

Authors and Disclosures

Nichola Johnson, Kate Walker, Lorna J. Gibson, Nick Orr, Elizabeth Folkard, Ben Haynes, Claire Palles, Ben Coupland, Minouk Schoemaker, Michael Jones, Peter Broderick, Elinor Sawyer, Michael Kerin, Ian P. Tomlinson, Marketa Zvelebil, Sarah Chilcott-Burns, Katarzyna Tomczyk, Gemma Simpson, Jill Williamson, Stephen G. Hillier, Gillian Ross, Richard S. Houlston, Anthony Swerdlow, Alan Ashworth, Mitch Dowsett, Julian Peto, Isabel dos Santos Silva and Olivia Fletcher

Breakthrough Breast Cancer Research Centre, Division of Breast Cancer Research, Institute of Cancer Research, London, UK (NJ, NO, BC, MZ, SC-B, KT, GS, JW, AA, OF); Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK (KW, LJG, JP, IdSS); The Academic Department of Biochemistry, The Royal Marsden Hospital, London, UK (EF, BH, MD); Molecular and Population Genetics, Wellcome Trust Centre for Human Genetics, Oxford, UK (CP, IPT); Oxford Comprehensive Biomedical Research Centre, Oxford, UK (CP, IPT); Division of Genetics and Epidemiology (MS, MJ, PB, RSH, AS) Institute of Cancer Research, Sutton, Surrey, UK; Biomedical Research Centre Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London, Great Maze Pond, London, UK (ES); Clinical Science Institute, University Hospital Galway, Galway, Ireland (MK); MRC Centre for Reproductive Health, The Queen's Medical Research Institute, The University of Edinburgh, Edinburgh, UK (SGH); The Royal Marsden NHS Foundation Trust, London, UK (GR)

Correspondence to

Olivia Fletcher, PhD, The Breakthrough Breast Cancer Research Centre, Division of Breast Cancer Research, Institute of Cancer Research, 237 Fulham Rd, London SW3 6JB, UK (e-mail: olivia.fletcher@icr.ac.uk).

N. Johnson and K. Walker contributed equally to this work. This study makes use of data generated by the Wellcome Trust Case–Control Consortium. A full list of the investigators who contributed to the generation of these data is available from www.wtccc.org.uk. We also used data from the Cancer Genetic Markers of Susceptibility (CGEMS) breast cancer study. A full list of the investigators who contributed to the generation of the CGEMS data is available from <http://cgems.cancer.gov/>. We are grateful to all the patients and control subjects for their participation. We thank the clinicians and other hospital staff, cancer registries, and study staff who contributed to the blood sample and data collection for the British Breast Cancer study, Breakthrough Generations Study, and Mammography Oestrogens and Growth Factors study. We thank Martha Urquhart et al. (Reproductive Medicine Laboratory, Royal Infirmary of Edinburgh) for technical support in the conduct of urinary steroid assays. The sponsors had no role in the design of the study; the collection, analysis and interpretation of the data; the writing of the article; and the decision to submit the article for publication.



From Journal of the National Cancer Institute

CYP3A Variation, Premenopausal Estrone Levels, and Breast Cancer Risk

Nichola Johnson; Kate Walker; Lorna J. Gibson; Nick Orr; Elizabeth Folkard; Ben Haynes; Claire Palles; Ben Coupland; Minouk Schoemaker; Michael Jones; Peter Broderick; Elinor Sawyer; Michael Kerin; Ian P. Tomlinson; Marketa Zvelebil; Sarah Chilcott-Burns; Katarzyna Tomczyk; Gemma Simpson; Jill Williamson; Stephen G. Hillier; Gillian Ross; Richard S. Houlston; Anthony Swerdlow; Alan Ashworth; Mitch Dowsett; Julian Peto; Isabel dos Santos Silva; Olivia Fletcher

Posted: 06/12/2012; J Natl Cancer Inst. 2012;104(9):657-669. © 2012 Oxford University Press

Abstract and Introduction

Abstract

Background Epidemiological studies have provided strong evidence for a role of endogenous sex steroids in the etiology of breast cancer. Our aim was to identify common variants in genes involved in sex steroid synthesis or metabolism that are associated with hormone levels and the risk of breast cancer in

premenopausal women.

Methods We measured urinary levels of estrone glucuronide (E1G) using a protocol specifically developed to account for cyclic variation in hormone levels during the menstrual cycle in 729 healthy premenopausal women. We genotyped 642 single-nucleotide polymorphisms (SNPs) in these women; a single SNP, rs10273424, was further tested for association with the risk of breast cancer using data from 10 551 breast cancer case patients and 17 535 control subjects. All statistical tests were two-sided.

Results rs10273424, which maps approximately 50 kb centromeric to the cytochrome P450 3A (*CYP3A*) gene cluster at chromosome 7q22.1, was associated with a 21.8% reduction in E1G levels (95% confidence interval [CI] = 27.8% to 15.3% reduction; $P = 2.7 \times 10^{-9}$) and a modest reduction in the risk of breast cancer in case patients who were diagnosed at or before age 50 years (odds ratio [OR] = 0.91, 95% CI = 0.83 to 0.99; $P = .03$) but not in those diagnosed after age 50 years (OR = 1.01, 95% CI = 0.93 to 1.10; $P = .82$).

Conclusions Genetic variation in noncoding sequences flanking the *CYP3A* locus contributes to variance in premenopausal E1G levels and is associated with the risk of breast cancer in younger patients. This association may have wider implications given that the most predominantly expressed *CYP3A* gene, *CYP3A4*, is responsible for metabolism of endogenous and exogenous hormones and hormonal agents used in the treatment of breast cancer.

Introduction

Epidemiological studies have provided strong evidence for a role of endogenous sex steroids in the etiology of breast cancer.^[1] Specifically, the risk of breast cancer is elevated among women with high postmenopausal levels of total estradiol (E2) and estrone (E1), as well as several androgenic precursors, including androstenedione, dehydroepiandrosterone (DHEA), and testosterone.^[2] In a collaborative reanalysis of nine prospective studies of endogenous sex hormones and breast cancer,^[2] the relative risks of breast cancer associated with a doubling of E2 levels and E1 levels in postmenopausal women were 1.29 (95% confidence interval [CI] = 1.15 to 1.44) and 1.46 (95% CI = 1.22 to 1.73), respectively. By contrast, there is evidence that postmenopausal levels of sex hormone-binding globulin (SHBG), a protein that binds to and restricts the biological activities of E2 and E1, are inversely associated with the risk of breast cancer.^[2]

Epidemiological studies also provide indirect evidence for an association between premenopausal hormone levels and the risk of breast cancer;^[3] they have established that risk factors for breast cancer include early age at menarche, nulliparity, older age at birth of first child, and older age at menopause. In addition, circulating levels of estrogens in premenopausal women are approximately 10 times (in the case of E2) and four to five times (in the case of E1) higher compared with those in postmenopausal women, suggesting that premenopausal estrogen levels may be more relevant with regard to breast cancer etiology. However, epidemiological studies that have examined associations between endogenous sex steroid levels and the risk of breast cancer in premenopausal women have produced inconsistent results,^[4–10] likely because the cyclical fluctuation of estrogen levels during the menstrual cycle makes capturing long-term average blood concentrations of estrogen in premenopausal women problematic.

Results of several studies^[11–13] have indicated that genetic factors influence the levels of endogenous sex steroids in both pre- and postmenopausal women. Thus, common variants in genes involved in sex steroid synthesis or metabolism are good candidates for breast cancer predisposition alleles.^[14] The statistical power to detect such associations is enhanced by measurement of intermediate quantitative phenotypes such as plasma or urinary hormone levels.^[15] Recent studies of candidate polymorphisms^[16] and tag single-nucleotide polymorphisms (SNPs)^[15,17,18] have revealed robust associations between SNPs at the aromatase (*CYP19A1*) locus and estrogen levels in postmenopausal women^[16–18] and between SNPs at the *SHBG* locus and SHBG levels in pre- and postmenopausal women.^[15,18,19] However, to our knowledge, none of these variants has been found to be associated directly with the risk of breast cancer.

To better understand the relationship between premenopausal hormone levels and the risk of breast cancer, we used a protocol that we previously developed for measuring endogenous hormone levels that takes into account the cyclic variation of estrogen and progesterone in premenopausal women, which involves measuring creatinine-adjusted urinary estrone glucuronide (E1G) and pregnanediol glucuronide (PG) levels in serial early morning urine samples on prespecified days of the menstrual cycle.^[20] The aim of this study was to identify genetic variants associated with levels of E1G, PG, plasma SHBG, and androgenic precursors in healthy premenopausal women and to examine whether these variants are associated with the risks of pre- or postmenopausal breast cancer.

Subjects and Methods

Study Subjects

Mammography, Oestrogens, and Growth Factors (MOG) Study. Full details of the MOG study have been published previously.^[20] Briefly, this is an observational study nested within a trial of annual mammography screening in young women that was conducted in Britain.^[21] 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 MOG 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. A total of 520 women from the MOG study provided urine samples for the cross-sectional analysis of hormones (Figure 1). All of the women included in this analysis reported being of Northern European ancestry, and none had been diagnosed with breast cancer at the time of study recruitment.

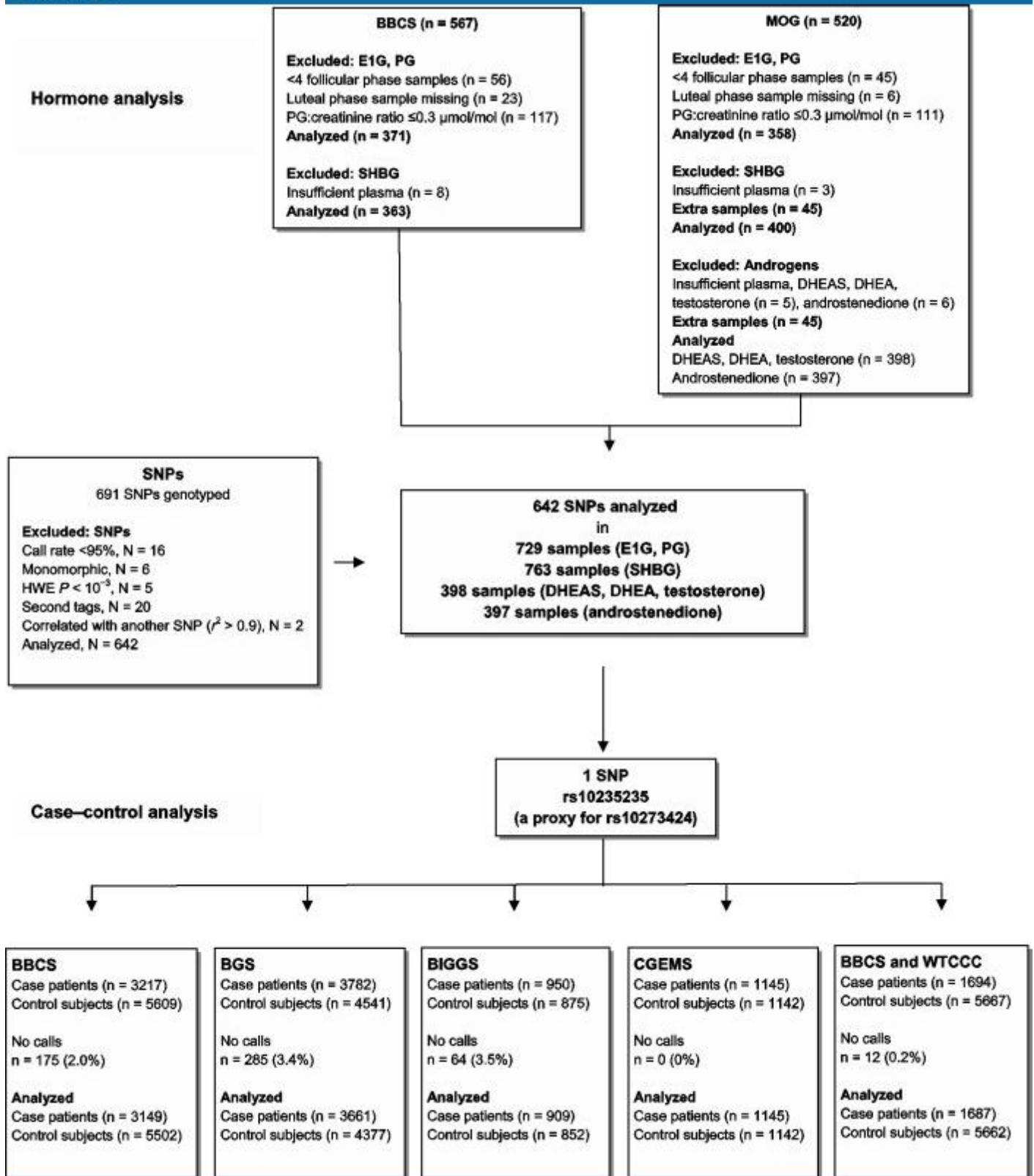


Figure 1. Sample and single-nucleotide polymorphism (SNP) exclusion schema. Numbers of samples (n) and numbers of SNPs (N) analyzed for the cross-sectional analysis of hormone levels and the case-control analysis are shown. Of 691 SNPs that were genotyped, 642 (92.9%) were tested for association with hormone levels in up to 763 samples. A single SNP, rs10235235, that tags rs10273424 ($r^2 = 1.0$) was taken forward to a case-control analysis in combined data from five case-control series totaling 10 551 breast cancer case patients and 17 535 control subjects. BBCS = British Breast Cancer study; BGS = Breakthrough Generations Study; BIGGS = Breast Cancer in Galway Genetic Study; CGEMS = Cancer Genetic Markers of Susceptibility Study; DHEA = dehydroepiandrosterone; DHEAS = dehydroepiandrosterone sulfate; E1G = creatinine-adjusted urinary estrone glucuronide; HWE = Hardy-Weinberg Equilibrium; MOG = Mammography Oestrogens and Growth Factors study; PG = creatinine-adjusted pregnanediol glucuronide; SHBG = sex hormone-binding globulin; SNPs = single-nucleotide polymorphisms; WTCCC = Wellcome Trust Case-Control Consortium.

British Breast Cancer Study (BBCS). Full details of the BBCS have been published previously.^[22] Briefly, the BBCS 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. Breast cancer patients younger than age 71 years at diagnosis (mean age at diagnosis = 52.4 years, SD = 9.2 years) were recruited between 2001 and 2010. Case patients 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. A total of 567 healthy women who were first-degree relatives, friends, or non-blood relatives of the case patients provided urine samples for the cross-sectional analysis of hormones; a total of 4911 case patients and 2045 friends and non-blood relatives of the case patients participated in the case-control analysis. All breast cancer case patients and control subjects reported being of Northern European ancestry, and all control subjects were free from breast cancer by self-report at the time of study recruitment.

Breast Cancer In Galway Genetic Study (BIGGS). BIGGS is a case-control study of breast cancer that is based in Western Ireland.^[23] Recruitment was between 1998 and 2008. Case patients were ascertained through the University College Hospital Galway, and control subjects were drawn from the same Western Irish population. The mean age at diagnosis of case patients was 52.3 years (SD = 11.3 years). All control subjects were aged 60 years or older and had no personal history of cancer and no family history of breast or ovarian cancer in a first- or second-degree relative. A total of 950 case patients and 875 control subjects from BIGGS were included in the case-control analysis (Figure 1). All breast cancer case patients and control subjects reported being of Northern European ancestry.

Breakthrough Generations Study (BGS). BGS 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.^[24] The study participants included prevalent breast cancer case patients who were diagnosed before entry into the study. Breast cancer diagnosis was self-reported. The mean age at breast cancer diagnosis was 50.8 years (SD = 9.4 years). A total of 3782 case patients and 4541 control subjects were included in the case-control analysis (Figure 1). All breast cancer case patients and control subjects reported being of Northern European ancestry and all control subjects were free from breast cancer by self-report at the time of recruitment.

Collection of blood samples and questionnaire information from all case patients and control subjects was undertaken with written informed consent and relevant ethical review board approval (MOG, BBCS, and BGS: South East Multicentre Research Ethics Committee; BIGGS Clinical Research Ethical Committee, University College Hospital, Galway, Ireland) in accordance with the tenets of the Declaration of Helsinki.

Cross-sectional Analysis of Hormone Levels

Eligible premenopausal women ascertained through the MOG study and relatives of case patients (one relative per case patient) and control subjects participating in the BBCS were invited to participate in this cross-sectional analysis by providing serial urine samples at prespecified days during their menstrual cycle. To be eligible, women had to be having regular menstrual cycles (ie, their usual cycle length had to be between 21 and 35 days), not using hormone replacement therapy or oral contraceptives, and not have been diagnosed with breast cancer at recruitment to the study. The protocol for measuring follicular and luteal phase urinary E1G and PG has been published previously.^[20] Briefly, E1G is a principal metabolite of the serum estrogens and, when measured in early morning urine samples, is highly correlated with the serum E2 curve, with a delay of 1–2 days (25,26). 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 cycle. On the basis of this estimate, women were asked to provide an early morning urine sample on

each of the 3 days immediately preceding their predicted day of ovulation, on their predicted ovulation day, and the first and second days after their predicted day of ovulation (follicular phase samples). They were also asked to provide a sample 7 days after the predicted day of ovulation (luteal phase sample). E1G and PG were measured using in-house enzyme-linked immunosorbent assays, full details of which have been published previously.^[27] Briefly, urine samples were incubated with labeled antigen (E1G [catalog number E1752] or PG [catalog number P3635]; Sigma-Aldrich, St Louis, MO) conjugated to horseradish peroxidase (catalog number P8375; Sigma-Aldrich) in the presence of rabbit antibodies against E1G and PG antibodies (RAB1 and RAB F 27/7/87, respectively; produced by the Medical Research Council/Agricultural and Food Research Council Comparative Physiology Research Group, Institute of Zoology, London, UK). Bound and free antigen were separated using solid-phase donkey anti-rabbit IgG (Scottish Antibody Production Unit, Lanarkshire, UK), and bound antigen was detected by incubation with α -phenylenediamine (catalog number P9029; Sigma-Aldrich). E1G and PG measurements were adjusted for urinary creatinine levels measured colorimetrically (creatinine color reagent, CAYM10005315-5; VWR International Ltd, Lutterworth, UK; creatinine standards C3488; Sigma-Aldrich) using the Jaffe method^[28] to account for urine concentration.

A total of 1087 eligible women (MOG [n = 520], BBCS [n = 567]) who had provided a blood sample also agreed to provide serial urine samples (Figure 1). Women who provided fewer than four follicular phase urine samples (n = 101 [9.3%]) or who did not provide a luteal phase sample (n = 29 [2.7%]) were excluded. PG was measured in luteal phase samples from the remaining 957 eligible women (88.0%; Figure 1). After measurement of PG, we excluded samples from women who had a luteal phase PG:creatinine ratio less than or equal to 0.3 $\mu\text{mol/mol}$ [the normal range for follicular phase PG:creatinine ratio is $\leq 0.3 \mu\text{mol/mol}$, the PG level then increases during the luteal phase (29)], which indicated that the woman had had an anovulatory cycle or that samples provided were unlikely to have been taken at the correct phase within her menstrual cycle, from further analysis (n = 228 [21.0%]). E1G and PG were analyzed in serial urine samples from the remaining 729 women (67.1%; Figure 1).

SHBG levels were measured in plasma samples from 718 of the 729 women included in the urinary E1G and PG analyses with the use of a commercially available radioimmunoassay (catalog number 68563; Orion Diagnostica, Espoo, Finland) as previously described (20); there was insufficient plasma for measurement of SHBG in 11 samples (Figure 1). To increase the statistical power of the analysis of plasma hormone levels, we included 45 women from the MOG study who were excluded from the analysis of urinary hormone levels because they had fewer than four follicular phase samples (n = 15), no luteal phase sample (n = 3), or a luteal phase PG:creatinine ratio less than or equal to 0.3 $\mu\text{mol/mol}$ (n = 27) (Figure 1). Plasma levels of DHEA, dehydroepiandrosterone sulfate (DHEAS), testosterone, and androstenedione were measured in women from the MOG study using commercially available radioimmunoassays (DHEA DSL-8900, DHEAS DSL-2700, and androstenedione DSL-4200 from Beckman Coulter, Webster, TX; TKTT2 Coat-a-Count [testosterone]; Siemens Healthcare Diagnostic Products Ltd, Gwynedd, UK) according to manufacturer's instructions. Sufficient plasma was available for measurement of DHEA, DHEAS, testosterone, and androstenedione in 398, 398, 398, and 397 samples, respectively (Figure 1).

Case-control Analysis of Breast Cancer Risk

The case-control series consisted of 950 case patients and 875 control subjects from BIGGS, 3782 case subjects and 4541 control subjects from the BGS, and 3217 case patients and 2045 control subjects from the BBCS. We also included additional British population-based control subjects (n = 3564), who were healthy individuals with no history of cancer ascertained through ongoing UK National Cancer Research Network initiatives (30,31), making a total of 5609 British population-based control subjects (Figure 1). A total of 128 of the control subjects from the BBCS study, who were included in the cross-sectional analysis of hormone levels, were also included in the case-control analysis. Relatives of the BBCS case patients and women ascertained through the MOG study were excluded from the case-control analysis. To increase the statistical power of the case-control analysis, we used publicly accessible British control data (n = 5667) from the Wellcome Trust Case Control Consortium [WTCCC, (32)] and data from both case (n = 1145) and control (n = 1142) subjects from the Cancer Genetic Markers of Susceptibility (CGEMS) study of breast cancer^[33] (Figure 1).

Selection of SNPs for Genotyping

For the cross-sectional analysis of hormone levels, we selected 42 genes that encode key regulatory components of the sex steroid synthesis and metabolism pathways (Supplementary Table 1, available online), several of which occur as gene clusters (ie, *GSTM1-GSTM3*, *HSD3B2-HSD3B1*, *CYP3A5-CYP3A7-CYP3A4*, *CYP1A1-CYP1A2*, *UGT2B*, and *AKR1C1-AKR1C2-AKR1C3-AKR1C4*) or as a cluster of 13 unique first

exons that are alternatively spliced to a set of common exons (ie, *UGT1A*). We selected 753 SNPs to capture the common genetic variation within the coding sequences of all 42 genes and at least 20 kb of 5' and 3' flanking sequence (single genes) or 50 kb of 5' and 3' flanking sequence (gene clusters). The SNPs were selected with the use of the Tagger program (<http://www.broadinstitute.org/mpg/tagger/>), which ensures that for each locus, all catalogued SNPs with a minor allele frequency (MAF) of at least 5% in the HapMap Centre d'Etude du Polymorphisme Humaine (CEPH) population are correlated ($r^2 \geq 0.8$) with at least one of the selected SNPs. SNPs were replaced if they were predicted to have a low genotyping success rate (using Illumina's in house algorithms, <http://www.illumina.com/>) or were less than 60 base pairs from another genotyped SNP. We biased the selection of SNPs toward nonsynonymous SNPs where possible, and where the linkage disequilibrium structure was such that a single selected tag SNP was correlated with large numbers of ungenotyped SNPs (a large tag bin), we selected two tag SNPs rather than a single tag SNP in case the original tag SNP failed genotyping, and we included an additional seven nonsynonymous SNPs that had been excluded because they had a MAF less than 5%. Our final list included 724 tag SNPs, 22 "second tags" (ie SNPs that were a second tag SNP for the largest tag bins), and seven additional rare nonsynonymous SNPs. A single SNP, rs10235235 (in the *CYP3A* gene cluster), was taken forward for the case-control analysis.

DNA Extraction and Genotyping

DNA was extracted in house from blood samples that were stored at -80°C using QIAamp DNA Blood mini kits (Qiagen, Crawley, West Sussex, UK) and commercially using a proprietary chloroform extraction, ethanol precipitation-based method (Tepnel Life Sciences, Manchester, UK) and quantified with the use of PicoGreen (Invitrogen, Grand Island, NY [BBCS and BGS samples]) or NanoDrop (Thermo Scientific, Waltham, MA [BIGGS samples]) quantification kits. Genotyping for the cross-sectional analysis was performed using customized Illumina Sentrix Bead Arrays (Illumina Inc, San Diego, CA) according to the manufacturer's protocols. Genotyping for the case-control analysis was carried out using competitive allele-specific polymerase chain reaction KASPar chemistry (KBiosciences Ltd, Hertfordshire, UK).

Post-Genotyping Quality Control

For the cross-sectional analysis, we genotyped 691 tag SNPs in 774 samples (729 samples included in the urinary hormones analysis and 45 additional samples included in the plasma analyses) (Figure 1). DNA samples with a GenCall score less than 0.25 at any locus were considered "no calls" (http://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf). There were no samples that were deemed to have failed (defined as generating genotypes at <90% of loci). We excluded 16 SNPs that had a call rate less than 95%, six SNPs that were monomorphic in our samples, five SNPs that showed deviation from Hardy-Weinberg equilibrium ($P < 10^{-3}$), and 20 SNPs that had been selected as second tags for large tag bins but for which the original tag passed post-genotyping quality control. To avoid colinearity in regression models that adjusted for multiple SNPs at a single locus, we excluded two additional SNPs that were highly correlated ($r^2 > 0.9$) with another SNP at the same locus in our data but not in HapMap CEPH data. After these exclusions, 642 SNPs were available for analysis in 729 samples that were tested for urinary E1G and PG; 763 samples that were tested for circulating SHBG; 398 samples that were tested for circulating DHEAS, DHEA, and testosterone; and 397 samples that were tested for circulating androstenedione (Figure 1). The mean call rate per sample was 99.5%, and the mean call rate per SNP was 99.9%. Full details of the genes and SNPs are given in Supplementary Table 2 (available online).

For the case-control analysis of rs10235235, call rates were 98.0% (BBCS), 96.6% (BGS), and 96.5% (BIGGS). Deviation of the genotype frequencies from those expected under Hardy-Weinberg equilibrium was assessed by a Fisher exact test in each control population; there was no evidence of departure from Hardy-Weinberg equilibrium in any of the control populations (BBCS: $P = .29$; BGS: $P = .92$; BIGGS: $P = .06$). For each study, we randomly selected a set of samples (1% of the BBCS samples, 5% of the BGS samples, and 5% of the BIGGS samples) to be re-genotyped to test for concordance between the two calls and found that duplicate concordance was 100%.

Statistical Analysis

Statistical analyses were performed using R (version 2.11, <http://cran.r-project.org>) and STATA software (version 11.0; College Station, TX). All reported P values are two-sided.

Cross-sectional Analysis. The percent change in hormone level per allele of each SNP was estimated using linear regression models of \log_e -transformed hormone levels that adjusted for 1) hormone measurement batch only, 2) hormone measurement batch and time between blood draw and analysis, or 3) hormone measurement

batch, age group (<35, 35–44, ≥45 years), body mass index at the time of interview (<25, 25 to <30, ≥30 kg/m²), and parity (0, 1–2, ≥3 full-term pregnancies). After adjusting for hormone measurement batch, adjustment for the additional covariates did not alter the estimates of effect size, and therefore the results presented are adjusted for hormone measurement batch only. We used *t* tests of the regression coefficient to estimate *P* values for linear trend. Quantile–quantile (Q–Q) plots of the *P* values were produced for each hormone. We applied a Bonferroni correction to establish a global statistical significance level of *P* less than 1×10^{-5} , based on our testing of eight hormone measures (E1G, PG, SHBG, DHEA, DHEAS, androstenedione, testosterone, and the ratio of testosterone to SHBG) and 642 SNPs. For the two loci that included at least one SNP that was associated with hormone levels ($P < 1 \times 10^{-5}$), haplotypes were imputed by using haplo.stats software (<http://cran.r-project.org/web/packages/haplo.stats>). Gaussian generalized linear models of log_e-transformed hormone level (adjusted for hormone batch) were fitted to estimate the percent change in hormone levels associated with common haplotypes (ie, those with a frequency >0.5% in this sample of women), taking into account the uncertainty in haplotype allocation.

Case–control Analysis. We calculated the per-allele odds ratios (ORs) and associated 95% confidence intervals (CIs) for each study and for all studies combined using unconditional logistic regression (adjusted for study in the combined analyses). *P* values were estimated using likelihood ratio tests with 1 *df*. We calculated the Cochran Q statistic to test for between-study heterogeneity and the *I*² statistic to quantify the proportion of the total variation due to between-study heterogeneity.^[34] All of the women who participated in the analysis of hormone levels were aged 53 years or younger and premenopausal; because menopausal status at diagnosis cannot be easily ascertained retrospectively, we used age at diagnosis as a proxy to test whether the association between genotype and risk of breast cancer differed according to menopausal status. We stratified the case patients by age at diagnosis (≤50 vs >50 years) and used case-only unconditional logistic regression to test for a statistical interaction between genotype and age at diagnosis. We also analyzed the data excluding breast cancer case patients who were aged 46–54 years at diagnosis (ie, potentially perimenopausal). We used unconditional logistic regression to calculate age group–specific odds ratios; case patients in each age group were compared with all control subjects, and the age group–specific odds ratio was adjusted for study in the combined analysis. *P* values were estimated using likelihood ratio tests.

Bioinformatics

Linkage disequilibrium metrics (ie, *r*² and *D'*) between SNPs reported in HapMap were based on release 27 of the National Center for Biotechnology Information genome Build 36.3 and were computed using the Tagzilla module as implemented in GLU version 1.0a6 (<http://code.google.com/p/glu-genetics/>). The structure of the amino-terminal G domain of SHBG, bound to several different ligands, has been solved by x-ray crystallography previously.^[35] To investigate the potential impact of rs6258 (P185L) on the structure of SHBG, we looked for evidence of cross-species conservation at this region using ClustalW (www.clustal.org/) and examined the structure of SHBG binding to estradiol (Protein Data Base code 1LHO) solved to 1.8 Å.^[36] We simulated the effect of a proline (P) to leucine (L) substitution at amino acid position 185 by mutating the residue within Discovery Studio (<http://accelrys.com/products/discovery-studio>) and subsequently performed a minimization followed by a short dynamics run using the YASARA program^[37] with the YAMBER3 force field at 300 K for 100 picoseconds.

Results

The first aim of this study was to identify genetic variants associated with sex hormone levels in healthy premenopausal women. We tested each of the 642 tag SNPs that we genotyped for association with E1G, PG, SHBG, the androgenic precursors DHEA, DHEAS, androstenedione, testosterone, and the testosterone:SHBG ratio and summarized the results using Q–Q plots (Supplementary Figure 1, available online)

Associations Between SNPs and Urinary E1G Levels

Based on the predefined statistical significance threshold of *P* less than 1×10^{-5} , there was evidence of an association between one SNP, rs10273424, and urinary E1G levels. The minor allele of rs10273424, which maps to 7q22.1 (99 034 009 bps), approximately 50 kb centromeric to *CYP3A5*, was associated with a per-allele change in E1G levels of –21.8% (95% CI = –27.8% to –15.3%; $P = 2.7 \times 10^{-9}$) (Table 1) and accounted for 4.9% of the variation in E1G levels. There was no evidence of an association between rs10273424 and levels of any of the other hormones measured.

Table 1. Associations between single-nucleotide polymorphisms (SNPs) spanning 262 kb mapping to 7q22.1 (cytochrome P450 3A cluster) and urinary estrone glucuronide levels in 729 premenopausal women*

SNP	Coordinate [†]	Major allele, minor allele	MAF	Geometric mean E1G level in $\mu\text{mol/mol}$ by genotype ^{‡,§}		Per-allele change in E1G [‡] , % (95% CI)	$P_{\text{trend }}$	Per-allele change in E1G [¶] , % (95% CI)	$P_{\text{trend }}$
				11 (n)	12 + 22 (n)				
rs10273424	99034009	T, A	0.07	17.8 (619)	13.9 (96)	-21.8 (-27.8 to -15.3)	2.7×10^{-9}	-26.2 (-32.7 to -19.0)	1.9×10^{-10}
rs7792939	99045812	T, C	0.13	16.6 (551)	17.7 (177)	6.3 (0.1 to 12.9)	.05	17.9 (5.5 to 31.7)	.004
rs776746	99108475	G, A	0.07	16.9 (629)	16.4 (100)	-2.6 (-10.1 to 5.4)	.51	4.9 (-13.4 to 27.1)	.62
rs2257401	99144621	G, C	0.09	16.9 (609)	16.7 (120)	-1.3 (-8.0 to 5.9)	.72	-5.6 (-20.6 to 12.2)	.51
rs6945984	99186264	T, C	0.11	16.9 (577)	16.7 (151)	-1.1 (-7.3 to 5.5)	.74	-3.4 (-26.2 to 26.4)	.80
rs4986910	99196460	T, C	0.01	16.8 (720)	20.6 (9)	22.3 (-5.5 to 58.2)	.13	3.3 (-12.4 to 21.9)	.70
rs2242480	99199402	C, T	0.09	16.9 (598)	16.7 (127)	-0.1 (-6.6 to 6.8)	.98	3.6 (-2.5 to 10.1)	.26
rs4646437	99203019	C, T	0.10	16.9 (591)	16.9 (138)	0.9 (-5.6 to 7.8)	.80	-0.9 (-21.6 to 25.3)	.94
rs680055	99295541	G, C	0.05	16.8 (659)	17.0 (70)	1.1 (-7.9 to 11.0)	.82	14.4 (-11.6 to 47.9)	.31

* CI = confidence interval; E1G = creatinine-adjusted mean follicular urinary estrone glucuronide; MAF = minor allele frequency; n = number of individuals with the given genotype.

† In base pairs, Build 36.3.

‡ Adjusted for hormone measurement batch only.

§ MAFs for all of the SNPs at this locus are low, and in each case, there were fewer than 20 women with the 22 genotype; accordingly, geometric mean E1G levels are presented for 12 and 22 genotypes combined. 11 = common allele homozygote, 12 = heterozygote, 22 = rare allele homozygote.

| Two-sided t test (1 df) for null hypothesis that the per-allele difference = 0%.

¶ Adjusted for hormone measurement batch and eight other SNPs genotyped at the CYP3A locus.

Among the 642 SNPs that we genotyped, an additional eight SNPs mapped to the 262-kb region spanning the *CYP3A5–CYP3A7–CYP3A4* cluster (Figure 2). None of these other SNPs was associated with urinary E1G levels (Table 1) or with the levels of any of the other hormones tested (data not shown). In an analysis that adjusted for the other eight SNPs in the *CYP3A* cluster, rs10273424 was associated with a per-allele change in E1G levels of -26.2% (95% CI = -32.7% to -19.0%; $P = 1.9 \times 10^{-10}$). In addition, there was some evidence that the minor allele of a second SNP (rs7792939) that maps 38 kb centromeric to *CYP3A5* (Figure 2) was associated with a per-allele increase in E1G levels of 17.9% (95% CI = 5.5% to 31.7%, $P = .004$), but this association was not statistically significant after Bonferroni correction for multiple testing. These nine SNPs defined 10 main haplotypes (frequency > 0.5% in this study), three of which carried the minor allele of rs10273424 (Supplementary Table 3, available online). The association between rs10273424 and lower urinary E1G level was confined to a single haplotype (frequency = 3.8%) that included the common alleles at each of the other eight SNPs. This haplotype was associated with a 32.3% reduction in urinary E1G level (95% CI = 38.9% to 24.9% reduction, $P = 3.5 \times 10^{-13}$).

Table 1. Associations between single-nucleotide polymorphisms (SNPs) spanning 262 kb mapping to 7q22.1 (cytochrome P450 3A cluster) and urinary estrone glucuronide levels in 729 premenopausal women*

SNP	Coordinate [†]	Major allele, minor allele	MAF	Geometric mean E1G level in $\mu\text{mol/mol}$ by genotype ^{‡,§}		Per-allele change in E1G [‡] , % (95% CI)	$P_{\text{trend }}$	Per-allele change in E1G [¶] , % (95% CI)	$P_{\text{trend }}$
				11 (n)	12 + 22 (n)				
rs10273424	99034009	T, A	0.07	17.8 (619)	13.9 (96)	-21.8 (-27.8 to -15.3)	2.7×10^{-9}	-26.2 (-32.7 to -19.0)	1.9×10^{-10}
rs7792939	99045812	T, C	0.13	16.6 (551)	17.7 (177)	6.3 (0.1 to 12.9)	.05	17.9 (5.5 to 31.7)	.004
rs776746	99108475	G, A	0.07	16.9 (629)	16.4 (100)	-2.6 (-10.1 to 5.4)	.51	4.9 (-13.4 to 27.1)	.62
rs2257401	99144621	G, C	0.09	16.9 (609)	16.7 (120)	-1.3 (-8.0 to 5.9)	.72	-5.6 (-20.6 to 12.2)	.51
rs6945984	99186264	T, C	0.11	16.9 (577)	16.7 (151)	-1.1 (-7.3 to 5.5)	.74	-3.4 (-26.2 to 26.4)	.80
rs4986910	99196460	T, C	0.01	16.8 (720)	20.6 (9)	22.3 (-5.5 to 58.2)	.13	3.3 (-12.4 to 21.9)	.70
rs2242480	99199402	C, T	0.09	16.9 (598)	16.7 (127)	-0.1 (-6.6 to 6.8)	.98	3.6 (-2.5 to 10.1)	.26
rs4646437	99203019	C, T	0.10	16.9 (591)	16.9 (138)	0.9 (-5.6 to 7.8)	.80	-0.9 (-21.6 to 25.3)	.94
rs680055	99295541	G, C	0.05	16.8 (659)	17.0 (70)	1.1 (-7.9 to 11.0)	.82	14.4 (-11.6 to 47.9)	.31

* CI = confidence interval; E1G = creatinine-adjusted mean follicular urinary estrone glucuronide; MAF = minor allele frequency; n = number of individuals with the given genotype.

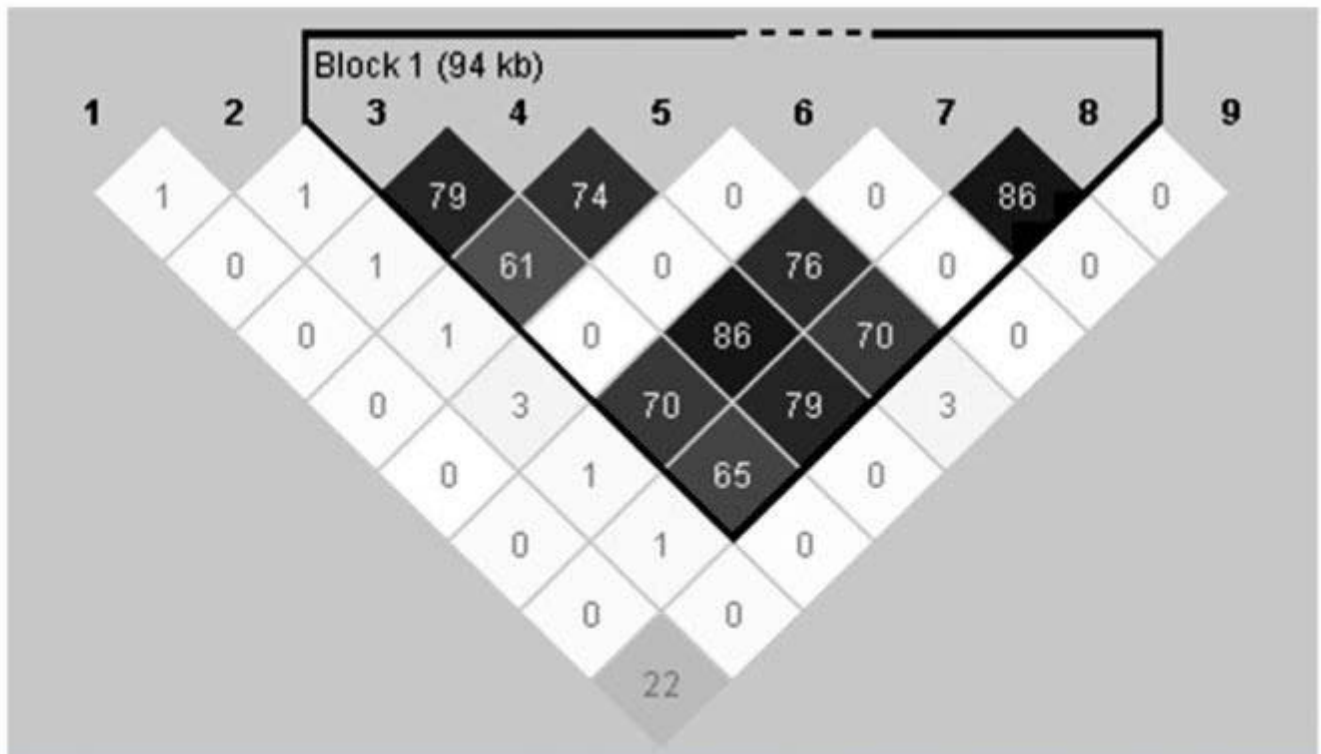
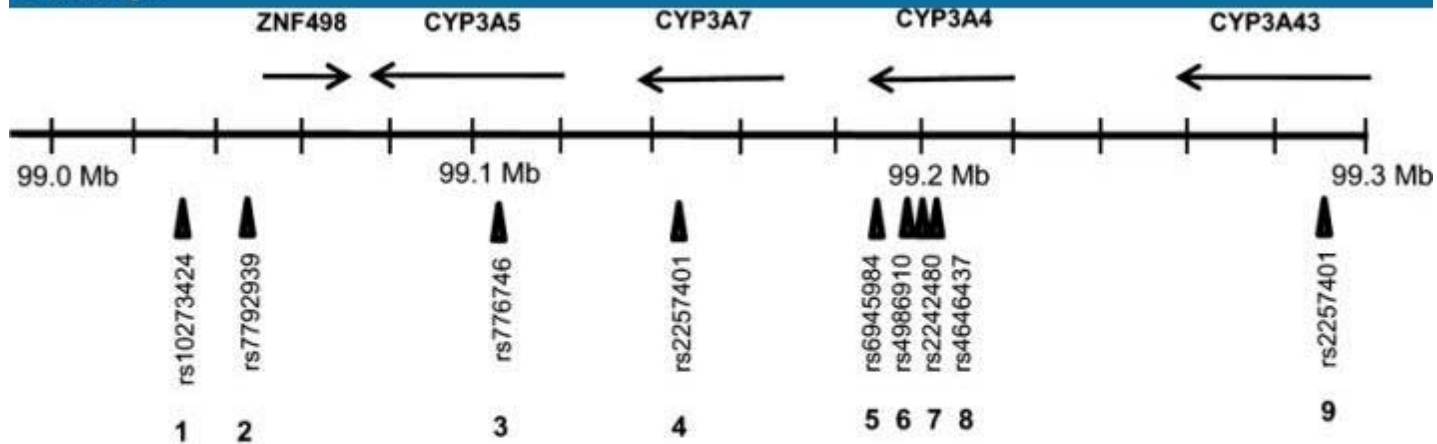
[†] In base pairs, Build 36.3.

[‡] Adjusted for hormone measurement batch only.

[§] MAFs for all of the SNPs at this locus are low, and in each case, there were fewer than 20 women with the 22 genotype; accordingly, geometric mean E1G levels are presented for 12 and 22 genotypes combined. 11 = common allele homozygote, 12 = heterozygote, 22 = rare allele homozygote.

| Two-sided t test (1 df) for null hypothesis that the per-allele difference = 0%.

[¶] Adjusted for hormone measurement batch and eight other SNPs genotyped at the CYP3A locus.



Source: J Natl Cancer Inst © 2012 Oxford University Press

Figure 2. Location and linkage disequilibrium (LD) structure of the cytochrome-P450 3A (*CYP3A*) gene cluster. The locations and the LD structure between single-nucleotide polymorphisms (SNPs) across the *CYP3A* cluster (7q22.1; 99 034 009 to 99 295 541 bps) based on data from 774 women of Northern European ancestry genotyped in this analysis are shown. The locations of all four *CYP3A* gene family members (*CYP3A5*, *CYP3A7*, *CYP3A4*, and *CYP3A43*) and one additional gene, zinc finger protein 498 (*ZNF498*), which maps to this region, based on NCBI build 36.3 (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap24_B36/) are shown above the LD plot. The LD plot shows estimates of the square of the correlation coefficient (r^2) in grayscale (**black** = 1, **white** = 0), calculated for each pair-wise comparison of SNPs derived by the Haploview program (www.broad.mit.edu/mpg/haploview/).

Associations Between SNPs and Plasma SHBG Levels

Two SNPs—rs6258 and rs727428—both of which map to 17p13 (*SHBG* locus), were statistically significantly associated with plasma SHBG levels ($P < 1.0 \times 10^{-5}$; Table 2 and Figure 3). rs6258 (7 475 403 bps) is a rare (MAF = 0.01) nonsynonymous SNP that results in a proline to leucine change at amino acid 185 (P185L). The rare allele of rs6258 (185L) was associated with a 59.9% reduction in plasma SHBG levels (95% CI = 69.6% to 47.2% reduction, $P = 1.3 \times 10^{-10}$) and accounted for 4.9% of the variation in SHBG levels. The proline at amino acid position 185 is highly conserved across species and falls within a 31-amino acid region of conserved sequence (Supplementary Figure 2, available online). This region is unlikely to influence estradiol binding (Figure 4); instead, these amino acids define a region on the surface of SHBG that binds to a calcium ion that is required for stabilizing the active SHBG dimer.^[38] In silico simulation analysis, in which the proline at position 185 was substituted with a leucine, caused the calcium ion to move by 0.52 Å, an alteration that may destabilize the SHBG dimer. In addition, the residue at position 185 is quite near to the dimer interface. Therefore, rs6258 may be causally associated with reduced plasma SHBG level. The minor allele of rs727428 (7 478 517 bps), a noncoding SNP that maps 1 kb 5' (telomeric) to *SHBG*, was associated with a per-allele reduction in plasma SHBG levels of 10.0% (95% CI = 14.0% to 5.8% reduction; $P = 7.0 \times 10^{-6}$). Neither rs727428 nor rs6258 was associated with levels of any of the other hormones tested or the ratio (data not shown).

Table 2. Associations between single-nucleotide polymorphisms (SNPs) spanning 16 kb mapping to 17p13.1 (sex hormone-binding globulin [SHBG] gene) and plasma SHBG levels in 763 premenopausal women*

SNP	Coordinate [†]	Major allele, minor allele	MAF [‡]	Geometric mean SHBG level in nmol/L by genotype ^{§,}			Per-allele change in SHBG [§] , % (95% CI)	$P_{\text{trend}}^{\dagger\dagger}$	Per-allele change in SHBG [#] , % (95% CI)	$P_{\text{trend}}^{\text{ }}$
				11 (n)	12 (n)	22 (n)				
rs13894	7470627	C, T	0.07	51.3 (624)	42.4 (104)	NA	-15.1 (-22.0 to -7.6)	1.6×10^{-4}	-11.4 (-19.6 to -2.4)	.01
rs858521	7470872	C, G	0.37	50.1 (284)	50.3 (342)	48.4 (102)	-1.3 (-5.8 to 3.5)	.59	-6.4 (-16.8 to 5.4)	.27
rs1799941	7474148	G, A	0.25	47.9 (412)	52.2 (272)	55.7 (45)	8.4 (3.0 to 14.2)	.002	-3.3 (-14.3 to 9.1)	.58
rs6258	7475403	C, T	0.01	50.7 (719)	20.3 (10)	NA	-59.9 (-69.6 to -47.2)	1.3×10^{-10}	-56.9 (-67.5 to -42.8)	7.9×10^{-9}
rs727428	7478517	G, A	0.45	54.9 (209)	49.9 (376)	44.4 (144)	-10.0 (-14.0 to -5.8)	7.0×10^{-6}	-12.1 (-18.6 to -5.2)	9.3×10^{-4}
rs2955617	7479510	T, G	0.35	52.1 (303)	49.3 (343)	45.3 (83)	-6.4 (-10.7 to -1.8)	.01	8.2 (-5.8 to 24.4)	.27
rs1641544	7480591	G, A	0.05	50.4 (650)	49.9 (79)	NA	-1.9 (-11.3 to 8.5)	.71	6.4 (-5.4 to 19.7)	.30
rs1641536	7486709	G, A	0.13	51.2 (546)	46.7 (183)	NA	-8.3 (-14.4 to -1.8)	.01	-13.4 (-21.1 to -4.9)	.003

* CI = confidence interval; MAF = minor allele frequency; n = number of individuals with the given genotype; NA = not applicable.

† In base pairs, Build 36.3.

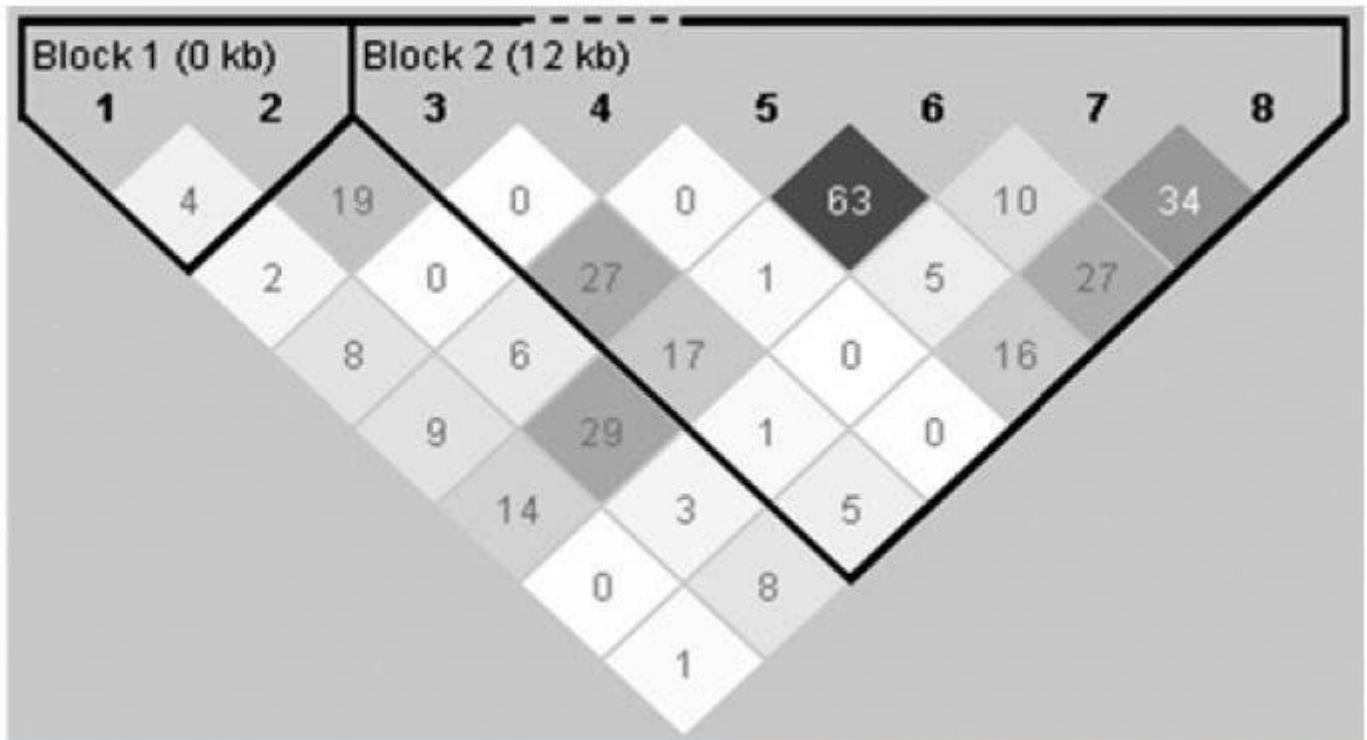
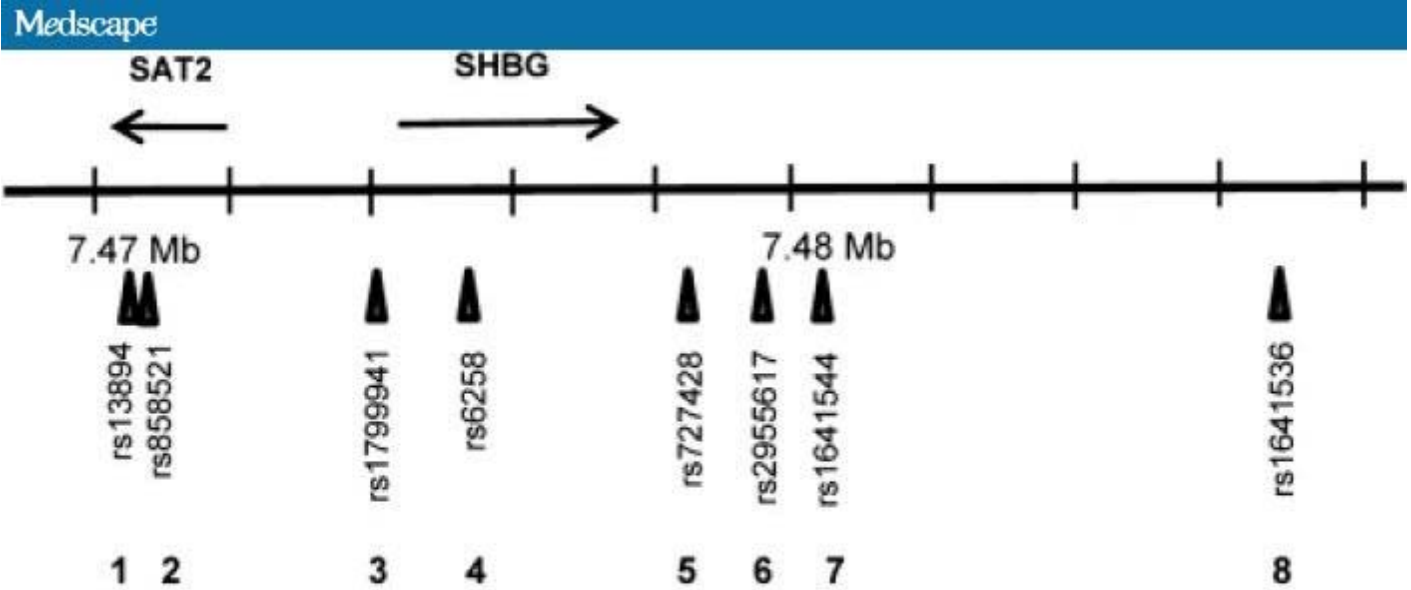
‡ MAFs are presented to two decimal places; the MAF for rs6258 was 0.007 and is rounded to 0.01 in the table.

§ Adjusted for hormone measurement batch only.

| Geometric means are presented for 12 and 22 genotypes combined when there were fewer than 20 women with the 22 genotype. 11 = common allele homozygote, 12 = heterozygote, 22 = rare allele homozygote.

¶¶ Two-sided t test (1 df) for null hypothesis that the per-allele difference = 0%.

Adjusted for hormone measurement batch and seven other SNPs genotyped at the SHBG locus.

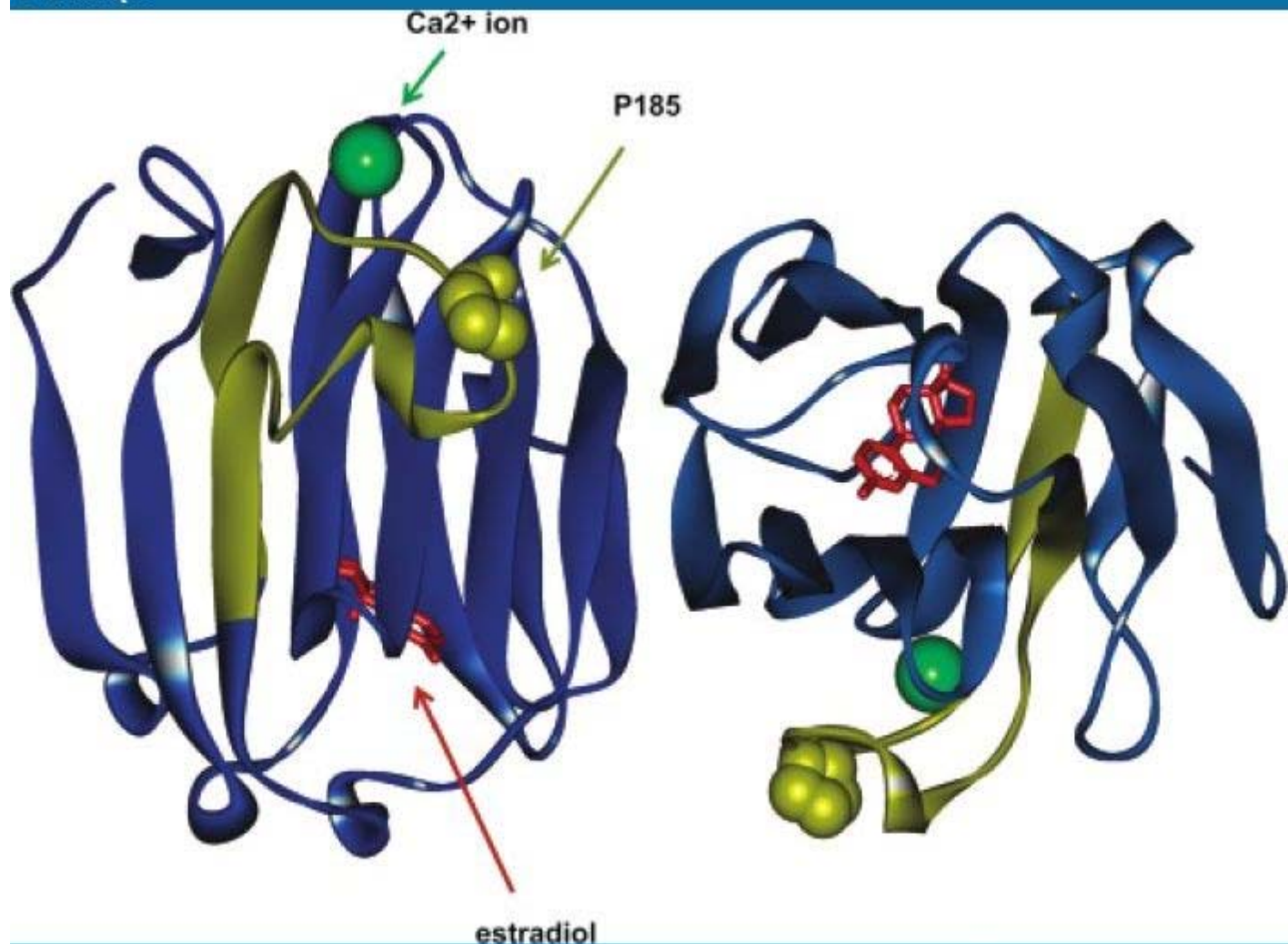


Source: J Natl Cancer Inst © 2012 Oxford University Press

Figure 3. Location and linkage disequilibrium (LD) structure of the sex hormone-binding globulin (*SHBG*) locus. The locations and the LD structure between single-nucleotide polymorphisms (SNPs) across the *SHBG* locus (17p13; 7 470 627 to 7 486 709 bps) based on data from 774 women of Northern European ancestry genotyped in this analysis are shown. The locations of *SHBG* and one additional gene, spermidine N1-acetyl transferase family member 2 (*SAT2*), which maps to this region, based on NCBI build 36.3 (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap24_B36/) are shown above the LD plot. The LD plot shows estimates of the square of the correlation coefficient (r^2) in grayscale (**black** = 1, **white** = 0), calculated

for each pair-wise comparison of SNPs derived by the Haploview program (www.broad.mit.edu/mpg/haploview/).

Medscape



Source: J Natl Cancer Inst © 2012 Oxford University Press

Figure 4. X-ray crystallographic structure of the sex hormone-binding globulin (SHBG) dimer binding to estradiol. X-ray crystallographic structure of the SHBG dimer binding to estradiol (Protein Data Bank code 1LHO) solved to 1.8 Å. (36) A conserved region (Supplementary Figure 2, available online) centered on proline 185 (indicated in **gold**) defines the surface that binds to a calcium ion (indicated in **green**) and which is required for stabilizing the active SHBG dimer. The estradiol (indicated in **red**) binding site lies distal to P185.

We successfully genotyped eight SNPs that mapped to the *SHBG* locus (7 470 627–7 486 709 bps); two of those SNPs—rs13894 (7 470 627 bps) and rs1799941 (7 474 148 bps)—showed some evidence of association with plasma SHBG levels, albeit not at the predefined statistical significance threshold of P less than 1×10^{-5} . The minor allele of rs13894 was associated with a per-allele reduction in SHBG levels of 15.1% (95% CI = 22.0% to 7.6% reduction; $P = 1.6 \times 10^{-4}$), and the minor allele of rs1799941 was associated with a per-allele increase in SHBG levels of 8.4% (95% CI = 3.0% to 14.2%; $P = .002$). However, rs727428 is weakly correlated with rs1799941 ($r^2 = 0.27$, $D' = 0.99$), and rs6258 is weakly correlated with rs13894 ($r^2 = 0.08$, $D' = 1.0$); in an analysis adjusting for all eight SNPs that were genotyped at this locus, only rs6258 remained statistically significantly associated with plasma SHBG levels (per-allele reduction in SHBG levels of 56.9%, 95% CI = 67.5% to 42.8% reduction; $P = 7.9 \times 10^{-9}$) (Table 2). These eight SNPs defined 10 haplotypes (frequency > 0.5% in this study); only the haplotype carrying the rare allele of rs6258 was statistically significantly associated with plasma SHBG levels (Supplementary Table 4, available online). This haplotype

was associated with a 56.7% reduction in plasma SHBG level (95% CI = 66.7% to 43.6% reduction, $P = 8.2 \times 10^{-10}$).

Table 2. Associations between single-nucleotide polymorphisms (SNPs) spanning 16 kb mapping to 17p13.1 (sex hormone-binding globulin [SHBG] gene) and plasma SHBG levels in 763 premenopausal women*

SNP	Coordinate [†]	Major allele, minor allele	MAF [‡]	Geometric mean SHBG level in nmol/L by genotype ^{§,}			Per-allele change in SHBG [§] , % (95% CI)	$P_{\text{trend}}^{\dagger\dagger}$	Per-allele change in SHBG [#] , % (95% CI)	$P_{\text{trend}}^{\#}$
				11 (n)	12 (n)	22 (n)				
rs13894	7470627	C, T	0.07	51.3 (624)	42.4 (104)	NA	-15.1 (-22.0 to -7.6)	1.6×10^{-4}	-11.4 (-19.6 to -2.4)	.01
rs858521	7470872	C, G	0.37	50.1 (284)	50.3 (342)	48.4 (102)	-1.3 (-5.8 to 3.5)	.59	-6.4 (-16.8 to 5.4)	.27
rs1799941	7474148	G, A	0.25	47.9 (412)	52.2 (272)	55.7 (45)	8.4 (3.0 to 14.2)	.002	-3.3 (-14.3 to 9.1)	.58
rs6258	7475403	C, T	0.01	50.7 (719)	20.3 (10)	NA	-59.9 (-69.6 to -47.2)	1.3×10^{-10}	-56.9 (-67.5 to -42.8)	7.9×10^{-9}
rs727428	7478517	G, A	0.45	54.9 (209)	49.9 (376)	44.4 (144)	-10.0 (-14.0 to -5.8)	7.0×10^{-6}	-12.1 (-18.6 to -5.2)	9.3×10^{-4}
rs2955617	7479510	T, G	0.35	52.1 (303)	49.3 (343)	45.3 (83)	-6.4 (-10.7 to -1.8)	.01	8.2 (-5.8 to 24.4)	.27
rs1641544	7480591	G, A	0.05	50.4 (650)	49.9 (79)	NA	-1.9 (-11.3 to 8.5)	.71	6.4 (-5.4 to 19.7)	.30
rs1641536	7486709	G, A	0.13	51.2 (546)	46.7 (183)	NA	-8.3 (-14.4 to -1.8)	.01	-13.4 (-21.1 to -4.9)	.003

* CI = confidence interval; MAF = minor allele frequency; n = number of individuals with the given genotype; NA = not applicable.

† In base pairs, Build 36.3.

‡ MAFs are presented to two decimal places; the MAF for rs6258 was 0.007 and is rounded to 0.01 in the table.

§ Adjusted for hormone measurement batch only.

|| Geometric means are presented for 12 and 22 genotypes combined when there were fewer than 20 women with the 22 genotype. 11 = common allele homozygote, 12 = heterozygote, 22 = rare allele homozygote.

†† Two-sided t test (1 *df*) for null hypothesis that the per-allele difference = 0%.

Adjusted for hormone measurement batch and seven other SNPs genotyped at the SHBG locus.

Associations Between SNPs and the Risk of Breast Cancer

Previous studies that used sex hormone levels as an intermediate phenotype for breast cancer have demonstrated strong associations between genetic variants and circulating hormone levels,^[15-17] but these same variants were not directly associated with the risk of breast cancer. This lack of association may, in part, have been due to a lack of statistical power. We used data from a combined reanalysis of nine prospective studies of postmenopausal women^[2] to estimate the numbers of case patients and control subjects required to detect an association between rs10273424 or rs6258 and risk of breast cancer. In the combined reanalysis,^[2] the odds ratio of breast cancer associated with a doubling of plasma E1 levels was 1.46 (95% CI = 1.22 to 1.73, $P = 1.0 \times 10^{-5}$). Based on these data, a 26% reduction in E1 levels (Table 1) would correspond to an odds ratio of 0.89 and an analysis of 9000 case patients and 9000 control subjects would have 80% power to detect an association between rs10273424 (*CYP3A* cluster) and the risk of breast cancer at a 5% level of statistical significance. In the same combined reanalysis,^[2] a doubling of plasma SHBG levels was associated with an odds ratio of 0.88 (95% CI = 0.76 to 1.03, $P = .062$). For rs6258 (*SHBG*), therefore, a 57% reduction in SHBG levels (Table 3) would correspond to an odds ratio of 1.15 and an analysis of 55 000 case patients and 55 000 control subjects would be required for similar statistical power.

Table 1. Associations between single-nucleotide polymorphisms (SNPs) spanning 262 kb mapping to 7q22.1 (cytochrome P450 3A cluster) and urinary estrone glucuronide levels in 729 premenopausal women*

SNP	Coordinate [†]	Major allele, minor allele	MAF	Geometric mean E1G level in $\mu\text{mol/mol}$ by genotype ^{‡,§}		Per-allele change in E1G [‡] , % (95% CI)	$P_{\text{trend }}$	Per-allele change in E1G [¶] , % (95% CI)	$P_{\text{trend }}$
				11 (n)	12 + 22 (n)				
rs10273424	99034009	T, A	0.07	17.8 (619)	13.9 (96)	-21.8 (-27.8 to -15.3)	2.7×10^{-9}	-26.2 (-32.7 to -19.0)	1.9×10^{-10}
rs7792939	99045812	T, C	0.13	16.6 (551)	17.7 (177)	6.3 (0.1 to 12.9)	.05	17.9 (5.5 to 31.7)	.004
rs776746	99108475	G, A	0.07	16.9 (629)	16.4 (100)	-2.6 (-10.1 to 5.4)	.51	4.9 (-13.4 to 27.1)	.62
rs2257401	99144621	G, C	0.09	16.9 (609)	16.7 (120)	-1.3 (-8.0 to 5.9)	.72	-5.6 (-20.6 to 12.2)	.51
rs6945984	99186264	T, C	0.11	16.9 (577)	16.7 (151)	-1.1 (-7.3 to 5.5)	.74	-3.4 (-26.2 to 26.4)	.80
rs4986910	99196460	T, C	0.01	16.8 (720)	20.6 (9)	22.3 (-5.5 to 58.2)	.13	3.3 (-12.4 to 21.9)	.70
rs2242480	99199402	C, T	0.09	16.9 (598)	16.7 (127)	-0.1 (-6.6 to 6.8)	.98	3.6 (-2.5 to 10.1)	.26
rs4646437	99203019	C, T	0.10	16.9 (591)	16.9 (138)	0.9 (-5.6 to 7.8)	.80	-0.9 (-21.6 to 25.3)	.94
rs680055	99295541	G, C	0.05	16.8 (659)	17.0 (70)	1.1 (-7.9 to 11.0)	.82	14.4 (-11.6 to 47.9)	.31

* CI = confidence interval; E1G = creatinine-adjusted mean follicular urinary estrone glucuronide; MAF = minor allele frequency; n = number of individuals with the given genotype.

† In base pairs, Build 36.3.

‡ Adjusted for hormone measurement batch only.

§ MAFs for all of the SNPs at this locus are low, and in each case, there were fewer than 20 women with the 22 genotype; accordingly, geometric mean E1G levels are presented for 12 and 22 genotypes combined. 11 = common allele homozygote, 12 = heterozygote, 22 = rare allele homozygote.

| Two-sided t test (1 df) for null hypothesis that the per-allele difference = 0%.

¶ Adjusted for hormone measurement batch and eight other SNPs genotyped at the CYP3A locus.

Table 3. Maternal and offspring characteristics at baseline and follow-up according to level of participation.

Study	All case subjects			Case subjects aged ≤ 50 y at diagnosis			Case subjects aged > 50 y at diagnosis		
	No. of case subjects/No. of control subjects	OR (95% CI)	P^{\dagger}	No. of case subjects/No. of control subjects	OR (95% CI)	P^{\dagger}	No. of case subjects/No. of control subjects	OR (95% CI)	P^{\dagger}
BBCS	3149/5502	0.97 (0.87 to 1.09)	.63	1318/5502	0.95 (0.82 to 1.11)	.53	1829/5502	0.99 (0.87 to 1.13)	.88
BGS	3661/4377	0.98 (0.87)	.66	1798/4377	0.92 (0.80)	.22	1855/4377	1.04 (0.91)	.60

		to 1.09)			to 1.05)			to 1.19)	
BIGGS	909/852	0.96 (0.75 to 1.21)	.71	403/852	0.80 (0.58 to 1.10)	.17	477/852	1.13 (0.87 to 1.48)	.37
CGEMS	1145/1142	1.00 (0.82 to 1.21)	.98	–	–	–	1145/1142	1.00 (0.82 to 1.21)	.98
BBCS/WTCCC	1687/5662	0.91 (0.79 to 1.04)	.17	917/5662	0.87 (0.72 to 1.04)	.14	764/5662	0.94 (0.78 to 1.14)	.55
All	10551/17 535	0.96 (0.90 to 1.02)	.22	4436/16 393	0.91 (0.83 to 0.99)	.03	6070/17 535	1.01 (0.93 to 1.10)	.82

Age at diagnosis was missing for 45 case subjects (eight in BBCS, eight in BGS, 29 in BIGGS). Tests for heterogeneity between studies (Cochran Q) and estimates of the proportion of total variation due to between-study heterogeneity (I^2) were: All case subjects: $Q = 0.90$, $df = 4$, $P = .92$; $I^2 = 0\%$; Case subjects aged ≤ 50 years at diagnosis: $Q = 1.19$, $df = 3$, $P = .76$; $I^2 = 0\%$; Case subjects aged > 50 years at diagnosis: $Q = 1.52$, $df = 4$, $P = .82$; $I^2 = 0\%$. BBCS = British Breast Cancer study; BGS = Breakthrough Generations study; BIGGS = Breast Cancer in Galway Genetic study; CGEMS = Cancer Genetic Markers of Susceptibility study; CI = confidence interval; OR = odds ratio; WTCCC = Wellcome Trust Case–Control Consortium; – = not applicable. † Two-sided likelihood ratio test for null hypothesis that OR = 1.0.

To examine whether rs10273424 (*CYP3A* cluster) was associated with the risk of breast cancer, we selected a proxy—rs10235235 [which is included on the Illumina Human 370K, 550K, and 1M arrays used in our genome-wide association scan,^[39] CGEMS,^[33] and WTCCC,^[32] respectively]—that tags rs10273424 ($r^2 = 1.0$ based on HapMap data, release 27 of the National Center for Biotechnology Information genome Build 36.3). We genotyped rs10273424 in study subjects from BBCS, BGS, and BIGGS; we pooled these data with genotype data for breast cancer case patients from our genome-wide association scan,^[39] for control subjects from the WTCCC,^[32] and case patients and control subjects from CGEMS.^[33] Together, these studies comprise genotype data on 10 551 breast cancer case patients (of whom 42% were diagnosed at age ≤ 50 years) and 17 535 control subjects (Table 3).

Table 3. Maternal and offspring characteristics at baseline and follow-up according to level of participation.

Study	All case subjects			Case subjects aged ≤ 50 y at diagnosis			Case subjects aged > 50 y at diagnosis		
	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger
BBCS	3149/5502	0.97 (0.87 to 1.09)	.63	1318/5502	0.95 (0.82 to 1.11)	.53	1829/5502	0.99 (0.87 to 1.13)	.88
BGS	3661/4377	0.98 (0.87 to 1.09)	.66	1798/4377	0.92 (0.80 to 1.05)	.22	1855/4377	1.04 (0.91 to 1.19)	.60
BIGGS	909/852	0.96	.71	403/852	0.80	.17	477/852	1.13	.37

		(0.75 to 1.21)			(0.58 to 1.10)			(0.87 to 1.48)	
CGEMS	1145/1142	1.00 (0.82 to 1.21)	.98	–	–	–	1145/1142	1.00 (0.82 to 1.21)	.98
BBCS/WTCCC	1687/5662	0.91 (0.79 to 1.04)	.17	917/5662	0.87 (0.72 to 1.04)	.14	764/5662	0.94 (0.78 to 1.14)	.55
All	10551/17 535	0.96 (0.90 to 1.02)	.22	4436/16 393	0.91 (0.83 to 0.99)	.03	6070/17 535	1.01 (0.93 to 1.10)	.82

Age at diagnosis was missing for 45 case subjects (eight in BBCS, eight in BGS, 29 in BIGGS). Tests for heterogeneity between studies (Cochran Q) and estimates of the proportion of total variation due to between-study heterogeneity (I^2) were: All case subjects: $Q = 0.90$, $df = 4$, $P = .92$; $I^2 = 0\%$; Case subjects aged ≤ 50 years at diagnosis: $Q = 1.19$, $df = 3$, $P = .76$; $I^2 = 0\%$; Case subjects aged > 50 years at diagnosis: $Q = 1.52$, $df = 4$, $P = .82$; $I^2 = 0\%$. BBCS = British Breast Cancer study; BGS = Breakthrough Generations study; BIGGS = Breast Cancer in Galway Genetic study; CGEMS = Cancer Genetic Markers of Susceptibility study; CI = confidence interval; OR = odds ratio; WTCCC = Wellcome Trust Case–Control Consortium; – = not applicable. † Two-sided likelihood ratio test for null hypothesis that OR = 1.0.

Joint analysis of combined data from all five case–control series provided no evidence of an association between rs10235235 and the risk of breast cancer overall (OR = 0.96; 95% CI = 0.90 to 1.02, $P = .22$) (Table 3). However, a case-only analysis that stratified case subjects by age at diagnosis (≤ 50 vs > 50 years) as a proxy for menopausal status showed modest evidence of effect modification by age at diagnosis ($P = .04$), and there was some evidence from a stratified case–control analysis that the minor allele of rs10235235 was associated with a 9% reduction in breast cancer risk in case subjects diagnosed at or before age 50 years (OR = 0.91, 95% CI = 0.83 to 0.99; $P = .03$) but not in those diagnosed after age 50 years (OR = 1.01, 95% CI = 0.93 to 1.10; $P = .82$) (Table 3). When we excluded from the analysis breast cancer case subjects who were aged 46–54 years at diagnosis (ie, potentially perimenopausal), the odds ratio for case subjects diagnosed at or before age 45 years (OR = 0.90, 95% CI = 0.80 to 1.00; $P = .05$) was marginally lower than that of case subjects diagnosed at or before age 50 years and the odds ratio in older case subjects (OR = 1.02, 95% CI = 0.93 to 1.13; $P = .64$) was marginally higher than that of those diagnosed after age 50 years. There was no evidence of between-study heterogeneity for all case subjects ($P = .92$), case subjects diagnosed at or before age 50 years ($P = .76$), or case subjects diagnosed after age 50 years ($P = .82$) (Table 3).

Table 3. Maternal and offspring characteristics at baseline and follow-up according to level of participation.

Study	All case subjects			Case subjects aged ≤ 50 y at diagnosis			Case subjects aged > 50 y at diagnosis		
	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger
BBCS	3149/5502	0.97 (0.87 to 1.09)	.63	1318/5502	0.95 (0.82 to 1.11)	.53	1829/5502	0.99 (0.87 to 1.13)	.88
BGS	3661/4377	0.98 (0.87 to 1.09)	.66	1798/4377	0.92 (0.80 to 1.05)	.22	1855/4377	1.04 (0.91 to 1.19)	.60

BIGGS	909/852	0.96 (0.75 to 1.21)	.71	403/852	0.80 (0.58 to 1.10)	.17	477/852	1.13 (0.87 to 1.48)	.37
CGEMS	1145/1142	1.00 (0.82 to 1.21)	.98	–	–	–	1145/1142	1.00 (0.82 to 1.21)	.98
BBCS/WTCCC	1687/5662	0.91 (0.79 to 1.04)	.17	917/5662	0.87 (0.72 to 1.04)	.14	764/5662	0.94 (0.78 to 1.14)	.55
All	10551/17 535	0.96 (0.90 to 1.02)	.22	4436/16 393	0.91 (0.83 to 0.99)	.03	6070/17 535	1.01 (0.93 to 1.10)	.82

Age at diagnosis was missing for 45 case subjects (eight in BBCS, eight in BGS, 29 in BIGGS). Tests for heterogeneity between studies (Cochran Q) and estimates of the proportion of total variation due to between-study heterogeneity (I^2) were: All case subjects: $Q = 0.90$, $df = 4$, $P = .92$; $I^2 = 0\%$; Case subjects aged ≤ 50 years at diagnosis: $Q = 1.19$, $df = 3$, $P = .76$; $I^2 = 0\%$; Case subjects aged >50 years at diagnosis: $Q = 1.52$, $df = 4$, $P = .82$; $I^2 = 0\%$. BBCS = British Breast Cancer study; BGS = Breakthrough Generations study; BIGGS = Breast Cancer in Galway Genetic study; CGEMS = Cancer Genetic Markers of Susceptibility study; CI = confidence interval; OR = odds ratio; WTCCC = Wellcome Trust Case–Control Consortium; – = not applicable. † Two-sided likelihood ratio test for null hypothesis that OR = 1.0.

Table 3. Maternal and offspring characteristics at baseline and follow-up according to level of participation.

Study	All case subjects			Case subjects aged ≤ 50 y at diagnosis			Case subjects aged >50 y at diagnosis		
	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger	No. of case subjects/No. of control subjects	OR (95% CI)	P^\dagger
BBCS	3149/5502	0.97 (0.87 to 1.09)	.63	1318/5502	0.95 (0.82 to 1.11)	.53	1829/5502	0.99 (0.87 to 1.13)	.88
BGS	3661/4377	0.98 (0.87 to 1.09)	.66	1798/4377	0.92 (0.80 to 1.05)	.22	1855/4377	1.04 (0.91 to 1.19)	.60
BIGGS	909/852	0.96 (0.75 to 1.21)	.71	403/852	0.80 (0.58 to 1.10)	.17	477/852	1.13 (0.87 to 1.48)	.37
CGEMS	1145/1142	1.00 (0.82 to 1.21)	.98	–	–	–	1145/1142	1.00 (0.82 to 1.21)	.98
BBCS/WTCCC	1687/5662	0.91 (0.79 to 1.04)	.17	917/5662	0.87 (0.72 to 1.04)	.14	764/5662	0.94 (0.78 to 1.14)	.55

All	10551/17 535	0.96 (0.90 to 1.02)	.22	4436/16 393	0.91 (0.83 to 0.99)	.03	6070/17 535	1.01 (0.93 to 1.10)	.82
-----	--------------	------------------------------	-----	-------------	------------------------------	-----	-------------	------------------------------	-----

Age at diagnosis was missing for 45 case subjects (eight in BBCS, eight in BGS, 29 in BIGGS). Tests for heterogeneity between studies (Cochran *Q*) and estimates of the proportion of total variation due to between-study heterogeneity (*I*²) were: All case subjects: *Q* = 0.90, *df* = 4, *P* = .92; *I*² = 0%; Case subjects aged ≤50 years at diagnosis: *Q* = 1.19, *df* = 3 *P* = .76; *I*² = 0%; Case subjects aged >50 years at diagnosis: *Q* = 1.52, *df* = 4, *P* = .82; *I*² = 0%. BBCS = British Breast Cancer study; BGS = Breakthrough Generations study; BIGGS = Breast Cancer in Galway Genetic study; CGEMS = Cancer Genetic Markers of Susceptibility study; CI = confidence interval; OR = odds ratio; WTCCC = Wellcome Trust Case–Control Consortium; – = not applicable. † Two-sided likelihood ratio test for null hypothesis that OR = 1.0.

Table 3. Maternal and offspring characteristics at baseline and follow-up according to level of participation.

Study	All case subjects			Case subjects aged ≤50 y at diagnosis			Case subjects aged >50 y at diagnosis		
	No. of case subjects/No. of control subjects	OR (95% CI)	<i>P</i> [†]	No. of case subjects/No. of control subjects	OR (95% CI)	<i>P</i> [†]	No. of case subjects/No. of control subjects	OR (95% CI)	<i>P</i> [†]
BBCS	3149/5502	0.97 (0.87 to 1.09)	.63	1318/5502	0.95 (0.82 to 1.11)	.53	1829/5502	0.99 (0.87 to 1.13)	.88
BGS	3661/4377	0.98 (0.87 to 1.09)	.66	1798/4377	0.92 (0.80 to 1.05)	.22	1855/4377	1.04 (0.91 to 1.19)	.60
BIGGS	909/852	0.96 (0.75 to 1.21)	.71	403/852	0.80 (0.58 to 1.10)	.17	477/852	1.13 (0.87 to 1.48)	.37
CGEMS	1145/1142	1.00 (0.82 to 1.21)	.98	–	–	–	1145/1142	1.00 (0.82 to 1.21)	.98
BBCS/WTCCC	1687/5662	0.91 (0.79 to 1.04)	.17	917/5662	0.87 (0.72 to 1.04)	.14	764/5662	0.94 (0.78 to 1.14)	.55
All	10551/17 535	0.96 (0.90 to 1.02)	.22	4436/16 393	0.91 (0.83 to 0.99)	.03	6070/17 535	1.01 (0.93 to 1.10)	.82

Age at diagnosis was missing for 45 case subjects (eight in BBCS, eight in BGS, 29 in BIGGS). Tests for heterogeneity between studies (Cochran *Q*) and estimates of the proportion of total variation due to between-study heterogeneity (*I*²) were: All case subjects: *Q* = 0.90, *df* = 4, *P* = .92; *I*² = 0%; Case subjects aged ≤50 years at diagnosis: *Q* = 1.19, *df* = 3 *P* = .76; *I*² = 0%; Case subjects aged >50 years at diagnosis: *Q* = 1.52, *df* = 4, *P* = .82; *I*² = 0%. BBCS = British Breast Cancer study; BGS = Breakthrough Generations study; BIGGS = Breast Cancer in Galway Genetic study; CGEMS = Cancer Genetic Markers of Susceptibility study; CI = confidence interval; OR = odds ratio; WTCCC = Wellcome Trust Case–Control Consortium; – = not applicable. † Two-sided likelihood ratio test for null hypothesis that OR = 1.0.

Discussion

The main finding of this study was a strong association between a tag SNP (rs10273424) mapping 50 kb 3' to *CYP3A5* and lower levels of urinary E1G in premenopausal women. This same SNP showed modest evidence of an association with a reduction in breast cancer risk in case subjects who were diagnosed at or before age 50 years, a finding that is consistent with lower circulating E2 levels in carriers of the rare allele of rs10273424. Although the association between urinary E1G level and circulating levels of E2 within the general population is well established (25,26), further work will be required to demonstrate that this relationship is not modified by genotype at the *CYP3A* locus. The *CYP3A* genes (*CYP3A5*, *CYP3A7*, *CYP3A4*, and *CYP3A43*) occur as a cluster on the long arm of chromosome 7 (7q22, Figure 2). *CYP3A4*, the major isoform in adults, is expressed in several tissues including breast but predominantly in adult liver, where it is the most abundant P-450 enzyme, accounting for 30% of total cytochrome P-450 proteins.^[40] *CYP3A4* is responsible for oxidation of a broad range of substrates, including endogenous hormones, exogenous hormones, and hormonal agents that are used in the treatment of breast cancer, such as tamoxifen (40,41).

We searched the literature for other studies that had tested for an association between SNPs that mapped to the *CYP3A* locus and risk of breast cancer. A nested case-control study of common variants in estrogen biosynthesis and metabolism-related genes conducted in US radiological technologists reported that a rare nonsynonymous *CYP3A4* SNP (rs4986910, M445T) was associated with reduced risk of breast cancer risk (OR = 0.3, 95% CI = 0.1 to 0.9; $P = .04$).^[42] However, we found no association between rs4986910 and urinary levels of E1G (Table 1) or any of the other hormones tested. The National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3) screen (18,43) included five *CYP3A4* SNPs, two of which were also genotyped in this study (rs2242480 and rs4646437, Table 1), and there have been several studies of a *CYP3A4* promoter variant (*CYP3A4*1B*, rs2740574) and predisposition to breast cancer.^[44-48] No associations between these SNPs and hormone levels or breast cancer risk were reported.

Table 1. Associations between single-nucleotide polymorphisms (SNPs) spanning 262 kb mapping to 7q22.1 (cytochrome P450 3A cluster) and urinary estrone glucuronide levels in 729 premenopausal women*

SNP	Coordinate [†]	Major allele, minor allele	MAF	Geometric mean E1G level in $\mu\text{mol/mol}$ by genotype ^{‡,§}		Per-allele change in E1G [‡] , % (95% CI)	$P_{\text{trend II}}$	Per-allele change in E1G [¶] , % (95% CI)	$P_{\text{trend II}}$
				11 (n)	12 + 22 (n)				
rs10273424	99034009	T, A	0.07	17.8 (619)	13.9 (96)	-21.8 (-27.8 to -15.3)	2.7×10^{-9}	-26.2 (-32.7 to -19.0)	1.9×10^{-10}
rs7792939	99045812	T, C	0.13	16.6 (551)	17.7 (177)	6.3 (0.1 to 12.9)	.05	17.9 (5.5 to 31.7)	.004
rs776746	99108475	G, A	0.07	16.9 (629)	16.4 (100)	-2.6 (-10.1 to 5.4)	.51	4.9 (-13.4 to 27.1)	.62
rs2257401	99144621	G, C	0.09	16.9 (609)	16.7 (120)	-1.3 (-8.0 to 5.9)	.72	-5.6 (-20.6 to 12.2)	.51
rs6945984	99186264	T, C	0.11	16.9 (577)	16.7 (151)	-1.1 (-7.3 to 5.5)	.74	-3.4 (-26.2 to 26.4)	.80
rs4986910	99196460	T, C	0.01	16.8 (720)	20.6 (9)	22.3 (-5.5 to 58.2)	.13	3.3 (-12.4 to 21.9)	.70
rs2242480	99199402	C, T	0.09	16.9 (598)	16.7 (127)	-0.1 (-6.6 to 6.8)	.98	3.6 (-2.5 to 10.1)	.26
rs4646437	99203019	C, T	0.10	16.9 (591)	16.9 (138)	0.9 (-5.6 to 7.8)	.80	-0.9 (-21.6 to 25.3)	.94
rs680055	99295541	G, C	0.05	16.8 (659)	17.0 (70)	1.1 (-7.9 to 11.0)	.82	14.4 (-11.6 to 47.9)	.31

* CI = confidence interval; E1G = creatinine-adjusted mean follicular urinary estrone glucuronide; MAF = minor allele frequency; n = number of individuals with the given genotype.

† In base pairs, Build 36.3.

‡ Adjusted for hormone measurement batch only.

§ MAFs for all of the SNPs at this locus are low, and in each case, there were fewer than 20 women with the 22 genotype; accordingly, geometric mean E1G levels are presented for 12 and 22 genotypes combined. 11 = common allele homozygote, 12 = heterozygote, 22 = rare allele homozygote.

| Two-sided t test (1 df) for null hypothesis that the per-allele difference = 0%.

¶ Adjusted for hormone measurement batch and eight other SNPs genotyped at the CYP3A locus.

Table 1. Associations between single-nucleotide polymorphisms (SNPs) spanning 262 kb mapping to 7q22.1 (cytochrome P450 3A cluster) and urinary estrone glucuronide levels in 729 premenopausal women*

SNP	Coordinate [†]	Major allele, minor allele	MAF	Geometric mean E1G level in $\mu\text{mol/mol}$ by genotype ^{‡,§}		Per-allele change in E1G [‡] , % (95% CI)	$P_{\text{trend }}$	Per-allele change in E1G [¶] , % (95% CI)	$P_{\text{trend }}$
				11 (n)	12 + 22 (n)				
rs10273424	99034009	T, A	0.07	17.8 (619)	13.9 (96)	-21.8 (-27.8 to -15.3)	2.7×10^{-9}	-26.2 (-32.7 to -19.0)	1.9×10^{-10}
rs7792939	99045812	T, C	0.13	16.6 (551)	17.7 (177)	6.3 (0.1 to 12.9)	.05	17.9 (5.5 to 31.7)	.004
rs776746	99108475	G, A	0.07	16.9 (629)	16.4 (100)	-2.6 (-10.1 to 5.4)	.51	4.9 (-13.4 to 27.1)	.62
rs2257401	99144621	G, C	0.09	16.9 (609)	16.7 (120)	-1.3 (-8.0 to 5.9)	.72	-5.6 (-20.6 to 12.2)	.51
rs6945984	99186264	T, C	0.11	16.9 (577)	16.7 (151)	-1.1 (-7.3 to 5.5)	.74	-3.4 (-26.2 to 26.4)	.80
rs4986910	99196460	T, C	0.01	16.8 (720)	20.6 (9)	22.3 (-5.5 to 58.2)	.13	3.3 (-12.4 to 21.9)	.70
rs2242480	99199402	C, T	0.09	16.9 (598)	16.7 (127)	-0.1 (-6.6 to 6.8)	.98	3.6 (-2.5 to 10.1)	.26
rs4646437	99203019	C, T	0.10	16.9 (591)	16.9 (138)	0.9 (-5.6 to 7.8)	.80	-0.9 (-21.6 to 25.3)	.94
rs680055	99295541	G, C	0.05	16.8 (659)	17.0 (70)	1.1 (-7.9 to 11.0)	.82	14.4 (-11.6 to 47.9)	.31

* CI = confidence interval; E1G = creatinine-adjusted mean follicular urinary estrone glucuronide; MAF = minor allele frequency; n = number of individuals with the given genotype.

† In base pairs, Build 36.3.

‡ Adjusted for hormone measurement batch only.

§ MAFs for all of the SNPs at this locus are low, and in each case, there were fewer than 20 women with the 22 genotype; accordingly, geometric mean E1G levels are presented for 12 and 22 genotypes combined. 11 = common allele homozygote, 12 = heterozygote, 22 = rare allele homozygote.

| Two-sided t test (1 df) for null hypothesis that the per-allele difference = 0%.

¶ Adjusted for hormone measurement batch and eight other SNPs genotyped at the CYP3A locus.

Multiple variants in *SHBG* have been found to be associated with plasma SHBG levels, and when more than one study has tested the same variant, the associations have been consistent (Supplementary Table 5, available online) (15,16,18). However, none of the other studies examined the nonsynonymous SNP rs6258 (P185L); the analysis by Thompson et al.^[15] included rs13894 (a noncoding variant that is moderately correlated with rs6258) and reported a statistically significant per-allele change in SHBG levels of -15% (95% CI = -20% to -9%), which is similar in magnitude to the reduction that we observed for rs13894 in this study (-15.1%,

95% CI = -22.0% to -7.6%). In this analysis, the association between the rare allele of rs13894 and a reduction in SHBG levels was weaker than the association between the rare allele of rs6258 and reduced SHBG levels (per-allele change = -59.9%, 95% CI = -69.6% to -47.2%), and the proportion of variance explained by the nonsynonymous SNP was much higher (rs6258: $r^2 = 4.9\%$; rs13894: $r^2 = 1.7\%$). However, to our knowledge, none of the SHBG variants tested to date has been found to be associated with the risk of breast cancer, possibly because the reported association between SHBG levels and breast cancer risk in postmenopausal women is very modest.^[2] Given the rarity of rs6258 (MAF = 0.01) and the modest association between SHBG and the risk of postmenopausal breast cancer, we did not test for an association between rs6258 and risk of breast cancer in this study.

We also searched the literature to determine whether there were any variants that mapped to loci included in this screen and for which there was consistent evidence of an association with circulating levels of one or more sex steroid hormones. Multiple investigators have reported analyses of SNPs at the *CYP19A1* and *CYP17A1* loci and circulating levels of E2 and E1 and E2:testosterone and E1:androstenedione ratios (16,49–52). Three variants in *CYP19A1*—a [TTTA]_n repeat in intron 4, a nearby 3-base pair (TCT) deletion (rs11575899), and a T/C SNP in the 3' untranslated region (rs10046)—have been studied extensively and reviewed.^[51] Conflicting results were reported for the [TTTA]_n repeat and rs115758, whereas all three studies that tested for an association between rs10046 and E2:testosterone and/or E1:androstenedione ratios (16,49,50) reported higher ratios in carriers of one or two T alleles. A subsequent study^[17] genotyped 19 tag SNPs selected to capture the common variation within a 181-kb region spanning the *CYP19A1* locus. That study showed a convincing trend of higher plasma levels of E2 and E1 and higher E2:testosterone and E1:androstenedione ratios in carriers of one or two T alleles. However, all of these analyses tested associations with hormone levels in postmenopausal women. In our analysis of hormone levels in premenopausal women, there was no association between rs10046 and urinary E1G or plasma androstenedione, nor was there an association with any of the other 40 *CYP19A1* SNPs that we tested (Supplementary Table 6, available online). Although our sample size (n = 729) was smaller than those of Dunning et al. [n = 1795^[6]] and Haiman et al. [n = 3325^[17]], we still had reasonable power to detect a per-allele difference of 5.5% as reported by Haiman et al.^[17] The lack of association between rs10046 and either urinary E1G or plasma androstenedione in our study may reflect a difference in E1 metabolism between pre- and postmenopausal women. In premenopausal women, the hypothalamic–pituitary–gonadal axis is subject to complex feedback controls that are not relevant in postmenopausal women, in whom the ovaries do not synthesize estrogens; these feedback controls may lead to a rebalancing of estrogen levels in the presence of SNPs that affect estrogen levels, and therefore the effect of the SNP may not be observable in steady-state measurements in premenopausal women. Further studies will be needed to determine whether this lack of association between rs10046 and E1G levels in premenopausal women reflects a difference in E1 metabolism between pre- and postmenopausal women or a chance observation.

The most commonly studied *CYP17A1* SNP—rs743572 (-34T>C)—occurs within the promoter sequence. Olson et al.^[51] identified 12 studies that tested for an association between rs743572 and hormone levels, of which two small studies (53,54) reported statistically significant associations between rs743572 and circulating levels of one or more of the hormones tested, whereas the majority of the studies found no statistically significant associations. Subsequent analyses from the BPC3 consortium (18,52) found no association between rs743572 and plasma levels of sex hormones in 3723 postmenopausal women. We genotyped 10 variants in *CYP17A1*, including rs6163, which is highly correlated with rs743572 ($r^2 = 1.0$); we found that the minor (A) allele of rs6163 was associated with a modest reduction in levels of urinary E1G and a modest increase in levels of plasma androstenedione in premenopausal women, but neither association was statistically significant after accounting for multiple testing (Supplementary Table 7, available online).

The strengths of this study include our use of E1G measurements that took into account the cyclic variation in E1G levels in more than 700 premenopausal women and our approach of tagging entire gene clusters, including relatively large (50 kb) regions of noncoding flanking sequences. Study limitations were that body mass index was measured at interview, not at sample collection, and thus may not be an accurate representation of the woman's body mass index at the time the hormone measurements were made, and that androgenic precursors of E2 and E1 were only measured in a subset of women (397 or 398). The fact that we found no association between DHEA, DHEAS, androstenedione, or testosterone and any of the variants that we genotyped may, therefore, simply reflect a lack of statistical power.

In conclusion, we have identified two novel associations: A nonsynonymous SNP in *SHBG* that may be causally associated with lower plasma SHBG levels and a noncoding variant at the *CYP3A* locus that is associated with lower urinary E1G levels and a reduced risk of breast cancer in younger women (ie, those aged

≤50 years). The implication of the latter finding is that we have identified a long-range regulatory element that is tagged by rs10273424 and maps 50 kb 3' to *CYP3A5*. It is entirely plausible that rs10273424 mediates an effect on E1G levels, and, hence, the risk of breast cancer through differential expression of one of the *CYP3A* genes, possibly the predominantly expressed *CYP3A4*. Although replication of our findings is required, our observation of an association between rs10273424 and urinary levels of E1G may have wider implications; *CYP3A4* is responsible for the oxidative metabolism of an estimated 50% of all clinically used drugs,^[55] including exogenous hormones that are used in postmenopausal hormone replacement therapy and hormonal agents that are used in the treatment of breast cancer. Hence, genotyping of rs10273424 may provide a biomarker for identifying a subgroup of women who will be disproportionately affected by such treatments.

Contexts and Caveats

Prior Knowledge

Epidemiological studies have provided indirect evidence for an association between premenopausal hormone levels and the risk of breast cancer. Genetic factors influence the levels of endogenous sex steroids in both pre- and postmenopausal women, suggesting that common single-nucleotide polymorphisms (SNPs) in genes involved in sex steroid synthesis or metabolism are good candidates for breast cancer predisposition alleles.

Study Design

Endogenous estrone glucuronide (E1G) and pregnanediol glucuronide (PG) levels were measured in urine samples from healthy premenopausal women using a protocol that takes into account the cyclic variation of estrogen and progesterone. The women were genotyped for 642 SNPs that tagged genes that might influence endogenous sex steroid levels. The association between a single SNP and the risk of breast cancer was examined using data from 10 551 breast cancer case patients and 17 535 control subjects.

Contribution

A tag SNP (rs10273424) mapping 50 kb 3' to *CYP3A5* was strongly associated with lower levels of urinary E1G in premenopausal women and showed a modest association with a reduced risk of breast cancer risk in women who were diagnosed at or before age 50 years.

Implications

A long-range regulatory element tagged by rs10273424 may mediate an effect on E1G levels, and, hence, the risk of breast cancer through differential expression of possibly the most predominantly expressed *CYP3A* gene, *CYP3A4*, which is responsible for metabolism of endogenous and exogenous hormones and hormonal agents used in the treatment of breast cancer.

Limitations

Body mass index was not measured at the same time that the hormone measurements were made. Androgenic precursors of estradiol and estrone were only measured in a subset of women.

From the Editors

References

1. Chen WY. Exogenous and endogenous hormones and breast cancer. *Best Pract Res Clin Endocrinol Metab.* 2008;22(4):573–585.
2. Key T, Appleby P, Barnes I, et al. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst.* 2002;94(8):606–616.
3. Pike MC, Krailo MD, Henderson BE, et al. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. *Nature.* 1983;303(5920):767–770.
4. Wysowski DK, Comstock GW, Helsing KJ, et al. Sex hormone levels in serum in relation to the development of breast cancer. *Am J Epidemiol.* 1987;125(5):791–799.
5. Helzlsouer KJ, Alberg AJ, Bush TL, et al. A prospective study of endogenous hormones and breast cancer. *Cancer Detect Prev.* 1994;18(2):79–85.
6. Rosenberg CR, Pasternack BS, Shore RE, et al. Premenopausal estradiol levels and the risk of breast cancer: a new method of controlling for day of the menstrual cycle. *Am J Epidemiol.* 1994;140(6):518–525.

7. Thomas HV, Key TJ, Allen DS, et al. A prospective study of endogenous serum hormone concentrations and breast cancer risk in premenopausal women on the island of Guernsey. *Br J Cancer*. 1997;75(7):1075–1079.
8. Kabuto M, Akiba S, Stevens RG, et al. A prospective study of estradiol and breast cancer in Japanese women. *Cancer Epidemiol Biomarkers Prev*. 2000;9(6):575–579.
9. Kaaks R, Berrino F, Key T, et al. Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J Natl Cancer Inst*. 2005;97(10):755–765.
10. Eliassen AH, Missmer SA, Tworoger SS, et al. Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. *J Natl Cancer Inst*. 2006;98(19):1406–1415.
11. Henderson BR, Gerkins V, Rosario I, et al. Elevated serum levels of estrogen and prolactin in daughters of patients with breast cancer. *N Engl J Med*. 1975;293(16):790–795.
12. Pike MC, Casagrande JT, Brown JB, et al. Comparison of urinary and plasma hormone levels in daughters of breast cancer patients and controls. *J Natl Cancer Inst*. 1977;59(5):1351–1355.
13. Stone J, Folkard E, Doody D, et al. Familial correlations in postmenopausal serum concentrations of sex steroid hormones and other mitogens: a twins and sisters study. *J Clin Endocrinol Metab*. 2009;94(12):4793–4800.
14. Hunter DJ, Riboli E, Haiman CA, et al. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat Rev Cancer*. 2005;5(12):977–985.
15. Thompson DJ, Healey CS, Baynes C, et al. Identification of common variants in the SHBG gene affecting sex hormone-binding globulin levels and breast cancer risk in postmenopausal women. *Cancer Epidemiol Biomarkers Prev*. 2008;17(12):3490–3498.
16. Dunning AM, Dowsett M, Healey CS, et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. *J Natl Cancer Inst*. 2004;96(12):936–945.
17. Haiman CA, Dossus L, Setiawan VW, et al. Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res*. 2007;67(5):1893–1897.
18. Beckmann L, Husing A, Setiawan VW, et al. Comprehensive analysis of hormone and genetic variation in 36 genes related to steroid hormone metabolism in pre- and postmenopausal women from the breast and prostate cancer cohort consortium (BPC3). *J Clin Endocrinol Metab*. 2011;96(2):E360–E367.
19. Atkinson C, Warren RM, Sala E, et al. Red-clover-derived isoflavones and mammographic breast density: a double-blind, randomized, placebo-controlled trial [ISRCTN42940165]. *Breast Cancer Res*. 2004;6(3):R170–R179.
20. Walker K, Fletcher O, Johnson N, et al. Premenopausal mammographic density in relation to cyclic variations in endogenous sex hormone levels, prolactin, and insulin-like growth factors. *Cancer Res*. 2009;69(16):6490–6499.
21. Moss SM, Cuckle H, Evans A, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet*. 2006;368(9552):2053–2060.
22. Johnson N, Fletcher O, Naceur-Lombardelli C, et al. Interaction between CHEK2*1100delC and other low-penetrance breast-cancer susceptibility genes: a familial study. *Lancet*. 2005;366(9496):1554–1557.
23. McInerney NM, Miller N, Rowan A, et al. Evaluation of variants in the CHEK2, BRIP1 and PALB2 genes in an Irish breast cancer cohort. *Breast Cancer Res Treat*. 2010;121(1):203–210.
24. Swerdlow AJ, Jones ME, Schoemaker MJ, et al. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. *Br J Cancer*. 2011;105(7):911–917.
25. Tanabe K, Susumu N, Hand K, et al. Prediction of the potentially fertile period by urinary hormone measurements using a new home-use monitor: comparison with laboratory hormone analyses. *Human Reprod*. 2001;16(8):1619–1624.
26. Catalan R, Castellanos JM, Palomino T, et al. Correlation between plasma estradiol and estrone-3-glucuronide in urine during the monitoring of ovarian induction therapy. *Int J Fertil*. 1989;34(4):271–275.
27. Feinberg DR, Jones BC, Law Smith MJ, et al. Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Horm Behav*. 2006;49(2):215–222.
28. Masson P, Ohlsson P, Bjorkhem I. Combined enzymic-Jaffe method for determination of creatinine in serum. *Clin Chem*. 1981;27(1):18–21.
29. Yong EL, Glasier A, Hillier H, et al. Effect of cyclofenil on hormonal dynamics, follicular development and cervical mucus in normal and oligomenorrhoeic women. *Hum Reprod*. 1992;7(1):39–43.

30. Penegar S, Wood W, Lubbe S, et al. National study of colorectal cancer genetics. *Br J Cancer*. 2007;97(9):1305–1309.
31. Eisen T, Matakidou A, Houlston R. Identification of low penetrance alleles for lung cancer: the GENetic Lung CANcer Predisposition Study (GELCAPS). *BMC Cancer*. 2008;8:244.
32. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.
33. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007;39(7):870–874.
34. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–560.
35. Hammond GL, Avvakumov GV, Muller YA. Structure/function analyses of human sex hormone-binding globulin: effects of zinc on steroid-binding specificity. *J Steroid Biochem Mol Biol*. 2003;85(2–5):195–200.
36. Grishkovskaya I, Avvakumov GV, Hammond GL, et al. Steroid ligands bind human sex hormone-binding globulin in specific orientations and produce distinct changes in protein conformation. *J Biol Chem*. 2002;277(35):32086–32093.
37. Krieger E, Joo K, Lee J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins*. 2009;77 suppl 9:114–122.
38. Bocchinfuso WP, Hammond GL. Steroid-binding and dimerization domains of human sex hormone-binding globulin partially overlap: steroids and Ca²⁺ stabilize dimer formation. *Biochemistry*. 1994;33(35):10622–10629.
39. Fletcher O, Johnson N, Orr N, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst*. 2011;103(5):425–435.
40. Keshava C, McCanlies EC, Weston A. CYP3A4 polymorphisms—potential risk factors for breast and prostate cancer: a HuGE review. *Am J Epidemiol*. 2004;160(9):825–841.
41. Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. *Oncogene*. 2006;25(11):1679–1691.
42. Sigurdson AJ, Bhatti P, Chang SC, et al. Polymorphisms in estrogen biosynthesis and metabolism-related genes, ionizing radiation exposure, and risk of breast cancer among US radiologic technologists. *Breast Cancer Res Treat*. 2009;118(1):177–184.
43. Canzian F, Cox DG, Setiawan VW, et al. Comprehensive analysis of common genetic variation in 61 genes related to steroid hormone and insulin-like growth factor-I metabolism and breast cancer risk in the NCI breast and prostate cancer cohort consortium. *Hum Mol Genet*. 2010;19(19):3873–3884.
44. Spurdle AB, Goodwin B, Hodgson E, et al. The CYP3A4*1B polymorphism has no functional significance and is not associated with risk of breast or ovarian cancer. *Pharmacogenetics*. 2002;12(5):355–366.
45. Le Marchand L, Donlon T, Kolonel LN, et al. Estrogen metabolism-related genes and breast cancer risk: the multiethnic cohort study. *Cancer Epidemiol Biomarkers Prev*. 2005;14(8):1998–2003.
46. Rebbeck TR, Troxel AB, Shatalova EG, et al. Lack of effect modification between estrogen metabolism genotypes and combined hormone replacement therapy in postmenopausal breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2007;16(6):1318–1320.
47. The MARIE-GENICA Consortium on Genetic Susceptibility for Menopausal Hormone Therapy-Related Breast Cancer Risk. Genetic polymorphisms in phase I and phase II enzymes and breast cancer risk associated with menopausal hormone therapy in postmenopausal women. *Breast Cancer Res Treat*. 2010;119(2):463–474.
48. Kato I, Cichon M, Yee CL, et al. African American-preponderant single nucleotide polymorphisms (SNPs) and risk of breast cancer. *Cancer Epidemiol*. 2009;33(1):24–30.
49. Haiman CA, Stram DO, Pike MC, et al. A comprehensive haplotype analysis of CYP19 and breast cancer risk: the Multiethnic Cohort. *Hum Mol Genet*. 2003;12(20):2679–2692.
50. Paynter RA, Hankinson SE, Colditz GA, et al. CYP19 (aromatase) haplotypes and endometrial cancer risk. *Int J Cancer*. 2005;116(2):267–274.
51. Olson SH, Bandera EV, Orlov I. Variants in estrogen biosynthesis genes, sex steroid hormone levels, and endometrial cancer: a HuGE review. *Am J Epidemiol*. 2007;165(3):235–245.
52. Setiawan VW, Schumacher FR, Haiman CA, et al. CYP17 genetic variation and risk of breast and prostate cancer from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiol Biomarkers Prev*. 2007;16(11):2237–2246.
53. Feigelson HS, Shames LS, Pike MC, et al. Cytochrome P450c17alpha gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations. *Cancer Res*. 1998;58(4):585–587.

54. Hong CC, Thompson HJ, Jiang C, et al. Association between the T27C polymorphism in the cytochrome P450 c17alpha (CYP17) gene and risk factors for breast cancer. *Breast Cancer Res Treat.* 2004;88(3):217–230.
55. Perera MA. The missing linkage: what pharmacogenetic associations are left to find in CYP3A? *Expert Opin Drug Metab Toxicol.* 2010;6(1):17–28.

Funding

This work was supported by Cancer Research UK (C150/A5660 and C1178/A3947 to JP and IdSS), Breakthrough Breast Cancer and Breast Cancer Campaign (2006NovPR33 to OF). ES is supported by the Department of Health via the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. We acknowledge National Health Service funding to the NIHR Royal Marsden Biomedical Research Centre and the National Cancer Research Network (NCRN).

J Natl Cancer Inst. 2012;104(9):657-669. © 2012 Oxford University Press