*CYP3A7*1C* allele: linking premenopausal oestrone and progesterone levels with risk of hormone receptor-positive breast cancers.

Nichola Johnson[1]*, Sarah Maguire[2]*, Anna Morra[3]*, Pooja Middha Kapoor[4, 5]*, Katarzyna Tomczyk[1], Michael E. Jones[6], Minouk J. Schoemaker[6], Clare Gilham[7], Manjeet K. Bolla[8], Qin Wang[8], Joe Dennis[8], Thomas U. Ahearn[9], Irene L. Andrulis[10, 11], Hoda Anton-Culver[12], Natalia N. Antonenkova[13], Volker Arndt[14], Kristan J. Aronson[15], Annelie Augustinsson[16], Caroline Baynes[17], Laura E. Beane Freeman[9], Matthias W. Beckmann[18], Javier Benitez[19, 20], Marina Bermisheva[21], Carl Blomqvist[22, 23], Bram Boeckx[24, 25], Natalia V. Bogdanova[13, 26, 27], Stig E. Bojesen[28-30], Hiltrud Brauch[31-33], Hermann Brenner[14, 33, 34], Barbara Burwinkel[35, 36], Daniele Campa[4, 37], Federico Canzian[38], Jose E. Castelao[39], Stephen J. Chanock[9], Georgia Chenevix-Trench[40], Christine L. Clarke[41], NBCS Collaborators[42-52], Don M. Conroy[17], Fergus J. Couch[53], Angela Cox[54], Simon S. Cross[55], Kamila Czene[56], Thilo Dörk[27], A. Heather Eliassen[57, 58], Christoph Engel[59, 60], D. Gareth Evans[61, 62], Peter A. Fasching[18, 63], Jonine Figueroa[9, 64, 65], Giuseppe Floris[66], Henrik Flyger[67], Manuela Gago-Dominguez[68, 69], Susan M. Gapstur[70], Montserrat García-Closas[9], Mia M. Gaudet[70], Graham G. Giles[71-73], Mark S. Goldberg[74, 75], Anna González-Neira[20], AOCS Group[76, 77], Pascal Guénel[78], Eric Hahnen[79, 80], Christopher A. Haiman[81], Niclas Håkansson[82], Per Hall[56, 83], Ute Hamann[84], Patricia A. Harrington[17], Steven N. Hart[85], Maartje J. Hooning[86], John L. Hopper[72], Anthony Howell[87], David J. Hunter[58, 88], ABCTB Investigators[89], kConFab Investigators[76, 90], Agnes Jager[86], Anna Jakubowska[91, 92], Esther M. John[93, 94], Rudolf Kaaks[4], Renske Keeman[3], Elza Khusnutdinova[21, 95], Cari M. Kitahara[96], Veli-Matti Kosma[97-99], Stella Koutros[9], Peter Kraft[58, 100], Vessela N. Kristensen[43, 101], Allison W. Kurian[93, 94], Diether Lambrechts[24, 25], Loic Le Marchand[102], Martha Linet[96], Jan Lubiński[91], Arto Mannermaa[97-99], Siranoush Manoukian[103], Sara Margolin[83, 104], John W.M. Martens[86], Dimitrios Mavroudis[105], Rebecca Mayes[17], Alfons Meindl[106], Roger L. Milne[71-73], Susan L. Neuhausen[107], Heli Nevanlinna[108], William G. Newman[61, 62], Sune F. Nielsen[28, 29], Børge G. Nordestgaard[28-30], Nadia Obi[109], Andrew F. Olshan[110], Janet E. Olson[85], Håkan Olsson[16], Ester Orban[111], Tjoung-Won Park-Simon[27], Paolo Peterlongo[112], Dijana Plaseska-Karanfilska[113], Katri Pylkäs[114, 115], Gad Rennert[116], Hedy S. Rennert[116], Kathryn J. Ruddy[117], Emmanouil Saloustros[118], Dale

P. Sandler[119], Elinor J. Sawyer[120], Rita K. Schmutzler[79, 80, 121], Christopher Scott[85], Xiao-Ou Shu[122], Jacques Simard[123], Snezhana Smichkoska[124], Christof Sohn[125], Melissa C. Southey[71, 73, 126], John J. Spinelli[127, 128], Jennifer Stone[72, 129], Rulla M. Tamimi[58, 130], Jack A. Taylor[119, 131], Rob A.E.M. Tollenaar[132], Ian Tomlinson[133, 134], Melissa A. Troester[110], Thérèse Truong[78], Celine M. Vachon[135], Elke M. van Veen[61, 62], Sophia S. Wang[136, 137], Clarice R. Weinberg[138], Camilla Wendt[104], Hans Wildiers[66], Robert Winqvist[114, 115], Alicja Wolk[82, 139], Wei Zheng[122], Argyrios Ziogas[12], Alison M. Dunning[17], Paul D.P. Pharoah[8, 17], Douglas F. Easton[8, 17], A. Forbes Howie[140], Julian Peto[7], Isabel dos-Santos-Silva[7], Anthony J. Swerdlow[6, 141], Jenny Chang-Claude[4, 111#], Marjanka K. Schmidt[3, 142#], Nick Orr[2#], Olivia Fletcher[1#]

[1] The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK.

[2] Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Ireland, UK.

[3] Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

[4] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[5] Faculty of Medicine, University of Heidelberg, Heidelberg, Germany.

[6] Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK.

[7] Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.

[8] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[9] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA.

[10] Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada.

[11] Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.

[12] Department of Epidemiology, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA.

[13] N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus.

[14] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[15] Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada.

[16] Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden.

[17] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.

[18] Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.

[19] Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain.

[20] Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain.

[21] Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia.

[22] Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland.

[23] Department of Oncology, Örebro University Hospital, Örebro, Sweden.

[24] VIB Center for Cancer Biology, Leuven, Belgium.

[25] Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium.

[26] Department of Radiation Oncology, Hannover Medical School, Hannover, Germany.

[27] Gynaecology Research Unit, Hannover Medical School, Hannover, Germany.

[28] Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.

[29] Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.

[30] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[31] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany.

[32] iFIT-Cluster of Excellence, University of Tübingen, Tübingen, Germany.

[33] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.

[34] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.

[35] Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[36] Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Germany.

[37] Department of Biology, University of Pisa, Pisa, Italy.

[38] Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[39] Oncology and Genetics Unit, Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain.

[40] Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

[41] Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia.

[42] Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway.

[43] Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway.

[44] Department of Research, Vestre Viken Hospital, Drammen, Norway.

[45] Section for Breast- and Endocrine Surgery, Department of Cancer, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Ullevål, Oslo, Norway.

[46] Department of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway.

[47] Department of Pathology, Akershus University Hospital, Lørenskog, Norway.

[48] Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway.

[49] Department of Oncology, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital-Radiumhospitalet, Oslo, Norway.

[50] National Advisory Unit on Late Effects after Cancer Treatment, Oslo University Hospital-Radiumhospitalet, Oslo, Norway.

[51] Department of Oncology, Akershus University Hospital, Lørenskog, Norway.

[52] Breast Cancer Research Consortium, Oslo University Hospital, Oslo, Norway.

[53] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.

[54] Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK.

[55] Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK.

[56] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[57] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

[58] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[59] Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany.

[60] LIFE - Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany.

[61] Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.

[62] North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

[63] David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA.

[64] Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK.

[65] Cancer Research UK Edinburgh Centre, The University of Edinburgh, Edinburgh, UK.

[66] Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium.

[67] Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.

[68] Fundación Pública Galega de Medicina Xenómica, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain.

[69] Moores Cancer Center, University of California San Diego, La Jolla, CA, USA.

[70] Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA.

[71] Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia.

[72] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.

[73] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia.

[74] Department of Medicine, McGill University, Montréal, QC, Canada.

[75] Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, Canada.

[76] Peter MacCallum Cancer Center, Melbourne, Victoria, Australia.

[77] Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, New South Wales, Australia.

[78] Center for Research in Epidemiology and Population Health (CESP), Team Exposome and Heredity, INSERM, University Paris-Saclay, Villejuif, France.

[79] Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.

[80] Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.

[81] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[82] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

[83] Department of Oncology, Södersjukhuset, Stockholm, Sweden.

[84] Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[85] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.

[86] Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands.

[87] Division of Cancer Sciences, University of Manchester, Manchester, UK.

[88] Nuffield Department of Population Health, University of Oxford, Oxford, UK.

[89] Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia.

[90] Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Victoria, Australia.

[91] Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland.

[92] Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland.

[93] Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA.

[94] Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA.

[95] Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia.

[96] Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.

[97] Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland.

[98] Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland.

[99] Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland.

[100] Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[101] Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway.

[102] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA.

[103] Unit of Medical Genetics, Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy.

[104] Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden.

[105] Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece.

[106] Department of Gynecology and Obstetrics, University of Munich, Campus Großhadern, Munich, Germany.

[107] Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA.

[108] Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland.

[109] Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

[110] Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[111] Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

[112] Genome Diagnostics Program, IFOM - the FIRC Institute of Molecular Oncology, Milan, Italy.

[113] Research Centre for Genetic Engineering and Biotechnology 'Georgi D. Efremov', MASA, Skopje, Republic of North Macedonia.

[114] Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland.

[115] Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland.

[116] Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel.

[117] Department of Oncology, Mayo Clinic, Rochester, MN, USA.

[118] Department of Oncology, University Hospital of Larissa, Larissa, Greece.

[119] Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.

[120] School of Cancer & Pharmaceutical Sciences, Comprehensive Cancer Centre, Guy's Campus, King's College London, London, UK.

[121] Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany.

[122] Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA.

[123] Genomics Center, Centre Hospitalier Universitaire de Québec - Université Laval Research Center, Québec City, QC, Canada.

[124] Ss. Cyril and Methodius University in Skopje, Medical Faculty, University Clinic of Radiotherapy and Oncology, Skopje, Republic of North Macedonia.

[125] National Center for Tumor Diseases, University Hospital and German Cancer Research Center, Heidelberg, Germany.

[126] Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia.

[127] Population Oncology, BC Cancer, Vancouver, BC, Canada.

[128] School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada.

[129] The Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and University of Western Australia, Perth, Western Australia, Australia.

[130] Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA.

[131] Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.

[132] Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands.

[133] Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK.

[134] Wellcome Trust Centre for Human Genetics and Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK.

[135] Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA.

[136] Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA, USA.

[137] City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA, USA.

[138] Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.

[139] Department of Surgical Sciences, Uppsala University, Uppsala, Sweden.

[140] MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh, UK.

[141] Division of Breast Cancer Research, The Institute of Cancer Research, London, UK.

[142] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands.

* These authors contributed equally to this work

[#] Joint senior authors

Correspondence to:

Nichola Johnson

Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK

Tel: +44 (0) 207 153 5487, Fax: +44 (0) 207 153 5340

E-mail: nichola.johnson@icr.ac.uk

ORCID ID: 0000-0002-8230-5662

Key words: premenopausal oestrogens, breast cancer risk, low penetrance risk alleles

Running title: Premenopausal hormones and breast cancer risk

ABSTRACT:

Background: Epidemiological studies provide strong evidence for a role of endogenous sex-hormones in the aetiology of breast cancer. The aim of this analysis was to identify genetic variants that are associated with urinary sex-hormone levels and breast cancer risk.

Methods: We carried out a genome-wide association study of urinary oestrone-3-glucuronide and pregnanediol-3-glucuronide levels in 560 premenopausal women, with additional analysis of progesterone levels in 298 premenopausal women. To test for association with breast cancer risk we carried out follow-up genotyping in 90 916 cases and 89 893 controls from the Breast Cancer Association Consortium. All women were of European ancestry.

Results: For pregnanediol-3-glucuronide, there were no genome-wide significant associations; for oestrone-3-glucuronide, we identified a single peak mapping to the *CYP3A* locus, annotated by rs45446698. The minor rs45446698-C allele was associated with lower oestrone-3-glucuronide (-49.2%, 95% CI -56.1% to -41.1%, P=3.1x10^{-18}); in follow-up analyses, rs45446698-C was also associated with lower progesterone (-26.7%, 95% CI -39.4% to -11.6%, P=0.001) and reduced risk of oestrogen and progesterone receptor-positive breast cancer (OR=0.86, 95% CI 0.82-0.91, P=6.9x10^{-8}).

Conclusions: The *CYP3A7*1C* allele is associated with reduced risk of hormone receptor-positive breast cancer possibly mediated via an effect on the metabolism of endogenous sex-hormones in premenopausal women.

BACKGROUND

Epidemiological studies provide strong evidence for a role of endogenous hormones in the aetiology of breast cancer [1,2]. Pooled analyses of data from prospective studies estimated that a doubling of circulating oestradiol or oestrone was associated with a 30-50% increase in breast cancer risk in postmenopausal women and a 20-30% increase in breast cancer risk in premenopausal women; there was no evidence that premenopausal progesterone levels were associated with breast cancer risk [2,3]. We have previously screened 642 SNPs tagging 42 genes involved in sex steroid synthesis or metabolism, and tested for association with premenopausal urinary oestrone glucuronide and pregnanediol-3-glucuronide levels, measured in urine samples collected at pre-specified days of the woman's menstrual cycle [4]. Oestrone-3-glucuronide and pregnanediol-3-glucuronide are urinary metabolites of oestrogen and progesterone, respectively [5,6], that are used in the context of reproductive medicine to monitor ovarian activity [7]. None of the variants that we tested was associated with urinary pregnanediol-3-glucuronide but a rare haplotype, defined by two SNPs spanning the cytochrome P450 family 3 subfamily A (*CYP3A*) gene cluster, was associated with a highly significant 32% difference in urinary oestrone-3-glucuronide [4]. Fine-scale mapping analyses, identified the SNP rs45446698 as a putative causal variant at this locus; rs45446698 is one of seven highly correlated SNPs that cluster within the *CYP3A7* promoter and comprise the *CYP3A7*1C* allele [8]. A genome-wide association study (GWAS) of postmenopausal plasma oestradiol levels found no association at this locus [9]. A subsequent GWAS of pre- and postmenopausal hormone levels similarly found no association with plasma oestradiol at this locus; they did however find associations at this locus with DHEAS and progesterone [10].

The *CYP3A* genes (*CYP3A5*, *CYP3A7*, and *CYP3A4*) encode enzymes that metabolize a diverse range of substrates [11]; in addition to a role in the oxidative metabolism of hormones, CYP3A enzymes metabolize around 50% of all clinically used drugs including many of the agents used in treating cancer [12]. CYP3A4, the major isoform in adults, is predominantly expressed in the liver, where it is

the most abundant P450, accounting for 30% of total CYP450 protein. *CYP3A7*, the major isoform in the foetus, is generally silenced shortly after birth [13]. In *CYP3A7*1C* carriers a region within the foetal *CYP3A7* promoter has been replaced with the equivalent region from the adult *CYP3A4* gene [14]; this results in adult expression of *CYP3A7* in *CYP3A7*1C* carriers and may influence metabolism of endogenous hormones, exogenous hormones used in menopausal hormone treatment and clinically prescribed drugs, including agents used in treating cancer, in these individuals [12,15]. In order to identify additional variants that are associated with premenopausal urinary hormone levels and to further characterise the associations at the *CYP3A* locus we carried out a GWAS of urinary oestrone-3-glucuronide and pregnanediol-3-glucuronide levels, using mid-luteal phase urine samples from women of European ancestry and followed-up by testing for an association with breast cancer risk in cases and controls from the Breast Cancer Association Consortium (BCAC). To determine whether the *CYP3A7*1C* allele influences metabolism of exogenous hormones we evaluated gene environment interactions with menopausal hormone treatment for breast cancer risk and to investigate whether adult expression of *CYP3A7* impacts on agents used in treating cancer, we analysed associations with breast cancer-specific survival.

METHODS

GWAS subjects

Generations Study

Full details of the Generations Study have been published previously [16]. Briefly, the Generations Study is a cohort study of more than 110 000 women from the UK general population, who were recruited beginning in 2003 and from whom detailed questionnaires and blood samples have been collected to investigate risk factors for breast cancer.

British Breast Cancer Study

342 Full details of the British Breast Cancer Study have been published previously [17]. Briefly, the British
343 Breast Cancer Study is a national case–control study of breast cancer in which cases of breast cancer
344 were ascertained through the cancer registries of England and Scotland and through the National
345 Cancer Research Network. Cases were asked to invite a healthy female first-degree relative with no
346 history of cancer and a female friend or non-blood relative to participate in the study.

347 **Mammography Oestrogens and Growth Factors Study**

348 Full details of the Mammography Oestrogens and Growth Factors study have been published previ-
349 ously [18]. Briefly, this is an observational study nested within a trial of annual mammography
350 screening in young women that was conducted in Britain [19]. Approximately 54 000 women aged 39–
351 41 years were randomly assigned to the intervention arm from 1991 to 1997 and offered annual
352 mammograms until age 48 years. From 2000 to 2003, women in the intervention arm who were still
353 participating in this trial were invited to participate in the Mammography Oestrogens and Growth
354 Factors study; they were asked to provide a blood sample and complete a questionnaire detailing
355 demographic, lifestyle, and reproductive factors. More than 8 000 women were enrolled in the
356 study.

357 GWAS subjects were drawn from the Generation Study (N=184), the British Breast Cancer Study
358 (N=284) and the Mammography Oestrogens and Growth Factors study (N=109). To be eligible for
359 the GWAS analysis of oestrone-3-glucuronide and pregnanediol-3-glucuronide levels, women had to
360 be having regular menstrual cycles (i.e. their usual cycle length had to be between 21 and 35 days)
361 and not using menopausal hormone therapy or oral contraceptives. All of the women included in
362 this analysis reported being of European ancestry, and none had been diagnosed with breast cancer
363 at the time of study recruitment.

364 **Measurement of Hormone Levels**

The protocol for collecting timed urine samples has been published previously [18]. Briefly, a woman's predicted date of ovulation was estimated from the date of the first day of her last menstrual period and her usual cycle length; ovulation was predicted to occur 14 days before the date of her next menstrual period. On this basis, women were asked to provide a series of early morning urine samples on pre-specified days of their cycle. For this analysis the mid-luteal phase sample, taken at seven days after the predicted day of ovulation was used. To confirm that ovulation had occurred, consistent with the predicted date of ovulation, pregnanediol-3-glucuronide was measured; to take account of differences in volume in early morning urine samples from different women we measured creatinine, a waste product of normal muscle and protein metabolism that is released at a constant rate by the body. Samples in which pregnanediol-3-glucuronide, adjusted for creatinine levels, was >0.3 umol/mol were taken forward for measurement of creatinine adjusted oestrone-3-glucuronide. Pregnanediol-3-glucuronide and oestrone-3-glucuronide were analysed by commercial competitive ELISA Kits (Arbor Assays, Ann Arbor, USA) according to the manufacturer's instructions. For pregnanediol-3-glucuronide, the lower limit of detection was determined as 0.64 nmol/l; intra and inter assay coefficients of variation were 3.7% and 5.2%, respectively. For oestrone-3-glucuronide the lower limit of detection was determined as 19.6 pmol/l; intra and inter assay coefficients of variation were 3.5% and 5.9%, respectively. Creatinine was determined using the creatininase/creatinase specific enzymatic method [20] using a commercial kit (Alpha Laboratories Ltd. Eastleigh, UK) adapted for use on a Cobas Fara centrifugal analyser (Roche Diagnostics Ltd, Welwyn Garden City, UK). For within run precision, the coefficient of variation was <3% while for intra-batch precision the coefficient of variation was <5%.

For 303 premenopausal women participating in the Generations Study (184 as above and an additional 119 for whom timed urine samples were accrued more recently) urinary progesterone levels were also measured using an "in house" ELISA. 96-well plates (Greiner Bio-One GmbH, Frickenhausen, Germany) were coated with 100µl of 5µg/ml GAM (Arbor Assays, Ann Arbor, USA) in ELISA coating buffer (100mM Na Bicarbonate, pH 9.6) covered and incubated in a fridge at 4 ^{0}C

overnight. Before use the plates were washed three times with wash buffer 0.05M Tris/HCl and 0.05% Tween 20, pH 7.4 (Tween® 20, Sigma-Aldrich, Inc., St. Louis, MO, USA). Standards, samples and controls (20µl per well) were added to each well, followed by 80µl of Progesterone 3-HRP conjugate (Astra Biotech GmbH, Berlin, Germany) at 1:10 000 in assay buffer (PBS pH 7.4 containing 0.1 %BSA and 250 ng/ml Cortisol), followed by 50 µl of monoclonal progesterone Ab (Astra Biotech GmbH, Berlin, Germany) 1:50 000 in assay buffer. Plates were incubated at room temperature for two hours on a microtitre plate shaker (IKA®, Schuttler MTS4, IKA Labortechnik, Staufen, Germany), then washed five times with assay wash buffer and 120µl of substrate solution (3,3,5,5-Tetramethylbenzidine, Millipore Corporation, Temecula, CA, USA) was added to each well. Plates were incubated at room temperature without shaking in the dark. After 20 min, the reaction was stopped by adding 80µl of 2N H_2SO_4 solution (Sigma-Aldrich Company Ltd., Dorset, UK). Finally, plates were read on a plate reader at 450nm. Standard curves were prepared with a total of eight different concentrations (16, 8, 4, 2, 1, 0.5, 0.25 0 ng/ml). Samples, standards and controls were included in duplicate. Inter- and intra-assay coefficients of variation were calculated from two controls of low and high progesterone in duplicate in each of eight assays. The inter-assay coefficients of variation for low and high pools respectively were 11.4 and 9.1%, the intra-assay coefficients of variation were 8.9 and 5.6%. The lower limit of detection was calculated at 0.1 ng/ml. Cross-reaction with other steroids was: oestrone: 0.17%, oestradiol: 0.28%, oestriol: 0.18%, dehydroepiandrosterone: 0.02%, testosterone: 0.36%, dihydrotestosterone: 0.15%, 17α-hydroxyprogesterone: 2.9%, androstenedione: 0.14%, 11-deoxycortisol: 0.46%, corticosterone: 0.18%, cortisone: 0.04% and cortisol: 0.04%.

GWAS genotyping and quality control

DNA from 577 women was genotyped using Illumina Infinium OncoArray 500K BeadChips. We excluded samples for which <95% of SNPs were successfully genotyped. Identity by descent analysis was used to identify closely related individuals enabling exclusion of first-degree relatives. We

applied SmartPCA [21] to our data and used phase II HapMap samples to identify individuals with non-Caucasian ancestry. The first two principal components for each individual were plotted and k-means clustering was used to identify samples separated from the main Caucasian cluster. SNPs with call rates <95% were excluded, as were SNPs with minor allele frequency (MAF) <2% and those whose genotype frequencies deviated from Hardy Weinberg proportions at $P<1x10^{-05}$. Following QC, 487 659 SNPs were successfully genotyped in 560 samples (Generations Study: N=179, British Breast Cancer Study: N=278, Mammography Oestrogens and Growth Factors study: N=103). Genome-wide imputation was performed using 1KGP Phase 3 reference data. Haplotypes were pre-phased using SHAPEIT2 [22]. Imputation was performed using IMPUTE2 [23]. Imputed SNPs with INFO scores <0.8 and MAFs<2% were excluded from subsequent analyses. After QC a set of 7 792 694 successfully imputed SNPs were available for association analysis.

Genotyping rs45446698 and sequencing of the *CYP3A7*1C* allele

For the 119 Generations Study women who were not included in the GWAS but for whom progesterone was subsequently measured, rs45446698 was genotyped by TaqMan (Thermo Fisher Scientific Ltd, UK). The call rate was 100% with 100% concordance between 12 duplicates. To confirm that rs45446698 tags the *CYP3A7*1C* allele we sequenced this region in 31 women selected on the basis of their rs45446698 genotype (9 common homozygotes and 22 carriers). A 370bp DNA region (chr7: 99 332 745-99 333 114; GRCh37/hg19) was amplified using Phusion High-Fidelity DNA Polymerase (New England Biolabs, UK) and primers CCATAGAGACAAGAGGAGA (Forward) and CTGAGTCTTTTTTTCAGCAGC (Reverse). The PCR product was purified using QIAquick Gel Extraction Kit (Qiagen) and Sanger sequenced using a commercially available service (Eurofins Genomics, Germany).

Statistical analysis of GWAS data

Tests of association between SNP genotypes and log-transformed creatinine-adjusted oestrone-3-glucuronide and pregnanediol-3-glucuronide adjusted for study, were performed using linear regression in SNPTEST v2.5 [24]. Test statistic inflation was assessed visually using a QQ Plot (Figure S1) and formally by calculating the inflation factor, λ. There was no evidence of systematic test statistic inflation (λ=1.01 for both oestrone-3-glucuronide and pregnanediol-3-glucuronide). For the single significant association (rs45446698) we used multivariate linear regression to adjust for potential confounders: age at menarche (<12, 12, 13, 14, >14 years), age at collection of urine samples (<35, 35-40, ≥40 years), body mass index (BMI; <18.5, 18.5 - <20.0, 20.0 - <25.0, 25.0 - <30.0, ≥ 30.0 kg/m^2) and parity (0, 1, 2, ≥3 live births).

Follow-up genotyping of rs45446698

Genotype data for rs45446698 were generated as part of iCOGS [25] and OncoArray [26]. Full details of SNP selection, array design, genotyping and post-genotyping QC have been published [25,26]. Participants genotyped in both collaborations were excluded from the iCOGS data sets with the exception of the GxE interaction analysis of menopausal hormone treatment, for which five studies (CPSII, PBCS, UKBGS, MCCS and pKARMA) were excluded from OncoArray, rather than iCOGS, in order to maximise the number of studies with sufficient cases and controls for analysis. We excluded cases with breast tumours of unknown invasiveness, or *in situ* disease and those for whom age at diagnosis was not known. After QC exclusions [26], the call rate for rs45446698 in OncoArray data was 99.66% and there was no evidence of deviation from Hardy–Weinberg equilibrium in controls (Table S1). In iCOGS data, rs45446698 was imputed using 1KGP Phase 3 reference data (info score=0.94); we used gene dosages (≤0.2=0, >0.8 & ≤1.2=1, >1.8=2) to call genotypes for 99.22% of samples.

Statistical analysis of rs45446698 and breast cancer risk

Due to the low MAF of rs45446698 (3.7%, 0.03% and 0.4% in individuals of European, Asian and African ancestry, respectively) we restricted our analyses to individuals of European ancestry and excluded studies with <50 cases or controls; there were 35 (iCOGS) and 56 (OncoArray) studies for the current case-control analysis (Tables S1 and S2).

We combined heterozygote and rare homozygote genotypes and estimated carrier ORs using logistic regression, adjusted for 15 principal components [25,26] and study. Stratum specific carrier ORs were estimated for a set of pre-specified prognostic variables (ER, PR, HER2, grade and stage). We excluded studies with <50 cases or controls in any individual stratum from stratified analyses. Interactions were assessed based on case-only models (oestrogen receptor (ER), progesterone receptor (PR), Her2, stage and grade). In the subset of studies for which covariate data was available, we used multivariable logistic regression to adjust for reference age (defined as age at diagnosis for cases and age at interview for controls), age at menarche, BMI and parity (as above). Finally, we stratified our analyses on menopausal status at reference age. When menopausal status was missing, the reference age was used as a surrogate (<54 premenopausal and ≥54 postmenopausal). To select the reference age that most accurately captured menopausal status in this group of studies, we generated AUC curves based on women who had reported natural menopause with different reference age cut-offs (50 to 56 years); on this basis a reference age of 54 was selected. *P* values were estimated using likelihood ratio tests with one degree of freedom. All *P* values reported, for all analyses, are two-sided. Statistical analyses were performed using STATA version 11.0 (StataCorp, College Station, TX, USA).

Statistical analysis of gene-environment interaction (GxE) with menopausal hormone treatment

Postmenopausal women from 13 (iCOGS) and 27 (OncoArray) studies provided data on menopausal hormone treatment. Menopausal status and postmenopausal hormone use were derived as of the reference date (defined as date of diagnosis for cases and interview for controls); women with unknown age at reference date were excluded from this analysis. All analyses were conducted only

in postmenopausal women. Carrier ORs for breast cancer risk were estimated using logistic regression stratified by current use of menopausal hormone treatment, oestrogen-progesterone therapy and oestrogen only therapy, respectively. Analyses were adjusted for study, ten principal components, reference age, age at menarche, parity, BMI, former use of menopausal hormone treatment, and use of any menopausal hormone treatment preparation other than the one of interest in analyses of current use of menopausal hormone treatment by type. To account for potential heterogeneity of main effects of menopausal hormone treatment/oestrogen-progesterone therapy/oestrogen only therapy by study design, we included an interaction term between the risk factor of interest and an indicator variable for study design (prospective cohorts/population-based case-control studies, non-population-based studies). Interactions between rs45446698 and current use of menopausal hormone treatment, oestrogen-progesterone therapy, and oestrogen only therapy were assessed using likelihood ratio tests, based on logistic regression models with and without interaction between rs45446698 and current use of menopausal hormone treatment, oestrogen-progesterone therapy, and oestrogen only therapy, respectively. Statistical analyses were performed using SAS 9.4 and R (version 3.4.4).

Statistical analysis of Breast Cancer-Specific Survival in cases

38 (iCOGS) and 63 (OncoArray) studies provided follow-up data for analysis of breast cancer-specific survival. Analysis of outcome was restricted to patients who were at least 18 years old at diagnosis and for whom vital status at, and date of last follow-up were known. Patients ascertained for a second tumour were excluded. Time-to-event was calculated from the date of diagnosis. For prevalent cases with study entry after diagnosis left truncation was applied, i.e. follow-up started at the date of study entry [27]. Follow-up was right censored at date of death (death known to be due to breast cancer was considered an event), date the patient was last known to be alive if death did not occur, or at 10 years after diagnosis, whichever came first. Follow-up was censored at 10 years due to limited data availability after this time. Hazard ratios (HR) for association of rs45446698 genotype

512 with breast cancer-specific survival were estimated using Cox proportional hazards regression
513 implemented in the R package survival (v. 2.43-3) stratified by country. iCOGS and OncoArray
514 estimates were combined using an inverse-variance weighted meta-analysis.

515 **RESULTS**

516 We tested 8 280 353 autosomal SNPs for association with luteal phase creatinine adjusted oestrone-
517 3-glucuronide and pregnanediol-3-glucuronide in 560 premenopausal women. For oestrone-3-
518 glucuronide, we identified a single peak mapping to the *CYP3A* locus at chromosome 7q22.1 (Figure
519 1, Table S3); conditioning on any of the top SNPs there were no additional independent signals. Four
520 of the SNPs that were significant at $P < 1 \times 10^{-8}$ comprise part of the seven SNP *CYP3A7*1C* allele [8,15]
521 including the top, directly genotyped SNP, rs45446698 (Table S3). The rare rs45446698-C allele
522 (MAF=0.035) was associated with a 49.2% reduction in luteal phase oestrone-3-glucuronide (95% CI -
523 56.1% to -41.1%, $P=3.1 \times 10^{-18}$; Table 1) and explained 11.5% of the variation in oestrone-3-
524 glucuronide in these premenopausal women. Since hormone levels may be influenced by both
525 demographic and reproductive factors, we adjusted for age at urine collection, age at menarche,
526 body mass index and parity; these adjustments did not alter the association (fully adjusted model:
527 44.8% reduction, 95% CI -53.3% to -34.8%, $P=2.1 \times 10^{-12}$; Table 1).

528 For pregnanediol-3-glucuronide, there were no associations that were significant at a threshold of
529 $P<1 \times 10^{-8}$ (Figure S2). An association between the *CYP3A* locus and progesterone levels has been
530 reported previously [10], accordingly we measured progesterone in addition to pregnanediol-3-
531 glucuronide in premenopausal women from the Generations Study. Progesterone was moderately
532 correlated with pregnanediol-3-glucuronide (r=0.37, $P=7.4 \times 10^{-12}$) but while there was no
533 association between the rs45446698-C allele and urinary pregnanediol-3-glucuronide levels (5.5%
534 reduction, 95% CI -24.2% to +17.7%, $P = 0.61$) in this group of women, the rs45446698-C allele was
535 associated with significantly lower luteal phase urinary progesterone levels (26.7% reduction, 95% CI

-39.4% to -11.6%, P=0.001; Table 1). Adjusting these analyses for covariates, as above, did not alter the results (Table 1).

To test for association between rs45446698 and breast cancer risk, we combined genotype data from 56 studies (OncoArray; Table S1) with imputed data from 35 studies (iCOGS; Table S2) in a total of 90 916 cases and 89 893 controls of European Ancestry. The rs45446698-C allele was associated with a reduction in breast cancer risk (OR=0.94, 95% CI 0.91-0.98, P=0.002; Table 2) with no evidence of heterogeneity between data sets (P_{het}=0.58). There was no evidence that the reduction in breast cancer risk associated with being a rs45446698-C carrier differed according to Her2 status, tumour grade, or stage (Table S4). Stratifying by ER status, the association was limited to ER-positive (ER+) breast cancers (OR=0.91, 95% CI 0.87-0.96, P=0.0002 and OR=1.03, 95% CI 0.95-1.11, P=0.50 for ER+ and ER- cancers, respectively; P_{int}=0.03; Table 2). Stratifying by ER and PR status, the association was limited to ER+/PR+ cancers (ER+/PR+: OR=0.86, 95% CI 0.82–0.91, P=6.9 x 10^{-8}; ER+/PR-: OR=1.06, 95% CI 0.96–1.16, P=0.25; P_{int}=0.0001; Table 2). Adjusting for demographic and reproductive factors in the subset of studies for which these additional covariates were available did not alter this association (Table S5). Defining reference age as age at diagnosis for cases and age at interview for controls and using this as a proxy for menopausal status (<54 or ≥54 years) we further stratified our analysis on menopausal status; there was little evidence that the association with ER+/PR+ breast cancer differed by menopausal status (premenopausal OR=0.94 95% CI 0.84-1.06, P=0.31, postmenopausal OR=0.86 95% CI 0.80-0.93, P=0.0001, P_{het}=0.28).

On the assumption that genetic variants that influence metabolism of endogenous hormones [5] may also impact on metabolism of exogenous hormones, we investigated whether menopausal hormone treatment modified the association between rs45446698 genotype and ER+/PR+ breast cancer risk in 17 831 postmenopausal breast cancer cases and 40 437 postmenopausal controls. The rs45446698-C carrier OR was lower (i.e. more protective) in current users of any menopausal hormone treatment but particularly in those who used combined oestrogen-progesterone therapy

(current users: OR=0.68, 95% CI 0.52-0.90, *P*=0.007; never users: OR=0.85, 95% CI 0.76-0.95, *P*=0.005; Table 3). This difference was not, however, statistically significant (P_{int}=0.15; Table 3).

Finally, to determine whether rs45446698 genotype could affect patient outcome by influencing metabolism of cytotoxic agents that are CYP3A substrates [15], we tested for association between rs45446698 genotype and 10-year breast cancer-specific survival in 91 539 breast cancer cases from 71 studies for whom follow-up data were available. There was no overall association between rs45446698 genotype and breast cancer-specific survival (HR=0.99, 95% CI 0.91–1.09, *P*=0.90, Table 4); nor was there any evidence of an association in analyses stratified by tumour characteristics (Table S6). Stratifying by treatment regimen we found no evidence that rs45446698 genotype influenced outcome in cases who were treated with a hormonal agent (i.e. tamoxifen or an aromatase inhibitor; Table 4). There was, however, some evidence that in cases who were treated with a taxane, carriers of the rs45446698-C allele had reduced breast cancer-specific survival compared with non-carriers (HR=1.46, 95% CI 1.08–1.97, *P*=0.01; Table 4).

DISCUSSION

This present GWAS identified a single, highly significant association between the *CYP3A7*1C* allele (tagged by rs4546698) and premenopausal urinary oestrone-3-glucuronide. This finding alone is not novel; we have previously reported an association between the *CYP3A7*1C* allele, parent oestrogens and several oestrogen metabolites [5]. What we have demonstrated for the first time is the extent to which this signal dominates the genetic architecture of hormone levels in premenopausal women of Northern European ancestry (Figure1; rs45446698 *P*=3.1 x 10^{-18}, all other signals P >1 x 10^{-8}) and we estimate that 11.5% of the variance in urinary oestrone-3-glucuronide levels is explained by this one allele.

Two previous GWAS of circulating oestrogen levels have been published, neither reported an association with the *CYP3A* locus [9,10]. This lack of replication may be explained by our choice of study

population. The first GWAS [9] was conducted in postmenopausal women (N=1 623) participating in the Nurses' Health Study and the Sisters in Breast Screening Study. The second was conducted within the Twins UK study (N=2 913) and included men as well as pre-, peri- and postmenopausal women. A strength or our GWAS is that all of the women were premenopausal and had regular menstrual cycles; circulating levels of oestrogens in premenopausal women are much higher compared with those in postmenopausal women [4,28]. For each woman we assayed a single urine sample taken in the mid-luteal phase of her cycle at exactly seven days after her predicted day of ovulation. Thus, although our study is relatively small (N=560), we may have had greater power to detect an association at the *CYP3A* locus than previous studies due to the very homogeneous premenopausal study population that we selected.

Our findings also demonstrate the potential significance of the choice of hormone or hormone metabolite; both of the previous GWAS assayed plasma oestradiol. In a targeted analysis of urinary oestrogen metabolites, we have previously shown that the association between the *CYP3A7*1C* allele and oestrone (45.3% lower levels in carriers, P=0.0005) is more pronounced than the association with oestradiol (26.7% lower levels, P=0.07) with the implication that measuring urinary oestrone-3-glucuronide (rather than plasma oestradiol) may have contributed to our positive findings. Similarly, by measuring pregnanediol-3-glucuronide and progesterone in premenopausal women from the Generations Study we were able to demonstrate a significant association of rs45446698 with progesterone (27% reduction, P=0.001) in the absence of an association with pregnanediol-3-glucuronide (6% reduction, P=0.61).

The fact that we measured a urinary oestrogen metabolite (oestrone-3-glucuronide) rather than serum or plasma oestrogens (oestradiol or oestrone) limits the interpretation of our results in terms of a causal association. Estimates of the association between circulating oestrogens and breast cancer risk are based on measurements of hormone levels in plasma or serum [3] and in a recent study that measured luteal phase serum oestrogens and urinary oestrogen metabolites in 249

611 premenopausal women [29], serum oestradiol and oestrone were only moderately correlated with
612 urinary oestrone (serum oestradiol: r=0.39, serum oestrone: r=0.48). Our analysis of rs45446698
613 genotypes in 90 916 cases and 89 893 controls from BCAC, however, provides robust evidence of an
614 association of the *CYP3A7*1C* allele with breast cancer risk overall (OR=0.94, *P*=0.002) and a more
615 pronounced protective effect on ER+/PR+ breast cancers (OR=0.86, $P=6.9 \times 10^{-8}$). The specificity of
616 this association (comparing ER+/PR- with ER+/PR+ cancers, P_{het}=0.001) and our replication of Ruth
617 and colleagues report of a signal at the *CYP3A* locus in their analysis of circulating progesterone
618 levels [10] raise the possibility that premenopausal progesterone levels might influence risk of ER+/PR+
619 breast cancers. This would be in contrast to the findings from Key and colleagues who reported no
620 evidence of an association between premenopausal progesterone levels and breast cancer risk
621 overall and no heterogeneity in estimates stratified by PR status [3]. However, the number of cases of
622 PR+ (N=158) and PR- (N=61) breast cancer was small, and this analysis may have lacked power to
623 detect modest associations in subgroups of cancers. Alternatively, the association of rs45446698
624 genotype with ER+/PR+ breast cancer risk specifically, may be due to the fact that PR is a marker for
625 an intact oestrogen signalling pathway [30] confirming a direct link between levels of oestrogen (or
626 oestrogen signalling) and proliferation in this subgroup of cancers.

627 Our analysis of the *CYP3A7*1C* allele, menopausal hormone treatment and breast cancer risk was
628 inconclusive; while the carrier ORs were consistent with a greater protective effect of this allele in
629 women taking exogenous hormones, particularly oestrogen-progesterone therapy, none of the
630 interactions was statistically significant. Overall, there were 14 119 ER+/PR+ breast cancer cases and
631 32 418 controls for this sub-group analysis but for what was, arguably, the most pertinent subgroup
632 (i.e. current oestrogen-progesterone therapy use) the number of cases who were current users was
633 relatively small (*CYP3A7*1C* carriers N=107, non-carriers N=1 498) and power was limited to detect
634 modest interactions. There are limitations to this analysis; we focussed on current menopausal
635 hormone treatment use (adjusted for past use) as it is for current use that the association with

breast cancer risk is strongest [31] but we did not have information on dose, duration or the formulation that was used.

Finally, we found no association between *CYP3A7*1C* carrier status and survival in patients treated with tamoxifen, a known CYP3A substrate. This may reflect the fact that compared to CYP3A4, CYP3A7 is a poor metaboliser of tamoxifen [32] or that standard doses of tamoxifen achieve high levels of oestrogen receptor saturation [33]. There was some evidence that breast cancer-specific survival was reduced in *CYP3A7*1C* carriers who were treated with a taxane, compared with non-carriers (P=0.01); this may, however, be a chance finding given the number of comparisons that were tested.

In conclusion we present strong evidence that the *CYP3A7*1C* allele impacts on the metabolism of endogenous hormones which in turn reduces the risk of hormone receptor positive breast cancer in carriers. Optimal strategies for breast cancer prevention in women at high risk of breast cancer and in the general population is an area of active research. In this context, *CYP3A7*1C* carriers represent a naturally occurring cohort in which the effects of reduced exposure to endogenous oestrogens and progesterones throughout a woman's premenopausal years can be further investigated. Our results regarding the impact of *CYP3A7*1C* carrier status on exogenous hormones and chemotherapeutic agents are preliminary but warrant further investigation, preferably in the setting of randomised trials.

ADDITIONAL INFORMATION

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. The COGS study would not have been possible without the contributions of the following: Kyriaki Michailidou, (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara

Benlloch (PRACTICAL), Antonis Antoniou, Lesley McGuffog and Ken Offit (CIMBA), Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, the staff of the CNIO genotyping unit, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility.

ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Rosemary Balleine, Robert Baxter, Stephen Braye, Jane Carpenter, Jane Dahlstrom, John Forbes, Soon Lee, Debbie Marsh, Adrienne Morey, Nirmala Pathmanathan, Rodney Scott, Allan Spigelman, Nicholas Wilcken, Desmond Yip. Samples are made available to researchers on a non-exclusive basis. BBCS thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, Rachel Lloyd. The BCINIS study would not have been possible without the contributions of Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. BIGGS thanks Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones. The BREOGAN study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREOGAN), José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS; Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain. BSUCH thanks Peter Bugert, Medical Faculty Mannheim. CBCS

686 thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer
687 Study, and project coordinators Agnes Lai and Celine Morissette. CCGP thanks Styliani Apostolaki,
688 Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou, Georgia Sevastaki,
689 Konstantinos Pompodakis. CGPS thanks staff and participants of the Copenhagen General Population
690 Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne
691 Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing
692 infrastructure for the collection of blood samples for the cases. CNIO-BCS thanks Guillermo Pita,
693 Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit
694 (CNIO). Investigators from the CPS-II cohort thank the participants and Study Management Group for
695 their invaluable contributions to this research. They also acknowledge the contribution to this study
696 from central cancer registries supported through the Centers for Disease Control and Prevention
697 National Program of Cancer Registries, as well as cancer registries supported by the National Cancer
698 Institute Surveillance Epidemiology and End Results program. The authors would like to thank the
699 California Teachers Study Steering Committee that is responsible for the formation and maintenance
700 of the Study within which this research was conducted. A full list of California Teachers Study team
701 members is available at https://www.calteachersstudy.org/team. We thank the participants and the
702 investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks
703 Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach. FHRISK thanks
704 NIHR for funding. GC-HBOC thanks Stefanie Engert, Heide Hellebrand, Sandra Kröber and LIFE -
705 Leipzig Research Centre for Civilization Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter,
706 Ronny Baber). The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology,
707 Stuttgart, and University of Tübingen, Germany [HB, Wing-Yee Lo, Reiner Hoppe], Department of
708 Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany
709 [Yon-Dschun Ko, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter
710 Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ),
711 Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the

712 German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany
713 [Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine
714 and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth].
715 HABCS thanks Michael Bremer. HEBCS thanks Sofia Khan, Johanna Kiiski, Kristiina Aittomäki, Rainer
716 Fagerholm, Kirsimari Aaltonen, Karl von Smitten, Irja Erkkilä. HMBCS thanks Peter Hillemanns, Hans
717 Christiansen and Johann H. Karstens. HUBCS thanks Shamil Gantsev. KARMA and SASBAC thank the
718 Swedish Medical Research Counsel. KBCP thanks Eija Myöhänen, Helena Kemiläinen. kConFab/AOCS
719 wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the
720 heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received
721 funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National
722 Institute of Health (USA)) for their contributions to this resource, and the many families who
723 contribute to kConFab. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and
724 Kathleen Corthouts. MABCS thanks Milena Jakimovska (RCGEB “Georgi D. Efremov”), Emilija
725 Lazarova (University Clinic of Radiotherapy and Oncology), Katerina Kubelka-Sabit, Mitko Karadjozov
726 (Adzibadem-Sistina Hospital), Andrej Arsovski and Liljana Stojanovska (Re-Medika Hospital) for their
727 contributions and commitment to this study. MARIE thanks Petra Seibold, Dieter Flesch-Janys, Judith
728 Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and
729 Stefan Nickels. MBCSG (Milan Breast Cancer Study Group): Paolo Radice, Bernard Peissel, Jacopo
730 Azzollini, Dario Zimbalatti, Daniela Zaffaroni, Bernardo Bonanni, Irene Feroce, Mariarosaria Calvello,
731 Aliana Guerrieri Gonzaga, Monica Marabelli, Davide Bondavalli and the personnel of the Cogentech
732 Cancer Genetic Test Laboratory. The MCCS was made possible by the contribution of many people,
733 including the original investigators, the teams that recruited the participants and continue working
734 on follow-up, and the many thousands of Melbourne residents who continue to participate in the
735 study. We thank the coordinators, the research staff and especially the MMHS participants for their
736 continued collaboration on research studies in breast cancer. MTLGEBCS would like to thank Martine
737 Tranchant (CHU de Québec – Université Laval Research Center), Marie-France Valois, Annie Turgeon

738 and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA
739 extraction, sample management and skilful technical assistance. J.S. is Chair holder of the Canada
740 Research Chair in Oncogenetics. The following are NBCS Collaborators: Kristine K. Sahlberg (PhD),
741 Lars Ottestad (MD), Rolf Kåresen (Prof. Em.) Dr. Ellen Schlichting (MD), Marit Muri Holmen (MD),
742 Toril Sauer (MD), Vilde Haakensen (MD), Olav Engebråten (MD), Bjørn Naume (MD), Alexander Fosså
743 (MD), Cecile E. Kiserud (MD), Kristin V. Reinertsen (MD), Åslaug Helland (MD), Margit Riis (MD),
744 Jürgen Geisler (MD), OSBREAC, Grethe I. Grenaker Alnæs (MSc) and Anne-Lise Børresen-Dale (PhD).
745 NBHS thank study participants and research staff for their contributions and commitment to the
746 studies. For NHS and NHS2 the study protocol was approved by the institutional review boards of
747 the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of
748 participating registries as required. We would like to thank the participants and staff of the NHS and
749 NHS2 for their valuable contributions as well as the following state cancer registries for their help:
750 AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH,
751 OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and
752 interpretation of these data. OBCS thanks Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri
753 Otsukka, Leena Keskitalo and Kari Mononen for their contributions to this study. The OFBCR thanks
754 Teresa Selander, Nayana Weerasooriya and Steve Gallinger. ORIGO thanks E. Krol-Warmerdam, and
755 J. Blom for patient accrual, administering questionnaires, and managing clinical information. The
756 LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCDOC)
757 with the help of Dr. J. Molenaar. PBCS thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-
758 Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. PROCAS thanks NIHR for
759 funding. The RBCS thanks Corine M. Beaufort, Jannet Blom, Renée Broeren – Foekens, Saskia
760 Pelders, Wendy J.C. Prager – van der Smissen, Kirsten Ruigrok – Ritstier, Anita M.A.C. Trapman –
761 Jansen, Michelle van der Vlugt – Daane, Vanja de Weerd, and the Erasmus MC Family Cancer Clinic.
762 SBCS thanks Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and
763 Malcolm W.R. Reed. We thank the SEARCH and EPIC teams. SZBCS thanks Ewa Putresza. UCIBCS

764 thanks Irene Masunaka. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for
765 support and funding of the Generations Study, and the study participants, study staff, and the
766 doctors, nurses and other health care providers and health information sources who have
767 contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical
768 Research Centre.

769 **Author contributions**

770 N.J., K.T., M.E.J., M.J.S., C. G., A.F.H., J.P., I.dS.S. and A.J.S. generated data for the GWAS reported in
771 this manuscript. N.J., S.M., A.M., P.M.K., J.C-C., M.K.S., N.O. and O.F. analysed data relating to this
772 manuscript and drafted the initial manuscript. D.F.E. coordinated the BCAC and led the iCOGS and
773 OncoArray genotyping. P.H. led the COGS collaboration, J.S. led the OncoArray collaboration. M.K.B.,
774 Q.W. and J.D. coordinated the BCAC database. The remaining authors led individual studies and
775 contributed to the design of the study, data collection, and revising the manuscript.

776 **Ethics approval and consent to participate**

777 Collection of blood samples, urine samples and questionnaire information were undertaken with
778 written informed consent and relevant ethical review board approval in accordance with the tenets
779 of the Declaration of Helsinki (Table S7).

780 **Data availability**

781 GWAS data and the complete dataset for follow-up genotyping will not be made publicly available
782 due to restraints imposed by the ethics committees of individual studies; requests for data can be
783 made to the corresponding author (GWAS data) or the Data Access Coordination Committee (follow-
784 up genotyping data) of BCAC (http://bcac.ccge.medschl.cam.ac.uk/). Summary results for all variants
785 genotyped by BCAC (including rs45446698) are available at http://bcac.ccge.medschl.cam.ac.uk/.

Competing interests

These authors declare the following conflict of interest.: Matthias W. Beckmann conducts research funded by Amgen, Novartis and Pfizer. Peter A. Fasching conducts research funded by Amgen, Novartis and Pfizer. He received Honoraria from Roche, Novartis and Pfizer. Allison W. Kurian's institution has received research funding from Myriad Genetics for an unrelated project (funding dates 2017-2019). Per Hall, Paul Pharoah, Olivia Fletcher, Kamila Czene and Angie Cox are members of the Editorial Board of BJC. All other authors declare no conflict of interest.

Funding

This work was supported by Programme Grants from Breast Cancer Now as part of Programme Funding to the Breast Cancer Now Toby Robins Research Centre to O.F. BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710,

809 C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the
810 National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19
811 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-
812 1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of
813 Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and
814 the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065.

815 The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the
816 National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the
817 views or policies of the National Cancer Institute or any of the collaborating centers in the Breast
818 Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or
819 organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also
820 supported by the National Health and Medical Research Council of Australia, the New South Wales
821 Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast
822 Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC)
823 Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was
824 supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]. The Australian Breast
825 Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of
826 Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The AHS study is
827 supported by the intramural research program of the National Institutes of Health, the National
828 Cancer Institute (grant number Z01-CP010119), and the National Institute of Environmental Health
829 Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of
830 the University Hospital of Erlangen. The BBCS is funded by Cancer Research UK and Breast Cancer
831 Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National
832 Cancer Research Network (NCRN). The BCEES was funded by the National Health and Medical
833 Research Council, Australia and the Cancer Council Western Australia and acknowledges funding
834 from the National Breast Cancer Foundation (JS). For BIGGS, ES is supported by NIHR Comprehensive

Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, United Kingdom. IT is supported by the Oxford Biomedical Research Centre. The BREast Oncology GAlician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. CCGP is supported by funding from the University of Crete. The CECILE study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The CNIO-BCS was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The California Teachers Study and the research reported in this publication were supported by the National Cancer Institute of the National Institutes of Health under award number U01-CA199277; P30-CA033572; P30-CA023100; UM1-CA164917; and R01-CA077398. The content is solely the responsibility of the authors and does not

861 necessarily represent the official views of the National Cancer Institute or the National Institutes of
862 Health. The collection of cancer incidence data used in the California Teachers Study was supported
863 by the California Department of Public Health pursuant to California Health and Safety Code Section
864 103885; Centers for Disease Control and Prevention's National Program of Cancer Registries, under
865 cooperative agreement 5NU58DP006344; the National Cancer Institute's Surveillance, Epidemiology
866 and End Results Program under contract HHSN261201800032I awarded to the University of
867 California, San Francisco, contract HHSN261201800015I awarded to the University of Southern
868 California, and contract HHSN261201800009I awarded to the Public Health Institute. The opinions,
869 findings, and conclusions expressed herein are those of the author(s) and do not necessarily reflect
870 the official views of the State of California, Department of Public Health, the National Cancer
871 Institute, the National Institutes of Health, the Centers for Disease Control and Prevention or their
872 Contractors and Subcontractors, or the Regents of the University of California, or any of its
873 programs. The coordination of EPIC is financially supported by the European Commission (DG-
874 SANCO) and the International Agency for Research on Cancer. The national cohorts are supported
875 by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale,
876 Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid,
877 German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF)
878 (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione
879 Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of
880 Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds,
881 Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund
882 (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to
883 Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country,
884 Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk;
885 C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-
886 Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The ESTHER study was supported by a

887 grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were
888 recruited in the context of the VERDI study, which was supported by a grant from the German
889 Cancer Aid (Deutsche Krebshilfe). FHRISK is funded from NIHR grant PGfAR 0707-10031. The GC-
890 HBOC (German Consortium of Hereditary Breast and Ovarian Cancer) is supported by the German
891 Cancer Aid (grant no 110837, coordinator: Rita K. Schmutzler, Cologne). This work was also funded
892 by the European Regional Development Fund and Free State of Saxony, Germany (LIFE - Leipzig
893 Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470,
894 14575/2470). The GENICA was funded by the Federal Ministry of Education and Research (BMBF)
895 Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch
896 Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for
897 Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the
898 Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine,
899 Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany; gefördert durch die
900 Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der
901 Länder - EXC 2180 – 390900677; German Cancer Consortium (DKTK) and German Cancer Research
902 Center (DKFZ). The GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the German
903 Cancer Research Center (DKFZ). The HABCS study was supported by the Claudia von Schilling
904 Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf
905 Bartling Foundation. The HEBCS was financially supported by the Helsinki University Hospital
906 Research Fund, the Finnish Cancer Society, and the Sigrid Juselius Foundation. The HMBCS was
907 supported by a grant from the Friends of Hannover Medical School and by the Rudolf Bartling
908 Foundation. The HUBCS was supported by a grant from the German Federal Ministry of Research
909 and Education (RUS08/017), B.M. was supported by grant 17-44-020498, 17-29-06014 of the Russian
910 Foundation for Basic Research, D.P. was supported by grant 18-29-09129 of the Russian Foundation
911 for Basic Research, E.K was supported by the program for support the bioresource collections №007-
912 030164/2, and the study was performed as part of the assignment of the Ministry of Science and

913 Higher Education of the Russian Federation (№AAAA-A16-116020350032-1). Financial support for
914 KARBAC was provided through the regional agreement on medical training and clinical research
915 (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The
916 Gustav V Jubilee foundation and Bert von Kantzows foundation. The KARMA study was supported by
917 Märit and Hans Rausings Initiative Against Breast Cancer. The KBCP was financially supported by the
918 special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo,
919 the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland.
920 kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the
921 National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer
922 Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of
923 Western Australia. Financial support for the AOCS was provided by the United States Army Medical
924 Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer
925 Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of
926 Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council
927 of Australia (NHMRC; 400413, 400281, 199600). G.C.T. and P.W. are supported by the NHMRC. RB
928 was a Cancer Institute NSW Clinical Research Fellow. LMBC is supported by the 'Stichting tegen
929 Kanker'. DL is supported by the FWO. The MABCS study is funded by the Research Centre for Genetic
930 Engineering and Biotechnology "Georgi D. Efremov", MASA. The MARIE study was supported by the
931 Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg
932 Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education
933 and Research (BMBF) Germany [01KH0402]. MBCSG is supported by grants from the Italian
934 Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the
935 5/1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori,
936 according to Italian laws (INT-Institutional strategic projects “5x1000”). The MCBCS was supported
937 by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research
938 Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a

939 generous gift from the David F. and Margaret T. Grohne Family Foundation. The Melbourne
940 Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council
941 Victoria. The MCCS was further augmented by Australian National Health and Medical Research
942 Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council
943 Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the
944 Australian Institute of Health and Welfare, including the National Death Index and the Australian
945 Cancer Database. The MEC was supported by NIH grants CA63464, CA54281, CA098758, CA132839
946 and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant,
947 Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation,
948 Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201,
949 CA140286 and CA177150. The work of MTLGEBCS was supported by the Quebec Breast Cancer
950 Foundation, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast
951 Cancer" program – grant # CRN-87521 and the Ministry of Economic Development, Innovation and
952 Export Trade – grant # PSR-SIIRI-701. The NBCS has received funding from the K.G. Jebsen Centre for
953 Breast Cancer Research; the Research Council of Norway grant 193387/V50 (to A-L Børresen-Dale
954 and V.N. Kristensen) and grant 193387/H10 (to A-L Børresen-Dale and V.N. Kristensen), South
955 Eastern Norway Health Authority (grant 39346 to A-L Børresen-Dale) and the Norwegian Cancer
956 Society (to A-L Børresen-Dale and V.N. Kristensen). The NBHS was supported by NIH grant
957 R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared
958 Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family
959 Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grants
960 U01CA164920 and U01CA167551 from the USA National Cancer Institute of the National Institutes
961 of Health. The content of this manuscript does not necessarily reflect the views or policies of the
962 National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry
963 (BCFR) or the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial
964 products, or organizations imply endorsement by the USA Government or the BCFR or CCFR. The

965 Carolina Breast Cancer Study (NCBCS) was funded by Komen Foundation, the National Cancer
966 Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer
967 Research Fund. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19
968 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. The OBCS
969 was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland
970 (grant number 250083, 122715 and Center of Excellence grant number 251314), the Finnish Cancer
971 Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support
972 Foundation and the special Governmental EVO funds for Oulu University Hospital-based research
973 activities. The ORIGO study was supported by the Dutch Cancer Society (RUL 1997-1505) and the
974 Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS was
975 funded by Intramural Research Funds of the National Cancer Institute, Department of Health and
976 Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of
977 the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is
978 supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics
979 and supported by contracts from the Division of Cancer Prevention, National Cancer Institute,
980 National Institutes of Health. PROCAS is funded from NIHR grant PGfAR 0707-10031. The RBCS was
981 funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The SASBAC study was
982 supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR),
983 the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The
984 SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now
985 Tissue Bank. SEARCH is funded by Cancer Research UK [C490/A10124, C490/A16561] and supported
986 by the UK National Institute for Health Research Biomedical Research Centre at the University of
987 Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the
988 East of England through the Clinical Academic Reserve. The Sister Study (SISTER) is supported by the
989 Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-
990 ES044005 and Z01-ES049033). The SMC is funded by the Swedish Cancer Foundation and the

Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The SZBCS was supported by Grant PBZ_KBN_122/P05/2004 and the program of the Minister of Science and Higher Education under the name "Regional Initiative of Excellence" in 2019-2022 project number 002/RID/2018/19 amount of financing 12 000 000 PLN. The TNBCC was supported by: a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), a grant from the Breast Cancer Research Foundation, a generous gift from the David F. and Margaret T. Grohne Family Foundation. The UCIBCS component of this research was supported by the NIH [CA58860, CA92044] and the Lon V Smith Foundation [LVS39420]. The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA.

Supplementary information is available at the British Journal of Cancer's website.

REFERENCES

1 Chen, W. Y. Exogenous and endogenous hormones and breast cancer. *Best Pract Res Clin Endocrinol Metab* **22**, 573-585 (2008).

2 Key, T., Appleby, P., Barnes, I., Reeves, G., Endogenous, H. & Breast Cancer Collaborative, G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* **94**, 606-616 (2002).

3 Key, T. J., Appleby, P. N., Reeves, G. K., Travis, R. C., Alberg, A. J., Barricarte, A. *et al.* Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of

individual participant data from seven prospective studies. *Lancet Oncol* **14**, 1009-1019 (2013).

4 Johnson, N., Walker, K., Gibson, L. J., Orr, N., Folkerd, E., Haynes, B. *et al.* CYP3A variation, premenopausal estrone levels, and breast cancer risk. *J Natl Cancer Inst* **104**, 657-669 (2012).

5 Sood, D., Johnson, N., Jain, P., Siskos, A. P., Bennett, M., Gilham, C. *et al.* CYP3A7*1C allele is associated with reduced levels of 2-hydroxylation pathway oestrogen metabolites. *Br J Cancer* **116**, 382-388 (2017).

6 CHANG, E., SLAUNWHITE, W. R., JR. & SANDBERG, A. A. BILIARY AND URINARY METABOLITES OF 4-C14-PROGESTERONE IN HUMAN SUBJECTS*. *The Journal of Clinical Endocrinology & Metabolism* **20**, 1568-1575 (1960).

7 Blackwell, L. F., Cooke, D. G. & Brown, S. The Use of Estrone-3-Glucuronide and Pregnanediol-3-Glucuronide Excretion Rates to Navigate the Continuum of Ovarian Activity. *Front Public Health* **6**, 153 (2018).

8 Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J. *et al.* Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genet* **27**, 383-391 (2001).

9 Prescott, J., Thompson, D. J., Kraft, P., Chanock, S. J., Audley, T., Brown, J. *et al.* Genome-wide association study of circulating estradiol, testosterone, and sex hormone-binding globulin in postmenopausal women. *PLoS One* **7**, e37815 (2012).

10 Ruth, K. S., Campbell, P. J., Chew, S., Lim, E. M., Hadlow, N., Stuckey, B. G. *et al.* Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. *Eur J Hum Genet* **24**, 284-290 (2016).

11 Rodriguez-Antona, C. & Ingelman-Sundberg, M. Cytochrome P450 pharmacogenetics and cancer. *Oncogene* **25**, 1679-1691 (2006).

12 Perera, M. A. The missing linkage: what pharmacogenetic associations are left to find in CYP3A? *Expert Opin Drug Metab Toxicol* **6**, 17-28 (2010).

13 Schuetz, J. D., Beach, D. L. & Guzelian, P. S. Selective expression of cytochrome P450 CYP3A mRNAs in embryonic and adult human liver. *Pharmacogenetics* **4**, 11-20 (1994).

14 Gonzalez, F. J. The molecular biology of cytochrome P450s. *Pharmacol Rev* **40**, 243-288 (1988).

15 Johnson, N., De Ieso, P., Migliorini, G., Orr, N., Broderick, P., Catovsky, D. *et al.* Cytochrome P450 Allele CYP3A7*1C Associates with Adverse Outcomes in Chronic Lymphocytic Leukemia, Breast, and Lung Cancer. *Cancer Res* 10.1158/0008-5472.CAN-15-1410 (2016).

16 Swerdlow, A. J., Jones, M. E., Schoemaker, M. J., Hemming, J., Thomas, D., Williamson, J. *et al.* The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. *Br J Cancer* **105**, 911-917 (2011).

17 Johnson, N., Fletcher, O., Naceur-Lombardelli, C., dos Santos Silva, I., Ashworth, A. & Peto, J. Interaction between CHEK2*1100delC and other low-penetrance breast-cancer susceptibility genes: a familial study. *Lancet* **366**, 1554-1557 (2005).

18 Walker, K., Fletcher, O., Johnson, N., Coupland, B., McCormack, V. A., Folkerd, E. *et al.* Premenopausal mammographic density in relation to cyclic variations in endogenous sex hormone levels, prolactin, and insulin-like growth factors. *Cancer Res* **69**, 6490-6499 (2009).

19 Moss, S. M., Cuckle, H., Evans, A., Johns, L., Waller, M., Bobrow, L. *et al.* Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* **368**, 2053-2060 (2006).

20 Borner, U., Szasz, G., Bablok, W. & Busch, E. W. [A specific fully enzymatic method for creatinine: reference values in serum (author's transl)]. *J Clin Chem Clin Biochem* **17**, 679-682 (1979).

21 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).

22 Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).

23 Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).

24 Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913 (2007).

25 Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-361, 361e351-352 (2013).

26 Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).

27 Azzato, E. M., Greenberg, D., Shah, M., Blows, F., Driver, K. E., Caporaso, N. E. *et al.* Prevalent cases in observational studies of cancer survival: do they bias hazard ratio estimates? *Br J Cancer* **100**, 1806-1811 (2009).

28 Mauriac, L. & Smith, I. Aromatase inhibitors in early breast cancer treatment. *Semin Oncol* **30**, 46-57 (2003).

29 Maskarinec, G., Beckford, F., Morimoto, Y., Franke, A. A. & Stanczyk, F. Z. Association of estrogen measurements in serum and urine of premenopausal women. *Biomark Med* **9**, 417-424 (2015).

30 Wu, V. S., Kanaya, N., Lo, C., Mortimer, J. & Chen, S. From bench to bedside: What do we know about hormone receptor-positive and human epidermal growth factor receptor 2-positive breast cancer? *J Steroid Biochem Mol Biol* **153**, 45-53 (2015).

31 Collaborative Group on Hormonal Factors in Breast, C. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. *Lancet* **394**, 1159-1168 (2019).

32 Williams, J. A., Ring, B. J., Cantrell, V. E., Jones, D. R., Eckstein, J., Ruterbories, K. *et al.* Comparative metabolic capabilities of CYP3A4, CYP3A5, and CYP3A7. *Drug Metab Dispos* **30**, 883-891 (2002).

33 Lash, T. L., Lien, E. A., Sorensen, H. T. & Hamilton-Dutoit, S. Genotype-guided tamoxifen therapy: time to pause for reflection? *Lancet Oncol* **10**, 825-833 (2009).

FIGURE LEGENDS

Figure 1: Manhattan plot of single nucleotide polymorphism (SNP) associations with luteal phase urinary oestrone-3-glucuronide levels in 560 premenopausal women. $-\log_{10}$ *P*-values for SNP associations are plotted against the genomic coordinates (hg19). The red line indicates the conventionally accepted threshold for genome wide significance ($P=1 \times 10^{-8}$).

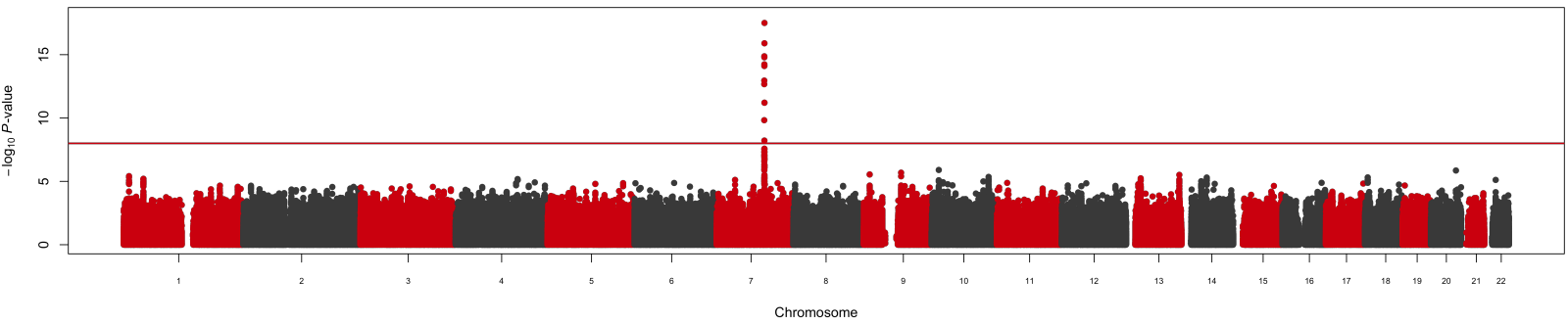
−log10 P-value
0
5
10
15
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Table 1: Association of rs45446698 with levels of oestrone-3-glucuronide, pregnanediol-3-glucuronide and progesterone in premenopausal women of European ancestry.

Hormone	Geometric mean by rs45446698 genotype (µmol/mol)		Unadjusted analysis		Adjusted analysis*	
	AA	AC/CC	% change	*P*	% change	*P*
GWAS (N=560)						
Oestrone-3-glucuronide	9.74	4.95	-49.2 (-56.1 to -41.1)	3.1×10^{-18}	-44.8 (-53.3 to -34.8)	2.1×10^{-12}
Pregnanediol-3-glucuroinde	0.78	0.70	-10.1 (-22.5 to 4.3)	0.16	-9.3 (-22.9 to 6.8)	0.24
Follow-up progesterone analysis (N=298)						
Oestrone-3-glucuronide	9.77	4.39	-55.0 (-63.1 to -45.1)	2.6×10^{-15}	-55.1 (-63.6 to -44.7)	4.0×10^{-14}
Pregnanediol-3-glucuroinde	0.79	0.75	-5.5 (-24.2 to 17.7)	0.61	-9.6 (-27.9 to 13.3)	0.38
Progesterone	29.37	21.51	-26.7 (-39.4 to -11.6)	0.001	-24.3 (-37.6 to -8.2)	0.005

* analysis was adjusted for age at menarche (<12, 12, 13, 14, >14 years), age at collection of urine samples (<35, 35-40, ≥40 years), body mass index (<18.5, 18.5 - <20.0, 20.0 - <25.0, 25.0 - <30.0, ≥ 30.0 kg/m^2) and parity (0, 1, 2, ≥3 live births).

Table 2: Association of rs45446698 among women of European ancestry overall and stratified by hormone receptor status.

	iCOGS				OncoArray				Combined			
	Cases	**Controls**	**OR (95% CI)**	P_1	**Cases**	**Controls**	**OR (95% CI)**	P_1	**Cases**	**Controls**	**OR (95% CI)**	P_1
All subjects	36 859	37 320	0.93 (0.87 to 0.98)	0.01	54 057	52 573	0.95 (0.91 to 1.00)	0.05	90 916	89 893	0.94 (0.91 to 0.98)	0.002 P_{het}=0.58
ER+	19 950	28 820	0.90 (0.84 to 0.97)	0.007	28 478	40 223	0.92 (0.87 to 0.98)	0.01	48 428	69 043	0.91 (0.87 to 0.96)	0.0002
ER-	4 298	28 820	1.04 (0.92 to 1.18)	0.56	6 592	40 223	1.03 (0.92 to 1.14)	0.64	10 890	69 043	1.03 (0.95 to 1.11)	0.50
NK	5 087				8 380							
Total	29 335	28 820		P_{int}=0.06	4 450	40 223		P_{int}=0.19				P_{int}=0.03
PR+	13 995	28 820	0.85 (0.78 to 0.93)	0.0002	21 500	40 223	0.87 (0.81 to 0.93)	0.0001	35 495	69 043	0.86 (0.82 to 0.91)	5.8×10^{-8}
PR-	6 154	28 820	1.06 (0.95 to 1.18)	0.33	10 058	40 223	1.05 (0.96 to 1.15)	0.26	16 212	69 043	1.05 (0.98 to 1.12)	0.18
NK	9 186				11 892							
Total	29 335	28 820		P_{int}=0.001	43 450	40 223		P_{int}=0.0004				P_{int}=1.3×10^{-6}
ER+, PR+	13 508	28 820	0.85 (0.78 to 0.93)	0.0003	20 624	40 223	0.87 (0.81 to 0.93)	0.0001	34 132	69 043	0.86 (0.82 to 0.91)	6.9×10^{-8}
ER+, PR-	2 890	28 820	1.03 (0.89 to 1.20)	0.66	4 597	40 223	1.07 (0.96 to 1.21)	0.21	7 487	69 043	1.06 (0.96 to 1.16)	0.25
NK	12 937				18 229							
Total	29 335	28 820		P_{int}=0.02	43 450	40 223		P_{int}=0.001				P_{int}=0.0001

P_1 = test of H_0 no association between rs45446698 and breast cancer risk. P_{int} = test of H_0 no difference between stratum specific estimates; P_{het} = test of H_0 no difference between iCOGS and OnocoArray data. NK = not known. Studies with less than 50 cases in any stratum were excluded from the stratified analyses leaving 16 studies for analysis in iCOGS data and 32 studies for analysis in OncoArray data.

Table 3: Association of rs45446698 genotype with ER+/PR+ breast cancer risk among women of European ancestry stratified by current use of postmenopausal hormone treatment.

	iCOGS				OncoArray				Combined			
	Cases	**Controls**	**OR (95% CI)**	P_1	**Cases**	**Controls**	**OR (95% CI)**	P_1	**Cases**	**Controls**	**OR (95% CI)**	P_1
MHT-	3 742	8 902	0.73 (0.62-0.86)	0.0002	5 961	15 128	0.95 (0.83-1.08)	0.45	9 703	24 030	0.86 (0.78-0.95)	0.003
MHT+	1 593	2 859	0.77 (0.59-0.99)	0.04	2 823	5 529	0.87 (0.71-1.06)	0.16	4 416	8 388	0.83 (0.70-0.97)	0.02
NK	622	1 793			3 090	6 226			3 712	8 019		
Total	5 957	13 554		P_{int}=0.81	11 874	26 883		P_{int}=0.26	17 831	40 437		P_{int}= 0.47
EPT-	3 736	7 826	0.77 (0.65-0.91)	0.002	4 173	8 566	0.93 (0.80-1.09)	0.38	7 909	16 392	0.85 (0.76-0.95)	0.005
EPT+	727	944	0.71 (0.48-1.05)	0.09	878	1 170	0.66 (0.45-0.97)	0.03	1 605	2 114	0.68 (0.52-0.90)	0.007
NK	1 494	4 784			6 823	17 147			8 317	21 931		
Total	5 957	13 554		P_{int}=0.72	11 874	26 883		P_{int}=0.09	17 831	40 437		P_{int}=0.15
ET-	3 840	7 710	0.75 (0.63-0.88)	0.0005	4 343	8 136	0.92 (0.79-1.07)	0.29	8 183	15 846	0.88 (0.79-0.97)	0.01
ET+	589	1 172	0.69 (0.45-1.06)	0.09	640	1 484	0.94 (0.62-1.42)	0.76	1 229	2 656	0.84 (0.64-1.10)	0.21
NK	1 528	4 672			6 891	17 263			8 419	21 935		
Total	5 957	13 554		P_{int}=0.83	11 874	26 883		P_{int}=0.85	17 831	40 437		P_{int}=0.78

MHT=menopausal hormone treatment, EPT=oestrogen-progesterone therapy, ET=oestrogen only therapy. P_1 = test of H_0 no association between rs45446698 and ER+/PR+ breast cancer risk; P_{int} = test of H_0 no difference between stratum specific estimates. NK = not known. Studies with less than 50 cases in any stratum were excluded from the stratified analyses leaving 13 studies for analysis in iCOGS data and 27 studies for analysis in OncoArray data. All models are adjusted for reference age, study, ten principal components, and former use of MHT. Additionally, when stratified by EPT or ET, models are

adjusted for use of any another type of MHT other than the one of interest. Further adjusting for age at menarche (<12, 12, 13, 14, >14), parity (0, 1, 2, ≥3 live births), BMI (<18.5, 18.5 - <20.0, 20.0 - <25.0, 25.0 - <30.0, ≥ 30.0 kg/m^2) did not alter these results.

Table 4. Association of rs45446698 with breast cancer specific survival in breast cancer cases of European Ancestry stratified by treatment regimen.

Group	iCOGS				OncoArray				Combined		
	Cases	Events	HR (95% CI)	P_1	Cases	Events	HR (95% CI)	P_1	HR (95% CI)	P_1	P_{het}
All breast cancer patients	32 743	2 580	0.93 (0.80 to 1.08)	0.35	58 796	3 799	1.04 (0.92 to 1.17)	0.57	0.99 (0.91 to 1.09)	0.90	0.28
Only patients that:											
Received Tamoxifen	9 766	825	1.22 (0.95 to 1.57)	0.13	7 803	746	0.95 (0.73 to 1.23)	0.68	1.08 (0.90 to 1.30)	0.41	0.18
Received Aromatase Inhibitor	3 794	246	0.94 (0.58 to 1.54)	0.82	5 460	247	1.03 (0.64 to 1.65)	0.91	0.99 (0.70 to 1.39)	0.94	0.81
Received CMF-like CT	919	99	0.30 (0.09 to 1.01)	0.05	1 692	229	0.88 (0.55 to 1.41)	0.60	0.77 (0.50 to 1.19)	0.24	0.11
Received Taxanes*	1 806	160	1.69 (0.96 to 2.99)	0.07	3 836	299	1.37 (0.96 to 1.96)	0.08	1.46 (1.08 to 1.97)	0.01	0.54
Received Anthracycline therapy	4 625	418	1.21 (0.83 to 1.75)	0.32	6 740	771	1.07 (0.82 to 1.38)	0.63	1.11 (0.90 to 1.37)	0.33	0.58

CMF=Cyclophosphamide Methotrexate Fluorouracil; CT=Chemotherapy. P_1 = test of H_0 no association between rs45446698 and breast cancer specific survival; P_{het} = test of H_0 no difference across genotyping platforms. 38 studies from iCOGS and 63 studies from OnocoArray provided follow up data for analysis of breast cancer specific survival. Results were censored at 10 years after diagnosis. HR for association of rs45446698 genotype with breast cancer-specific survival were estimated using Cox proportional hazards regression stratified by country. * To test for statistical interaction between rs45446698 genotype and treatment with a taxane we additionally compared the association in cases who received chemotherapy including a taxane to that in cases who received chemotherapy that did not include a taxane (P_{int}=0.02; the association in the latter group was in the opposite direction and not significant HR=0.88, 95% CI 0.67-1.15, P=0.34).