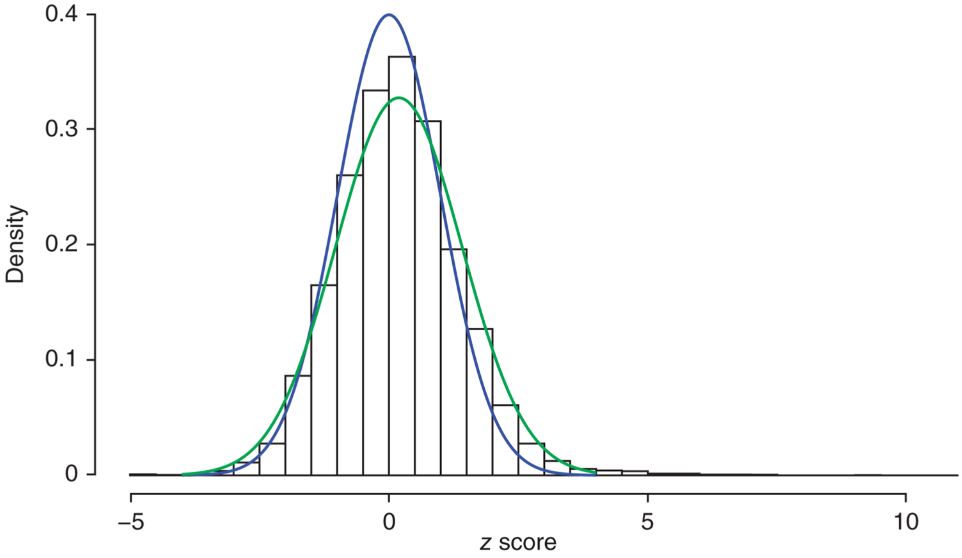

Figure 2.
Distribution of normalized effect sizes (*z* scores) in the iCOGS stage, with the direction of effect determined by the direction in the combined GWAS. The blue curve represents the standard normal distribution. The green curve represents the best-fit normal distribution (mean = 0.19, s.d. = 1.22).

Table 1

Summary of SNPs by level of statistical significance in the iCOGS stage

	Combined GWAS (n = 29,807)		Non-BCAC[a] (n = 126,360)		
Significance	**SNPs**	**Observed/expected**	**SNPs**	**Observed/expected**	**Relative excess**
$<1 \times 10^{-7}$	142	47,639.8	7	554.0	86.0
1×10^{-7}–1×10^{-6}	62	2080.0	13	102.9	20.2
1×10^{-6}–1×10^{-5}	108	362.3	25	19.8	18.3
1×10^{-5}–1×10^{-4}	157	52.7	136	10.8	4.9
1×10^{-4}–1×10^{-3}	360	12.1	348	2.8	4.3

[a] All SNPs excluding those proposed by BCAC and those in four common regions selected for fine mapping (Online Methods).

Table 2

Results for 41 SNPs for which association $P < 5 \times 10^{-8}$ in combined GWAS and iCOGS analysis

Lead SNP	Chr.[a]	Position[b]	Alleles[c]	MAF[d]	GWAS OR (95% CI)[e]	GWAS P^e	*iCOGS OR (95% CI)[f]*	iCOGS P^e	*Combined GWAS and iCOGS P* [e]	Genes
rs616488	1	10488802	A/G	0.33	0.94 (0.90–0.98)	0.0017	0.94 (0.92–0.96)	3.0×10^{-8}	2.0×10^{-10}	*PEX14*
rs11552449	1	114249912	C/T	0.17	1.08 (1.02–1.14)	0.0042	1.07 (1.04–1.09)	1.1×10^{-6}	1.8×10^{-8}	*PTPN22-BCL2L15-AP4B1-DCLRE1B-HIPK1*
rs4849887	2	120961592	C/T	0.098	0.90 (0.84–0.96)	0.0017	0.91 (0.88–0.94)	5.6×10^{-9}	3.7×10^{-11}	None
rs2016394	2	172681217	G/A	0.48	0.95 (0.92–0.99)	0.014	0.95 (0.93–0.97)	2.7×10^{-7}	1.2×10^{-8}	*METAP1D-DLX1-DLX2*
rs1550623	2	173921140	A/G	0.16	0.91 (0.86–0.96)	0.00027	0.94 (0.92–0.97)	1.2×10^{-5}	3.0×10^{-8}	*CDCA7*
rs16857609	2	218004753	C/T	0.26	1.09 (1.05–1.14)	4.5×10^{-5}	1.08 (1.06–1.10)	4.4×10^{-12}	1.1×10^{-15}	*DIRC3*
rs6762644	3	4717276	A/G	0.40	1.06 (1.02–1.11)	0.0016	1.07 (1.04–1.09)	3.5×10^{-10}	2.2×10^{-12}	*ITPR1-EGOT*
rs12493607	3	30657943	G/C	0.35	1.04 (1.00–1.09)	0.049	1.06 (1.03–1.08)	1.4×10^{-7}	2.3×10^{-8}	*TGFBR2*
rs9790517	4	106304227	C/T	0.23	1.09 (1.04–1.14)	0.00027	1.05 (1.03–1.08)	1.6×10^{-5}	4.2×10^{-8}	*TET2*
rs6828523	4	176083001	C/A	0.13	0.89 (0.83–0.94)	0.00011	0.90 (0.87–0.92)	6.6×10^{-13}	3.5×10^{-16}	*ADAM29*
rs10472076	5	58219818	T/C	0.38	1.06 (1.02–1.11)	0.005	1.05 (1.03–1.07)	1.6×10^{-6}	2.9×10^{-8}	*RAB3C*
rs1353747	5	58373238	T/G	0.095	0.90 (0.84–0.96)	0.0020	0.92 (0.89–0.95)	2.7×10^{-6}	2.5×10^{-8}	*PDE4D*
rs1432679	5	158176661	T/C	0.43	1.06 (1.02–1.10)	0.0023	1.07 (1.05–1.09)	2.1×10^{-12}	2.0×10^{-14}	*EBF1*
rs11242675	6	1263878	T/C	0.39	0.97 (0.93–1.01)	0.12	0.94 (0.92–0.96)	1.2×10^{-8}	7.1×10^{-9}	*FOXQ1*
rs204247	6	13830502	A/G	0.43	1.06 (1.02–1.10)	0.0057	1.05 (1.03–1.07)	4.2×10^{-7}	8.3×10^{-9}	*RANBP9*
rs720475	7	143705862	G/A	0.25	0.93 (0.89–0.98)	0.0024	0.94 (0.92–0.96)	7.8×10^{-9}	7.0×10^{-11}	*ARHGEF5-NOBOX*
rs9693444	8	29565535	C/A	0.32	1.07 (1.03–1.12)	0.00086	1.07 (1.05–1.09)	2.6×10^{-11}	9.2×10^{-14}	None
rs6472903	8	76392856	T/G	0.18	0.88 (0.84–0.93)	2.0×10^{-6}	0.91 (0.89–0.93)	8.4×10^{-13}	1.7×10^{-17}	None
rs2943559	8	76580492	A/G	0.07	1.17 (1.09–1.26)	1.2×10^{-5}	1.13 (1.09–1.17)	6.0×10^{-11}	5.7×10^{-15}	*HNF4G*
rs11780156	8	129263823	C/T	0.16	1.13 (1.07–1.19)	2.2×10^{-6}	1.07 (1.04–1.10)	5.0×10^{-7}	3.4×10^{-11}	*MIR1208*
rs10759243	9	109345936	C/A	0.39	1.07 (1.02–1.12)	0.0084	1.06 (1.03–1.08)	4.0×10^{-7}	1.2×10^{-8}	None
rs7072776	10	22072948	G/A	0.29	1.11 (1.07–1.16)	1.3×10^{-6}	1.07 (1.05–1.09)	1.6×10^{-9}	4.3×10^{-14}	*MLLT10-DNAJC1*
rs11814448	10	22355849	A/C	0.020	1.35 (1.17–1.56)	3.7×10^{-5}	1.26 (1.18–1.35)	3.6×10^{-12}	9.3×10^{-16}	*DNAJC1*
rs7904519	10	114763917	A/G	0.46	1.06 (1.02–1.10)	0.0059	1.06 (1.04–1.08)	1.5×10^{-8}	3.1×10^{-8}	*TCF7L2*
rs11199914	10	123083891	C/T	0.32	0.94 (0.89–0.98)	0.0030	0.95 (0.93–0.97)	1.5×10^{-6}	1.9×10^{-8}	None
rs3903072	11	65339642	G/T	0.47	0.92 (0.89–0.96)	5.1×10^{-5}	0.95 (0.93–0.96)	2.0×10^{-8}	8.6×10^{-12}	*DKFZp761E198-OVOL1-SNX32-CFL1-MUS81*

Lead SNP	Chr.^{a}	Position^{b}	Alleles^{c}	MAF^{d}	GWAS OR (95% CI)^{e}	GWAS P^{e}	*iCOGS OR (95% CI)^{f}*	iCOGS P^{e}	*Combined GWAS and iCOGS P ^{e}*	Genes
rs11820646	11	128966381	C/T	0.41	0.93 (0.90–0.97)	0.00068	0.95 (0.93–0.97)	3.2×10^{-7}	1.1×10^{-9}	None
rs12422552	12	14305198	G/C	0.26	1.11 (1.05–1.16)	4.2×10^{-5}	1.05 (1.03–1.07)	2.9×10^{-5}	3.7×10^{-8}	None
rs17356907	12	94551890	A/G	0.30	0.89 (0.85–0.93)	1.7×10^{-6}	0.91 (0.89–0.93)	1.4×10^{-17}	1.8×10^{-22}	*NTN4*
rs11571833	13	31870626	A/T	0.008	1.39 (1.13–1.71)	0.0016	1.26 (1.14–1.39)	5.7×10^{-6}	4.9×10^{-8}	*BRCA2-N4BP2L1-N4BP2L2*
rs2236007	14	36202520	G/A	0.21	0.88 (0.83–0.93)	2.0×10^{-5}	0.93 (0.91–0.95)	4.4×10^{-10}	1.7×10^{-13}	*PAX9-SLC25A21*
rs2588809	14	67730181	C/T	0.16	1.07 (1.01–1.13)	0.017	1.08(1.05–1.11)	2.3×10^{-9}	1.4×10^{-10}	*RAD51L1*
rs941764	14	90910822	A/G	0.34	1.05 (1.00–1.09)	0.043	1.06 (1.04–1.09)	2.3×10^{-9}	3.7×10^{-10}	*CCDC88C*
rs17817449	16	52370868	T/G	0.40	0.95 (0.91–0.99)	0.010	0.93 (0.91–0.95)	1.3×10^{-12}	6.4×10^{-14}	*MIR1972-2-FTO*
rs13329835	16	79208306	A/G	0.22	1.14 (1.09–1.19)	9.2×10^{-8}	1.08 (1.05–1.10)	5.8×10^{-11}	2.1×10^{-16}	*CDYL2*
rs527616	18	22591422	G/C	0.38	0.91 (0.87–0.95)	3.0×10^{-5}	0.95 (0.93–0.97)	3.1×10^{-7}	1.6×10^{-10}	None
rs1436904	18	22824665	T/G	0.4	0.93 (0.9–0.97)	0.0008	0.96 (0.94–0.98)	6.9×10^{-6}	3.2×10^{-8}	*CHST9*
rs4808801	19	18432141	A/G	0.35	0.94 (0.90–0.98)	0.0027	0.93 (0.91–0.95)	3.9×10^{-13}	4.6×10^{-15}	*SSBP4-ISYNA1-ELL*
rs3760982	19	48978353	G/A	0.46	1.06 (1.02–1.10)	0.0022	1.06 (1.04–1.08)	2.5×10^{-8}	2.1×10^{-10}	*C19orf61-KCNN4-LYPD5-ZNF283*
rs132390	22	27951477	T/C	0.036	1.36 (1.19–1.54)	3.0×10^{-6}	1.12 (1.07–1.18)	5.9×10^{-6}	3.1×10^{-9}	*EMID1-RHBDD3-EWSR1*
rs6001930	22	39206180	T/C	0.11	1.17 (1.11–1.25)	2.9×10^{-7}	1.12 (1.09–1.16)	2.0×10^{-13}	8.8×10^{-19}	*MKL1*

Results for the SNPs showing the strongest association in each region are given.

[a] Chromosome.

[b] Build 36 position.

[c] Major/minor allele, based on the forward strand and minor allele frequency in Europeans.

[d] Mean minor allele frequency over all European controls in iCOGS.

[e] One-degree-of-freedom P_{trend}.

[f] Per-allele OR for the minor allele relative to the major allele.