

Identification of Common Variants in the SHBG Gene Affecting Sex Hormone-Binding Globulin Levels and Breast Cancer Risk in Postmenopausal Women

Deborah J. Thompson,¹ Catherine S. Healey,² Caroline Baynes,² Bolot Kalmyrzaev,² Shahana Ahmed,² Mitch Dowsett,⁴ Elizabeth Folkerd,⁴ Robert N. Luben,³ David Cox,⁵ Dennis Ballinger,⁵ Paul D.P. Pharoah,² Bruce A.J. Ponder,² Alison M. Dunning,² Douglas F. Easton,¹ and The Studies in Epidemiology and Risks of Cancer Heredity Team^{1,2}

¹Cancer Research UK Genetic Epidemiology Unit, ²Department of Oncology, and ³European Prospective Investigation of Cancer, University of Cambridge, Cambridge, United Kingdom; ⁴Academic Department of Biochemistry, Royal Marsden Hospital, London, United Kingdom; and ⁵Perlegen Sciences, Mountain View, California

Abstract

Background: Circulating levels of sex hormone-binding globulin (SHBG) are inversely associated with breast cancer risk in postmenopausal women. Three polymorphisms within the SHBG gene have been reported to affect SHBG levels, but there has been no systematic attempt to identify other such variants.

Methods: We looked for associations between SHBG levels in 1,134 healthy, postmenopausal women and 11 tagging single nucleotide polymorphisms (SNP) in or around the SHBG gene. Associations between SHBG SNPs and breast cancer were tested in up to 6,622 postmenopausal breast cancer cases and 6,784 controls. **Results:** Ten SNPs within or close to the SHBG gene were significantly associated with SHBG levels as was the (TAAA)_n polymorphism. The best-fitting combination of rs6259, rs858521, and rs727428 and body mass index, waist, hip, age, and smoking status accounted for 24% of the variance in SHBG levels

(natural logarithm transformed). Haplotype analysis suggested that rs858518, rs727428, or a variant in linkage disequilibrium with them acts to decrease SHBG levels but that this effect is neutralized by rs6259 (D356N). rs1799941 increases SHBG levels, but the previously reported association with (TAAA)_n repeat length appears to be a consequence of linkage disequilibrium with these SNPs. One further SHBG SNP was significantly associated with breast cancer (rs6257, per-allele odds ratio, 0.88; 95% confidence interval, 0.82-0.95; *P* = 0.002).

Conclusion: At least 3 SNPs showed associations with SHBG levels that were highly significant but relatively small in magnitude. rs6257 is a potential breast cancer susceptibility variant, but relationships between the genetic determinants of SHBG levels and breast cancer are complex. (Cancer Epidemiol Biomarkers Prev 2008;17(12):3490-8)

Introduction

It is well established that close relatives of women with breast cancer are themselves at increased risk of the disease, and evidence from twin studies suggests that a considerable proportion of this excess risk can be attributed to inherited genetic factors (e.g., ref. 1). Rare, high-risk mutations in the BRCA1, BRCA2, and TP53 genes are known to account for up to 20% of the familial

relative risk (2), and recent family-based case-control studies and genome-wide association studies have detected new susceptibility genes that together account for a further ~6% (3, 4). However, the majority of the excess familial risk of breast cancer remains unexplained. Many association studies have inadequate statistical power to detect variants conferring modest increases in risk [e.g., odds ratios (OR), 1.1-1.5]. For a fixed sample size, power can be improved by instead looking for variants that are associated with a common quantitative trait that in turn affects the risk of disease. In this way, the search for breast cancer predisposition alleles can be approached via intermediate quantitative phenotypes, such as endogenous serum hormone levels.

In human blood, circulating sex steroid hormones such as testosterone and estradiol are predominantly bound to sex hormone-binding globulin (SHBG), a 373-amino acid glycoprotein produced mainly by the liver. By binding these hormones, SHBG restricts their bioavailability, limiting the amount of "free" circulating androgens and estrogens. It has recently been reported that higher prediagnosis levels of androgens and estrogens are associated with an increased risk of breast cancer in

Received 8/11/08; revised 9/2/08; accepted 9/15/08.

Grant support: Cancer Research UK. D.F. Easton is a principal fellow, and P.D.P. Pharoah is a senior clinical research fellow of Cancer Research UK.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

The SEARCH study team are currently: Jean Abraham, Shahana Ahmed, Antonis Antoniou, Caroline Baynes, Patrick Benusiglio, Fiona Blows, Arancha Cebrian, Don Conroy, Bridget Curzon, Gary Dew, Kristy Driver, Helen Field, Maya Ghousaini, Patricia Harrington, Catherine Healey, Sue Irvine, Bolot Kalmyrzaev, Clare Jordan, Fabienne Lesueur, Craig Luccarini, Rebecca Mayes, Melanie Maranian, Jonathan Morrison, Hannah Munday, Barbara Perkins, Karen Pooley, Karen Redman, Serena Scollen, Danielle Shadforth, Mital Shah, Anabel Simpson, Anne Stafford, Deborah Thompson, Jonathan Tyrer, Paula Smith, and Judy West.

Requests for reprints: Deborah J. Thompson, Cambridge University, Strangeways Research Laboratories, Worts Causeway, Cambridge CB1 8RN, United Kingdom. Phone: 44-1223-740164; Fax: 44-1223-740159. E-mail: deborah.thompson@phpc.com.ac.uk

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-08-0734

postmenopausal women, whereas the association between SHBG levels and breast cancer is in the opposite direction (5-7). It is plausible that any genetic variant that lowered the levels of SHBG would have the effect of allowing more "free" estrogens to circulate and hence could increase the risk of postmenopausal breast cancer. In addition, it is thought that SHBG levels are inversely associated with type 2 diabetes, particularly in women (8). Twin and sibling studies have estimated the heritability of SHBG levels to be between one-third and three-quarters, indicating that such genetic variants should be identifiable (9-11).

We have reported previously associations between two single nucleotide polymorphisms (SNP) in the SHBG gene (a 5'-untranslated region g-a substitution, now named rs1799941, and the exon 8 D356N amino acid change rs6259, referred to elsewhere as D327N) and raised SHBG levels in postmenopausal women (12). The 356N variant is thought to be associated with a reduced rate of clearance of SHBG from circulation (13). Higher SHBG levels among carriers of the N allele have been reported by French and Chinese groups, but no significant difference in levels was reported among postmenopausal women in the California/Hawaii Multiethnic Cohort (14-16). The same Chinese group also reported a lower risk of postmenopausal breast cancer among carriers of the N allele (15), but a study of Nordic and Polish women found no differences in genotype frequencies between breast cancer cases and controls (17). Furthermore, there have been various reports that a pentanucleotide (TAAAA)_n repeat in the 5'-promoter of the SHBG gene is associated with SHBG levels, but the exact nature of the association is yet to be established (14, 16, 18-20). A recent genome-wide study of SHBG levels found an association with rs6761, over 115 kb distal to the SHBG gene, which they attributed to linkage disequilibrium (LD) with rs1799941, although no attempt was made to establish whether other SHBG SNPs could better explain the finding (21).

Following on from our earlier study, we have now genotyped the SHBG 5'-untranslated region TAAAA repeat polymorphism and a comprehensive set of 11 tagging SNPs in or close to the SHBG gene in 1,134 healthy postmenopausal women with measured SHBG levels, and in a case-control study comprising 2,271 postmenopausal breast cancer cases and 2,280 controls.

Subjects and Methods

Healthy Postmenopausal Women. Sex hormone levels were measured on 2,115 women, selected from the ~25,000 participants in the European Prospective Investigation of Cancer-Norfolk cohort study (ref. 22 for details of the design and recruitment of the European Prospective Investigation of Cancer-Norfolk cohort). The 2,115 women were selected at random from the subset of participants considered to be postmenopausal based on being >55 years, not having menstruated in the last year, and having not taken hormone replacement therapy for at least 3 months before sampling (12). The plasma and serum samples collected from these women had been stored at -70°C until analysis, and their whole blood samples had been stored at -30°C before DNA extraction. Anthropometric measurements [body mass index

(BMI), waist and hip circumference, and bra cup size] and other information (age at sample collection, menarche and menopause, smoking status, parity, time since last hormone replacement therapy use, and history of oral contraceptive use) were also collected for these women as part of the European Prospective Investigation of Cancer study.

This group of women partially overlaps with a group of 2,280 control subjects [referred to as "Studies in Epidemiology and Risks of Cancer Heredity (SEARCH) Set01 controls"; see below] who have been genotyped for SNPs in candidate breast cancer susceptibility genes in the course of ongoing case-control studies conducted in this laboratory (23), resulting in a group of 1,151 postmenopausal women (see above definition) for whom both SHBG measurements and SNP genotypes were available. Seventeen women whose self-reported age at menopause was <3 years before sample collection were excluded to further ensure the exclusion of any premenopausal or perimenopausal women, leaving 1,134 women for analysis. Relevant characteristics of the set are given in Supplementary Table S1.

Ethical approval was obtained from the Norwich Local Research Ethics Committee, LREC 98CN01. All study participants provided written informed consent.

Breast Cancer Case Patients and Control Subjects.

Breast cancer cases were drawn from SEARCH (Breast Cancer), an ongoing population-based study (3, 24).⁶ Cases for this study are ascertained through the East Anglian Cancer Registry. All patients diagnosed with invasive breast cancer ages <55 years since 1991 and still alive in 1996 (prevalent cases, median age, 48 years) together with all those diagnosed ages <70 years between 1996 and the present (incident cases, median age, 54 years) are eligible to take part. As of August 1, 2005, there were 12,767 eligible patients. Of these, 2,284 were not contacted because their general practitioner did not respond or thought that it would be inappropriate to contact the patient. Of the 10,583 patients who were contacted, 67% have returned a questionnaire and 64% provided a blood sample for DNA analysis. Eligible patients who did not take part in the study were similar to participants, except that, as might be expected, the proportion of clinical stage III/IV cases was somewhat higher in nonparticipants (10% versus 5%). A fuller description of the cases is given in ref. 25. Controls were randomly selected from the European Prospective Investigation of Cancer-Norfolk participants, representing the same geographic region from which the cases were recruited. Controls are not matched to cases but are broadly similar in age (42-81 years). The ethnic background of both cases and controls as reported on the questionnaires is similar, with >98% being White. The samples have been split into three sets to save DNA and reduce genotyping costs: Set01 (2,271 cases and 2,280 controls) is genotyped for all candidate SNPs, whereas Set02 (2,203 cases and 2,280 controls) and Set03 (2,287 cases and 2,257 controls) are reserved for those SNPs that show marginally significant associations in Set01. The study is approved

⁶ <http://www.srl.cam.ac.uk/search/Homepage.htm>

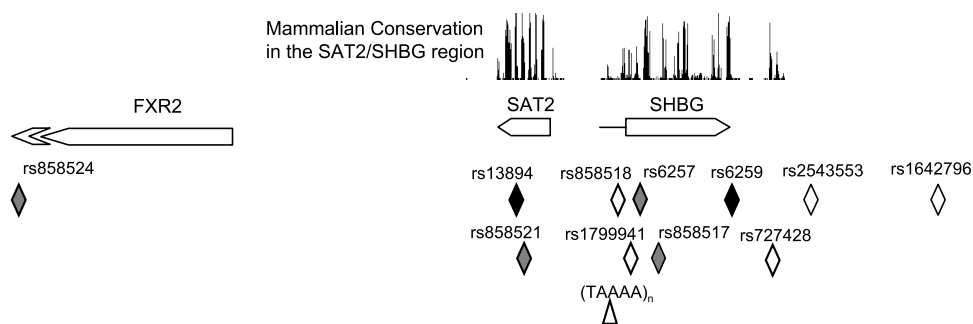


Figure 1. Positions of the 11 tested SHBG SNPs and the (TAAAA)_n repeat polymorphism shown with evolutionary conservation information for the region around the SHBG gene. *Black diamonds*, coding SNPs; *gray diamonds*, intronic SNPs; *unfilled diamonds*, other SNPs; *triangles*, repeat polymorphisms. Conservation information is from Vertebrate Multiz Alignment & PhastCons Conservation (17 placental mammals; <http://genome.ucsc.edu/cgi-bin/hgTracks>).

by the Eastern Region Multicentre Research Ethics Committee, and all patients gave written informed consent.

SHBG Measurements. For the first 151 women, sex hormone measurements were made from stored serum samples. After confirming that there were only minor differences between values obtained from plasma or serum from the same individual, the remaining measurements were made on plasma samples. SHBG levels were measured using a liquid-phase immunoradiometric kit (Orion Diagnostica). Within- and between-batch coefficients of variation were 2.1% and 7.4%, respectively, at a concentration of 11 nmol/L, and the sensitivity limit was 0.5 nmol/L (12).

SNP Selection and Genotyping. DNA for genotyping was extracted from the whole blood samples of all subjects by Whatman Biosciences. All samples were genotyped using the ABI PRISM 7900 sequence detection system or "TaqMan" (Applied Biosystems). PCR was carried out on whole genome-amplified genomic DNA (10 ng) using TaqMan Universal PCR Master Mix (Applied Biosystems), forward and reverse primers, and 6-carboxyfluorescein- and VIC-labeled probes designed by Applied Biosystems (ABI Assay-by-Designs) in a 5 μ L reaction. The completed PCRs were read on an ABI PRISM 7900 Sequence Detector in endpoint mode using the Allelic Discrimination Sequence Detector Software (Applied Biosystems). For each set, cases and controls were arrayed together in twelve 384-well plates and a 13th plate contained eight duplicate samples from each of the 12 plates to ensure a good quality of genotyping. Each 384-well plate included two non-template controls. Concordance for duplicate samples was >98% for all assays. Failed genotypes were not repeated (the rate for failed genotypes did not exceed 4.0% for any SNP). Deviations of the genotype distributions from those expected under Hardy-Weinberg equilibrium were tested in controls.

Only one polymorphic variant in SHBG with minor allele frequency (MAF) $\geq 5\%$ is included in Release 22 of the HapMap project (rs6259), so SNP selection was based on data from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium project.⁷ The Breast and Prostate Cancer Cohort Con-

sortium provides access to phased genotypes from 88 Caucasian trios for SNPs reported by public databases or identified by resequencing a panel of germ-line DNA samples from 190 breast or prostate cancer patients from five ethnic groups, having >85% power to detect SNPs with MAF $\geq 5\%$ in any population group (26). Twenty-eight SNPs within 20 kb of the SHBG gene (including 7 within SHBG) had been genotyped as part of the Breast and Prostate Cancer Cohort Consortium project. Using the program "Tagger" (27),⁸ a set of 9 tagging SNPs (Fig. 1) were required to provide either a pairwise or 2-SNP haplotype tag with $r^2 \geq 0.8$ for all 28 SNPs. An additional 3 SNPs were also genotyped to ensure that each SNP within the SHBG gene had either been genotyped directly or was in perfect LD (pairwise $r^2 = 1$) with a tagging SNP.

There was significant deviation from Hardy-Weinberg equilibrium for rs9898876 ($P < 0.001$), which was excluded from the analysis. It was not possible to find a tag for this SNP (maximum $r^2 = 0.43$ with rs6259). For all other SNPs, the Hardy-Weinberg equilibrium test was nonsignificant at the 5% level.

Genotyping of the (TAAAA)_n Polymorphism. The Set01 samples were genotyped for the 5'-untranslated region (TAAAA)_n polymorphic repeat. PCR amplification of the polymorphic fragment was generated using primers flanking the repeat region, and forward primer was 5'-labeled with a fluorescent dye 6-carboxyfluorescein: forward 5'-6-carboxyfluorescein-GAGAGGCA-GAGGCAGCAGT and reverse 5'-CAGGGCCTAAAC-AGTCTAGCA. PCR conditions were as follows: (a) 95°C for 12 min; (b) 40 cycles: 94°C for 15 s, 60°C for 15 s, and 72°C for 30 s; (c) 72°C for 12 min; and (d) 4°C hold. The separation and detection of PCR products were accomplished with ABI PRISM 3100 Genetic Analyser (Applied Biosystems) following manufacturer's protocol, and allele calling was carried out using GeneMapper Software version 3.5 (Applied Biosystems). The number of TAAAA repeats was determined by sizing of the PCR products with an internal standard GS500 ROX (Applied

⁷ <http://snp500cancer.nci.nih.gov>

⁸ <http://www.broad.mit.edu/mpg/tagger/>

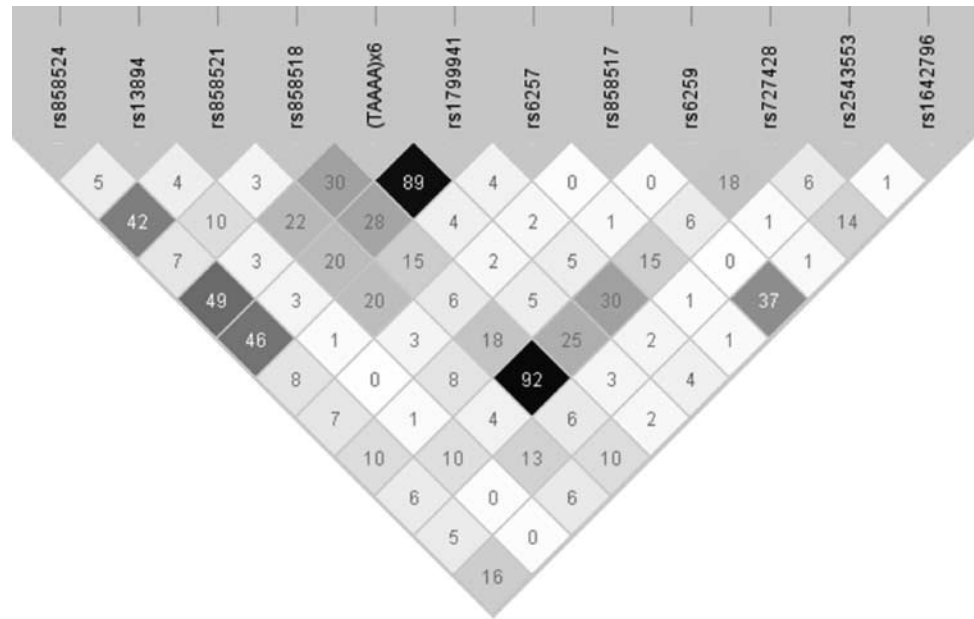


Figure 2. LD, as measured by the pairwise r^2 coefficient, for the 11 tested SHBG SNPs and the (TAAAA)_n repeat polymorphism.

Biosystems). This was additionally confirmed by sequencing four DNA samples, each homozygous for a different common allele.

Statistical Methods. The SHBG levels were log transformed to give a distribution more similar to a normal distribution. Multiple linear regression was used to select anthropometric and/or other variables that were significantly associated with SHBG levels. Linear regression analyses of SNP genotypes on log SHBG levels were adjusted for these covariates (see Results).

The standard r^2 metric of LD between pairs of SNPs was computed using the HaploView program (28). Genotype frequencies were compared between breast cancer cases and controls using unconditional logistic regression for both the 1 *df* Cochran-Armitage trend test and the general 2 *df* heterogeneity test. The Haplo Stats package (version 1.2.0; ref. 29)⁹ was used to estimate haplotype frequencies and to test for associations between haplotypes and SHBG levels or breast cancer status. All statistical tests were two-sided.

Analyses were done using Stata 8.2 (StataCorp) unless specified.

Results

SNPs within the Region of the SHBG Gene. We analyzed 11 tag SNPs within the region of the SHBG gene in the 1,134 postmenopausal control women from Set01 of the SEARCH study for whom SHBG levels were available. Women were ages between 55 and 81 years (mean, 65.2 years) and were sampled on average 15.3 years [SE (mean) = 2.0 years] after their menopause. SHBG levels were in the 6.7 to 205 nmol/L range [geometric mean, 39.3 nmol/L; 95% confidence interval

(95% CI), 38.2-40.4 nmol/L]. Higher SHBG levels were significantly associated with lower BMI, smaller waist and hip circumferences, smaller bra cup size, older age, and former oral contraceptive use (Supplementary Table S1). The best-fitting multivariate model included terms for BMI, waist and hip circumferences, age, and smoking status and accounted for 18% of the variance of log SHBG. The analyses were adjusted for these variables.

The MAFs of the 11 genotyped tagging SNPs within or near to the SHBG gene (Fig. 1) varied between 0.05 and 0.45. Under the log-additive model, 10 of the 11 SNPs were significantly associated with SHBG levels (Table 2). The associations were particularly significant for rs727428 (exponentiated regression coefficient, $e^{\beta} = 0.89$; 95% CI, 0.86-0.92; $P = 5 \times 10^{-12}$, accounting for 4.2% of the variance in log SHBG not accounted for by the nongenetic covariates), rs858518 (in strong LD with rs727428; $r^2 = 0.92$), rs6257 ($e^{\beta} = 0.87$; 95% CI, 0.82-0.91; $P = 2 \times 10^{-7}$), and, in the opposite direction, rs1799941 ($e^{\beta} = 1.13$; 95% CI, 1.08-1.17; $P = 3 \times 10^{-9}$).

Stratifying by BMI above or below 26 (the median) did not affect the results (data not shown). In multiple regression analysis, 3 of the 11 SNPs remained independently significant: rs6259 ($e^{\beta} = 1.15$; 95% CI, 1.08-1.21; $P = 2 \times 10^{-6}$), rs858521 ($e^{\beta} = 0.94$; 95% CI, 0.91-0.98; $P = 0.003$), and rs727428 ($e^{\beta} = 0.85$; 95% CI, 0.82-0.88; $P = 1 \times 10^{-17}$). The best-fitting model accounted for 7.3% of the residual variance in log SHBG. The 2-SNP haplotype comprising the common alleles at rs858521 and rs727428 tags rs1799941 with a $r^2 = 0.85$, whereas the haplotype comprising the rare alleles at both SNPs tags rs6257 ($r^2 = 0.95$); hence, the four most strongly associated SNPs are tagged by this minimal set of 3 SNPs (Supplementary Table S2). Each of these 3 SNPs is also a pairwise tag for an ungenotyped SNP in the region: rs727428 tags rs1624085 with $r^2 = 0.83$, rs858518 tags rs858520 with $r^2 = 1$, and rs6259 tags rs3760213 with $r^2 = 1$.

⁹ <http://mayoresearch.mayo.edu/mayo/research/biostat/schaid.cfm>

SHBG (TAAAA)_n Repeat Polymorphism. The (TAAAA)_n repeat polymorphism in the SHBG promoter region was successfully genotyped in 2,263 controls and 2,279 cases from the SEARCH Set01 case-control panel, including 1,129 of the controls for whom SHBG levels were available. The allele distribution was similar to that reported previously in other White populations (refs. 14, 18; Supplementary Fig. S2); allele lengths ranged from 5 to 12, with the 8, 6, and 9 alleles being the most common, accounting for ~90% of alleles. SHBG levels, adjusted as in the SNP analysis, were inversely correlated with the total number of repeats carried ($e^{\beta} = 0.97$; 95% CI, 0.63-0.98; $P < 0.001$) and also with the length of the shorter and longer allele carried (data not shown). Examining the median SHBG levels for each genotype suggested dichotomizing the repeat between the 6 allele (no length 5 alleles were observed in women with SHBG levels) and those of length ≥ 7 . Under the additive model, each copy of the 6 allele increased the SHBG level by 11% (95% CI, 7-16%; $P < 0.001$).

SHBG Variant Haplotypes and SHBG Levels. There was limited LD between the genotyped SHBG SNPs, as measured by the pairwise r^2 metric (Fig. 2), so that these SNPs defined 11 common haplotypes (labeled A-K in Table 2). When the (TAAAA)_n repeat polymorphism was also considered, 4 of the 11 SNP-based haplotypes were seen with more than one allele of the repeat, giving 18 haplotypes with frequency $>1\%$.

As expected, the residual log SHBG levels were significantly associated with haplotypes ($P = 10^{-12}$, global score test). The most common haplotype [A(6),

25.6%] was associated with the highest SHBG levels. This haplotype is defined by 6 (TAAAA) repeats and the rare allele of rs1799941, which are strongly correlated ($r^2 = 0.89$). Of the remaining 17 haplotypes, 4 were clearly associated with lower SHBG levels [F(8), F(10), E(8), and J(8)]. All of these carry the minor alleles at both rs727428 and rs858518. One haplotype, H(6), containing the minor allele at rs727428 but not rs858518 was also associated with lower levels ($P = 0.004$), consistent with the observation that rs727428 exhibits the stronger association as a single SNP. Two haplotypes, B(8) and I(8), were not associated with reduced SHBG levels despite carrying the rare alleles at rs727428 and rs858518 [e.g., $P = 9 \times 10^{-7}$ and $P = 0.06$, respectively, compared with E(8) as a baseline]. The former is completely correlated with the minor allele of rs6259.

For clarity, the haplotype analysis was repeated using just rs858521, rs6259, and rs727428, the 3 SNPs in the best-fitting multivariate model (Supplementary Table S2). Compared with the baseline (CCT) haplotype, the CCC and GCC are associated with similar reduced SHBG levels, whereas the CTC haplotype is associated with a similar level to the baseline. The GCT haplotype is associated with an intermediate risk.

SHBG Variants and Breast Cancer Risk. All 11 SHBG SNPs and the (TAAAA)_n polymorphism were examined for their association with breast cancer risk in the SEARCH Set01 breast cancer case-control study (Table 1). rs858524, rs858518, and rs6257 were associated with breast cancer risk at $P < 0.1$ on the trend test and were therefore genotyped in Set02. The most significant

Table 1. Geometric mean circulating SHBG levels by genotype for the associated SNPs and the effects of these SNPs on breast cancer risk

Variant	MAF	n	Geometric mean SHBG (adjusted), by genotype			e^{β} * (95% CI)	P (additive model)*	r^2 , the % variance explained by SNP	No. cases/controls	Breast cancer, OR (95% CI; trend)	P_{trend}
			00	01	11						
SHBG rs858524	0.42	1,127	37.9	39.4	42.5	1.06 (1.02-1.09)	0.003	0.8	4,366/4,548	1.07 (1.01-1.13)	0.033
SHBG rs13894	0.076	1,129	40.3	34.9	28.3	0.85 (0.80-0.91)	3×10^{-6}	1.9	2,174/2,262	0.96 (0.82-1.13)	0.64
SHBG rs858521	0.37	1,116	41.2	38.3	38.1	0.95 (0.92-0.99)	0.009	0.6	2,174/2,259	0.96 (0.88-1.04)	0.33
SHBG rs858518	0.44	1,133	43.3	38.9	34.5	0.89 (0.86-0.92)	1×10^{-10}	3.6	4,353/4,542	0.95 (0.90-1.01)	0.13
SHBG rs1799941	0.26	1,134	36.9	41.7	46.6	1.13 (1.08-1.17)	3×10^{-9}	3.0	2,190/2,277	1.03 (0.94-1.13)	0.49
SHBG rs6257	0.12	1,128	40.6	35.4	29.3	0.87 (0.82-0.91)	2×10^{-7}	2.4	6,622/6,784	0.88 (0.82-0.95)	0.002
SHBG rs858517	0.05	1,126	39.7	35.9	26.7	0.89 (0.83-0.97)	0.005	0.7	2,155/2,263	1.16 (0.95-1.40)	0.14
SHBG rs6259	0.12	1,134	38.7	41.1	45.0	1.07 (1.01-1.13)	0.019	0.5	2,188/2,279	0.91 (0.80-1.04)	0.16
SHBG rs727428	0.45	1,129	43.8	38.8	34.3	0.89 (0.86-0.92)	5×10^{-12}	4.2	2,186/2,274	0.93 (0.86-1.02)	0.11
SHBG rs2543553	0.08	1,128	39.3	39.7	40.0	1.01 (0.95-1.08)	0.80	0.0	2,159/2,269	1.05 (0.91-1.22)	0.50
SHBG rs1642796	0.11	1,125	40.2	36.6	31.1	0.90 (0.86-0.96)	3×10^{-4}	1.1	2,181/2,258	1.11 (0.98-1.27)	0.10
SHBG (TAAAA) ₆ vs (TAAAA) _{>6}	0.28	1,125	37.2	40.9	46.7	1.11 (1.07-1.15)	7×10^{-8}	2.5	2,279/2,263	1.02 (0.93-1.12)	0.62

*Exponentiated regression coefficient for ln(SHBG levels) adjusted for BMI, waist, hip, age, and smoking status (see Supplementary Table S1); the unadjusted means were very similar. There were no significant interactions between genotypes and BMI (data not shown).

Table 2. Haplotype analysis of the 11 SNPs in and around the SHBG gene and the (TAAAA)_n repeat polymorphism, for circulating SHBG levels and for the breast cancer case-control study

Haplotype	(TAAAA) _n	rs858524 (G>A)	rs13894 (C>T)	rs858521 (C>G)	rs858518 (T>C)	rs1799941 (G>A)	rs6257 (T>C)	rs858517 (A>G)	rs6259 (G>A)	rs727428 (G>A)	rs2543553 (C>A)	rs1642796 (T>C)	SHBG levels exp(β)* (95% CI)	P	Est. freq in 2,280 controls [†] (%)	Est. freq in 2,192 cases [†] (%)	Odds ratio (95% CI)	P
A(6) (base)	6	1	0	0	0	1	0	0	0	0	0	0			25.5	25.8		
B(8)	8	0	0	0	1	0	0	0	1	1	0	0	0.97 (0.91-1.03)	0.27	12.4	10.9	0.86 (0.74-1.00)	0.054
C(8)	8	0	0	1	0	0	0	0	0	0	0	0	0.92 (0.82-1.04)	0.19	2.4	3.2	1.34 (1.00-1.78)	0.047
C(9)	9												0.95 (0.89-1.01)	0.12	9.2	9.3	1.01 (0.86-1.19)	0.89
C(10)	10												0.96 (0.89-1.04)	0.34	5.4	5.1	0.93 (0.76-1.13)	0.46
D(7)	7	0	0	1	0	0	0	0	0	0	1	0	0.92 (0.78-1.08)	0.31	1.2	1.3	1.09 (0.74-1.61)	0.67
D(8)	8												0.89 (0.80-1.00)	0.043	2.7	2.6	0.93 (0.70-1.23)	0.61
D(9)	9												0.96 (0.86-1.07)	0.46	2.8	3.1	1.06 (0.82-1.39)	0.65
E(8)	8	0	0	1	1	0	1	0	0	1	0	0	0.81 (0.76-0.86)	10⁻¹¹	11.2	9.0	0.79 (0.68-0.92)	0.0033
F(8)	8	0	1	0	1	0	0	0	0	1	0	0	0.69 (0.57-0.83)	10 ⁻⁴	1.1	1.3	1.19 (0.78-1.81)	0.42
F(9)	9												0.87 (0.80-0.95)	0.003	4.7	4.0	0.82 (0.66-1.04)	0.10
F(10)	10												0.66 (0.56-0.77)	10 ⁻⁷	1.4	1.5	1.07 (0.73-1.56)	0.74
G(7)	7	1	0	0	0	0	0	0	0	0	0	0	0.98 (0.87-1.10)	0.73	2.4	2.2	0.91 (0.68-1.22)	0.54
H(6)	6	1	0	0	0	0	0	0	0	1	0	1	0.79 (0.67-0.93)	0.0038	1.5	1.6	0.99 (0.70-1.40)	0.77
I(8)	8	1	0	0	1	0	0	0	0	1	0	0	0.92 (0.80-1.06)	0.26	1.8	2.1	1.13 (0.82-1.56)	0.45
J(8)	8	1	0	0	1	0	0	0	0	1	0	1	0.74 (0.63-0.86)	10 ⁻⁴	1.4	1.6	1.20 (0.83-1.74)	0.34
J(9)	9												0.91 (0.82-1.02)	0.10	3.0	2.9	1.09 (0.84-1.43)	0.50
K(8)	8	1	0	0	1	0	0	1	0	1	0	1	0.85 (0.78-0.93)	10 ⁻⁴	4.1	4.5	1.10 (0.88-1.37)	0.40

NOTE: Global haplotype score test for SHBG levels, $\chi^2_{18} = 95.1$, $P = 10^{-12}$, Global haplotype score test for breast cancer status; $\chi^2_{18} = 38.2$, $P = 0.004$.

Letters A-K denote SNP-haplotypes, with the number in brackets specify the number of (TAAAA) repeats. For example, SNP-haplotype F occurs with 8, 9, or 10 (TAAAA) repeats, designated haplotypes F(8), F(9), and F(10) respectively.

*Exponentiated regression coefficient for log SHBG levels, adjusted for BMI, waist, and hip circumferences, age; and smoking status.

[†] Haplotype frequencies below 1% not shown. Major alleles are represented by 0s; minor alleles represented by 1s in bold type.

SNP, rs6257, was also genotyped in Set03. No other SNPs were significant using the trend test or under the dominant or recessive models and adjusting the analyses for BMI did not materially alter the results (data not shown). Based on 6,622 breast cancer cases and 6,784 controls in Set01/Set02/Set03, the minor allele of rs6257 was associated with a significant reduction in risk (OR, 0.88; 95% CI, 0.82-0.95; $P = 0.002$). There was no evidence for departure from a per-allele model. Although there was some suggestion of a stronger effect for women in the lowest quartile of BMI (<23.1 kg/m²; OR, 0.70; 95% CI, 0.57-0.86; $P = 0.0006$), the interaction was not significant ($P = 0.15$).

The combined analysis of Set01/Set02 gave a borderline significant result for rs858524 (OR, 1.07; 95% CI, 1.01-1.13; $P = 0.033$), which was not significant when rs6257 was included in the model. rs858518 was nonsignificant in combined Set01/Set02 ($P = 0.73$).

There was no evidence for an association between the (TAAAA)_n polymorphism and breast cancer risk either considered as a continuous covariate (OR, 0.99; 95% CI, 0.96-1.02; $P = 0.63$ considering an individual's total number of repeats) or dichotomized into (TAAAA)₆ versus (TAAAA)_{>6} (per-allele OR, 1.02; 95% CI, 0.93-1.12; $P = 0.62$).

The effects of the SHBG variant haplotypes on breast cancer risk were evaluated, assuming a multiplicative model (Table 2). Although the global score test was significant ($P = 0.004$), only haplotype, E(8), uniquely defined by rs6257, showed a significant association with breast cancer, with each copy conferring a 21% reduction in risk (OR, 0.79; 95% CI, 0.68-0.92; $P = 0.003$).

Discussion

SNP Associations with SHBG Levels. Analysis of 11 SNPs within or close to the gene, including SNPs specifically chosen for their ability to tag other common variants in the gene, gave a best-fitting model, which includes just 3 SNPs that together account for >7% of the variance in log SHBG levels not accounted for by BMI, waist, hip, age, or smoking status. The linear combination of these five nongenetic variables with the 3 SNPs accounts for 24% of the variance. Inclusion of the (TAAAA)_n repeat polymorphism in the SHBG promoter region did not improve the fit of the model.

rs727428 was the single most significant SNP ($P = 10^{-11}$), accounting for 4.2% of the variance, with the minor allele being associated with lower SHBG levels. This SNP lies 1.1 kb beyond the 3'-end of the gene, very close to a small peak in the level of interspecies conservation, as reported on University of California-Santa Cruz Genome Browser¹⁰ (see Fig. 1). Areas of high conservation beyond the end of a gene may be relevant for the binding of transcription factors, suggesting a possible functional role for rs727428. rs727428 is in strong LD with rs858518 (within the promoter region, upstream of the regions suggested as being associated with maximal promoter activity; ref. 30); hence, the association between SHBG levels and rs858518 is almost as strong (3.6% of the variance; $P = 10^{-10}$). Neither SNP has been

reported previously in the context of SHBG levels. The only haplotype carrying the C allele of rs727428 without the G allele of rs858518 was rare, giving limited power to distinguish between the two variants. It is also possible that both SNPs are functionally important in determining levels. Alternatively, because the region has not been exhaustively resequenced, it remains possible that the association is due to another as yet unidentified SNP or SNPs in strong LD with rs727428, although the Breast and Prostate Cancer Cohort Consortium sequencing on which the tag SNP selection was based had at least an 85% power to detect SNPs within the gene's coding regions with MAF >5%.

The T allele of rs6259 (encoding the amino acid substitution D356N) has been reported previously as increasing SHBG levels (e.g., ref. 14). In fact, this allele only occurs on a single haplotype, which also carries the minor alleles of rs858518 and rs727428, with rs6259 cancelling out the reduced levels otherwise associated with these SNPs rather than actually increasing levels. The N residue (but not the D) is thought to be the anchor site for an additional N-linked carbohydrate group, which reduces the rate of clearance of SHBG from circulation (13).

In our previous analysis of hormone levels, we reported a positive association between rs1799941 (referred to as 5'-untranslated region g-a) and higher SHBG levels (12), an association also identified in the replication stage of a recent genome-wide study of SHBG levels (21). The effect of this SNP was also clear in the present study and was tagged by the combination of two of the SNPs in the best-fitting model. rs1799941 and (TAAAA)_n are both located in the SHBG promoter region and are in strong LD; rs1799941 is located 8 nucleotides upstream of the transcription initiation site, just outside the FP1 region suggested as regulating expression (30). The 6-repeat allele of (TAAAA)_n was associated with higher SHBG levels, consistent with a study of Swedish men, and with a reported tendency toward higher SHBG levels in hirsute women homozygous for shorter repeats (14, 18), although this appears counter to earlier evidence of lower transcriptional activity in promoters containing only 6 repeats (19). Given their location, both rs1799941 and (TAAAA)_n are plausible functional variants and either or both may explain some of the associations with SHBG levels, but the stronger associations with other markers indicate that they would not be sufficient, even in combination with the amino acid substitution rs6259, to explain all the associations. Thus, at least one further functional variant must exist. Although we cannot reliably distinguish the apparent effect of the short repeat allele from that of rs1799941, haplotype analysis suggests that it is the SNP rather than the repeat polymorphism that has the stronger association and is therefore more likely to be causal.

A recent genome-wide study of SHBG levels in 1,200 European men and women of mixed ages found no associated SNPs outside the region of the SHBG gene and concluded that the best observed association (rs6761, >115 kb distal to SHBG) was driven by weak LD with rs1799941 (21). Based on the Illumina Infinium Human-Hap550 genotyping chip, the study had 65% coverage of SNPs with MAF > 1% located 300 kb either side of the

¹⁰ <http://genome.ucsc.edu/cgi-bin/hgTracks>

SHBG gene, which may explain why it failed to identify the other associated SNPs presented here.

Furthermore, 1,555 of our set of 2,115 women with measured SHBG levels were among the control subjects in stage II of a recent genome-wide breast cancer association study (3) and therefore had genotypes for 12,020 SNPs (~5% of the stage I SNPs, chosen based on their association with breast cancer in stage I), allowing a partial genome-wide association study of SHBG levels. The strongest association was with rs4511593, located 80 kb distal to the SHBG gene. This effect was not significant when adjusted for the SHBG SNPs, suggesting that this too was the consequence of LD. No other SNPs showed significant evidence of association with SHBG levels (see Supplementary Fig. S1 for more details).

To date, no other loci associated with circulating SHBG levels have been identified through either candidate gene association or genome-wide analysis. Nevertheless, approximately three-quarters of the variance in log SHBG levels remains unexplained by the SHBG locus and the other known risk factors considered in this study, and twin and family studies suggest that much of this variance is determined by genetic factors (9-11). Genome scans with more comprehensive coverage may identify additional loci related to SHBG levels. Alternatively, additional resequencing of the SHBG gene and surrounding region may reveal novel SNPs, which act to control SHBG levels either alone or in combination with known variants.

Breast Cancer Association. Previous studies have reported an inverse association between SHBG levels and breast cancer risk. The Endogenous Hormones and Breast Cancer Collaborative Group (6) reported a breast cancer relative risk of 0.88 (95% CI, 0.76-1.03) for a doubling of SHBG levels. Likewise, a study of 677 European postmenopausal women who went on to develop breast cancer and 1,309 controls found marginally lower SHBG levels in cases than controls (geometric mean, 31.5 versus 33.4 nmol/L; $P = 0.04$; ref. 5), whereas a prospective analysis of 297 postmenopausal breast cancer cases and 563 age-matched controls found that cases had 12% lower SHBG levels at baseline than controls ($P < 0.001$), with the risk of breast cancer being 49% lower for those in the highest quintile of SHBG levels compared with the lowest quintile (7).

One aim of our study was to use SHBG levels as a step toward detecting SNPs associated with breast cancer. However, it appears that the relationship between these two phenotypes is not straightforward. The SNP most significantly associated with SHBG levels, rs727428, had no effect on breast cancer risk ($P = 0.11$), whereas rs6257, the SNP with the strongest evidence of an association with breast cancer, had no effect on SHBG levels once rs727428 and rs6259 were taken into account ($P = 0.30$). Given the number of variants tested, it is possible that the association between rs6257 and breast cancer was a false-positive result. A recent case-control study found no association between ovarian cancer risk and SHBG SNPs, including rs6257 (31). On the other hand, rs6257 lies within intron 1 of SHBG and displays very limited pairwise LD with its neighbors, which may explain why it was not identified by previous SNP tagging breast cancer studies. It is therefore worthy of further study.

The apparent lack of an association between the SNPs and breast cancer risk might be taken as evidence that the

reported association between SHBG levels and risk is not causal but due to confounding or chance. However, the effects of the SNPs on SHBG levels, although highly significant, are small in magnitude; hence, the predicted effect on breast cancer will also be small. For example, rs727428 was the SNP showing the strongest univariate evidence of an association with SHBG levels, yet the median SHBG levels for carriers of the common and rare homozygous rs727428 genotypes were 43 and 35 nmol/L, respectively, corresponding to the second and third quintiles of the distribution presented by Zeleniuch-Jacquotte et al. between which the risk of breast cancer was not reported to differ (unadjusted OR, 0.98; ref. 7). Based on the estimated effect of SHBG levels on breast cancer risk from the Endogenous Hormones and Breast Cancer Collaborative Group study, rs727428 would be predicted to be associated with a per-allele OR of 1.02 (95% CI, 1.00-1.05) within the 95% confidence interval for the observed association. This prediction may be an underestimate given that SHBG levels are not measured precisely, but nevertheless it is not surprising that no effect of rs727428 on risk could be identified. Thus, although the current study provides no support for the hypothesis that genetic variants influence breast cancer risk through SHBG levels as an intermediate marker, it would require a much larger case-control study, or the identification of genetic variants with larger effects on levels, to confirm or refute this hypothesis definitively.

Even where a risk association was found, the direction was not that which would have been anticipated. Scaling the Endogenous Hormones and Breast Cancer Collaborative Group (6) estimate of the effect of a doubling of SHBG levels to the size of the rs6257 effect, we would predict rs6257 heterozygotes to have a breast cancer relative risk of 1.02 (95% CI, 0.99-1.06) and homozygotes a relative risk of 1.05 (95% CI, 0.98-1.14). In contrast, the observed heterozygote and homozygote breast cancer ORs were 0.79 (95% CI, 0.68-0.92) and 0.59 (95% CI, 0.35-1.01), respectively, in the opposite direction to the predicted results. In fact, rs6259 and rs858517 were the only 2 of the 11 tested SNPs for which the estimated effect on breast cancer risk was in the direction compatible with its predicted effect on SHBG levels.

In summary, we have identified highly significant associations between at least 3 SNPs within the SHBG gene and circulating SHBG levels in healthy postmenopausal women. However, the magnitudes of the effects were relatively small, and given that the reported associations between SHBG levels and breast cancer risk are also modest, it is not surprising that the most strongly associated SNPs were not associated with breast cancer risk. The genotyping of rs6257 in further large series of breast cancer cases and controls will be necessary to confirm or refute its association with disease risk.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank the European Prospective Investigation of Cancer management team for access to control DNA (K-T. Khaw, S. Bingham, and N. Wareham) and the Eastern Cancer Registry and Intelligence Unit.

References

- Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.
- Easton DF. How many more breast cancer predisposition genes are there? *Breast Cancer Res* 1999;1:14–7.
- Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087–93.
- Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 2007;39:165–7.
- Kaaks R, Rinaldi S, Key TJ, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition. *Endocr Relat Cancer* 2005;12:1071–82.
- Key T, Appleby P, Barnes J, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* 2002;94:606–16.
- Zeleniuch-Jacquotte A, Shore RE, Koenig KL, et al. Postmenopausal levels of oestrogen, androgen, and SHBG and breast cancer: long-term results of a prospective study. *Br J Cancer* 2004;90:153–9.
- Ding EL, Song Y, Malik VS, Liu S. Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA* 2006;295:1288–99.
- Bogaert V, Taes Y, Konings P, et al. Heritability of blood concentrations of sex-steroids in relation to body composition in young adult male siblings. *Clin Endocrinol (Oxf)* 2008;69:129–35.
- Jaquish CE, Blangero J, Haffner SM, Stern MP, MacCluer JW. Quantitative genetics of serum sex hormone-binding globulin levels in participants in the San Antonio Family Heart Study. *Metabolism* 1997;46:988–91.
- Ring HZ, Lessov CN, Reed T, et al. Heritability of plasma sex hormones and hormone binding globulin in adult male twins. *J Clin Endocrinol Metab* 2005;90:3653–8.
- Dunning AM, Dowsett M, Healey CS, et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. *J Natl Cancer Inst* 2004;96:936–45.
- Cousin P, Dechaud H, Grenot C, Lejeune H, Hammond GL, Pugeat M. Influence of glycosylation on the clearance of recombinant human sex hormone-binding globulin from rabbit blood. *J Steroid Biochem Mol Biol* 1999;70:115–21.
- Cousin P, Calemard-Michel L, Lejeune H, et al. Influence of SHBG gene pentanucleotide TAAAA repeat and D327N polymorphism on serum sex hormone-binding globulin concentration in hirsute women. *J Clin Endocrinol Metab* 2004;89:917–24.
- Cui Y, Shu XO, Cai Q, et al. Association of breast cancer risk with a common functional polymorphism (Asp³²⁷Asn) in the sex hormone-binding globulin gene. *Cancer Epidemiol Biomarkers Prev* 2005;14:1096–101.
- Haiman CA, Riley SE, Freedman ML, Setiawan VW, Conti DV, Le Marchand L. Common genetic variation in the sex steroid hormone-binding globulin (SHBG) gene and circulating SHBG levels among postmenopausal women: the Multiethnic Cohort. *J Clin Endocrinol Metab* 2005;90:2198–204.
- Forsti A, Jin Q, Grzybowska E, et al. Sex hormone-binding globulin polymorphisms in familial and sporadic breast cancer. *Carcinogenesis* 2002;23:1315–20.
- Eriksson AL, Lorentzon M, Mellstrom D, et al. SHBG gene promoter polymorphisms in men are associated with serum SHBG, androgen and androgen metabolite levels and hip BMD. *J Clin Endocrinol Metab* 2006;91:5029–37.
- Hogeveen KN, Talikka M, Hammond GL. Human sex hormone-binding globulin promoter activity is influenced by a (TAAAA)_n repeat element within an Alu sequence. *J Biol Chem* 2001;276:36383–90.
- Xita N, Tsatsoulis A, Chatzikiyakidou A, Georgiou I. Association of the (TAAAA)_n repeat polymorphism in the sex hormone-binding globulin (SHBG) gene with polycystic ovary syndrome and relation to SHBG serum levels. *J Clin Endocrinol Metab* 2003;88:5976–80.
- Melzer D, Perry JR, Hernandez D, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* 2008;4:e1000072.
- Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer. Br J Cancer* 1999;80 Suppl 1:95–103.
- Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA. Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet* 2007;3:e42.
- Pooley KA, Healey CS, Smith PL, et al. Association of the progesterone receptor gene with breast cancer risk: a single-nucleotide polymorphism tagging approach. *Cancer Epidemiol Biomarkers Prev* 2006;15:675–82.
- Azzato EM, Driver KE, Lesueur F, et al. Effects of common germline genetic variation in cell cycle control genes on breast cancer survival: results from a population-based cohort. *Breast Cancer Res* 2008;10:R47.
- 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;5:977–85.
- de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet* 2005;37:1217–23.
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–34.
- Janne M, Hammond GL. Hepatocyte nuclear factor-4 controls transcription from a TATA-less human sex hormone-binding globulin gene promoter. *J Biol Chem* 1998;273:34105–14.
- Garcia-Closas M, Brinton LA, Lissowska J, et al. Ovarian cancer risk and common variation in the sex hormone-binding globulin gene: a population-based case-control study. *BMC Cancer* 2007;7:60.