CYP17 Genetic Variation and Risk of Breast and Prostate Cancer from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3)

CYP17 Genetic Variation and Risk of Breast and Prostate Cancer from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3)

Veronica Wendy Setiawan,[1] Fredrick R. Schumacher,[2,3] Christopher A. Haiman,[1] Daniel O. Stram,[1] Demetrius Albanes,[4] David Altshuler,[5] Göran Berglund,[6] Julie Buring,[7] Eugenia E. Calle,[10] Françoise Clavel-Chapelon,[11] David G. Cox,[2,3] J. Michael Gaziano,[8,12] Susan E. Hankinson,[3,9] Richard B. Hayes,[4] Brian E. Henderson,[1] Joel N. Hirschhorn,[5] Robert Hoover,[4] David J. Hunter,[2,3] Rudolf Kaaks,[13] Laurence N. Kolonel,[14] Peter Kraft,[2,3] Jing Ma,[9] Loïc Le Marchand,[14] Jakob Linseisen,[13] Eiliv Lund,[15] Carmen Navarro,[16] Kim Overvad,[17] Domenico Palli,[18] Petra H.M. Peeters,[19] Malcolm C. Pike,[1] Elio Riboli,[20] Meir J. Stampfer,[3] Michael J. Thun,[10] Ruth C. Travis,[21] Dimitrios Trichopoulos,[3] Meredith Yeager,[4] Regina G. Ziegler,[4] Heather Spencer Feigelson,[10] and Stephen J. Chanock[4]

[1]Department of Preventive Medicine, University of Southern California, Los Angeles, California; [2]Program in Molecular and Genetic Epidemiology, [3]Department of Epidemiology, Harvard School of Public Health, Boston Massachusettes; [4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; [5]Broad Institute of Harvard and MIT, Cambridge, Massachusetts; [6]Department of Clinical Sciences, Malmö University Hospital, Malmö, Sweden; Divisions of [7]Preventive Medicine and [8]Aging and the [9]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [10]American Cancer Society, Atlanta, Georgia; [11]INSERM, Institut Gustave Roussy, Villejuif, France; [12]VA Boston Healthcare System, Boston, Massachusetts; [13]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; [14]Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, Hawaii; [15]Institute of Community Medicine, University of Tromsö, Tromsö, Norway; [16]Department of Epidemiology, Murcia Health Council, Murcia, and CIBER Epidemiologia y Salud Publica (CIBERESP), Spain; [17]Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark; [18]Molecular and Nutritional Epidemiology Unit, SCPO-Scientific Institute of Tuscany, Florence, Italy; [19]Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands; [20]Imperial College, London, United Kingdom; and [21]Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom

Abstract

***CYP17* encodes cytochrome p450c17α, which mediates activities essential for the production of sex steroids. Common germ line variation in the *CYP17* gene has been related to inconsistent results in breast and prostate cancer, with most studies focusing on the nonsynonymous single nucleotide polymorphism (SNP) *T27C* (rs743572). We comprehensively characterized variation in *CYP17* by direct sequencing of exons followed by dense genotyping across the 58 kb region around *CYP17* in five racial/ethnic populations. Two blocks of strong linkage disequilibrium were identified and nine haplotype-tagging SNPs, including *T27C*, were chosen to predict common haplotypes ($R^2_h \geq 0.85$). These haplotype-tagging SNPs were genotyped in 8,138 prostate cancer cases and 9,033 controls, and 5,333 breast cancer cases and 7,069 controls from the Breast and Prostate Cancer Cohort Consortium. We observed borderline significant associations with prostate cancer for rs2486758 [*TC* versus *TT*, odds ratios (OR), 1.07; 95% confidence intervals (95% CI), 1.00-1.14; *CC* versus *TT*, OR, 1.09; 95% CI, 0.95-1.26; *P* trend = 0.04] and rs6892 (*AG* versus *AA*, OR, 1.08; 95% CI, 1.00-1.15; *GG* versus *AA*, OR, 1.11; 95% CI, 0.95-1.30; *P* trend = 0.03). We also observed marginally significant associations with breast cancer for rs4919687 (*GA* versus *GG*, OR, 1.04; 95% CI, 0.97-1.12, *AA* versus *GG*, OR, 1.17; 95% CI, 1.03-1.34; *P* trend = 0.03) and rs4919682 (*CT* versus *CC*, OR, 1.04; 95% CI, 0.97-1.12; *TT* versus *CC*, OR, 1.16; 95% CI, 1.01-1.33; *P* trend = 0.04). Common variation at *CYP17* was not associated with circulating sex steroid hormones in men or postmenopausal women. Our findings do not support the hypothesis that common germ line variation in *CYP17* makes a substantial contribution to postmenopausal breast or prostate cancer susceptibility.** (Cancer Epidemiol Biomarkers Prev 2007;16(11):2237–46)

Introduction

A large body of evidence suggests that genetic susceptibility plays an important role in breast and prostate cancer etiology (1). Given the importance of steroid hormones in breast and prostate carcinogenesis (2), we

Received 6/27/07; revised 8/13/07; accepted 8/29/07.

Grant support: 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. National Cancer Institute Career Development Award CA116543 (V.W. Setiawan).

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.

Requests for reprints: Veronica Wendy Setiawan, Department of Preventive Medicine, USC/Norris Comprehensive Cancer Center, 1441 Eastlake Avenue, Room 4425, Los Angeles, CA 90033. Phone: 323-865-0411; Fax: 323-865-0127. E-mail: vsetiawa@usc.edu

doi:10.1158/1055-9965.EPI-07-0589

hypothesized that common germ line variation in genes related to sex steroid hormone biosynthesis and metabolism could alter the function of these genes and the proteins they encode, thus altering breast and prostate cancer risk.

Genetic variation in one particular gene in this pathway, *CYP17*, has been studied extensively in relation to gonadal development and the synthesis of androgens and estrogens. *CYP17* encodes cytochrome p450c17α, which mediates both steroid 17α-hydroxylase and 17,20-lyase activities that are essential for the production of glucocorticoids and sex steroids. Mutations in *CYP17* cause 17-hydroxylase deficiency, a rare form of congenital adrenal hyperplasia, and sexual infantilism (3). Approximately 40 different mutations in *CYP17* have been reported to cause 17-hydroxylase deficiency (3, 4). The absence of 17,20-lyase activity results in impaired production of gonadal sexual steroids and high levels of basal progesterone (which is a substrate of 17α-hydroxylase; ref. 4). These severe clinical phenotypes have led researchers to hypothesize that *CYP17* may also play a role in polycystic ovarian syndrome, a hormone-dependent disorder characterized by hyperandrogenism and chronic annovulation (5).

Genetic variation in *CYP17* has been studied extensively in relation to breast and prostate cancer (6-10), and most of these studies have focused on the single nucleotide polymorphism (SNP) rs743572 (denoted *T27C* or A1/A2), located 34 bp upstream from the translation start site of the gene (5). The *T27C* SNP has been associated with endogenous hormone levels and cancers of the breast and prostate in some but not all studies (reviewed in refs. 6-8, 11-13). This SNP was originally thought to create an additional Sp-1 binding site in the *CYP17* promoter (5) and to influence steroid hormone exposure and risk of hormone-related cancer through increasing *CYP17* expression. A later study, however, showed that this variant allele does not influence Sp-1 binding (14). It is unclear whether this SNP marks important variation located in the vicinity of the gene because a comprehensive assessment of common genetic variation in *CYP17* in association with endogenous steroid hormone levels and risk of breast and prostate cancer has, thus far, not been done.

Here, we did a systematic evaluation of common genetic variation in *CYP17* using a combination of approaches that included resequencing of the coding region to identify common and rare missense variation, and a haplotype-based analysis to characterize common patterns of genetic variation across the entire locus, including the regulatory and noncoding regions. Subsequent tests of association with risk of breast and prostate cancer and circulating levels of steroid hormones were done in a large multicenter collaborative study, the Breast and Prostate Cancer Cohort Consortium (BPC3).

Materials and Methods

Study Population. The rationale and background of BPC3 have been described elsewhere (1), as well as the details about the participating cohorts, case-control selection procedures, and characteristics (15-17). Briefly, the breast cancer study includes five case-control studies nested within the following cohorts: the American Cancer Society Cancer Prevention Study II, the European Prospective Investigation into Cancer and Nutrition (EPIC), the Harvard Nurses' Health Study, the Women's Health Study, and the Hawaii-Los Angeles Multiethnic Cohort (MEC). The prostate cancer study includes seven case-control studies nested within these cohorts: Cancer Prevention Study II, the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, EPIC, the Health Professionals Follow-up Study, the MEC, the Physicians Health Study, and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO). 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). With the exception of the MEC (which also consists of African Americans, Native Hawaiian, Japanese Americans, and Latinos) and PLCO (which also includes African Americans), the majority of the participants (>90%) in the other cohorts were Whites of European descent. Thus, our study focused on genetic variation at the *CYP17* locus common among Whites of European ancestry. A total of 5,333 breast cancer cases and 7,069 controls, and 8,138 prostate cancer cases and 9,033 controls were included in the present analysis. The majority of breast cancer cases were postmenopausal. This study was approved by the Institutional Review Boards at all institutions.

SNP Discovery and Haplotype-Tagging SNP Selection. We used a combination of genomic approaches that included resequencing of the coding regions to identify possible causal alleles directly and a haplotype-based analysis to characterize common patterns of genetic variation across the *CYP17* locus. The eight exons and neighboring exon-intron junctions of *CYP17* were resequenced in a panel of 190 advanced breast (stage ≥ 2) and advanced prostate cancer (Gleason >7 or stage ≥ 2) cases from the MEC (19 of each cancer from each of the five ethnic groups: African American, Native Hawaiian, Japanese American, Latino, and White). Bidirectional sequence analysis was conducted at the Broad Institute (Cambridge, MA). No common (minor allele frequency >5%) novel missense SNP was identified from resequencing. Linkage disequilibrium (LD) structure of the *CYP17* locus was characterized in a multiethnic panel of 349 subjects from the MEC (including 70 Whites) by genotyping 29 common SNPs (minor allele frequency >5% among Whites). These SNPs were distributed over a 58-kb region across the *CYP17* locus, from 32 kb 5′ of exon 1 through 19 kb 3′ of the 3′ untranslated region (the average spacing of one SNP per 2 kb). The minor allele frequencies for all SNPs used for LD and haplotype characterization are provided in Supplemental Table S1.

The $|D'|$ and r^2 statistics were used to assess pairwise LD between the 29 common SNPs. Within regions of strong LD (18), haplotype frequencies estimates were constructed from genotype data among Whites within blocks using the expectation-maximization algorithm of Excoffier and Slatkin (19). The squared correlation (R^2_h) between the true haplotypes (h) and their estimates were then calculated as described by Stram et al. (20). Haplotype-tagging SNPs (htSNP) for the BPC3 case-control samples were then chosen by finding the minimum set of SNPs that would have $R^2_h \geq 0.7$ for all common haplotypes with an estimated frequency of ≥5%.

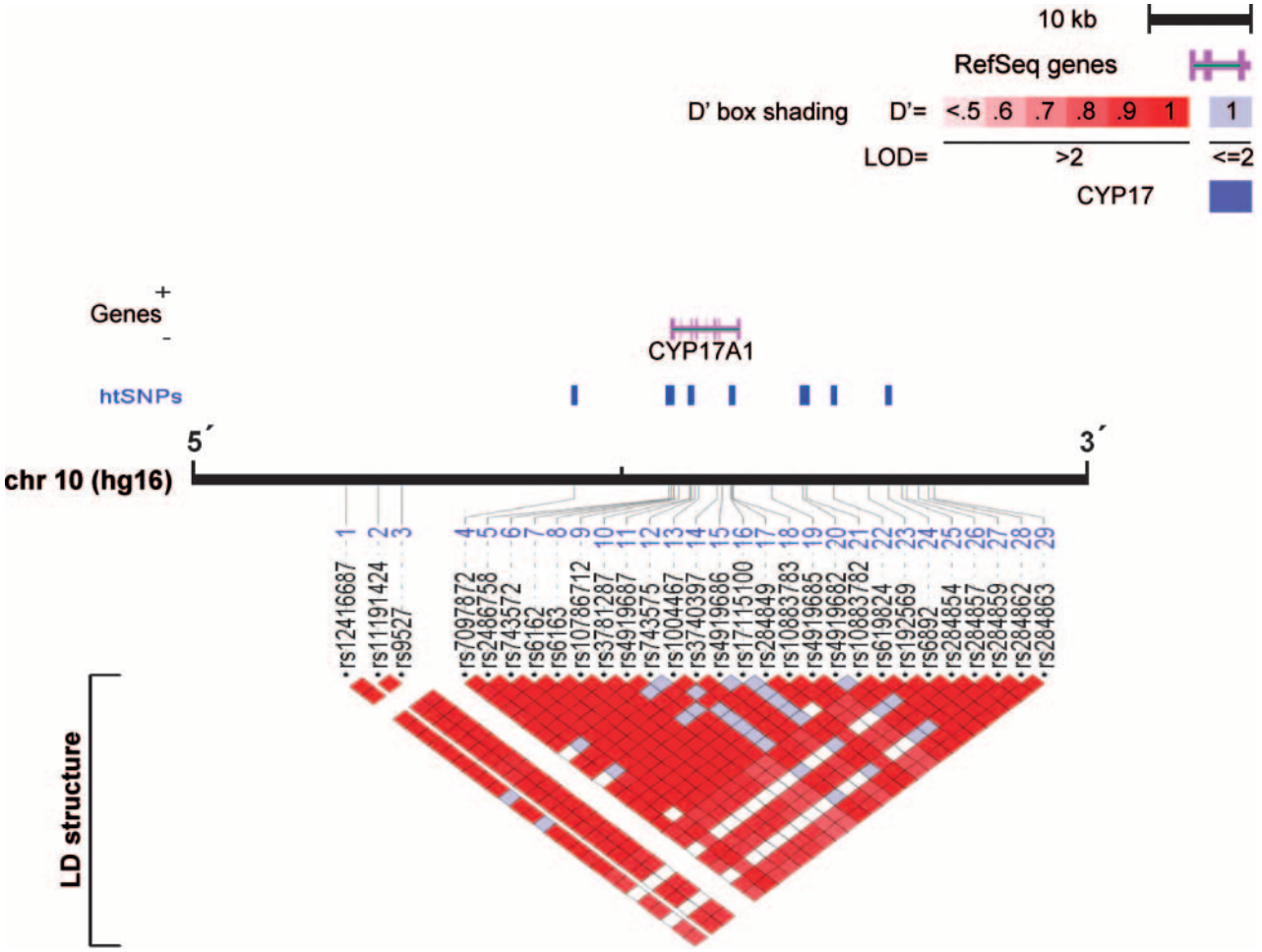

Figure 1. SNPs and LD structure (among Whites) across the *CYP17* locus. Twenty-nine SNPs were selected covering a 58 kb region around *CYP17* at an average density of one SNP per 2 kb.

Genotyping of htSNPs in Case-Control Samples. Genotyping of htSNPs in the case-control samples was done in six laboratories [University of Southern California (Los Angeles, CA), University of Hawaii (Honolulu, HI), Harvard School of Public Health (Boston, MA), Core Genotyping Facility, National Cancer Institute (Bethesda, MD), IARC (Lyon, France), and Cambridge University (Cambridge, United Kingdom)] using a fluorescent 5 endonuclease assay and the ABI-PRISM 7900 for sequence detection (TaqMan; Applied Biosystems, Inc.). Assay information is available in the MEC Genetics web site.[22] Sequence validation for each SNP assay was done and 100% concordance was observed in the 102 samples of the SNP500 Cancer project[23] (21). To assess interlaboratory variation, each genotyping center ran assays on a designated set of 94 samples from the Coriell Biorepository (Camden, NJ) drawn from the SNP500 Cancer set. The internal quality of genotype data at each genotyping center was assessed by typing 5% to 10% blinded samples in duplicate or triplicate (depending on the study). The genotyping concordance was >99% among and within centers for blinded quality control samples. The average genotyping success rate across cohorts was 96.7% (ranging from 93.0% to 97.7%). No deviation from fitness for Hardy-Weinberg proportion was observed (at the $P < 0.01$ level) across more than one study for any given SNP.

Hormone Assay. Levels of androstenedione, testosterone, estrone, and estradiol were measured in prospectively collected samples from women (n = 3,723) as part of previous studies within the EPIC (22), MEC (23), and the Nurses' Health Study (24). Levels of androstenedione and testosterone were measured in men (n = 4,960) within the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (25), EPIC (26), Health Professionals Follow-up Study (27), Physicians Health Study (28), and PLCO. Levels of estradiol had been measured in men (n = 2,174) within the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (25), the Health Professionals Follow-up Study (27), and the Physicians Health Study (28). Hormone levels were measured by either direct or indirect RIA depending on the study. Details of hormone

[22] http://www.uscnorris.com/MECgenetics
[23] http://snp500cancer.nci.nih.gov

Table 1. Associations of *CYP17* common haplotypes with breast cancer in the BPC3

	Haplotype copies	Cases (n = 5,333)*	Controls (n = 7,069)*	OR† (95% CI)	P heterogeneity‡
Block 1					
CTAGG	0	2,119	2,806	1.00 (ref.)	0.60
	1	2,435	3,277	1.00 (0.93-1.07)	
	2	734	946	1.01 (0.91-1.13)	
				P trend = 0.89	
TTGAG	0	2,847	3,832	1.00 (ref.)	0.99
	1	2,029	2,706	1.03 (0.96-1.11)	
	2	412	491	1.17 (1.02-1.33)	
				P trend = 0.04	
CCAGT	0	3,455	4,571	1.00 (ref.)	0.61
	1	1,625	2,178	0.96 (0.89-1.03)	
	2	208	280	0.98 (0.83-1.17)	
				P trend = 0.38	
TTGGG	0	4,041	5,285	1.00 (ref.)	0.50
	1	1,136	1,606	0.93 (0.86-1.01)	
	2	111	137	0.97 (0.78-1.20)	
				P trend = 0.13	
Global test block 1: χ^2 = 6.18, 4 *df*, P = 0.186					
Block 2					
TAAA	0	2,943	3,958	1.00 (ref.)	0.97
	1	1,960	2,616	1.04 (0.97-1.11)	
	2	385	456	1.17 (1.02-1.35)	
				P trend = 0.03	
CACA	0	3,296	4,432	1.00 (ref.)	0.25
	1	1,763	2,298	1.02 (0.95-1.10)	
	2	229	299	1.01 (0.85-1.19)	
				P trend = 0.61	
CACG	0	3,585	4,749	1.00 (ref.)	0.41
	1	1,521	2,039	0.96 (0.89-1.03)	
	2	183	241	1.00 (0.83-1.19)	
				P trend = 0.41	
CAAA	0	3,461	4,595	1.00 (ref.)	0.59
	1	1,588	2,119	0.99 (0.92-1.06)	
	2	239	314	0.95 (0.82-1.11)	
				P trend = 0.54	
CGCA	0	3,786	4,904	1.00 (ref.)	0.81
	1	1,378	1,955	0.94 (0.87-1.02)	
	2	124	170	0.96 (0.77-1.20)	
				P trend = 0.17	
Global test block 2: χ^2 = 6.71, 5 *df*, P = 0.243					

*Numbers do not sum to total due to missing genotype data.
† The ORs and 95% CIs were estimated using conditional (on matching factors: age, ethnicity, and country within EPIC) logistic regression and adjusted for cohort.
‡ P value for heterogeneity in ORs across cohorts.

assays used in each study have been previously published.

Statistical Analysis. The statistical methods used have been described previously (16, 17). Briefly, we used the method of Zaykin et al. (29) to test for the association between haplotypes and cancer risk. We estimated haplotype frequencies using the combined set of cases and controls and the expected counts carrying zero, one, or two copies of each haplotype for each individual, conditional on that individual's genotype data. This was calculated separately for each cohort (and country within EPIC or ethnicity in the MEC and PLCO). These expectations were substituted for true counts and indicator variables in conditional logistic regression to form score tests (of no haplotype-specific risk) and to estimate odds ratios (OR) and 95% confidence intervals (CI). ORs were estimated for each common haplotype using all other haplotypes as the reference group. The matching factors in the conditional logistic regression were age, ethnicity, and country within EPIC. For SNP analyses, ORs were estimated using conditional logistic regression using the more common genotype as the reference group.

To assess the consistency of genetic effects across ethnicity or cohorts, we first tested for heterogeneity across ethnicity or cohorts prior to pooling genetic data. These tests were done using a likelihood ratio test following the inclusion of an interaction term between each haplotype (or SNP) and ethnicity or cohort in the logistic regression model. If tests for heterogeneity were not statistically significant, we pooled genetic data across ethnicity and cohorts.

To test the global null hypothesis of no association between variation in *CYP17* haplotypes and risk of breast or prostate cancer, we used a likelihood ratio test comparing a model with additive effects for each common haplotype (treating the most common haplotype as the reference group) to the intercept-only model.

We adjusted for known risk factors for breast and prostate cancer in logistic models. The risk factors for breast cancer included age at menarche, age at

menopause, menopausal status, parity, body mass index (BMI), and use of postmenopausal hormones, family history of breast cancer, and personal history of benign breast disease. The risk factors for prostate cancer included BMI and family history of prostate cancer. Because further adjustment for the above risk factors did not affect our results, we only present results from the conditional analyses.

We also assessed possible interaction effects between *CYP17* SNPs and haplotypes and several risk factors using the likelihood ratio test as described above. Finally, we evaluated the association of *CYP17* variation with advanced breast cancer (metastatic and regional disease), high-stage prostate cancer (stage $\geq T_{3b}$, N_1, or M_1), and high-grade prostate cancer (Gleason score ≥ 8).

In the hormone analyses, all hormone values were natural log-transformed to provide approximately normal distributions. Geometric mean hormone levels according to haplotypes or SNPs were calculated using linear regression analysis while adjusting for age at blood draw, assay batch, BMI, ethnicity, country within EPIC, cohort, and case-control status. We performed a hormone-haplotype global test within each *CYP17* block for all common haplotypes. The global test for hormone-haplotype effect was done by using an *F* test to compare the sum of the squared residuals for a full model with all haplotypes within a block and a nested model without haplotypes. All analyses were conducted in SAS 9.0 (SAS Institute), and all statistical tests were two-sided.

Table 2. Associations of *CYP17* htSNPs with breast cancer in the BPC3

	Cases (n = 5,333)*	Controls (n = 7,069)*	OR (95% CI)†	P heterogeneity‡
Block 1				
rs7097872				
CC	1,709	2,249	1.00 (ref.)	0.70
CT	2,415	3,219	1.00 (0.93-1.08)	
TT	842	1,101	1.04 (0.94-1.15)	
			P trend = 0.53	
rs2486758				
TT	3,177	4,203	1.00 (ref.)	0.95
TC	1,716	2,284	0.97 (0.90-1.04)	
CC	251	346	0.96 (0.82-1.12)	
			P trend = 0.36	
rs743572 (*T27C*)				
AA	1,869	2,474	1.00 (ref.)	0.66
AG	2,445	3,338	0.99 (0.92-1.06)	
GG	833	1,070	1.05 (0.95-1.16)	
			P trend = 0.49	
rs4919687				
GG	2,765	3,742	1.00 (ref.)	0.85
GA	1,992	2,644	1.04 (0.97-1.12)	
AA	409	486	1.17 (1.03-1.34)	
			P trend = 0.03	
rs284849				
GG	3,383	4,472	1.00 (ref.)	0.55
GT	1,594	2,142	0.96 (0.89-1.03)	
TT	208	283	0.96 (0.81-1.14)	
			P trend = 0.30	
Block 2				
rs4919682				
CC	2,855	3,833	1.00 (ref.)	0.90
CT	1,912	2,528	1.04 (0.97-1.12)	
TT	379	453	1.16 (1.01-1.33)	
			P trend = 0.04	
rs10883782				
AA	3,650	4,722	1.00 (ref.)	0.59
AG	1,358	1,913	0.95 (0.88-1.03)	
GG	123	170	0.96 (0.77-1.19)	
			P trend = 0.21	
rs619824				
CC	1,526	2,082	1.00 (ref.)	0.86
CA	2,562	3,456	1.01 (0.94-1.10)	
AA	1,070	1,359	1.08 (0.98-1.18)	
			P trend = 0.14	
rs6892				
AA	3,489	4,611	1.00 (ref.)	0.33
AG	1,482	1,998	0.96 (0.89-1.03)	
GG	184	238	1.01 (0.85-1.21)	
			P trend = 0.46	

*Numbers do not sum to total due to missing genotype data.
† The ORs and 95% CIs were estimated using conditional (on matching factors: age, ethnicity, and country within EPIC) logistic regression and adjusted for cohort.
‡ P value for heterogeneity in ORs across cohorts.

Table 3. Associations of *CYP17* common haplotypes with prostate cancer in the BPC3

	Haplotype copies	Cases (n = 8,138)*	Controls (n = 9,033)*	OR (95% CI)†	P heterogeneity‡
Block 1					
CTAGG	0	3,287	3,586	1.00 (ref.)	0.33
	1	3,696	4,123	0.99 (0.93-1.06)	
	2	1,156	1,324	0.98 (0.89-1.08)	
				P trend = 0.27	
TTGAG	0	4,457	4,874	1.00 (ref.)	0.99
	1	3,128	3,497	0.98 (0.92-1.05)	
	2	553	662	0.91 (0.80-1.03)	
				P trend = 0.16	
CCAGT	0	5,279	6,011	1.00 (ref.)	0.25
	1	2,504	2,657	1.06 (0.99-1.13)	
	2	355	364	1.06 (0.91-1.23)	
				P trend = 0.02	
TTGGG	0	6,171	6,876	1.00 (ref.)	0.95
	1	1,804	1,979	1.02 (0.94-1.10)	
	2	163	178	1.03 (0.82-1.29)	
				P trend = 0.65	
Global test block 1: χ^2 = 3.62, 4 *df*, P = 0.460					
Block 2					
TAAA	0	4,617	5,048	1.00 (ref.)	0.84
	1	3,006	3,375	0.98 (0.92-1.05)	
	2	515	610	0.91 (0.80-1.04)	
				P trend = 0.17	
CACA	0	4,968	5,518	1.00 (ref.)	0.69
	1	2,775	3,065	1.01 (0.94-1.08)	
	2	395	450	0.96 (0.83-1.11)	
				P trend = 0.92	
CACG	0	5,408	6,182	1.00 (ref.)	0.29
	1	2,389	2,528	1.07 (0.99-1.14)	
	2	340	323	1.14 (0.97-1.34)	
				P trend = 0.002	
CAAA	0	5,318	5,907	1.00 (ref.)	0.92
	1	2,396	2,601	1.04 (0.97-1.12)	
	2	424	525	0.93 (0.80-1.07)	
				P trend = 0.53	
CGCA	0	6,031	6,655	1.00 (ref.)	0.33
	1	1,928	2,172	0.98 (0.91-1.05)	
	2	179	206	1.00 (0.81-1.23)	
				P trend = 0.50	
Global test block 2: χ^2 = 7.59, 5 *df*, P = 0.180					

*Numbers do not sum to total due to missing genotype data.
† The ORs and 95% CIs were estimated using conditional (on matching factors: age, ethnicity, and country within EPIC) logistic regression and adjusted for cohort.
‡ P value for heterogeneity in ORs across cohorts.

Results

Characterization of *CYP17* Genetic Variation and htSNP Selection. The LD pattern across the *CYP17* locus is shown in Fig. 1. We identified two regions of strong LD: block 1 (SNPs 1-19, size = 42 kb) and block 2 (SNPs 20-29, size = 13 kb; Supplemental Table S1). The interblock distance was only 3 kb. In each block, four to five common (≥ 5%) haplotypes were observed and they accounted for at least 91% of all chromosomes in Whites (Supplemental Table S2). Nine htSNPs that strongly predicted these common haplotypes were selected; the minimum R^2_h in each block was ≥0.85. We also genotyped the htSNPs in the HapMap CEU trios to assess the coverage of common variation in relation to the HapMap database (HapMap Data Rel 19/phase II Oct 05, on NCBI B34 assembly, dbSNP b124). These nine htSNPs predict, with a minimum $r^2 \geq 0.7$, 70% of all common SNPs (n = 46) genotyped in the HapMap CEU population across this region. Thus, we believe that these selected htSNPs adequately predict the vast majority for common genetic variation at this locus.

The common haplotype frequencies within each LD block among the White controls in each cohort are provided in Supplemental Table S2; the haplotype frequencies were consistent across cohorts.

Breast Cancer Analysis. We observed no significant heterogeneity in risk associated with any haplotype across cohorts ($P \geq 0.25$) or across ethnicity ($P \geq 0.26$); thus, here we report results from only pooled analyses (Table 1). Cohort- and ethnic-specific results are available in Supplemental Tables S3 and S4. The global test for comparison of haplotype frequencies in cases and controls in either block was not statistically significant (block 1, P = 0.19; block 2, P = 0.24), however, having two copies of *TTGAG* in block 1 (OR, 1.17; 95% CI, 1.02-1.33) and *TAAA* in block 2 (OR, 1.17; 95% CI, 1.02-1.35) were significantly associated with breast cancer.

The htSNP-specific ORs did not show significant heterogeneity across cohorts ($P \geq 0.33$) or across

Table 4. Associations of *CYP17* htSNPs with prostate cancer in the BPC3

	Cases (n = 8,138)*	Controls (n = 9,033)*	OR† (95% CI)	P heterogeneity‡
Block 1				
rs7097872				
CC	2,573	2,899	1.00 (ref.)	0.59
CT	3,560	4,054	1.00 (0.93-1.07)	
TT	1,189	1,425	0.94 (0.85-1.03)	
			P trend = 0.26	
rs2486758				
TT	4,824	5,558	1.00 (ref.)	0.24
TC	2,655	2,821	1.07 (1.00-1.14)	
CC	435	435	1.09 (0.95-1.26)	
			P trend = 0.04	
rs743572 (*T27C*)				
AA	2,901	3,197	1.00 (ref.)	0.85
AG	3,811	4,230	1.00 (0.94-1.07)	
GG	1,236	1,407	0.97 (0.88-1.06)	
			P trend = 0.57	
rs4919687				
GG	4,359	4,760	1.00 (ref.)	0.96
GA	3,080	3,432	0.98 (0.92-1.05)	
AA	547	656	0.90 (0.80-1.02)	
			P trend = 0.17	
rs284849				
GG	5,195	5,899	1.00 (ref.)	0.23
GT	2,485	2,622	1.06 (0.99-1.13)	
TT	351	362	1.05 (0.90-1.22)	
			P trend = 0.14	
Block 2				
rs4919682				
CC	4,524	4,934	1.00 (ref.)	0.93
CT	2,968	3,332	0.98 (0.92-1.04)	
TT	520	616	0.91 (0.80-1.03)	
			P trend = 0.17	
rs10883782				
AA	5,856	6,451	1.00 (ref.)	0.28
AG	1,904	2,143	0.98 (0.91-1.05)	
GG	181	207	0.99 (0.80-1.21)	
			P trend = 0.62	
rs619824				
CC	2,446	2,594	1.00 (ref.)	0.77
CA	3,789	4,265	0.96 (0.89-1.03)	
AA	1,648	1,878	0.96 (0.87-1.04)	
			P trend = 0.27	
rs6892				
AA	5,306	6,073	1.00 (ref.)	0.28
AG	2,369	2,485	1.08 (1.00-1.15)	
GG	338	329	1.11 (0.95-1.30)	
			P trend = 0.03	

*Numbers do not sum to total due to missing genotype data.
† The ORs and 95% CIs were estimated using conditional (on matching factors: age, ethnicity, and country within EPIC) logistic regression and adjusted for cohort.
‡ P value for heterogeneity in ORs across cohorts.

ethnicity ($P \geq 0.29$). Table 2 shows the pooled htSNP associations with breast cancer. Borderline significant associations were observed with the following SNPs: rs4919687 (*GA* versus *GG*, OR, 1.04; 95% CI, 0.97-1.12; *AA* versus *GG*, OR, 1.17; 95% CI, 1.03-1.34; P for trend = 0.03) and rs4919682 (*CT* versus *CC*, OR, 1.04; 95% CI, 0.97-1.12; *TT* versus *CC*, OR, 1.16; 95% CI, 1.01-1.33; P for trend = 0.04). The previously studied SNP, rs743572 (*T27C*), was not associated with breast cancer (*AG* versus *AA*, OR, 0.99; 95% CI, 0.92-1.06; *GG* versus *AA*, OR, 1.05; 95% CI, 0.95-1.16; P for trend = 0.49). Analyses among advanced cases only (n = 1,315) did not reveal stronger associations with rs4919687 and rs4919682 (P for trend = 0.03 for both SNPs; Supplemental Table S9).

Data on hormone receptor status were available from four of the five participating cohorts (all but EPIC), including 2,010 ER+/PR+ cases and 480 ER−/PR− cases. The associations between *CYP17* SNPs and haplotypes and breast cancer did not differ by hormone receptor status ($P \geq 0.35$; Supplemental Tables S6 and S10).

We also tested for statistical interaction between *CYP17* and the following breast cancer risk factors: age at menarche (≤12, 13-14, ≥15), parity (nulliparous, parous), age at first birth (nulliparous, ≤24, >24), age at menopause (<40, 40-44, 45-49, 50-54, ≥55), BMI (<25, 25-29, ≥30 kg/m^2), use of postmenopausal hormones (ever, never), first-degree family history of breast cancer (yes, no), and history of benign breast disease (yes, no). We did not find any evidence of interaction with any of

the examined risk factors for any of the common haplotypes or SNPs (P interaction ≥ 0.11).

Prostate Cancer Analysis. The haplotype-specific ORs for prostate cancer did not show significant heterogeneity across cohorts ($P \geq 0.25$) or across ethnicity ($P \geq 0.11$). Results from the pooled analyses are shown in Table 3. The global test for comparison of haplotype frequencies in cases and controls was not statistically significant (block 1, $P = 0.46$; block 2, $P = 0.18$) and none of the common haplotypes showed a significant association with prostate cancer risk.

We observed no significant heterogeneity in SNP-specific ORs across cohorts ($P \geq 0.23$) or ethnicity ($P \geq 0.45$). Table 4 shows the pooled htSNP associations with prostate cancer. Borderline significant associations were observed with the following SNPs: rs2486758 (*TC* versus *TT*, OR, 1.07; 95% CI, 1.00-1.14; *CC* versus *TT*, OR, 1.09; 95% CI, 0.95-1.26; P for trend = 0.04) and rs6892 (*AG* versus *AA*, OR, 1.08; 95% CI, 1.00-1.15; *GG* versus *AA*, OR, 1.11; 95% CI, 0.95-1.30; P for trend = 0.03).

Data on tumor stage and grade were available from six of the seven participating cohorts (all but PLCO), including 1,079 high-stage cases and 950 high-grade cases. The associations of rs2486758 (*TC* versus *TT*, OR, 1.18; 95% CI, 1.01-1.37; *CC* versus *TT*, OR, 1.15; 95% CI, 0.85-1.54; P for trend = 0.06) and rs6892 (*AG* versus *AA*, OR, 1.26; 95% CI, 1.07-1.47; *GG* versus *AA*, OR, 1.23; 95% CI, 0.88-1.70; P for trend = 0.008) with prostate cancer seem to be stronger among high-grade cases and no evidence of heterogeneity in ORs was observed across studies (Supplemental Table S15).

Using a dominant model, the ORs for rs2486758 and rs6892 were 1.17 (95% CI, 1.01-1.36; $P = 0.03$) and 1.25 (95% CI, 1.08-1.46; $P = 0.003$) in high-grade tumors, respectively. Rs2486758 (*C* allele) and rs6982 (*G* allele) were modestly correlated with each other ($r^2 = 0.57$). When these SNPs were modeled concurrently (global test $P = 0.007$), only rs6892 remained significantly associated with high-grade prostate cancer (*AG/GG* versus *AA*, OR, 1.32; 95% CI, 1.02-1.71). No evidence of heterogeneity across cohorts ($P = 0.35$) or ethnicity ($P = 0.93$) was observed.

Finally, we tested for statistical interaction between *CYP17* with the following risk factors: age at the time of diagnosis (<65, ≥65), first-degree family history of prostate cancer (yes, no), and BMI (<25, 25-29, ≥30 kg/m^2). None of the tests for statistical interaction for haplotypes was significant (P interaction ≥ 0.11). However, a marginally significant interaction was observed between age at diagnosis and the following SNPs in block 1: rs7097872 ($P = 0.03$) and rs2486758 ($P = 0.04$). For men age ≥65, those with *TT* genotype of rs7097872 compared with *CC* genotype had OR of 0.88 (95% CI, 0.79-0.99); for men age ≥65 with *CT* or *CC* genotype of rs2486758 had 1.10-fold (95% CI, 1.02-1.19) and 1.18-fold (95% CI, 1.00-1.39) increased risk of prostate cancer as compared to men with the *TT* genotype.

Steroid Hormone Analysis. None of the global tests for haplotype effects for any of the measured postmenopausal hormones were statistically significant ($P \geq 0.15$). None of the htSNPs were associated with hormone levels, including the two SNPs that were modestly associated with breast cancer (Supplemental Table S17). It has been hypothesized that the effect of polymorphisms on the levels of circulating hormone levels is more pronounced in lean postmenopausal women in whom peripheral conversions of androgens in the adipose tissue would be minimal (30); to test this hypothesis, we restricted our analysis to postmenopausal women with a BMI <25 kg/m^2 and found no significant association in this subgroup of women (Supplemental Table S18).

The two htSNPs (rs2486758 and rs6892) found to be related to prostate cancer were not associated with circulating hormones in men (Supplemental Table S17). The global test for haplotype effects on hormone levels by block was not significant ($P \geq 0.99$).

Discussion

With a key role in steroid hormone biosynthesis, variation in *CYP17* has been hypothesized to be related with the risk of developing breast and prostate cancer. In the present study, we did a systematic analysis of common variation across the *CYP17* locus, considering the possibility that *CYP17* variants other than the *T27C* polymorphism might contribute to cancer susceptibility and influence sex steroid hormone levels. We directly surveyed variation in the coding region of the *CYP17* gene to search for common missense variants, and did a detailed haplotype-based analysis to uncover the effects of functional variation in noncoding regions. Our results indicate that major single-gene effects for either both breast or prostate cancer or hormone levels in *CYP17* could be excluded among Whites.

We did however observe modest nominal significant positive associations between prostate cancer and two SNPs (rs2486758 and rs6892) which seem to be stronger among aggressive prostate cancer. A Swedish study with 2,826 prostate cancer cases and 1,705 controls recently showed that multiple SNPs in *CYP17* were associated with prostate cancer (9). In this study, Lindstrom et al. found that compared with noncarriers, men carrying one variant allele of the rs2486758 SNP located in the promoter region of *CYP17*, were associated with a 15% elevated risk of prostate cancer. In BPC3, we found that the variant allele for this SNP was associated with a smaller (7%) increase in prostate cancer risk. Lindstrom et al. found two variants (rs10883782 and rs619824) at the 3′ untranslated region of the gene to be associated with a decreased risk; findings not seen in the BPC3. Instead, we observed that the rs6892 variant, also in the 3′ untranslated region, was significantly associated with increased risk of prostate cancer and that the association seems to be stronger among aggressive prostate cancer. When we modeled rs2486758 and rs6892 together, only rs6892 remained significantly associated with high-grade prostate cancer. Interestingly, the region around rs6892 is highly conserved among mammals, and microarray expression data showed that the transcript is up-regulated in several prostate cancer lines.[24] Furthermore, this SNP is located in the 3′ untranslated region of a flanking gene, *OPAL1*, suggested to be important in the prognosis of pediatric acute lymphoblastic leukemia (31).

[24] http://genome.ucsc.edu/

Thus, it is possible that rs6892 might have some functional significance or mark a causal variant in the region. However, given the lack of association between this SNP and hormone levels and many tests done in this study, we cannot rule out that the association we observed was due to chance. In light of the small magnitude of the estimated effects and the number of tests conducted, replication in other large prostate cancer studies is clearly warranted.

We also observed modest nominal significant positive associations between breast cancer and two haplotypes (*TTGAG* in block 1 and *TAAA* in block 2) and the two SNPs (rs4919687 and rs4919682) that tagged these haplotypes. The relationship between *T27C* SNP and breast cancer risk have been investigated extensively (6, 7, 12), but no studies have looked at other SNPs in the *CYP17* locus. Consistent with our results, in a review of published studies, Feigelson et al. did not find a consistent association between the *T27C* SNP with breast cancer (32). Given the lack of association between rs4919687 and rs4919682 SNPs and hormone levels and the number of tests done here, it is possible that the associations we observed were due to chance. It has been suggested that *CYP17* may act as an effect modifier for breast cancer risk factors, in particular, age at menarche (13). In the BPC3, we did not observe any significant interaction between *CYP17* and age at menarche.

We found no association between *CYP17* polymorphisms and postmenopausal endogenous hormones. A marginally significant association between *CYP17 T27C* and serum estrogen levels in postmenopausal women was first reported in 1999 (33), but subsequent studies of disease-free postmenopausal women that attempted to replicate the original finding found no association (34-36).

The strengths of the BPC3 include its large sample size and comprehensive characterization of variation around the *CYP17* locus. Our analysis provides evidence against a strong main effect between the overall risk of breast and prostate cancer and circulating sex steroid hormone levels and variants in *CYP17* that are common among Whites. The weak associations between several *CYP17* SNPs and prostate and breast cancer which we observed should be evaluated in future studies.

Acknowledgments

We gratefully thank the cohort participants and the expert contributions of Laure Dossus (EPIC) and Loreall Pooler (USC).

References

1. 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.
2. Henderson BE, Feigelson HS. Hormonal carcinogenesis. Carcinogenesis 2000;21:427–33.
3. Costa-Santos M, Kater CE, Auchus RJ. Two prevalent CYP17 mutations and genotype-phenotype correlations in 24 Brazilian patients with 17-hydroxylase deficiency. J Clin Endocrinol Metab 2004;89:49–60.
4. Martin RM, Lin CJ, Costa EM, et al. P450c17 deficiency in Brazilian patients: biochemical diagnosis through progesterone levels confirmed by CYP17 genotyping. J Clin Endocrinol Metab 2003;88:5739–46.
5. Carey AH, Waterworth D, Patel K, et al. Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. Hum Mol Genet 1994;3:1873–6.
6. Dunning AM, Healey CS, Pharoah PD, et al. A systematic review of genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1999;8:843–54.
7. Feigelson HS, McKean-Cowdin R, Henderson BE. Concerning the CYP17 MspA1 polymorphism and breast cancer risk: a meta-analysis. Mutagenesis 2002;17:445–6; author reply 447–8.
8. Ntais C, Polycarpou A, Ioannidis JP. Association of the CYP17 gene polymorphism with the risk of prostate cancer: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2003;12:120–6.
9. Lindstrom S, Wiklund F, Adami HO, et al. Germ-line genetic variation in the key androgen-regulating genes androgen receptor, cytochrome P450, and steroid-5-α-reductase type 2 is important for prostate cancer development. Cancer Res 2006;66:11077–83.
10. Lindstrom S, Zheng SL, Wiklund F, et al. Systematic replication study of reported genetic associations in prostate cancer: strong support for genetic variation in the androgen pathway. Prostate 2006;66:1729–43.
11. Ntais C, Polycarpou A, Tsatsoulis A. Molecular epidemiology of prostate cancer: androgens and polymorphisms in androgen-related genes. Eur J Endocrinol 2003;149:469–77.
12. Ye Z, Parry JM. The CYP17 MspA1 polymorphism and breast cancer risk: a meta-analysis. Mutagenesis 2002;17:119–26.
13. Sharp L, Cardy AH, Cotton SC, Little J. CYP17 gene polymorphisms: prevalence and associations with hormone levels and related factors. a HuGE review. Am J Epidemiol 2004;160:729–40.
14. Nedelcheva Kristensen V, Haraldsen EK, Anderson KB, et al. CYP17 and breast cancer risk: the polymorphism in the 5′ flanking area of the gene does not influence binding to Sp-1. Cancer Res 1999;59:2825–8.
15. Cox DG, Blanche H, Pearce CL, et al. A comprehensive analysis of the androgen receptor gene and risk of breast cancer: results from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3). Breast Cancer Res 2006;8:R54.
16. Feigelson HS, Cox DG, Cann HM, et al. Haplotype analysis of the HSD17B1 gene and risk of breast cancer: a comprehensive approach to multicenter analyses of prospective cohort studies. Cancer Res 2006;66:2468–75.
17. Kraft P, Pharoah P, Chanock SJ, et al. Genetic variation in the HSD17B1 gene and risk of prostate cancer. PLoS Genet 2005;1:e68.
18. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
19. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 1995;12:921–7.
20. Stram DO, Haiman CA, Hirschhorn JN, et al. Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. Hum Hered 2003;55:27–36.
21. Packer BR, Yeager M, Burdett L, et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. Nucleic Acids Res 2006;34:D617–21.
22. 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.
23. Setiawan VW, Haiman CA, Stanczyk FZ, Le Marchand L, Henderson BE. Racial/ethnic differences in postmenopausal endogenous hormones: the Multiethnic Cohort study. Cancer Epidemiol Biomarkers Prev 2006;15:1849–55.
24. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. J Natl Cancer Inst 2004;96:1856–65.
25. Hartman TJ, Dorgan JF, Virtamo J, et al. Association between serum α-tocopherol and serum androgens and estrogens in older men. Nutr Cancer 1999;35:10–5.
26. Kaaks R, Lukanova A, Rinaldi S, et al. Interrelationships between plasma testosterone, SHBG, IGF-I, insulin and leptin in prostate cancer cases and controls. Eur J Cancer Prev 2003;12:309–15.
27. Platz EA, Leitzmann MF, Rifai N, et al. 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 2005;14:1262–9.
28. Haiman CA, Stampfer MJ, Giovannucci E, et al. The relationship between a polymorphism in CYP17 with plasma hormone levels and prostate cancer. Cancer Epidemiol Biomarkers Prev 2001;10:743–8.
29. Zaykin DV, Westfall PH, Young SS, et al. Testing association of

statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 2002;53:79–91.
30. Feigelson HS, McKean-Cowdin R, Coetzee GA, et al. Building a multigenic model of breast cancer susceptibility: CYP17 and HSD17B1 are two important candidates. Cancer Res 2001;61:785–9.
31. Holleman A, den Boer ML, Cheok MH, et al. Expression of the outcome predictor in acute leukemia 1 (OPAL1) gene is not an independent prognostic factor in patients treated according to COALL or St Jude protocols. Blood 2006;108:1984–90.
32. Feigelson H, Pearce CL, Diver WR, McKean-Cowdin R. CYP17 as a candidate gene for hormone-related cancers among women: a review of the evidence. Int J Cancer Prev 2005;2:25–35.
33. Haiman CA, Hankinson SE, Colditz GA, Hunter DJ, De Vivo I. A polymorphism in CYP17 and endometrial cancer risk. Cancer Res 2001;61:3955–60.
34. 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.
35. Travis RC, Churchman M, Edwards SA, et al. No association of polymorphisms in CYP17, CYP19, and HSD17-1 with plasma estradiol concentrations in 1,090 British women. Cancer Epidemiol Biomarkers Prev 2004;13:2282–4.
36. Tworoger SS, Chubak J, Aiello EJ, et al. Association of CYP17, CYP19, CYP1B1, and COMT polymorphisms with serum and urinary sex hormone concentrations in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2004;13:94–101.

CYP17 Genetic Variation and Risk of Breast and Prostate Cancer from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3)

Veronica Wendy Setiawan, Fredrick R. Schumacher, Christopher A. Haiman, et al.

Cancer Epidemiol Biomarkers Prev 2007;16:2237-2246.