

Quantitative trait loci predicting circulating sex steroid hormones in men from the NCI-Breast and Prostate Cancer Cohort Consortium (BPC3)

Jiyoung Ahn^{1,2}, Fredrick R. Schumacher³, Sonja I. Berndt¹, Ruth Pfeiffer¹, Demetrius Albanes¹, Gerald L. Andriole⁴, Eva Ardanaz^{5,6}, Heiner Boeing⁷, Bas Bueno-de-Mesquita⁸, Stephen J. Chanock¹, Françoise Clavel-Chapelon^{9,10}, W. Ryan Diver¹¹, Heather Spencer Feigelson¹¹, J. Michael Gaziano^{12,14}, Edward Giovannucci^{12,14}, Christopher A. Haiman³, Brian E. Henderson³, Robert N. Hoover¹, Laurence N. Kolonel¹⁵, Peter Kraft¹³, Jing Ma^{12,14}, Loïc Le Marchand¹⁵, Kim Overvad¹⁶, Domenico Palli¹⁷, Pär Stattin¹⁸, Meir Stampfer^{12,14}, Daniel O. Stram³, Gilles Thomas¹, Michael J. Thun¹¹, Ruth C. Travis¹⁹, Dimitrios Trichopoulos^{12,14}, Jarmo Virtamo²⁰, Stephanie J. Weinstein¹, Meredith Yeager¹, Rudolf Kaaks²¹, David J. Hunter¹³ and Richard B. Hayes^{1,2,*}

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, ²Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, 650 First Ave, New York, NY 10016, USA, ³Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, ⁴Division of Urologic Surgery, Washington University School of Medicine, St. Louis, MO, USA, ⁵Public Health Institute of Navarra, Pamplona, Spain, ⁶CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain, ⁷Department of Epidemiology, German Institute of Human Nutrition, Potsdam, Germany, ⁸National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, ⁹Institut National de la Santé et de la Recherche Médicale (Inserm), ERI 20, Paris EA 4045, France, ¹⁰Institut Gustave Roussy, Villejuif F-94805, France, ¹¹Department of Epidemiology and Surveillance Research, American Cancer Society, Williams Street, NW, Atlanta, GA 30303-1002, USA, ¹²Departments of Nutrition and Epidemiology, and ¹³Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA, ¹⁴Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, ¹⁵Cancer Research Center, University of Hawaii, Honolulu, Hawaii, ¹⁶Department of Clinical Epidemiology and Cardiology, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark, ¹⁷Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy, ¹⁸Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Umeå, Sweden, ¹⁹Cancer Epidemiology Unit, University of Oxford, Oxford, UK, ²⁰Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki FIN-00300, Finland and ²¹Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

Received March 10, 2009; Revised and Accepted June 29, 2009

Twin studies suggest a heritable component to circulating sex steroid hormones and sex hormone-binding globulin (SHBG). In the NCI-Breast and Prostate Cancer Cohort Consortium, 874 SNPs in 37 candidate genes in the sex steroid hormone pathway were examined in relation to circulating levels of SHBG ($N = 4720$), testosterone ($N = 4678$), 3α -androstane diol-glucuronide ($N = 4767$) and 17β -estradiol ($N = 2014$) in Caucasian men. rs1799941 in *SHBG* is highly significantly associated with circulating levels of SHBG ($P = 4.52 \times 10^{-21}$), consistent with previous studies, and testosterone ($P = 7.54 \times 10^{-15}$), with mean difference of 26.9

*To whom correspondence should be addressed. Tel: +1 2122639242; Fax: +1 2122638570; Email: richard.b.hayes@nyumc.org

and 14.3%, respectively, comparing wild-type to homozygous variant carriers. Further noteworthy novel findings were observed between SNPs in *ESR1* with testosterone levels (rs722208, mean difference = 8.8%, $P = 7.37 \times 10^{-6}$) and *SRD5A2* with 3α -androstenediol-glucuronide (rs2208532, mean difference = 11.8%, $P = 1.82 \times 10^{-6}$). Genetic variation in genes in the sex steroid hormone pathway is associated with differences in circulating SHBG and sex steroid hormones.

INTRODUCTION

Twin studies suggest a heritable component to circulating sex steroid hormones and sex hormone-binding globulin (SHBG). The sex steroids and SHBG have critical roles in human development, behavior (1) and potentially, prostate carcinogenesis (2). With the completion of the Human Genome (3) and the HapMap Projects (4) providing the genomic basis and the assembly of a large epidemiologic research resource in the NCI-Breast and Prostate Cancer Cohort Consortium (BPC3) (5), we were able to identify quantitative trait loci predicting sex steroid levels in Caucasian men, by pooling data from five independent cohorts in the BPC3.

RESULTS

The NCI-BPC3 is a large multicenter association study (5) that systematically explores the role of genetic variation in genes encoding for the sex steroid hormones in the etiology of breast and prostate cancer, including the American Cancer Society Cancer Prevention Study II (CPS-II), the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study, the European Prospective Investigation into Cancer and Nutrition (EPIC), the Health Professionals Follow-Up Study (HPFP), the Hawaii-Los Angeles Multi-ethnic Cohort (MEC) Study, the Physicians Health Study (PHS); and the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. For 37 genes in the sex steroid hormone pathway (Fig. 1; Supplementary Material, Table S1 for the list of genes), we genotyped 874 SNPs and related them to sex steroid hormone levels, and imputed an additional 3191 SNPs, for supplemental analyses. The relationships between SNPs and sex steroids (SHBG [nmol/l, $N = 4720$], testosterone [ng/ml, $N = 4678$], 3α -androstenediol-glucuronide [ng/ml, $N = 4767$] and 17β -estradiol [pg/ml, $N = 2014$]) were assessed using linear regression, after log transformation of hormone values, adjusting for body mass index (BMI), age, prostate cancer case-control status and cohort. Age and BMI were associated with SHBG, testosterone and 3α -androstenediol-glucuronide levels, consistent with previous studies (6,7), although case-control status was unrelated to these steroid levels (Fig. 2; inter-quartile ranges of sex steroids according to each covariate). There was some heterogeneity in the measured sex steroid levels between cohorts, because the steroid hormones for cohorts were assayed in different laboratories at different points in time (Supplementary Material, Table S2; hormone assay methods).

Several SNPs in *SHBG* on chromosome 17p13 were associated with circulating SHBG (Fig. 3A), most strongly marked by genotyped SNP rs1799941 (Fig. 4, blue circles), located 67 base pairs upstream from the transcription start site of *SHBG*

and conferring a 25% mean difference in blood levels between variant and wild-type homozygotes ($P = 4.52 \times 10^{-21}$; geometric means for wild-type, heterozygote, homozygote = 42.63, 49.01 and 54.10 nmol/l, respectively; Table 1). The relationship of rs1799941 with circulating SHBG was consistent across strata of five independent cohorts (Table 1), age, BMI and case-control status (Supplementary Material, Table S3; all P for heterogeneity >0.05). Although eight other genotyped SNPs in *SHBG* were also strongly associated with circulating SHBG (Fig. 4; Supplementary Material, Table S4; $P < 10^{-5}$), when these eight SNPs [which were not strongly correlated with rs1799941 ($r^2 < 0.7$)] were simultaneously included in the multivariable model with rs1799941, three SNPs remained significantly associated with circulating SHBG: rs1799941 ($P = 3.66 \times 10^{-7}$), rs6257 ($P = 0.005$) and rs9913778 ($P = 7.98 \times 10^{-9}$). In a log-additive genetic model, rs1799941 and rs9913778 explain 2.0 and 1.0% of the variance in log-transformed circulating SHBG, respectively.

SHBG is a glycoprotein that regulates steroid bioavailability by binding with highest affinity to dihydrotestosterone (DHT), followed by testosterone and then estradiol (2). Interestingly, circulating SHBG was also predicted by rs10754396 ($P = 7.12 \times 10^{-6}$; Fig. 3A), a SNP located in the 3'untranslated region of the *HSD3B2* gene, which encodes for hydroxy-delta-5-steroid dehydrogenase, an enzyme irreversibly inactivating DHT. When rs10754396 (*HSD3B2*) and rs1799441 (*SHBG*) were simultaneously included in the multivariable model, both SNPs remained independently associated with circulating SHBG concentrations ($P = 4.54 \times 10^{-6}$ for rs10754396; $P = 1.08 \times 10^{-20}$ for rs1799941) and there was no evidence of interaction ($P_{\text{interaction}} = 0.45$).

In addition to its association with circulating SHBG, rs1799941 in *SHBG* was also strongly associated with circulating levels of testosterone [$P = 7.54 \times 10^{-15}$; geometric means = 4.56, 5.00 and 5.21 ng/ml; Figs 3B and 4 (red diamonds), and Table 1). Consistent with our findings, a recent genome-wide association study reported an association of rs1799941 with circulating SHBG levels ($P = 3.08 \times 10^{-7}$) (8) and a candidate gene study reported an association of rs1799941 with testosterone ($P < 0.01$), in addition to SHBG levels (9).

In our study, circulating SHBG and testosterone were correlated (Spearman partial correlation $r = 0.48$, P -value <0.001 ; age, BMI and study controlled) and, since the clearance and bioavailability of testosterone are related to SHBG levels (2), genetic determinants of SHBG could be expected to also exhibit an association with testosterone. In order to disentangle the SHBG and testosterone interrelationship, we examined associations of rs1799941 in relation to SHBG adjusted-testosterone concentration, calculated by residuals of testosterone regressed on SHBG, showing that rs1799941 also tended

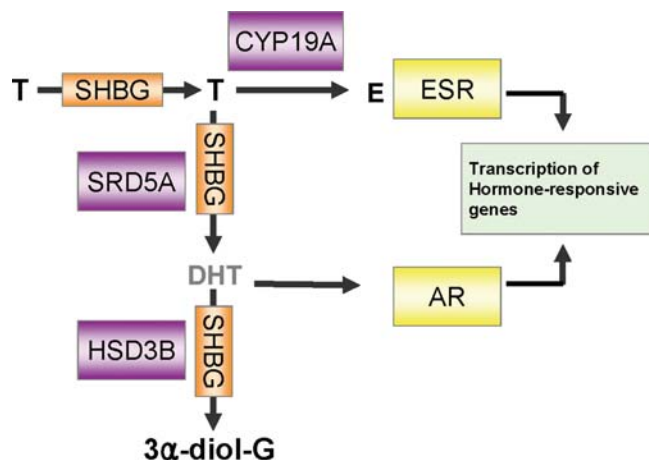


Figure 1. Sex steroid metabolism. In males, testosterone (T), carried by SHBG, is ultimately converted to the active metabolite dihydrotestosterone (DHT) by SRD5A (steroid-5- α -reductase), which binds to the androgen receptor (AR) with greater affinity than testosterone, leading to the activation of the transcription factor function of the androgen receptor. DHT is inactivated in the prostate by HSD3B (hydroxy- δ -5-steroid dehydrogenase) to 3 α -androstane- α -diol-glucuronide (3 α -diol-G). In the prostate, testosterone is converted by aromatase (CYP19A) to estradiol (E), which further interacts with estrogen receptors (ESR).

to predict SHBG adjusted-testosterone levels ($P = 1.33 \times 10^{-3}$). Nonetheless, measurement error and misclassification from a single measurement of these analyses may have precluded complete adjustment for underlying physiologic (versus genetic) correlations.

We also observed notable novel associations of circulating testosterone with several SNPs in *ESR1*, a gene encoding for estrogen receptor- α (Fig. 3B), with the strongest association for the genotyped SNP rs722208 ($P = 7.37 \times 10^{-6}$), conferring -8.8% mean difference in testosterone between variant and wild-type homozygotes (Table 1). This association was consistent across studies (Table 1) and strata of age, BMI and case-control status (Supplementary Material, Table S4; all P for heterogeneity >0.05). Three other genotyped SNPs (rs3020411, rs926778 and rs2348078), located in the same linkage disequilibrium block (Fig. 5), spanning exon 5 and 7 which encode for the estrogen binding domain of estrogen receptor- α (10), were also strongly associated with circulating testosterone (Supplementary Material, Table S4; P -value $<10^{-5}$); however, on simultaneous multivariable adjustment of these SNPs, none were more strongly associated than the others with testosterone, due to high linkage disequilibrium ($r^2 > 0.7$). The statistical association of SNPs in *ESR1* with testosterone may be due to physiologic correlations, as *ESR1* null mutations (11) are thought to impact circulating testosterone through feedback regulation via the hypothalamic-pituitary-gonadal axis (12). Estrogen receptor- α is essential for prostate development (13). Also, an *in vitro* reporter gene assay suggests that an *ESR1* splice variant missing exon 7, resulting from a SNP in intron 6 (rs2273207), confers decreased levels of transcription in the presence of estrogen compared with wild-type (14).

Since both rs722208 in *ESR1* and rs1799941 in *SHBG* predicted circulating testosterone levels, we evaluated joint effect of these two SNPs. Having greater cumulative number

of alleles of rs722208 (A) and rs1799941 (A) was associated with increasing testosterone levels (mean testosterone level = 4.22, 4.56, 4.73, 5.16 and 5.22 ng/ml for 0, 1, 2, 3 and 4 allele carriers, respectively; $P = 8.11 \times 10^{-6}$), with each SNP contributing additively to the association, as indicated by the test for multiplicative interaction ($P_{\text{interaction}} = 0.75$).

SRD5A2 on chromosome 2p23 encodes for steroid-5- α -reductase, which converts testosterone to the androgenically more potent DHT primarily in the prostate (2). Because DHT has rapid turnover (2), its downstream metabolite 3 α -androstane- α -diol-glucuronide is considered a circulating biomarker for intraprostatic DHT synthesis. In our study, rs2208532, in intron 1 of *SRD5A2*, was not associated with testosterone ($P = 0.94$), but was associated with circulating 3 α -androstane- α -diol-glucuronide ($P = 8.51 \times 10^{-7}$; -11.8% mean difference; Table 1; Fig. 3C). A previous study reported rs9282858 SNPs in exon 1 (A49T) of *SRD5A2* to be a predictor of 3 α -androstane- α -diol-glucuronide ($P = 0.001$) (15); this rare SNP (1% MAF in SNP500) was not included in our study.

Rs727479, located in intron 2 of *CYP19A1*, encoding for aromatase which converts androgens to estrogens, predicted blood estradiol levels ($P = 5.06 \times 10^{-5}$; Table 1; Fig. 3D). Consistent with our data, this SNP and another nearby SNPs (rs749292) spanning the coding and proximal 5'untranslated region of *CYP19A1* were shown to be significantly associated with a 10–20% decrease in estradiol levels in postmenopausal women (16).

To further explore the loci identified with hormone levels, we imputed the common SNPs in these gene regions from HapMap. Generally, similar associations were noted for imputed SNPs (Fig. 3A–D, shown as empty circles). Two imputed SNPs, rs7755185 in *ESR1* and rs559555 in *SRD5A2*, showed marginally stronger associations than genotyped SNPs in relation to testosterone [$P = 3.78 \times 10^{-6}$ for rs7755185 (imputed) versus $P = 7.37 \times 10^{-6}$ for rs3020411 (genotyped) in *ESR1*] and 3 α -androstane- α -diol-glucuronide [$P = 4.79 \times 10^{-7}$ for rs559555 (imputed) versus $P = 1.82 \times 10^{-6}$ for rs2208532 (genotyped) in *SRD5A2*], respectively; however, the imputed SNPs were highly correlated with nearby genotyped SNPs ($r^2 = 0.91$ for rs7755185 and rs3020411 in *ESR1*; $r^2 = 0.95$ for rs559555 and rs2208532 in *SRD5A2*). Results for SNPs and serum sex steroid levels were also similar when we used cohort-specific standardized Z scores for sex steroids, instead of absolute values (data not shown).

DISCUSSION

In a large consortium based study, we found a significant impact of genetic factors on circulating SHBG and sex steroid hormone levels, independent of age and BMI. We identified SNPs in *SHBG* associated with circulating SHBG and testosterone levels at high levels of statistical confidence. Notable novel associations were also identified relating SNPs in *ESR1* and *SRD5A2* with testosterone and 3 α -androstane- α -diol-glucuronide, respectively.

This is the first comprehensive characterization of genetic variation in the sex steroid pathway with circulating hormone levels. The strengths of the study include its large

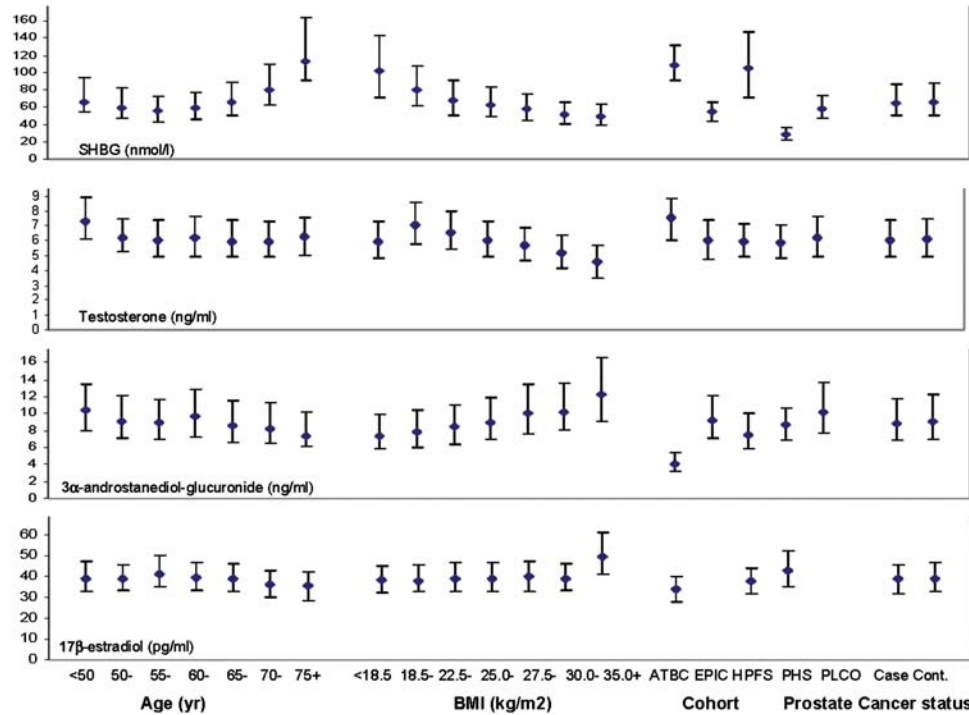


Figure 2. Median (inter-quartile range) levels of SHBG and sex steroids according to age, BMI, cohort and prostate cancer case–control status in the NCI-BPC3.

sample size, with findings confirmed in multiple independent populations. A potential limitation of our study is measurement of only a single circulating sex steroid measurement; hormone measures at multiple time points would have resulted in more precise estimates of circulating levels.

The underlying mechanisms for these gene-phenotype associations are unknown; they may mark for directly operating genetic causes in some cases, as is likely with gene variants in *SHBG* (e.g. as marked by rs1799941) and circulating SHBG, or indirectly through physiologic pathways, as may be the case for the *ESR1* SNP and circulating testosterone and for *SRD5A2* SNP and circulating androstane-diol-glucuronide.

In summary, genetic variation in genes in the sex steroid hormone pathway is associated with differences in circulating SHBG and sex steroid hormones.

MATERIALS AND METHODS

Study population

The rationale and background of BPC3 have been described elsewhere (5,17). Briefly, the prostate cancer study includes seven case–control studies nested within seven cohorts: CPS-II, the ATBC Study, the EPIC, the HPFS, the MEC, the PHS and the PLCO Cancer Screening Trial (see supplement methods for the details about the participating cohorts). Controls were matched to cases by age and ethnicity, and in some cohorts, additional matching criteria were employed (i.e. EPIC matched on country of residence). Caucasians of European descent are the predominant ethnic group in all cohorts, except the MEC. In this study, we limited our

analysis of genetic variation in the sex steroid hormone pathway (Supplementary Material, Table S1; gene list) to Caucasian men with hormone measurements available (ATBC, EPIC, HPFS, PHS and PLCO). This study was approved by the institutional review boards at all institutions.

Selection of SNPs and genotyping

A comprehensive characterization of genetic variation in steroid hormone genes with the sequencing, high density genotyping panels, and genotyping data generated in BPC3 are publicly available at both the USC/Norris Comprehensive Cancer website (<http://www.uscnorris.com/MECGenetics>) and the NCI Core Genotyping facility (<http://cgf.nci.gov/cohort.cfm>). As described previously (17), haplotype-tags were selected for the initial steroid hormone genes based on the correlation (r^2_{HT}) between the observed haplotypes and the haplotypes predicted by selected SNPs (18) using a high density genotyping panel and genotyping conducted using TaqMan assays in four cohort-specific laboratories. Genotyping of 169 haplotype tagging SNPs in 20 genes were conducted in five separate batches between June 2003 and February 2007. Information about the primers and probes for each TaqMan assay, and sample Taqman scatter plots from each genotyping center can be found on the University of Southern California website (<http://www.uscnorris.com/MECGenetics>). In order to substantially expand the density of genotyping and to add additional newly hypothesized genes, we selected tag SNPs using a pairwise r^2 procedure (19), which was programmed to capture alleles with an $r^2 \geq 0.8$ and a minor allele frequency $>2\%$ from the high density genotyping panel and $>5\%$ from the phase II

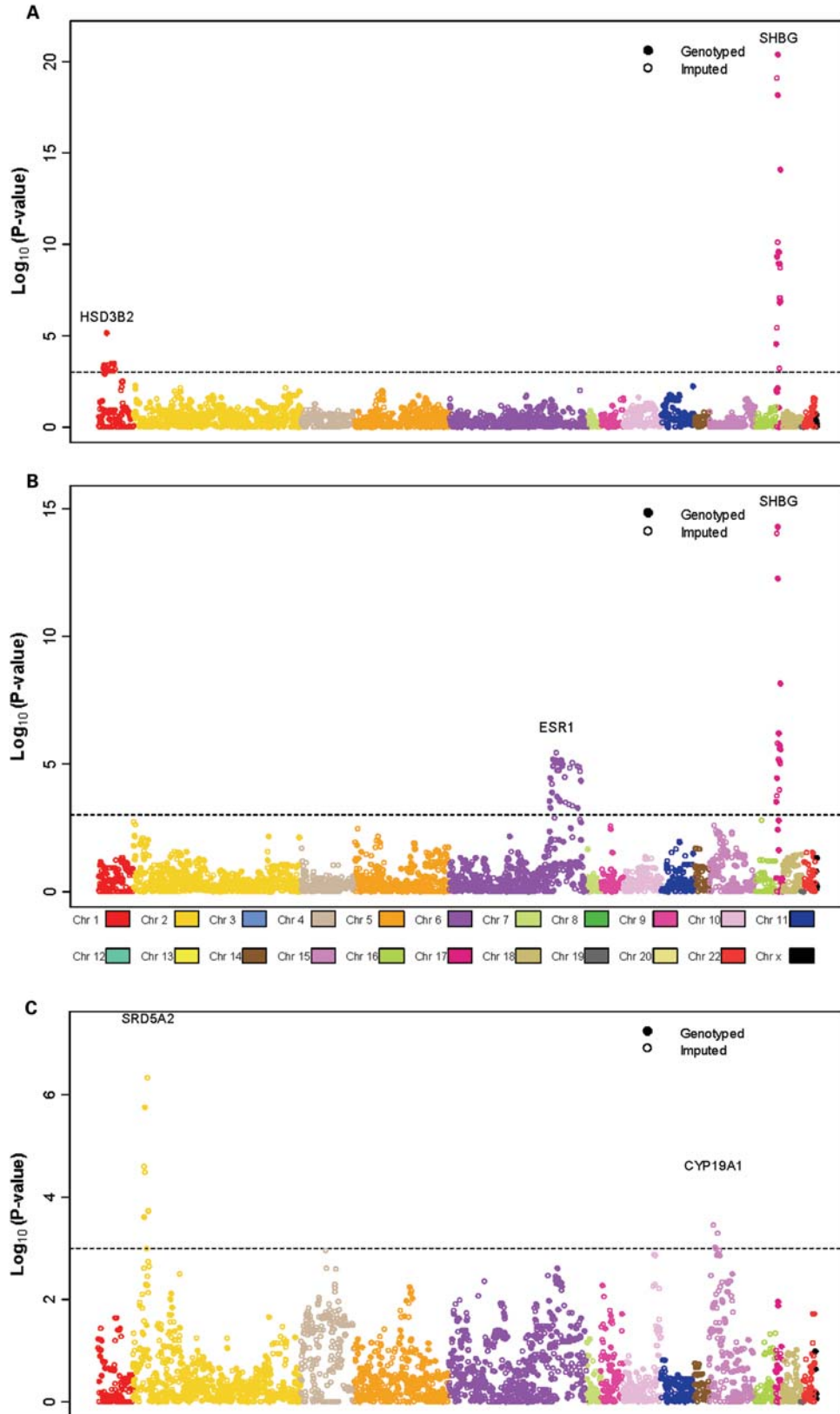


Figure 3. Associations of SNPs in sex hormone metabolizing genes with blood sex hormone concentrations in the NCI-BPC3. *P*-values (1 df test) for association of SNPs with circulating sex steroid levels SHBG (A), testosterone (B), 3 α -androstane-20-one-3 β -ol-20-one glucuronide (C) and 17 β -estradiol (D). The closed circles represent genotyped SNPs and the empty circles represent imputed SNPs.

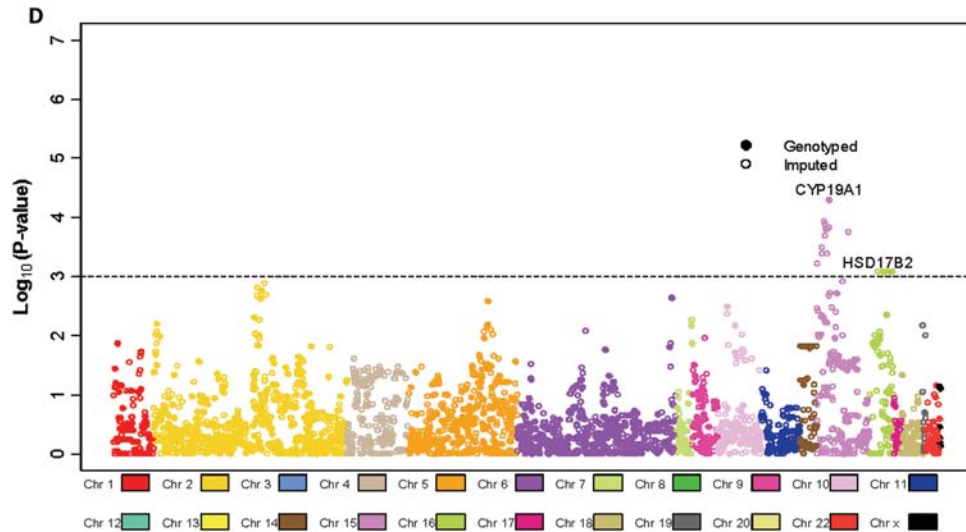


Figure 3. Continued

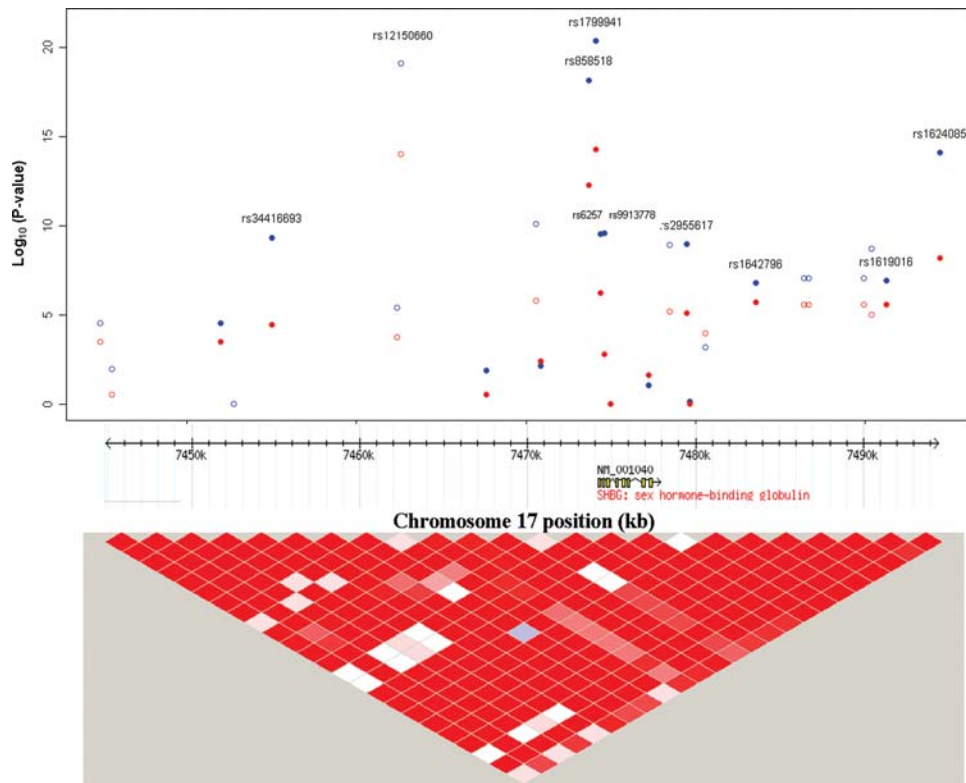


Figure 4. Associations of SNPs in the *SHBG* region with circulating SHBG and testosterone concentrations in the NCI-BPC3. *P*-values (1 df test) for association of *SHBG* SNPs with circulating SHBG (blue circles) and testosterone (red diamonds) levels. The closed circles or diamonds represent genotyped SNPs and the empty circles or diamonds represent imputed SNPs. LD plot at the bottom is based on the measure *D'* and LOD (bright red, *D'* = 1 and LOD score ≥ 2 ; shades of pink/red, *D'* < 1 and LOD score ≥ 2 ; white, *D'* < 1 and LOD score < 2; blue, *D'* = 1 and LOD < 2).

HapMap CEU panel. The selection of tag SNPs was limited to those with an Illumina design score >0.4 and genotyping was carried out with the Golden Gate Illumina BeadArray platform in four laboratories (USC, NCI, UK and Harvard). Thirty HapMap CEU trios were genotyped in all labs to examine inter-lab reproducibility and the concordance was 99.5%.

Within each study, blinded duplicate samples (~5%) were included with concordance estimates of 97.2–99.9% across the studies. Approximately 94% of the SNPs attempted yielded reliable genotypes. Greater than 95% of the polymorphisms yielded genotyping success rates greater than 90%.

Table 1. Association of blood SHBG and sex steroid hormone levels with sex-hormone polymorphisms in men from NCI-BPC3

Gene	dbSNP ^a (LOC ^b)	Sex steroid hormone		N	Risk allele	Mean wild-type	Mean hetero	Mean homo	% Difference (homo versus wt)	P _{value} (1df) ^c	P _{heterogeneity} ^d
SHBG	rs1799941	SHBG (nmol/l)	Total	4195	A (0.25)	42.63	49.01	54.10	26.91	4.52E-21	0.26
			HPFS	1240	A (0.24)	50.76	62.35	62.93	23.97	2.36E-06	
			PHS	403	A (0.25)	20.22	22.55	22.89	13.21	0.04	
			ATBC	112	A (0.22)	83.91	87.70	95.55	13.88	0.26	
			EPIC	1064	A (0.25)	41.95	48.43	56.18	33.92	1.08E-05	
			PLCO	1376	A (0.27)	41.15	45.11	51.93	26.20	4.13E-08	
SHBG	rs1799941	Testosterone (ng/ml)	Total	4153	A (0.25)	4.56	5.00	5.21	14.25	7.54E-15	0.46
			HPFS	1238	A (0.24)	4.07	4.52	4.95	21.62	6.59E-08	
			PHS	402	A (0.25)	4.20	4.41	4.63	10.24	0.07	
			ATBC	111	A (0.22)	5.66	5.90	6.68	18.02	0.21	
			EPIC	1031	A (0.25)	4.79	5.37	5.43	13.36	1.72E-05	
			PLCO	1371	A (0.27)	4.48	4.82	4.98	11.16	0.0006	
ESR1	rs722208	Testosterone (ng/ml)	Total	4214	G (0.28)	4.89	4.75	4.46	-8.79	7.16E-06	0.97
			HPFS	1277	G (0.27)	4.30	4.22	3.93	-8.60	0.023	
			PHS	415	G (0.26)	4.48	4.42	4.05	-9.60	0.13	
			ATBC	111	G (0.33)	5.91	5.78	5.19	-12.18	0.29	
			EPIC	1038	G (0.27)	5.13	4.97	4.84	-5.65	0.10	
			PLCO	1373	G (0.29)	4.81	4.61	4.37	-9.15	0.0057	
SRD5A2	rs2208532	3 α -androstane- diol-glucuronide (ng/ml)	Total	4282	G (0.43)	5.84	5.40	5.15	-11.77	1.82E-06	0.75
			HPFS	1257	G (0.41)	5.85	5.34	5.51	-5.84	0.12	
			PHS	416	G (0.44)	7.58	6.80	6.45	-14.94	0.01	
			ATBC	102	G (0.45)	2.36	2.84	1.92	-18.64	0.30	
			EPIC	1131	G (0.43)	6.91	6.53	5.76	-16.69	0.0005	
			PLCO	1376	G (0.43)	7.30	6.56	6.54	-10.49	0.0083	
CYP19A1	rs727479	17 β -estradiol (pg/ml)	Total	1914	C (0.35)	31.62	30.63	28.67	-9.34	5.06E-05	0.38
			HPFS	1243	C (0.37)	32.23	31.14	29.74	-7.74	0.0019	
			PHS	543	C (0.35)	31.05	29.76	27.09	-12.75	0.018	
			ATBC	128	C (0.32)	27.98	29.41	22.81	-18.48	0.16	
			EPIC	NA	C (0.35)	NA	NA	NA	NA	NA	
			PLCO	NA	C (0.34)	NA	NA	NA	NA	NA	

^aSNP identifier based on NCBI dbSNP; rs1799441, rs3020411, rs2208532 and rs727479 were all genotyped.

^bChromosomal location based on NCBI Human Genome Build 35 coordinates.

^cThe result of 1 df test based on generalized linear regression, using a log additive model in men, after log transformation of hormone concentrations—adjusted for age, BMI, prostate cancer case and control status, and cohort/study center.

^dWe tested for heterogeneity in trend estimates by cohorts and by disease status based on Q statistics.

Date filtering and imputation

Data from the Taqman and Golden Gate platforms were filtered separately. Any sample where greater than 25% of the SNPs attempted on a given platform failed was removed from the data set. Data were then filtered by study to remove poorly performing SNPs: all SNPs that failed on 25% or more samples were excluded from the data set, as were all SNPs that showed statistically significant ($P < 10^{-5}$) deviations from HWE genotype frequencies among European-ancestry controls, and all SNPs with MAF < 1%. Any SNP that was missing or excluded in more than three studies or exhibited large differences in European-ancestry allele frequencies across cohorts ($F_{st} > 0.02$) was excluded from further analysis. Average 90% of SNPs per gene after QC were included.

We imputed SNPs that were polymorphic in the 37 gene regions in any of the HapMap reference panels using observed genotypes from the BPC3 subjects and phased haplotypes from HapMap samples (release #21) using the MACH2QTL

program (Y. Li *et al.*, submitted for publication, 20). Imputation was performed stratified by study and ethnicity, and the imputed data were filtered by study and ethnicity to remove poorly imputed SNPs. SNPs with an estimated correlation between the imputed and true genotypes of less than 30% were excluded from analysis (Y. Li *et al.*, submitted for publication). Any imputed SNP that was excluded in more than three European-ancestry strata was excluded from analysis.

Sex steroid hormone measurement

Levels of androstane- α -diol-glucuronide (ng/ml), testosterone (ng/ml), estradiol (pg/ml) and sex hormone binding globulin (SHBG; nmol/l) were previously measured in blood samples of prostate cancer cases and control subjects from prospective cohort studies [SHBG (nmol/l, $N = 4720$), testosterone (ng/ml, $N = 4678$), 3 α -androstane- α -diol-glucuronide (ng/ml, $N = 4767$) and 17 β -estradiol (pg/ml, $N = 2014$)] (7,21–24). The blood

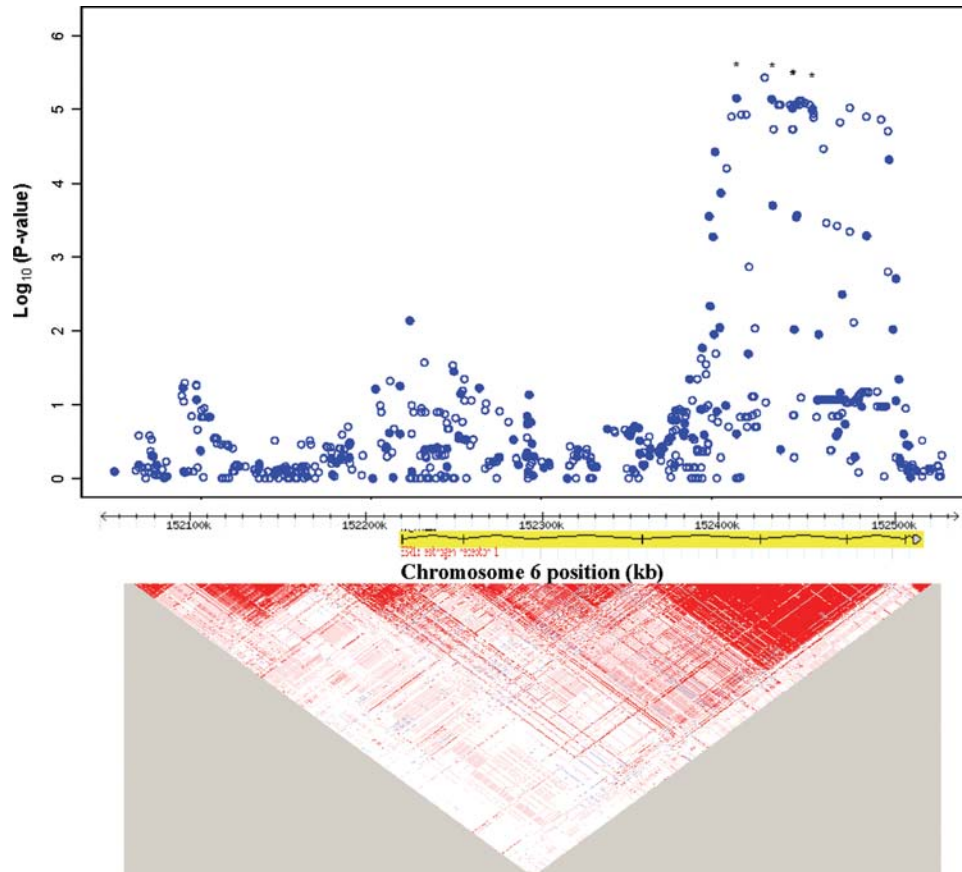


Figure 5. Associations of tagSNPs in the *ESRI* region with circulating testosterone concentrations in the NCI-BPC3. The closed circles represent genotyped SNPs and the empty circles represent imputed SNPs. LD plot at the bottom is based on the measure D' and LOD (bright red, $D' = 1$ and LOD score ≥ 2 ; shades of pink/red, $D' < 1$ and LOD score ≥ 2 ; white, $D' < 1$ and LOD score < 2 ; blue, $D' = 1$ and LOD < 2). *Four strongest genotyped SNPs, by P -value, were rs722208, rs3020411, rs926778 and rs2348078 (in order from left to right).

samples were collected at study baseline in these cohorts, before prostate cancers developed. Detailed information on sex steroid hormone assay methods and coefficients of variations are shown in the Supplementary Material, Table S2.

Data analysis

A total of 874 genotyped SNPs and an additional 3191 imputed SNPs from 37 genes in the sex steroid hormone pathway were available for the data analysis (Supplementary Material, Table S1). Analysis was limited to Caucasian men with sex steroid assays completed. The relationships between SNPs and sex steroids were assessed using linear regression, after log transformation of hormone values, adjusting for BMI, age, prostate cancer case–control status and cohort. The log transformation both improved normality of the distribution of the assay values and helped to stabilize the variances of the assays performed using different methodologies. We tests for heterogeneity among cohorts, based on Q statistics (25). To disentangle the SHBG and testosterone correlation, we calculated SHBG adjusted-testosterone levels, by creating residuals of testosterone regressed on SHBG, and examined associations of SHBG variants with the SHBG adjusted-testosterone level. Pairwise

linkage disequilibrium measures (D' and r^2) were estimated using the program, Haploview (<http://www.broad.mit.edu/personal/jcbarret/haploview/>). All statistical analyses were conducted using SAS version 9.1, software (SAS Institute, Inc., Cary, NC, USA) and PLINK (<http://pngu.mgh.harvard.edu/purcell/plink/>). *R* was used to generate the figures.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

FUNDING

National Cancer Institute cooperative agreements UO1-CA98233, UO1-CA98710, UO1-CA98216 and UO1-CA98758 and by the Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics.

REFERENCES

1. Bagatell, C. and Bremner, W.J. (2003) *Androgens in Health and Disease (Contemporary Endocrinology)*. Humana Press, Totowa, NJ.
2. Henderson, B.E., Ponder, B. and Ross, R.K. (2003) *Hormones, Genes, and Cancer*. Oxford University Press, New York, NY.
3. Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F. *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.
4. The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
5. Hunter, D.J., Riboli, E., Haiman, C.A., Albanes, D., Altshuler, D., Chanock, S.J., Haynes, R.B., Henderson, B.E., Kaaks, R., Stram, D.O. *et al.* (2005) A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat. Rev. Cancer*, **5**, 977–985.
6. Mohr, B.A., Bhasin, S., Link, C.L., O'Donnell, A.B. and McKinlay, J.B. (2006) The effect of changes in adiposity on testosterone levels in older men: longitudinal results from the Massachusetts Male Aging Study. *Eur. J. Endocrinol.*, **155**, 443–452.
7. Travis, R.C., Key, T.J., Allen, N.E., Appleby, P.N., Roddam, A.W., Rinaldi, S., Egevad, L., Gann, P.H., Rohrmann, S., Linseisen, J. *et al.* (2007) Serum androgens and prostate cancer among 643 cases and 643 controls in the European Prospective Investigation into Cancer and Nutrition. *Int. J. Cancer*, **121**, 1331–1338.
8. Melzer, D., Perry, J.R., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J.R., Paolisso, G. *et al.* (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet.*, **4**, e1000072.
9. Eriksson, A.L., Lorentzon, M., Mellstrom, D., Vandenput, L., Swanson, C., Andersson, N., Hammond, G.L., Jakobsson, J., Rane, A., Orwoll, E.S. *et al.* (2006) SHBG gene promoter polymorphisms in men are associated with serum sex hormone-binding globulin, androgen and androgen metabolite levels, and hip bone mineral density. *J. Clin. Endocrinol. Metab.*, **91**, 5029–5037.
10. Gennari, L., Merlotti, D., De, P.V., Calabro, A., Becherini, L., Martini, G. and Nuti, R. (2005) Estrogen receptor gene polymorphisms and the genetics of osteoporosis: a HuGE review. *Am. J. Epidemiol.*, **161**, 307–320.
11. Grumbach, M.M. and Auchus, R.J. (1999) Estrogen: consequences and implications of human mutations in synthesis and action. *J. Clin. Endocrinol. Metab.*, **84**, 4677–4694.
12. Hayes, F.J., DeCruz, S., Seminara, S.B., Boepple, P.A. and Crowley, W.F. Jr. (2001) Differential regulation of gonadotropin secretion by testosterone in the human male: absence of a negative feedback effect of testosterone on follicle-stimulating hormone secretion. *J. Clin. Endocrinol. Metab.*, **86**, 53–58.
13. Prins, G.S., Birch, L., Couse, J.F., Choi, I., Katzenellenbogen, B. and Korach, K.S. (2001) Estrogen imprinting of the developing prostate gland is mediated through stromal estrogen receptor alpha: studies with alphaERKO and betaERKO mice. *Cancer Res.*, **61**, 6089–6097.
14. Weickert, C.S., Miranda-Angulo, A.L., Wong, J., Perlman, W.R., Ward, S.E., Radhakrishna, V., Straub, R.E., Weinberger, D.R. and Kleinman, J.E. (2008) Variants in the estrogen receptor alpha gene and its mRNA contribute to risk for schizophrenia. *Hum. Mol. Genet.*, **17**, 2293–2309.
15. Hayes, V.M., Severi, G., Padilla, E.J., Morris, H.A., Tilley, W.D., Southey, M.C., English, D.R., Sutherland, R.L., Hopper, J.L., Boyle, P. and Giles, G.G. (2007) 5alpha-Reductase type 2 gene variant associations with prostate cancer risk, circulating hormone levels and androgenetic alopecia. *Int. J. Cancer*, **120**, 776–780.
16. Haiman, C.A., Dossus, L., Setiawan, V.W., Stram, D.O., Dunning, A.M., Thomas, G., Thun, M.J., Albanes, D., Altshuler, D., Ardanaz, E. *et al.* (2007) Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res.*, **67**, 1893–1897.
17. Kraft, P., Pharoah, P., Chanock, S.J., Albanes, D., Kolonel, L.N., Hayes, R.B., Altshuler, D., Andriole, G., Berg, C., Boeing, H. *et al.* (2005) Genetic variation in the HSD17B1 gene and risk of prostate cancer. *PLoS Genet.*, **1**, e68.
18. Stram, D.O. (2004) Tag SNP selection for association studies. *Genet. Epidemiol.*, **27**, 365–374.
19. Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L. and Nickerson, D.A. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.*, **74**, 106–120.
20. de Bakker, P.I., Burt, N.P., Graham, R.R., Guiducci, C., Yelensky, R., Drake, J.A., Bersaglieri, T., Penney, K.L., Butler, J., Young, S. *et al.* (2006) Transferability of tag SNPs in genetic association studies in multiple populations. *Nat. Genet.*, **38**, 1298–1303.
21. Platz, E.A., Leitzmann, M.F., Rifai, N., Kantoff, P.W., Chen, Y.C., Stampfer, M.J., Willett, W.C. and Giovannucci, E. (2005) Sex steroid hormones and the androgen receptor gene CAG repeat and subsequent risk of prostate cancer in the prostate-specific antigen era. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 1262–1269.
22. Gann, P.H., Hennekens, C.H., Ma, J., Longcope, C. and Stampfer, M.J. (1996) Prospective study of sex hormone levels and risk of prostate cancer. *J. Natl. Cancer Inst.*, **88**, 1118–1126.
23. Dorgan, J.F., Albanes, D., Virtamo, J., Heinonen, O.P., Chandler, D.W., Galmarini, M., McShane, L.M., Barrett, M.J., Tangrea, J. and Taylor, P.R. (1998) Relationships of serum androgens and estrogens to prostate cancer risk: results from a prospective study in Finland. *Cancer Epidemiol. Biomarkers Prev.*, **7**, 1069–1074.
24. Weiss, J.M., Huang, W.Y., Rinaldi, S., Fears, T.R., Chatterjee, N., Hsing, A.W., Crawford, E.D., Andriole, G.L., Kaaks, R. and Hayes, R.B. (2008) Endogenous sex hormones and the risk of prostate cancer: a prospective study. *Int. J. Cancer*, **122**, 2345–2350.
25. Higgins, J.P. and Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Stat. Med.*, **21**, 1539–1558.