**ARTICLE**

**Epidemiology**

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

Nichola Johnson et al.

**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 the 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%, \(P = 3.1 \times 10^{-18}\)); 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, \(P = 6.9 \times 10^{-4}\)).

**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.

**British Journal of Cancer** (2021) 124:842–854; https://doi.org/10.1038/s41416-020-01185-w

---

**BACKGROUND:** Epidemiological studies provide strong evidence for a role of endogenous hormones in the aetiology of breast cancer. 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; there was no evidence that premenopausal progesterone levels were associated with breast cancer risk. We have previously screened 642 SNPs tagging 42 genes involved in sex steroid synthesis or metabolism, and tested for the 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. Oestrone-3-glucuronide and pregnanediol-3-glucuronide are urinary metabolites of oestrogen and progesterone, respectively, that are used in the context of reproductive medicine to monitor ovarian activity. 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. 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. A genome-wide association study (GWAS) of postmenopausal plasma oestradiol levels found no association at this locus. 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.

The CYP3A genes (CYP3A5, CYP3A7 and CYP3A4) encode enzymes that metabolise a diverse range of substrates; in addition to a role in the oxidative metabolism of hormones, CYP3A enzymes metabolise ~50% of all clinically used drugs, including many of the agents used in treating cancer. 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. In CYP3A7*1C carriers, a region within...
the foetal CYP3A7 promoter has been replaced with the equivalent region from the adult CYP3A4 gene; 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. 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. 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. Full details of the British Breast Cancer Study have been published previously. Briefly, the British Breast Cancer Study is a national case–control study of breast cancer, in which cases of breast cancer were ascertained through the cancer registries of England and Scotland and through the National Cancer Research Network. Cases were asked to invite a healthy female first-degree relative with no history of cancer and a female friend or non-blood relative to participate in the study.

Mammography Oestrogens and Growth Factors study. Full details of the Mammography Oestrogens and Growth Factors study have been published previously. Briefly, this is an observational study nested within a trial of annual mammography screening in young women that was conducted in Britain. Approximately 54,000 women aged 39–41 years were randomly assigned to the intervention arm from 1991 to 1997 and offered annual mammograms until age 48 years. From 2000 to 2003, women in the intervention arm who were still participating in this trial were invited to participate in the Mammography Oestrogens and Growth Factors study; they were asked to provide a blood sample and complete a questionnaire detailing demographic, lifestyle and reproductive factors. More than 8000 women were enrolled in the study.

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

Measurement of hormone levels

The protocol for collecting timed urine samples has been published previously. 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 7 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 the 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 µmol/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 mmol/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 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 <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. In all, 96-well plates (Greiner Bio-One GmbH, Frickenhausen, Germany) were coated with 100 µl of 5 µg/ml GAM (Arbor Assays, Ann Arbor, USA) in ELISA coating buffer (100 mM Na Bicarbonate, pH 9.6) covered and incubated in a fridge at 4°C overnight. Before use, the plates were washed three times with wash buffer 0.05 M Tris/HCl and 0.05% Tween 20, pH 7.4 (Tween® 20, Sigma-Aldrich, Inc., St. Louis, MO, USA). Standards, samples and controls (20 µl per well) were added to each well, followed by 80 µl of progesterone 3-HRP conjugate (Astra Biotech GmbH, Berlin, Germany) at 1:10,000 in assay buffer with 50 µl of monochloroprogesterone Ab (Astra Biotech GmbH, Berlin, Germany) 1:50,000 in assay buffer. Plates were incubated at room temperature for 2 h on a microtitre plate shaker (IKA®, Schuttler MTS4, IKA Labortechnik, Staufen, Germany), then washed five times with assay wash buffer and 120 µl 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 80 µl of 2 N H2SO4 solution. In all, 96-well plates (Greiner Bio-One GmbH, Frickenhausen, Germany) were coated with 100 µl of 5 µg/ml GAM (Arbor Assays, Ann Arbor, USA) in ELISA coating buffer (100 mM Na Bicarbonate, pH 9.6) covered and incubated in a fridge at 4°C overnight. Before use, the plates were washed three times with wash buffer 0.05 M Tris/HCl and 0.05% Tween 20, pH 7.4 (Tween® 20, Sigma-Aldrich, Inc., St. Louis, MO, USA). Standards, samples and controls (20 µl per well) were added to each well, followed by 80 µl 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 monochloroprogesterone Ab (Astra Biotech GmbH, Berlin, Germany) 1:50,000 in assay buffer. Plates were incubated at room temperature for 2 h on a microtitre plate shaker (IKA®, Schuttler MTS4, IKA Labortechnik, Staufen, Germany), then washed five times with assay wash buffer and 120 µl 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 80 µl of 2 N H2SO4 solution (Sigma-Aldrich Company Ltd., Dorset, UK). Finally, the plates were read on a plate reader at 450 nm. Standard curves were prepared with a total of eight different concentrations (16, 8, 4, 2, 1, 0.5, 0.25 and 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 500 K 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 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 \( P < 1 \times 10^{-5} \).

Following QC, 487,659 SNPs were successfully genotyped in 560 samples (Generations Study: \( N = 179 \), British Breast Cancer Study: \( N = 278 \) and 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. Imputation was performed using IMPUTE2. Imputed SNPs with INFO scores <.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 370-bp DNA region (chr7: 99 333 114; GRCh37/hg19) was amplified using Phusion High-Fidelity DNA Polymerase (New England Biolabs, UK) and primers CCATAGACAAAGGGAGA (forward) and CTGAGTCTTTTTTTCAG-CAGC (reverse). The PCR product was purified using QIAquick Gel Extraction Kit (Qiagen) and Sanger sequenced using a commercially available service (Europins 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 v2.2. Test statistic inflation was assessed visually using a QQ Plot (Supplementary Fig. S1) and formally by calculating the inflation factor, \( \lambda \). There was no evidence of systematic test statistic inflation (\( \lambda = 1.01 \) 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 and >14 years), age at collection of urine samples (<35, 35–40 and ≥40 years), body mass index (BMI: <18.5, 18.5–<20.0, 20.0–<25.0, 25.0–<30.0 and ≥30.0 kg/m²) and parity (0, 1, 2 and ≥3 live births).

Follow-up genotyping of rs45446698
Genotype data for rs45446698 were generated as part of iCOGS and OncoArray. Full details of SNP selection, array design, genotyping and post-genotyping QC have been published.

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 (CPS-II, 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 exclusion, the call rate for rs45446698 in OncoArray data was 99.66% and there was no evidence of deviation from Hardy–Weinberg equilibrium in controls (Supplementary Table S1). In iCOGS data, rs45446698 was imputed using 1KGP Phase 3 reference data (info score = 0.94); we used gene dosages (≤0.2 = 0.4) and ≤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 (Supplementary Tables S1 and S2).

We combined heterozygote and rare homozygote genotypes and estimated carrier ORs using logistic regression, adjusted for 15 principal components and study. Stratum-specific carrier ORs were estimated for a set of pre-specified prognostic variables (oestrogen receptor (ER), progesterone receptor (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 (ER, PR, Her2, stage and grade). In the subset of studies for which covariate data were 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 (50–56 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 the 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 the 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
In total, 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 the 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. Follow-up was right-censored at the date of death (death known to be due to breast cancer considered an event), the 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 rs4544698 genotype with breast cancer-specific survival were estimated using Cox proportional hazards regression implemented in the R package survival (v. 2.43–3) stratified by country. iCOGS and OncoArray estimates were combined using an inverse-variance-weighted meta-analysis.

RESULTS
We tested 8,280,353 autosomal SNPs for association with luteal-phase creatinine-adjusted oestrogen-3-glucuronide and pregnenadiol-3-glucuronide in 560 premenopausal women. For oestrogen-3-glucuronide, we identified a single peak mapping to the CYP3A locus at chromosome 7q22.1 (Fig. 1 and Supplementary Table S3); conditioning on any of the top SNPs, there were no additional independent signals. Four of the SNPs that were significant at \( P < 1 \times 10^{-8} \) comprise part of the seven-SNP CYP3A7*1C allele, including the top, directly genotyped SNP, rs45446698 (Supplementary Table S3). The rare rs45446698-C allele (MAF = 0.035) was associated with a 49.2% reduction in luteal-phase oestrogen-3-glucuronide (95% CI –24.2% to +17.7%, \( P = 0.61 \)) in this group of women, the rs45446698-C allele was 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 the association between rs45446698 and breast cancer risk, we combined genotype data from 56 studies (OncoArray; Supplementary Table S1) with imputed data from 35 studies (iCOGS; Supplementary 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, evidence that the association with ER+/PR+ cancers (HER2+) breast cancers (OR = 0.91, 95% CI 0.87–0.96, \( P = 0.002 \) 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, \( P = 6.9 \times 10^{-3} \); 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 (Supplementary 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, \( P = 0.0001 \); \( P_{int} = 0.28 \)).

On the assumption that genetic variants that influence metabolism of endogenous hormones 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, 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, we tested for the association between rs45446698 genotype and 10-year breast cancer-specific survival in 91,539 breast cancer cases from 71 studies.

Fig. 1 Manhattan plot of single-nucleotide polymorphism (SNP) associations with luteal-phase urinary oestrogen-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 (\( P = 1 \times 10^{-8} \)).
for whom follow-up data were available. There was neither 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 (Supplementary Table S6). Stratifying by treatment regimen, we found no evidence that rs45446698 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 rs45446698) 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. 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 (Fig. 1; rs45446698 $P = 3.1 \times 10^{-18}$; all other signals $P > 1 \times 10^{-8}$) 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. This lack of replication may be explained by our choice of study population. The first GWAS was conducted in postmenopausal women (N = 1623) participating in the Nurses’ Health Study and the Sisters in Breast Screening Study. The second was conducted within the Twins UK study (N = 2913) and included men as well as pre-, peri- and postmenopausal women. A strength of 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. For each woman, we assayed a single urine sample taken in the mid-luteal phase of her cycle at exactly 7 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.

**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 × 10⁻⁸ | −44.8 (−53.3 to −34.8) | 2.1 × 10⁻¹² |
| Pregnanediol-3-glucuronide | 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-glucuronide | 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 and >14 years), age at collection of urine samples (<35, 35–40 and ≥40 years), body mass index (<18.5, 18.5–<20.0, 20.0–<25.0, 25.0–<30.0 and ≥30.0 kg/m²) and parity (0, 1, 2 and ≥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) | P₁ | Cases | Controls | OR (95% CI) | P₁ | Cases | Controls | OR (95% CI) | P₁ |
| All subjects | 36,859 | 37,320 | 0.93 (0.87–0.98) | 0.01 | 54,057 | 52,573 | 0.95 (0.91–1.00) | 0.05 | 90,916 | 89,893 | 0.94 (0.91–0.98) | 0.002 |
| ER+ | 19,950 | 28,820 | 0.90 (0.84–0.97) | 0.007 | 28,478 | 40,223 | 0.92 (0.87–0.98) | 0.01 | 48,428 | 69,043 | 0.91 (0.87–0.96) | 0.0002 |
| ER− | 4298 | 28,820 | 1.04 (0.92–1.18) | 0.56 | 6592 | 40,223 | 1.03 (0.92–1.14) | 0.64 | 10,890 | 69,043 | 1.03 (0.95–1.11) | 0.50 |
| NK | 5087 | 8380 | | | | | | | |
| PR+ | 29,335 | 28,820 | | 0.06 | 4450 | 40,223 | | 0.19 | 35,495 | 69,043 | 0.86 (0.82–0.91) | 5.8 × 10⁻⁶ |
| PR− | 13,995 | 28,820 | 0.85 (0.78–0.93) | 0.0002 | 21,500 | 40,223 | 0.87 (0.81–0.93) | 0.0001 | 16,212 | 69,043 | 1.05 (0.98–1.12) | 0.18 |
| NK | 9186 | 11,892 | | | | | | | |
| ER+, PR+ | 13,508 | 28,820 | 0.85 (0.78–0.93) | 0.0003 | 20,624 | 40,223 | 0.87 (0.81–0.93) | 0.0001 | 34,132 | 69,043 | 0.86 (0.82–0.91) | 6.9 × 10⁻⁶ |
| ER−, PR− | 2890 | 28,820 | 1.03 (0.89–1.20) | 0.66 | 4597 | 40,223 | 1.07 (0.96–1.21) | 0.21 | 7487 | 69,043 | 1.06 (0.96–1.16) | 0.25 |
| NK | 12,937 | 18,229 | | | | | | | |
| Total | 29,335 | 28,820 | | 0.02 | 43,450 | 40,223 | | 0.0001 | 41,192 | 69,043 | 1.05 (0.98–1.12) | 0.18 |

$P₁$ test of $H₀$ no association between rs45446698 and breast cancer risk, $P_{\text{het}}$ test of $H₀$ no difference between stratum-specific estimates, $P_{\text{int}}$ test of $H₀$ no difference between iCOGS and OncoArray 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.
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 pregnenediol-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, and in a recent study that measured luteal-phase serum oestrogens and urinary oestrogen metabolites in 249 premenopausal women, serum oestradiol and oestrone were only moderately correlated with urinary oestrone (serum oestradiol: \( r = 0.39 \), serum oestrone: \( r = 0.48 \)). Our analysis of rs45446698 genotypes in 90,916 cases and \( 1.06 \) = 0.16 7803 746 0.95 (0.73–1.23) 0.68 1.08 (0.90–1.30) 0.41 0.18

Received aromatase inhibitor 3794 246 0.94 (0.58–1.54) 0.82 5460 247 1.03 (0.64–1.65) 0.91 0.99 (0.70–1.39) 0.94 0.81

Received CMF-like CT 919 99 0.30 (0.09–1.01) 0.05 1692 229 0.88 (0.55–1.41) 0.60 0.77 (0.50–1.19) 0.24 0.11

Received taxanes* 1806 160 1.69 (0.96–2.99) 0.07 3836 299 1.37 (0.96–1.96) 0.08 1.46 (1.08–1.97) 0.01 0.54

Received anthracycline therapy 4625 418 1.21 (0.83–1.75) 0.32 6740 771 1.07 (0.82–1.38) 0.63 1.11 (0.90–1.37) 0.33 0.58

CMF cyclophosphamide methotrexate fluorouracil; CT chemotherapy; \( P_{\text{int}} \) test of \( H_0 \) no difference across genotyping platforms.

In total, 38 studies from iCOGS and 63 studies from OncoArray provided follow-up data for analysis of breast cancer-specific survival. The results were censored at 10 years after diagnosis. HR for association of rs45446698 genotype with breast cancer-specific survival was 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 (\( P_{\text{int}} = 0.02 \); the association in the latter group was in the opposite direction and not significant: \( HR = 0.88, 95\% \text{ CI} 0.67–1.15, P = 0.34 \)).

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.

| Genotype | iCOGS Cases | iCOGS Control | OncoArray Cases | OncoArray Control | Combined Cases | Combined Control | \( P_{\text{int}} \) |
|----------|-------------|---------------|----------------|-------------------|----------------|-----------------|--------------|
| MHT−     | 3742        | 8902          | 0.73 (0.62–0.86) | 0.0002 | 5961 | 15,128 | 0.95 (0.83–1.08) | 0.45 | 9703 | 24,030 | 0.86 (0.78–0.95) | 0.003 |
| MHT +    | 1593        | 2859          | 0.77 (0.59–0.99) | 0.04 | 2823 | 5529 | 0.87 (0.71–1.06) | 0.16 | 4416 | 8388 | 0.83 (0.70–0.97) | 0.02 |
| NK       | 622         | 1793          | | | 3009 | 6226 | | | 3712 | 8019 | |
| Total    | 5957        | 13,554        | | | 11,874 | 26,883 | | | 17,831 | 40,437 | 0.07 |

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

| Group | iCOGS Cases | iCOGS Events | HR (95% CI) | \( P_{\text{int}} \) |
|-------|-------------|--------------|-------------|--------------|
| All breast cancer patients | 32,743 | 2580 | 0.93 (0.80–1.08) | 0.35 |
| Only patients that received taxamofin | 9766 | 825 | 1.22 (0.95–1.57) | 0.13 |
| Only patients that received aromatase inhibitor | 3794 | 246 | 0.94 (0.58–1.54) | 0.82 |
| Only patients that received CMF-like CT | 919 | 99 | 0.30 (0.09–1.01) | 0.05 |
| Only patients that received taxanes* | 1806 | 160 | 1.69 (0.96–2.99) | 0.07 |
| Only patients that received anthracycline therapy | 4625 | 418 | 1.21 (0.83–1.75) | 0.32 |

\( \text{CMF cyclophosphamide methotrexate fluorouracil; CT chemotherapy; } P_{\text{int}} \text{ test of } H_0 \text{ no difference across genotyping platforms.} \)

In total, 38 studies from iCOGS and 63 studies from OncoArray provided follow-up data for analysis of breast cancer-specific survival. The results were censored at 10 years after diagnosis. HR for association of rs45446698 genotype with breast cancer-specific survival was 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 \( P_{\text{int}} = 0.02 \); the association in the latter group was in the opposite direction and not significant: HR = 0.88, 95% CI 0.67–1.15, \( P = 0.34 \).

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 pregnenediol-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, and in a recent study that measured luteal-phase serum oestrogens and urinary oestrogen metabolites in 249 premenopausal women, serum oestradiol and oestrone were only moderately correlated with urinary oestrone (serum oestradiol: \( r = 0.39 \), serum oestrone: \( r = 0.48 \)). Our analysis of rs45446698 genotypes in 90,916 cases and 89,893 controls from BCAC, however, provides robust evidence of an association of the CYP3A7*1C allele with breast cancer risk overall (OR = 0.94, \( P = 0.002 \)) and a more pronounced protective effect on ER+ /PR+ breast cancers (OR = 0.86, \( P = 6.9 \times 10^{-8} \)). The specificity of this association (comparing ER+/PR+ with ER+/PR+ cancers, \( P_{\text{int}} = 0.001 \) and our replication of Ruth and colleagues report of a signal at the CYP3A locus in their analysis of circulating progesterone levels raise the possibility that premenopausal progesterone levels might influence risk of ER+ /PR+ breast cancers. This would be in contrast to the findings from Key et al. who reported no evidence of an association between premenopausal progesterone levels and...
breast cancer risk overall and no heterogeneity in estimates stratified by PR status. However, the number of cases of PR+ (N = 158) and PR− (N = 61) breast cancer was small, and this analysis may have lacked power to detect modest associations in subgroups of cancers. Alternatively, the association of rs45446698 genotype with ER + PR + breast cancer risk, specifically, may be due to the fact that PR is a marker for an intact oestrogen signalling pathway confirming a direct link between the levels of oestrogen (or oestrogen signalling) and proliferation in this subgroup of cancers.

Our analysis of the CYP3A7*1C allele, menopausal hormone treatment and breast cancer risk was inconclusive; while the carrier ORs were consistent with a greater protective effect of this allele in women taking exogenous hormones, particularly oestrogen–progesterone therapy, none of the interactions was statistically significant. Overall, there were 14,119 ER + PR + breast cancer cases and 32,418 controls for this subgroup analysis, but for what was, arguably, the most pertinent subgroup (i.e., current oestrogen–progesterone therapy use), the number of cases who were current users was relatively small (CYP3A7*1C carriers N = 107, non-carriers N = 1498) and power was limited to detect modest interactions. There are limitations to this analysis; we focussed on current menopausal hormone treatment use (adjusted for past use) as it is for current use that the association with breast cancer risk is the strongest, 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 CYP3A4 substrate. This may reflect the fact that compared to CYP3A4, CYP3A7 is a poor metaboliser of tamoxifen, or that standard doses of tamoxifen achieve high levels of oestrogen receptor saturation. 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 are 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 progesterogens 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.
CYP3A7*1C allele: linking premenopausal oestrogen... N Johnson et al.
Vigo, Spain. BSUCH thanks Peter Bugert, Medical Faculty Mannheim. CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. CCGBP thanks Styliani Apostolaki, Anna Margioulakis, Georgios Ntios, Maria Perraki, Georgia Saloustrou, Georgiia Sevastaki, Konstantinos Pompidakis. CGS thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance from Maria Schünke, Maria Brandt, and Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. CNIO-BCS thanks Guillermo Pita, Chao Alfonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO). Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They acknowledge the contribution to this study from cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. The authors would like to thank the California Teachers Study Steering Committee that is responsible for the formation and maintenance of the Study within which this research was conducted. A full list of California Teachers Study team members is available at https://www.calteachersstudy.org/team. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks Hartwig Ziegler, Sonja Wolf, Volker Herrmann, Christina Stegmaier, Katja Butterbach. FHRISK thanks NIHR for funding. GC-HBC thanks Stefanie Engert, Heide Hitzebrand, Sandra Kröber and LEI - Leipzig Research Centre for Civilization Diseases (Markus Lorenf, Joachim Thiery, Matthias Nüchter, Ronny Baber). The GENICA Network. Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany (HB, Wing-Yee Lo, Reiner Hoppe), Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanner Krankenhaus, Bonn, Germany (Yon-Dschun Ko, Christian Baisch), Institute of Pathology, University of Bonn, Germany (Hans-Peter Fischer), Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany (Ute Hamann), Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany (Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz); and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany (Volker Harth). HABCs thanks Michael Bremer. HBBCS thanks Sofia Khan, Johanna Kiski, Kristina Attromlou, Rainer Fagerholm, Kismarud Aaltosen, Karl von Smitten, Ida Erkilia. HNBBCS thanks Piers Hillelmann, Hans Christiansen and Johann H. Karstens. HUBCS thanks Shamal Gantsev, KARMA and SASSBAC thank the Swedish Medical Research Counsel. KBCP thanks Eija Myöhänen, Helena Kemiläinen. KConFab/ACOS wish to thank Heather Thorne, Eveline Niedermayer, all the KConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to KConFab. LMBC thanks Giulian Petimer, Thomas Van Brussel, EvyVanderheyden and Kathleen Couthouts. MA2BCS thanks Milena Jakimovska (RGCBE “Georgi D. Efremov”), Emilija Lazarova (University Clinic of Radiology and Oncology), Katerina Kubelka-Sabit, Mirko Karadzic (Adzibadz-Sistina Hospital), Andrej Anorski and Liljana Stojanovska (Re-Medika Hospital) for their commitment and support to this study and thank Petra Sedlak, Dietmar Flesch-Janys, Judith Heinz, Nadia Obi, Alina Vrielig, Sabine Behrens, Ursula Elber, Muhbabet Celik, Til Olchers and Stefan Nickels. MBCSG (Milan Breast Cancer Study Group): Paolo Radice, Bernardo Peissel, Jacopo Azzollini, Dario Zambellati, Daniela Zaffaroni, Bernardo Bonanni, Irene Feroco, Mariarosaria Calvello, Aliana Guerrieri Gonzaga, Monica Marabelli, Davide Bondavalli and the personnel of the Cogentech Cancer Genetic Test Laboratory. The MCCS was made possible by the contributions of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. MLTLEGBCS would like to thank Martine Tranchant (CHU de Québec – Université Laval Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management and skilful technical assistance. J.S. is Chair holder of the Canada Research Chair in Oncogenetics. The following are NBCS Collaborators: Anne-Lise Berresen-Dale, Grethe J. Grenaker Alnæs, Kristine K. Sahling (PhD), Lars Ottasted (MD), Rolf Kåre Eckman, Ellen Schlichting (MD), Mari Mønun Holmén, Toril Sauerland (MD), Vilde Haakensten (MD), Olav Engebretsen (MD), Bjørn Naume (MD), Alexander Fossa (MD), Cecile E. Kiserud (MD), Kristin V. Reiret (MD), Åslaug Helland (MD), Margit Riis (MD), Jürgen Geiser (MD) and OSBREAC. NBHS thank study participants and research staff for their contributions and commitment to the studies. For NHHS and NHSS the study protocol was approved by the institutional review boards of the Brigham and Women’s Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We would like to thank the participants and staff of the NHS and NHS2 for their valuable contributions as well as the following state cancer registries for their help: 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, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. OBCS thanks Arja Jukkola-Vuorinen, Mirvii Grip, Salla Kaupilla, Meri Otti, Tarmo Kinnunen (MD) and Mämmi Märramäe (MD) for their contributions to this study. The OFBCR thanks Teresa Selander, Nayana Weerasooriya and Steve Gullanger. ORGO thanks E. Krol-Warlemann, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. The LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCOCO) with the help of Dr. J. Molenaar. PBCS thanks Louise Brinton, Mark Sherman, Neolima Szeszies-Dabrowskia, Beata Peplonska, Wiolet Zatonski, Pui Chau, Michael Stagner. PROCAS thanks NIHR for funding. The RBCS thanks Corine M. Beaufort, Jannet Blom, Renée Broeren—Foevens, Saskia Pelders, Wendy J.C. Prager—van der Smissen, Kirsten Ruigrok —Ritstter, Anita M.A.C. Trapman—Jansen, Michelle van der Vlugt—Daane, Vanja de Weerd, and the Erasmus MC Family Cancer Clinic. SBCS thanks Sue Higham, Helen Cramp, Dan Connelly, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed. We thank the SEARCH and EPIC teams. SZBCS thanks Eva Putrezecka. UCBCS thanks Irene Masunaka. UKBCS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the study participants, staff study, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

AUTHOR CONTRIBUTIONS
N.J.T., K.EJ., M.J.S., C.G., A.F.H., J.P., Ld.S.S. and A.J.S. generated the data for the GWAS reported in this paper. N.J.T., S.M., A.M., P.K.M., J.C.C., M.K.S., N.O. and O.F. analyzed the data relating to this paper and drafted the initial version. D.F.E. coordinated the BCAC and led the iCOGS and OncoArray genotyping. P.H. led the CCOG collection. J.S. led the OncoArray collaboration. M.K.B., Q.W. and J.D. coordinated the BCAC database. The remaining authors led individual studies and contributed to the design of the study, data collection and revising the paper.

ADDITIONAL INFORMATION
Ethics consent and permission to participate. Collection of blood samples, urine samples and questionnaire information was undertaken with written informed consent and relevant ethical review board approval in accordance with the tenets of the Declaration of Helsinki (Supplementary Table S7).

Consent to publish Not applicable.

Data availability GWAS data and the complete dataset for follow-up genotyping will not be made publicly available due to restraints imposed by the ethics committees of individual studies; requests for data can be made to the corresponding author (GWAS data) or the Data Access Coordination Committee (follow-up genotyping data) of BCAC (http://bcac.ccge.medschl.cam.ac.uk/). Summary results for all variants genotyped by BCAC (including rs45466698) are available at http://bcac.ccge.medschl.cam.ac.uk/.

Competing interests M.W.B. conducts research funded by Amgen, Novartis and Pfizer. P.A.F. conducts research funded by Amgen, Novartis and Pfizer. He received honoraria from Roche, Novartis and Pfizer. A.W.K.’s institution has received research funding from Myriad Genetics for an unrelated project (funding dates 2017–2019). P.H., J.P.D.P.P., O.F., K.C. and A.C. are members of the Editorial Board of BJC. The remaining authors declare no competing interests.

Funding information This work was supported by Programme Grants from Breast Cancer Now as part of Programme Funding to the Breast Cancer Now Toby Robbins 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 BRIDGE and B-CAST respectively) and by the European Community’s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (iCOGS). 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’Economie,
of Education and Research (BMWf) Germany (01KH0402). MBGCS is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment in support of the Fondazione IRRCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects “5x1000”). The MCBCS was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201) and the Breast Cancer Community Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria, which used their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Cancer Index and the Australian Cancer Database. The MEC was supported by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kramlard Foundation, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA111621, CA120826 and CA177150. The work of MTGGBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the "CHIR Team in Familial Risks of Breast Cancer" program—grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade—grant # PSR-SRI-701. The NBCS has received funding from the K.G. Jebsen Centre for Breast Cancer Research; the Research Council of Norway grant 193387/50 (to A.-L. Barresen-Dale and T.V. Kristensen) and grant 193387/110 (to A.-L. Barresen-Dale and V.N. Kristensen), South Eastern Norway Health Authority (grant 39346 to A.-L. Barresen-Dale) and the Norwegian Cancer Society (to A.-L. Barresen-Dale and V.N. Kristensen). The NBHS was supported by NIH grant R01 CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68483. The Northern California Breast Cancer Family Registry (NC-B CFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grants U01 CA164920 and U001CA167551 from the USA National Cancer Institute of the National Institutes of Health. The content of this paper does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the Breast Cancer Family Registry (BCFR) or the Colon Cancer Family Registry (CCFR) nor does mention of trade names, commercial products, or organisations imply endorsement by the USA Government or the BC FR CCFR. The Carolina Breast Cancer Study (NCBBS) was funded by Komen Foundation, the National Cancer Institute (P05 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, UM1 CA161107, and U19 CA14085. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA14805. The OBCS was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant number 250083, 122715 and Center of Excellence grant number 251314), the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation and the special Governmental EVO funds for Oulu University Hospital-based research activities. The ORIO study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL, C16). The PBCS was funded by Intramural Research Funds of the National Institutes of Health, National Cancer Institute, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES045003). The PBCS is supported by the Swedish Cancer Society via a grant from the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The S2BCS was supported by Grant PBZ_KBN_122/POS/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 NBHCS 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 UCBCS component of this research was supported by the NIH (CA58860, CA92044) and the Lon V Smith Foundation (LV539420). 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 The online version contains supplementary material available at https://doi.org/10.1038/s41416-020-01185-w.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Chen, W. Y. Exogenous and endogenous hormones and breast cancer. Best. Pr. 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., Folkers, E., Haynes, B. et al. CYP3A variation, postmenopausal estrone levels, and breast cancer risk. J. Natl. Cancer Inst. 104, 657–669 (2012).
5. Soon, D. 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. Going and urinary metabolites of 4-C14-progestin in human subjects. J. Clin. Endocrinol. Metab. 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., Assm, S., Schuetz, J. et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A expression. Nat. Genet. 27, 383–391 (2001).
9. Prescott, J., Thompson, D. J., Kraft, P., Channock, 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-Antonia, 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. Pharmacogenet. 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. 76, 1485–1493 (2016).
16. Swedlow, 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. & Petru, J. Interaction between CHEK2*I1100delC 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., Folkers, 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).
Nichola Johnson 1, Sarah Maguire 2, Anna Morra 3, Pooja Middha Kapoor 4,5,6, Katarzyna Tomczyk 1, Michael E. Jones 6, Rita K. Schmutzler 6,6,7, Christopher Scott 7, Melissa A. Troester 9,6, John L. Hopper 6,1,1, Celine M. Vachon 12,1, Elke M. van Veen 5,5, Sophia S. Wang 12,1,2, Clarice R. Weinberg 12, Camilla Wendt 9,3, Jose E. Castelao 39, Stephen J. Chanock 9, Georgia Chenevix-Trench 40, Christine L. Clarke 41, Peter Kraft 47,85, Vessela N. Kristensen 8,87, Allison W. Kurian 113,114, Peter A. Fasching 1168, Rebecca S. piled. 8,17, Sune F. Nielsen 28,29, Børge G. Nordestgaard 28,29,30, Nadia Obi 95, Andrew F. Olshan 96, Janet E. Olson 72, Håkan Olsson 16, Ester Orban 97, Minouk J. Schoemaker 1, Irene L. Andrulis 10,11, Esteban Lozano 853, Nick Orr 3, and Olivia Fletcher 1

1The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK; 2Department of Cancer Research and Cell Biology, Queen’s University Belfast, Belfast, Ireland, UK; 3Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands; 4Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; 5Faculty of Medicine, University of Heidelberg, Heidelberg, Germany; 6Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK; 7Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK; 8Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 9Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 10Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 11Department of Public Health Sciences, Cancer Research Institute, Queen’s University, Kingston, Canada; 12Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden; 13Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 14Department of Gynecology and Obstetrics, Medical Radiology, Minsk, Belarus; 15Division of Clinical Cancer Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; 16Department of Public Health Sciences, Cancer Research Institute, Queen’s University, Kingston, Canada; 17Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden; 18Department of Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 19Department of Gynecology and Obstetrics,
