Haplotype Analysis of the *HSD17B1* Gene and Risk of Breast Cancer: A Comprehensive Approach to Multicenter Analyses of Prospective Cohort Studies

Heather Spencer Feigelson,[1] David G. Cox,[2] Howard M. Cann,[6] Sholom Wacholder,[7] Rudolf Kaaks,[8] Brian E. Henderson,[10] Demetrius Albanes,[7] David Altshuler,[12] Goran Berglund,[14] Franco Berrino,[15] Sheila Bingham,[16] Julie E. Buring,[2,4] Noel P. Burtt,[12] Eugenia E. Calle,[1] Stephen J. Chanock,[17] Francoise Clavel-Chapelon,[18] Graham Colditz,[2,5] W. Ryan Diver,[1] Matthew L. Freedman,[13] Christopher A. Haiman,[10] Susan E. Hankinson,[2,5] Richard B. Hayes,[4] Joel N. Hirschhorn,[12] David Hunter,[2,5] Laurence N. Kolonel,[19] Peter Kraft,[3] Loic LeMarchand,[19] Jakob Linseisen,[20] William Modi,[17] Carmen Navarro,[21] Petra H. Peeters,[22] Malcolm C. Pike,[10] Elio Riboli,[9] V. Wendy Setiawan,[10] Daniel O. Stram,[11] Gilles Thomas,[6] Michael J. Thun,[1] Anne Tjonneland,[23] and Dimitrios Trichopoulos[24]

[1]Department of Epidemiology and Surveillance Research, American Cancer Society, National Home Office, Atlanta, Georgia; [2]Department of Epidemiology and [3]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health; [4]Division of Preventive Medicine and [5]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; [6]Foundation Jean Dausset, CEPH, Paris, France; [7]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; [8]Hormones and Cancer Group and [9]Unit of Nutrition and Cancer, IARC, Lyon, France; [10]Keck School of Medicine and [11]Division of Biostatistics and Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; [12]Broad Institute at Harvard and Massachusetts Institute of Technology; [13]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts; [14]Department of Medicine, Lund University, Malmö, Sweden; [15]Epidemiology Unit, National Cancer Institute, Milan, Italy; [16]Medical Research Council Dunn Nutrition Unit, Cambridge, United Kingdom; [17]Core Genotyping Facility, National Cancer Institute, Gaithersburg, Maryland; [18]Institut National de la Sante et de la Recherche Medicale, Institut Gustave Roussy, Villejuif, France; [19]Cancer Research Center, University of Hawaii, Honolulu, Hawaii; [20]Division of Clinical Epidemiology, Deutsches Krebsforschungszentrum, Heidelberg, Germany; [21]Epidemiology Department, Murcia Health Council, Murcia, Spain; [22]Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands; [23]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark; and [24]Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece

Abstract

The 17β-hydroxysteroid dehydrogenase 1 gene (*HSD17B1*) encodes 17HSD1, which catalyzes the final step of estradiol biosynthesis. Despite the important role of *HSD17B1* in hormone metabolism, few epidemiologic studies of *HSD17B1* and breast cancer have been conducted. This study includes 5,370 breast cancer cases and 7,480 matched controls from five large cohorts in the Breast and Prostate Cancer Cohort Consortium. We characterized variation in *HSD17B1* by resequencing and dense genotyping a multiethnic sample and identified haplotype-tagging single nucleotide polymorphisms (htSNP) that capture common variation within a 33.3-kb region around *HSD17B1*. Four htSNPs, including the previously studied SNP rs605059 (*S312G*), were genotyped to tag five common haplotypes in all cases and controls. Conditional logistic regression was used to estimate odds ratios (OR) for disease. We found no evidence of association between common *HSD17B1* haplotypes or htSNPs and overall risk of breast cancer. The OR for each haplotype relative to the most common haplotype ranged from 0.98 to 1.07 (omnibus test for association: $X^2 = 3.77$, $P = 0.58$, 5 degrees of freedom). When cases were subdivided by estrogen receptor (ER) status, two common haplotypes were associated with ER-negative tumors (test for trend, Ps = 0.0009 and 0.0076; $n = 353$ cases). *HSD17B1* variants that are common in Caucasians are not associated with overall risk of breast cancer; however, there was an association among the subset of ER-negative tumors. Although the probability that these ER-negative findings are false-positive results is high, these findings were consistent across each cohort examined and warrant further study. (Cancer Res 2006; 66(4): 2468-75)

Introduction

The 17β-hydroxysteroid dehydrogenase 1 gene (*HSD17B1* [MIM 109684]) encodes 17HSD1, whose primary function is to catalyze the final step of estradiol biosynthesis, converting estrone to the more biologically active estradiol (1). Its complement, 17HSD2 (encoded by *HSD17B2*) is the enzyme that predominates in the reverse reaction, the oxidation of estradiol to estrone. The balance of these two enzymes, in part, regulates estrogen concentrations in breast tissue. In normal breast tissue, 17HSD2 activity predominates, whereas 17HSD1 activity predominates in malignant breast tissue (2). *HSD17B1* amplification in breast tumors correlates with poorer prognosis, especially among women with estrogen receptor (ER)–positive tumors (2). Furthermore, breast cancer patients with tumors expressing 17HSD1 mRNA or protein had significantly shorter overall ($P = 0.001$) and disease-free ($P = 0.015$) survival than other patients, and detection of 17HSD1 mRNA in the tumor was an independent prognostic marker in multivariate analyses (1).

Despite these intriguing data, the relation of germ line variation in *HSD17B1* to breast cancer incidence has not been well investigated in epidemiologic studies. To date, only four epidemiologic

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

H.S. Feigelson, D.G. Cox, H.M. Cann, S. Wacholder, R. Kaaks, and B.E. Henderson were the writing committee for this article.

Requests for reprints: Heather Spencer Feigelson, Department of Epidemiology and Surveillance Research, American Cancer Society, 1599 Clifton Road Northeast, Atlanta, GA 30329. Phone: 404-929-6815; Fax: 404-327-6450; E-mail: heather.feigelson@cancer.org.

doi:10.1158/0008-5472.CAN-05-3574

studies of *HSD17B1* variation and breast cancer have been conducted, the largest of which included ~1,000 cases (3–6). Although several sequence variations in *HSD17B1* have been identified (3, 7, 8), three of these studies (4–6) focused only on a single polymorphism in exon 6, designated *S312G* (rs605059). This single nucleotide polymorphism (SNP) results in an amino acid change from serine (allele *A*) to glycine (allele *G*) but does not seem to affect the catalytic or immunologic properties of the enzyme (9). The results of these studies have been inconclusive but have provided some evidence that *HSD17B1* may influence risk of breast cancer.

We report here the results from an analysis of *HSD17B1* haplotypes and breast cancer risk from a large, collaborative study (The Breast and Prostate Cancer Cohort Consortium, or BPC3), which includes data from five cohorts from the United States and Europe (10). The large size of this study enables us to detect modest genetic effects, explore gene-environment interactions, and examine potentially important subclasses of tumors, such as those defined by stage or hormone receptors.

Materials and Methods

Study population. The BPC3 has been described in detail elsewhere (10). Briefly, the consortium includes large well-established cohorts (or consortia of smaller cohorts) assembled in the United States and Europe that have DNA for genotyping and extensive questionnaire data from cohort members. Written informed consent was obtained from all subjects, and each cohort has been approved by the appropriate institutional review board. This analysis includes 5,370 cases of invasive breast cancer and 7,480 matched controls from five cohorts: the American Cancer Society Cancer Prevention Study II (CPS-II; ref. 11), the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort (12), the Harvard Nurse's Health Study (NHS; ref. 13) and Women's Health Study (WHS; ref. 14), and the Hawaii-Los Angeles Multiethnic Cohort (MEC; ref. 15). With the exception of MEC, most women in these cohorts are Caucasians. The MEC includes U.S. Caucasians, African Americans, Latinos, Japanese, and native Hawaiians (15). Cases were identified in each cohort by self-report, with subsequent confirmation of the diagnosis from medical records or tumor registries and/or linkage with population-based tumor registries. Information on ER and progesterone receptor (PR) status was abstracted from medical records and/or tumor registries and was not available from the EPIC cohort. ER and PR receptor data was reported as positive, negative, borderline, or not available. In all cohorts, questionnaire data was collected prospectively before the diagnosis of cancer. Blood samples were collected before diagnosis in all cohorts except for the MEC and CPS-II cohorts, where most of the blood specimens were collected after diagnosis (10, 11, 16). Cases classified as carcinoma *in situ* were not included in this analysis. Controls were matched to cases by ethnicity and age, and in some cohorts, additional matching criteria were employed (e.g., EPIC matched on country of residence.) These analyses include a portion of the previously published data from the NHS cohort (3) and MEC (5).

Haplotype discovery and haplotype-tagging SNP selection. The BPC3 adopted a two-stage approach to comprehensively measure genetic variation in and around *HSD17B1* among cases and controls. The first stage consists of comprehensive haplotype discovery followed by haplotype-tagging SNP (htSNP) selection. The second stage is genotyping the htSNPs in all the BPC3 cases and controls.

In the first stage, we genotyped a dense set of SNPs spanning the region of interest in five population samples to identify regions of high linkage disequilibrium and low haplotype diversity using the algorithm of Gabriel et al. (17) as implemented in Haploview.[25] Novel SNPs were identified by systematically resequencing the *HSD17B1* exons in 95 cases of advanced prostate cancer and 95 advanced breast cancer cases from five population groups (equal numbers of U.S. Caucasians, Latinos, Japanese, native Hawaiians, and African Americans) in the MEC (resequencing details at MEC web site).[26] Then, SNPs were selected from public databases to cover the introns and flanking regions around *HSD17B1*.

In total, 26 SNPs were selected covering a 42-kb region around *HSD17B1* at an average density of one SNP per 1.6 kb. All but one of these SNPs (rs7208557 in the 3′ region) had a minor allele frequency >5% among Caucasians. The SNPs extended in the 5′ direction of *HSD17B1* into the adjoining *N*-acetylglucosaminidase-α (*NAGLU*) gene and pseudogene for *HSD17B1* (*HSD17BP1*), and in the 3′ direction into the genes for CoA synthase (*COASY*) and transcription factor-like 4 (*TCFL4*).

To identify regions of high linkage disequilibrium, these 26 SNPs were genotyped in a multiethnic panel of 349 unrelated women from the MEC with no history of cancer. Tagging SNPs were then chosen using the partition-ligation EM algorithm implemented in the program TAGSNPS.[27] Selection of htSNPs is based on R_H^2, a measure of the correlation between observed haplotypes and those predicted based on htSNP genotypes (18). This haplotype-tagging approach is based on the observation that within blocks of high linkage disequilibrium, there is limited haplotype diversity and that common variation in the region is highly correlated with the common haplotype patterns (17). Nineteen SNPs fall into a block of high linkage disequilibrium that could be characterized efficiently with four htSNPs: rs676387, rs598126, rs2010750, and the *S312G* SNP in exon 6 (rs605059), which we required to be in the htSNP set. These four htSNPs were genotyped in all the BPC3 cases and controls.

Genotyping. The 26 SNPs used for haplotype construction were genotyped in a reference panel made up of MEC populations at the Broad Institute using Sequenom and Illumina platforms. Genotyping the four htSNPs in the breast cancer cases and controls was done in four laboratories using a fluorescent 5′ endonuclease assay and the ABI-PRISM 7900 for sequence detection (Taqman). Initial quality control checks of the SNP assays were done at the manufacturer (ABI, Foster City, CA); an additional 500 test reactions were run by the BPC3. Assay characteristics for the four htSNPs for *HSD17B1* are available on a public web site.[28] Sequence validation for each SNP assay was done, and 100% concordance was observed.[29]

To assess interlaboratory variation, each genotyping center ran assays on a designated set of 94 samples from the Coriell Biorepository (Camden, NJ), showing completion and concordance rates of >99% (19). 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 study); resulting concordance was >99%.

Statistical analysis. We used conditional multiple logistic regression to estimate odds ratios (OR) for disease in subjects with a linear (additive) scoring for 0, 1, or 2 copies of the minor allele of each SNP. We also used conditional logistic regression with additive scoring and the most common haplotype as the reference to estimate haplotype-specific ORs using an expectation substitution approach to assign haplotypes based on the unphased genotype data (20). It has been shown (21, 22) that this methods performs well despite the uncertainty in assignment (20, 21). Haplotype frequencies and expected subject-specific haplotype indicators were calculated separately for each cohort (and country within EPIC). To test the global null hypothesis of no association between variation in *HSD17B1* haplotypes and htSNPs and risk of breast cancer (or subtypes defined by receptor status), we used a likelihood ratio test comparing a model with additive effects for each common haplotype (treating the most common haplotype as the referent) to the intercept-only model. We combined rare haplotypes (those with estimated individual frequencies <1%) into a single category, which comprised <0.5% of the controls. To test for heterogeneity

[25] http://www.broad.mit.edu/mpg/haploview/index.php.

[26] http://www.uscnorris.com/MECGenetics/.
[27] http://www-rcf.usc.edu/~stram/tagSNPs.html.
[28] http://www.uscnorris.com/mecgenetics/CohortGCKView.aspx.
[29] http://snp500cancer.nci.nih.gov.

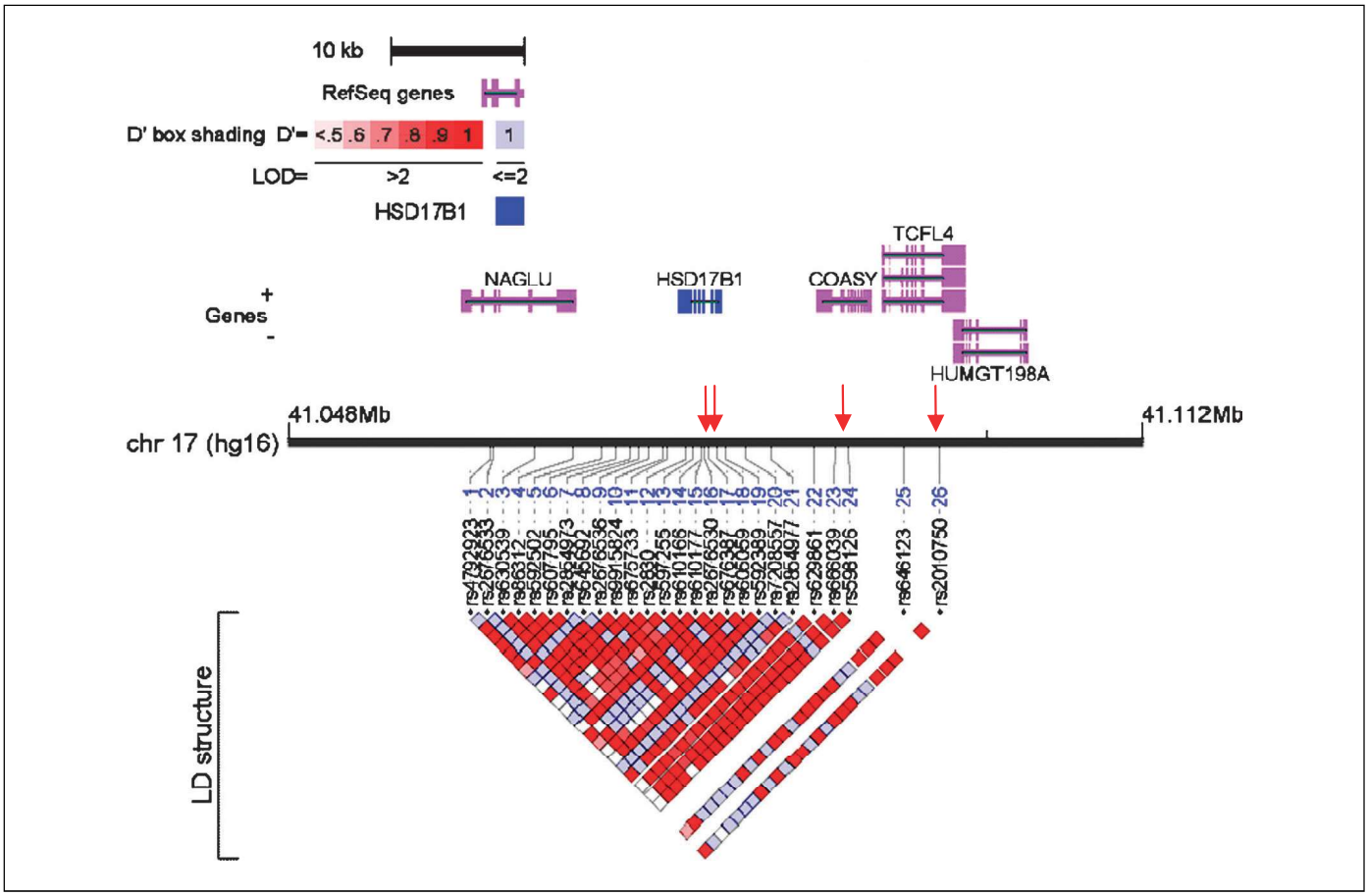

Figure 1. SNPs and linkage disequilibrium structure (among Caucasians) across the *HSD17B1* locus. Twenty-six SNPs were selected covering a 42-kb region around *HSD17B1* at an average density of one SNP per 1.6 kb. Haplotype tagging SNPs (*red arrows*).

across cohorts and ethnic groups, we used the Wald χ^2 for htSNPs and a likelihood ratio test for the haplotypes.

We considered conditional models both without adjustment and with adjustment for known breast cancer risk factors, including age at menarche, menopausal status, age at menopause, parity, age at first birth, history of benign breast disease, body mass index (BMI, in deciles), first-degree family history of breast cancer, and use of postmenopausal hormones. Because the results remained essentially unchanged regardless of the model used, we present results from the unadjusted conditional model. We evaluated these same covariates for possible interaction effects and also tested whether the association between *HSD17B1* and breast cancer differed by stage (localized versus regional or distant metastasis) or hormone receptor (ER and PR) status.

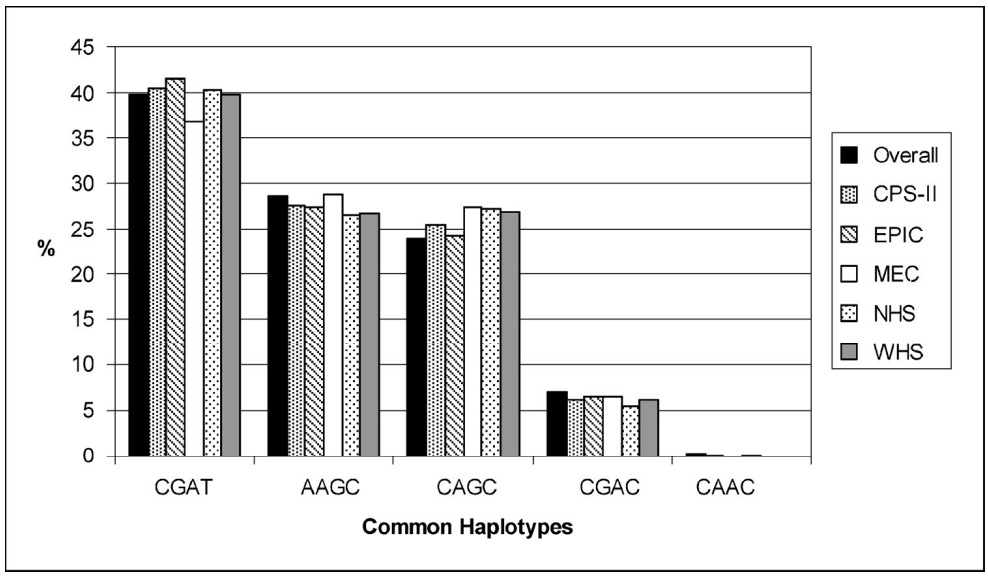

Figure 2. Haplotype frequencies by subcohort among Caucasians in the BPC3.

Table 1. Descriptive characteristics of invasive breast cancer cases (n = 5,370) and controls (n = 7,480) by cohort in the BPC3

	CPS-II		EPIC		MEC		NHS		WHS	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Number	395	505	1,610	2,844	1,601	1,962	1,059	1,464	705	705
Ethnicity										
White	98	99	100	100	25	22	94	94	96	96
Hispanic	1	0			21	20	0	0	1	1
African American	1	1			21	22	1	1	1	1
Asian	0	0			26	21	0	0	1	1
Hawaiian	0	0			7	15	0	0	0	0
Other/missing	0	0			0	0	5	5	2	2
Age at diagnosis (mean)	70.2		57.8		65.1		63.2		60.3	
Menopausal status (%)										
Premenopausal			23	28	11	16	19	17	22	21
Postmenopausal	100	100	68	63	87	82	71	75	63	60
Unknown/missing			9	9	2	2	9	8	15	19
Age at menarche (%)										
≤12	45	45	34	35	53	49	52	48	57	52
13-14	44	46	46	44	35	38	39	43	37	43
≥15	9	8	15	18	10	12	8	8	6	6
Unknown/missing	2	1	4	3	2	1	1	1	0	0
Age at menopause*										
≤44	20	22	11	14	30	35	22	22	16	22
45-49	20	28	22	27	26	27	28	28	29	30
50-54	44	39	37	37	32	28	44	44	41	35
≥55	14	10	8	8	9	7	6	6	7	8
Unknown/missing	3	2	22	15	3	3	0	0	6	6
Parity (%)										
Nulliparous	8	9	13	13	14	11	8	7	15	14
≤2 children	36	27	55	52	36	35	33	31	40	34
≥3 children	53	62	26	30	48	53	59	62	45	52
Unknown/missing	3	2	6	5	1	1	1	1	0	0
First-degree family history (%)										
Yes	20	15			17	11	19	14	20	16
No	78	81			83	89	81	86	79	83
Unknown	2	4	100	100	0	0	0	0	2	1
Hormone replacement therapy use (%)*										
Never	34	41	54	64	36	41	23	28	30	38
Ever	66	59	42	33	64	58	77	72	65	57
Unknown/missing	0	0	4	3	0	1	0	0	5	5
ER (%)										
Positive	62				62		68		80	
Negative	8				16		18		14	
Borderline	1				1		1		0	
Not available	29		100		21		13		5	
PR (%)										
Positive	53				51		58		72	
Negative	15				23		26		21	
Borderline	1				1		1		1	
Not available	31		100		25		15		6	

*Percentages for age at menopause and hormone replacement therapy use include postmenopausal women only.

Logistic models to examine associations for specific hormone receptor subtypes (ER positive, ER negative, PR positive, PR negative) included only cases classified as receptor positive or receptor negative, controlled for age and cohort, and were stratified by ethnicity because there was statistical evidence of heterogeneity by ethnicity. Cases without hormone receptor data or with receptor status of "borderline" were not included in the hormone receptor analyses. Controls in these models included all controls from each cohort that provided data on receptor status. Tests for heterogeneity by receptor status were obtained from case-only models comparing receptor-negative cases with receptor-positive cases.

Finally, we established a range of prior probabilities that variation in *HSD17B1* is related to breast cancer based on existing epidemiologic and

laboratory data to evaluate the false-positive or false-negative report probabilities (23). Based on existing evidence (3–6, 24, 25), we assumed a prior probability of 1%, with a range of 10% to 0.1% for an association between *HSD17B1* and overall breast cancer. Although we did not specify prior probabilities a priori for associations with ER-positive and ER-negative tumors, the prior probability for ER-positive tumors should be about the same as for overall breast cancer and perhaps 10-fold lower for ER-negative tumors, where the role of estrogen is less clear. The prior probability for any given haplotype or SNP in the gene is also somewhat less than the prior probability for the gene (23).

Results

The genomic structure of the region around *HSD17B1* is shown in Fig. 1. The four htSNPs chosen capture most of the variation, known and unknown, of all common haplotypes in this block (frequencies > 0.05, in at least one ethnic group). In each ethnic group, the four htSNPs predicted common haplotypes with a minimum R_H^2 above 80%. However, among African Americans, the cumulative frequency of the common haplotypes was only 62% (Supplementary Table 1S). A cumulative frequency of ≥70% among African Americans would be achieved only by genotyping three additional htSNPs, tagging two additional haplotypes, each with a frequency of just under 5%.[30] Among the controls in the BPC3 (n = 7,480), the five common htSNP haplotypes account for 99% of all haplotypes. One haplotype (*CAAC*) is common only among African Americans (frequency = 6.1% among African-American controls). Haplotype frequencies by cohort are shown in Fig. 2. (Haplotype frequencies by population in the MEC are shown in Supplementary Fig. 1S.).

The genotyping success rate was ≥94% for each of the four htSNPS at each genotyping center. No deviation from Hardy-Weinberg equilibrium was observed among the controls in each cohort (at the P < 0.01 level) or across more than one cohort for any given assay.

Study characteristics of each cohort are provided in Table 1. Case and control characteristics were comparable across cohorts. The majority of cases were postmenopausal. The mean age of diagnosis ranged from 57.8 in EPIC to 70.2 in CPS-II, reflecting differences in the age and length of follow-up in these cohorts. The percentage of women who reported age at menarche over 14 years was higher among European women in EPIC (18% of controls) than in the North American cohorts (6-12% of controls). European women also reported less hormone replacement therapy use than U.S. women.

As there was no heterogeneity in results for any of the main effects by cohort, we report results from only pooled analyses here; cohort-specific results are available in supplemental tables. Table 2 presents the pooled haplotype and SNP associations for *HSD17B1* and breast cancer. There were no significant differences in the haplotype frequencies between cases and controls (P of the omnibus test for association with breast cancer: X^2 = 3.77 with 5 degrees of freedom, P = 0.58). None of the individual SNPs, including the *S312G* SNP (rs605059), were associated with breast cancer. There was also no evidence that any of the individual rare haplotypes (defined as any haplotype with a frequency of <5%) were over represented in the cases (data not shown). Cohort-specific and population-specific results are shown in Supplementary Tables 2S-3S.

[30] http://www.uscnorris.com/MECGenetics/HSD17B1.htm.

We tested for statistical interaction with *HSD17B1* and the following breast cancer risk factors: age at menarche, menopausal status, age at menopause, parity, age at first birth, history of benign breast disease, BMI (in deciles), first-degree family history of breast cancer, and use of postmenopausal hormones (estrogen-only therapy and combination estrogen-progesterone therapy). We did not find any consistent evidence of effect modification with any of the examined risk factors for any of the SNPs or common haplotypes, nor did the association between *HSD17B1* variation and breast cancer differ by tumor stage at diagnosis (data not shown).

Data on receptor status were available from four of the five participating cohorts, including 353 cases of ER-negative tumors, 1,723 cases of ER-positive tumors, and 352 unclassified tumors among U.S. Caucasian women (Table 1). Missing data varied by cohort. The WHS had ER receptor data available on all but 5% of cases, whereas CPS-II had missing ER data for 29% of cases; no receptor data was available from the EPIC cohort. When the data were stratified based on ER status of the tumors, we found statistical evidence of heterogeneity for two htSNPs and one haplotype (at P < 0.05). As shown in Table 3, each of the four

Table 2. *HSD17B1* haplotype and SNP associations with invasive breast cancer among U.S. and European women in BPC3

	Cases (%*), n = 5,370	Controls (%*), n = 7,480	OR† (95% confidence interval)
Haplotypes‡			
CGAT (reference)	4,228.5 (39)	5,919.5 (40)	1.00
AAGC	3,048.2 (28)	4,242.9 (28)	1.03 (0.96-1.09)
CAGC	2,479.0 (23)	3,535.8 (24)	0.98 (0.92-1.05)
CGAC	793.5 (7)	1,052.9 (7)	1.05 (0.95-1.17)
CAAC§	50.9 (1)	59.9 (0)	1.07 (0.71-1.61)
All rare	59.9 (1)	62.9 (0)	1.26 (0.88-1.80)
rs676387‖			
C/C	2,666 (50)	3,740 (50)	1.00
C/A	2,088 (39)	2,931 (39)	1.01 (0.94-1.10)
A/A	457 (9)	618 (8)	1.08 (0.95-1.24)
S312G (rs605059)			
A/A	1,406 (26)	2,011 (27)	1.00
A/G	2,588 (48)	3,610 (48)	1.02 (0.93-1.11)
G/G	1,139 (21)	1,580 (21)	1.01 (0.91-1.12)
rs598126			
G/G	1,399 (26)	1,968 (28)	1.00
G/A	2,600 (48)	363 (49)	1.00 (0.92-1.09)
A/A	1,182 (22)	1,642 (22)	0.98 (0.89-1.09)
rs2010750			
C/C	1,854 (35)	2,593 (35)	1.00
C/T	2,453 (46)	3,475 (46)	0.98 (0.90-1.06)
T/T	834 (15)	1,169 (16)	0.98 (0.88-1.09)

*Percentages do not sum to 100% because of missing genotype data.
†Matched for age, cohort, and ethnicity.
‡Alleles listed in 5′ to 3′ order; global test for haplotype association with breast cancer: X^2 = 3.77, with 5 degrees of freedom for P = 0.58.
§Only common among African Americans (frequency of 6% among African-American controls).
‖htSNPs listed in 5′ to 3′ order.

htSNPs is statistically significantly associated with ER-negative tumors but not with ER-positive tumors. Each of the corresponding haplotypes that carry all high-risk alleles (*CGAT*) or all low-risk alleles (*AAGC*) for each htSNP is also associated with ER-negative tumors (P_{trend} = 0.0076 and 0.0009, respectively). All these htSNP and haplotype associations show evidence of a dose-response relationship for ER-negative tumors, with stronger associations for homozygotes than heterozygotes. Furthermore, analysis of haplotype combinations (Table 4) shows that every common haplotype combination that includes *AAGC* is associated with a reduced risk of ER-negative breast cancer. This association with the *AAGC* haplotype and ER-negative tumors is present among U.S. Caucasians in each of the four cohorts that contributed information on receptor status. (Cohort specific associations by ER status and specific numbers of cases and controls from each cohort are shown in Supplementary Tables 4S-5S).

Similar but nonsignificant associations were observed among U.S. Caucasian women with PR-negative tumors (Supplementary Table 6S). Statistical testing indicated heterogeneity by ethnicity (Supplementary Table 7S), but no single population group other than Caucasians was of sufficient size to examine the association by ER status.

The false-positive report probability (FPRP) values for the ER-negative tumors for prior probabilities of 0.01, 0.001, and 0.0001 are 0.23, 0.75, and 0.97, respectively, with statistical power near 1.0 for a trend test (26), assuming that carriers of the second most

Table 3. Association between HSD17B1 and invasive breast cancer among U.S. Caucasian women by ER status

	ER-positive*				ER-negative*				P_{het}§
	Cases (%)†, *n* = 1,737	Controls (%)†, *n* = 2,982	OR (95% confidence interval)	P_{trend}‡	Cases (%)†, *n* = 354	Controls (%)†, (*n* = 2,982)	OR (95% confidence interval)	P_{trend}‡	
rs676387‖									
C/C	909 (52)	1,529 (51)	1.00		217 (61)	1,529 (51)	1.00		0.018
C/A	652 (38)	1,143 (38)	0.97 (0.85 -1.10)		111 (31)	1,143 (38)	0.67 (0.53-0.86)		
A/A	122 (7)	210 (7)	0.99 (0.78-1.26)	0.73	18 (5)	210 (7)	0.61 (0.37-1.01)	0.0009	
S312G (rs605059)									
A/A	485 (28)	819 (27)	1.00		77 (22)	819 (27)	1.00		0.032
A/G	802 (46)	1,431 (48)	0.94 (0.82-1.09)		171 (48)	1,431 (48)	1.28 (0.97-1.71)		
G/G	357 (21)	573 (19)	1.04 (0.87-1.24)	0.78	91 (26)	573 (19)	1.71 (1.24-2.37)	0.0013	
rs598126									
G/G	482 (28)	802 (27)	1.00		78 (22)	802 (27)	1.00		0.059
G/A	822 (47)	1,452 (49)	0.94 (0.82-1.09)		170 (48)	1,452 (49)	1.22 (0.92-1.62)		
A/A	372 (21)	599 (20)	1.03 (0.86-1.22)	0.86	93 (26)	599 (20)	1.63 (1.18-2.25)	0.0028	
rs2010750									
C/C	593 (34)	1,000 (34)	1.00		102 (29)	1,000 (34)	1.00		0.12
C/T	792 (46)	1,408 (47)	0.94 (0.82-1.08)		170 (48)	1,408 (47)	1.18 (0.91-1.53)		
T/T	264 (15)	445 (15)	1.00 (0.83-1.20)	0.75	67 (19)	445 (15)	1.50 (1.07-2.08)	0.020	
Haplotype hCGAT									
None	633.0 (36)	1,063.0 (36)	1.00		106.3 (30)	1,063.0 (36)	1.00		0.063
One	821.2 (47)	1,449.4 (49)	0.94 (0.83-1.08)		177.5 (50)	1,449.4 (49)	1.22 (0.94-1.58)		
Two	268.8 (15)	451.6 (15)	1.00 (0.83-1.20)	0.77	69.2 (20)	451.6 (15)	1.57 (1.13-2.17)	0.0076	
Haplotype hAAGC									
None	936.2 (54)	1,579.2 (53)	1.00		220.8 (62)	1,579.2 (53)	1.00		0.024
One	665.4 (38)	1,169.4 (39)	0.97 (0.85-1.10)		113.9 (32)	1,169.4 (39)	0.68 (0.53-0.87)		
Two	121.4 (7)	215.4 (7)	0.96 (0.75-1.22)	0.60	18.3 (5)	215.4 (7)	0.60 (0.36-1.00)	0.0009	
Haplotype hCAGC									
None	917.4 (53)	1,584.5 (53)	1.00		187.4 (53)	1,584.5 (53)	1.00		0.91
One	677.6 (39)	1,171.9 (39)	1.00 (0.88-1.14)		142.6 (40)	1,171.9 (39)	1.03 (0.81-1.30)		
Two	128.0 (7)	207.6 (7)	1.06 (0.83-1.35)	0.75	23.0 (7)	207.6 (7)	0.93 (0.58-1.49)	0.95	
Haplotype hCGAC									
None	1,501.1 (86)	2,616.5 (88)	1.00		306.2 (87)	2,616.5 (88)	1.00		0.57
One	216.7 (12)	339.1 (11)	1.11 (0.92-1.34)		44.6 (13)	339.1 (11)	1.15 (0.81-1.62)		
Two	5.2 (<1)	8.5 (<1)	1.03 (0.33-3.19)	0.30	2.1 (1)	8.5 (<1)	2.15 (0.46-10.09)	0.29	

NOTE: Data on receptor status was available from the following cohorts: CPS-II, NHS, WHS, MEC.

*Global test of haplotype association with ER-positive breast cancer: X^2 = 1.39 with 5 degrees of freedom for P = 0.925; global test of haplotype association with ER-negative breast cancer: X^2 = 14.61 with 5 degrees of freedom for P = 0.012.

† Percentages do not add to 100% because of missing genotype data; same controls were used for both ER-positive and ER-negative models.

‡ *P*s for test of linear trend.

§ *P*s for etiologic heterogeneity by ER status were obtained from case-only models comparing ER-negative cases to ER-positive cases.

‖ htSNPs listed in 5′ to 3′ order.

Table 4. Association between HSD17B1 haplotype combinations and invasive breast cancer among U.S. Caucasian women by ER status

Observed combination*	ER-positive†			ER-negative†		
	Cases (%), *n* = 1,737	Controls (%), *n* = 2,982	OR (95% confidence interval)	Cases (%), *n* = 354	Controls (%), *n* = 2,982	OR (95% confidence interval)
CGAT-CGAT‡	270 (16)	451 (15)	1.00	69 (19)	451 (15)	1.00
CGAT-AAGC	344 (20)	639 (22)	0.90 (0.74-1.10)	64 (18)	639 (22)	0.64 (0.44-0.92)
CAGC-AAGC	257 (15)	431 (14)	1.00 (0.81-1.25)	41 (12)	431 (14)	0.59 (0.39-0.90)
CGAC-AAGC	59 (3)	87 (3)	1.14 (0.79-1.65)	7 (2)	87 (3)	0.50 (0.22-1.14)
AAGC-AAGC	118 (7)	207 (7)	0.96 (0.73-1.26)	18 (5)	207 (7)	0.56 (0.33-0.97)
CGAT-CAGC	373 (21)	645 (22)	0.96 (0.79-1.17)	90 (25)	645 (22)	0.89 (0.63-1.25)
CGAT-CGAC	97 (6)	>142 (5)	1.10 (0.82-1.49)	22 (6)	>142 (5)	0.96 (0.57-1.62)
CAGC-CAGC	124 (7)	198 (7)	1.03 (0.79-1.36)	23 (7)	198 (7)	0.75 (0.45-1.24)
CAGC-CGAC	50 (3)	97 (3)	0.88 (0.60-1.28)	13 (4)	97 (3)	0.95 (0.50-1.80)
CGAC-CGAC	5 (<1)	8 (<1)	1.00 (0.32-3.12)	2 (1)	8 (<1)	1.55 (0.32-7.53)
Other combinations	15 (1)	28 (1)	0.87 (0.45-1.67)	2 (1)	28 (1)	0.48 (0.11-2.06)

NOTE: Data on receptor status was available from the following cohorts: CPS-II, NHS, WHS, MEC.

*Calculated by rounding the estimated haplotype frequencies to whole numbers. Observations with rounded haplotypes that did not total two copies were excluded, as well as people with unknown genotype for three of the four SNPs.

†Global test of diplotype association with ER-positive breast cancer: X^2 = 4.20 with 10 degrees of freedom for P = 0.937; global test of diplotype association with ER-negative breast cancer: X^2 = 15.47 with 10 degrees of freedom for a P = 0.116.

‡The reference group is made up of observations with two copies of the most common haplotype (*CGAT*).

common haplotype had risk of 1.2 for each copy of the haplotype and the risks associated with other haplotypes were all equal to each other. Similarly, the FPRP values for the htSNPs that were significant at 0.05 level in ER-negative tumors were below 0.2 for priors of 0.01 but not for priors of 0.001 or below (data not shown).

Discussion

This is the first study of *HSD17B1* and breast cancer that comprehensively characterizes the variation in and around *HSD17B1*. Despite the large sample size and comprehensive approach for identifying common SNPs and haplotypes, we found no evidence of association overall between breast cancer and variants of *HSD17B1* that are common in Caucasians. However, among the subset of U.S. Caucasian cases with ER-negative tumors, we found evidence of an association with each of the four htSNPs, and with the corresponding haplotypes that carry all high-risk alleles (*CGAT*) or all low-risk alleles (*AAGC*). Analysis of haplotype combinations showed that every common haplotype combination that included *AAGC* was associated with a reduced risk of ER-negative breast cancer. However, the FPRP calculations suggest that these associations may be due to chance and thus should be considered preliminary until they can be confirmed in additional studies that have a large number of ER-negative tumors.

Results from previous epidemiologic studies (3–6) have found no overall association with *HSD17B1* and breast cancer but have suggested that an association may exist in specific subgroups defined by BMI (3, 5), advanced stage (5), menopausal status (6), or parity (6). We did not find evidence of any association in these subgroups, despite the large size of our study and comprehensive evaluation of the gene.

No previous study has reported an association with *HSD17B1* in a subgroup of ER-negative tumors, but it would be difficult to detect in a smaller study because ER-negative tumors typically make up <25% of all breast cancers in Western countries (27, 28). Given our a priori knowledge about the activity of this gene in breast tissue (2, 29), this finding was unexpected. However, etiologic differences between ER-positive and ER-negative tumors is a topic of considerable debate and active research (27, 30). Clinical, epidemiologic, and laboratory data show that ER-positive and ER-negative breast tumors have important differences (31). Epidemiologic studies suggest that risk factors differ by receptor status (30, 32, 33), and ER-positive and ER-negative tumors display different gene expression profiles (34, 35). Our data suggest that germ line variation may also influence ER status.

Further investigation is needed to confirm this association with ER-negative breast cancer, and, if confirmed, isolate the causal variant responsible for this association with *HSD17B1*-containing haplotypes that define a high linkage disequilibrium block spanning 33 kb, including *HSD17B1* and its 5′ and 3′ regions. There is evidence that several upstream regions that lie well within this block participate in the regulation of *HSD17B1* expression (8, 29). This region also includes two other genes: *NAGLU* (5′) and *TCFL4* (3′) (Fig. 1). Although there is no a priori reason to suspect that these other genes are associated with breast cancer, they cannot be definitively excluded as possible candidates until further characterization of this region is complete.

The strengths of the BPC3 include its unprecedented sample size and comprehensive characterization of variation around the *HSD17B1* locus. Our analysis provides powerful null evidence against a main effect association between the overall risk of breast cancer and variants in *HSD17B1* that are common among

Caucasians and in subgroups defined by common breast cancer risk factors. The subgroup association that we did observe among ER-negative tumors should be viewed as preliminary and evaluated in future studies.

Acknowledgments

Received 10/4/2005; revised 11/22/2005; accepted 12/9/2005.

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

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 participants in the component cohort studies and the expert contributions of Hardeep Ranu, Craig Labadie, Lisa Cardinale, Shamika Ketkar (Harvard University), Merideth Yeager, Robert Welch, Cynthia Glaser, Laurie Burdett (National Cancer Institute), Loreall Pooler (University of Southern Califonia), Antonia Trichopoulou, H. Bas Bueno de Mesquita, Heiner Boeing, Domenico Palli, Salvatore Panico, Rosario Tumino, Paolo Vineis, Carlos A. Gonzalez, Carmen Martinez-Garcia, Miren Dorronsoro, and Goran Hallmans (EPIC).

References

1. Oduwole O, Li Y, Isomaa V, et al. 17Beta-hydroxysteroid dehydrogenase type 1 is an independent prognostic marker in breast cancer. Cancer Res 2004;64:7604–9.
2. Gunnarsson C, Ahnstrom M, Kirschner K, et al. Amplification of HSD17B1 and ERBB2 in primary breast cancer. Oncogene 2003;22:34–40.
3. Setiawan V, Hankinson S, Colditz G, Hunter D, DeVivo I. HSD17B1 gene polymorphisms and risk of endometrial and breast cancer. Cancer Epidemiol Biomarkers Prev 2004;13:213–9.
4. Mannermaa A, Peltoketo H, Winqvist R, et al. Human familial and sporadic breast cancer: analysis of the coding regions of the 17beta-hydroxysteroid dehydrogenase 2 gene (EDH17B2) using a single-strand conformation polymorphism assay. Hum Genet 1994;93:319–24.
5. Feigelson H, McKean-Cowdin R, Coetzee G, Stram D, Kolonel L, Henderson B. Building a multigenic model of breast cancer susceptibility: CYP17 and HSD17B1 are two important candidates. Cancer Res 2001;61:785–9.
6. Wu A, Seow A, Arakawa K, Berg DVD, Lee H-P, Yu M. HSD17B1 and CYP17 polymorphisms and breast cancer risk among Chinese women in Singapore. Int J Cancer 2003;104:450–7.
7. Normand T, Narod S, Labrie F, Simard J. Detection of polymorphisms in the estradiol 17beta-hydroxysteroid dehydrogenase 2 gene at the EDH17B2 locus on 17q11–21. Hum Mol Genet 1993;2:479–83.
8. Peltoketo H, Piao Y, Mannermaa A, et al. A point mutation in the putative TATA box, detected in nondiseased individuals and patients with hereditary breast cancer, decreases promoter activity in the 17beta-hydroxysterpod dehydrogenase type 1 gene 2 (EDH17B2) *in vitro*. Genomics 1994;23:250–2.
9. Puranen T, Poutanen M, Peltoketo H, Vihko P, Vihko R. Site-directed mutagenesis of the putative active site of human 17beta-hydroxysteroid dehydrogenase type 1. Biochem J 1994;304:289–93.
10. Hunter D, Riboli E, Haiman C, et al. The NCI Cohort Consortium on breast and prostate cancer: rationale and design. Nat Rev Cancer 2005;5:977–85.
11. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Nutrition Cohort: rationale, study design and baseline characteristics. Cancer 2002;94:2490–501.
12. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5:1113–24.
13. Hankinson SE, Willett WC, Manson JE, et al. Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. J Natl Cancer Inst 1998;90:1292–9.
14. Rexrode K, Lee I, Cook N, Hennekens C, Buring J. Baseline characteristics of participants in the Women's Health Study. J Womens Health Gend Based Med 2000;9:19–27.
15. Kolonel L, Henderson B, Hankin J, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151:346–57.
16. Kolonel L, Altshuler D, Henderson B. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. Nat Rev Cancer 2004;4:519–27.
17. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
18. Stram D, Haiman C, Hirschhorn J, 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.
19. Packer BR, Yeager M, Staats B, et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. Nucleic Acids Res 2004;32 Database issue:D528–32.
20. Zaykin D, Westfall P, Young S, Karnoub M, Wagner M, Ehm M. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 2002;53:79–91.
21. Kraft P, Cox D, Paynter R, Hunter D, De Vivo I. Accounting for haplotype uncertainty in association studies: a comparison of simple and flexible techniques. Genet Epidemiol 2005;28:261–72.
22. Xie R, Stram D. Asymptotic equivalence between two score tests for haplotype-specific risk in general linear models. Genet Epidemiol 2005;29:166–70.
23. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004;96:434–42.
24. Feigelson HS, Ross RK, Yu MC, Coetzee GA, Reichardt JK, Henderson BE. Genetic susceptibility to cancer from exogenous and endogenous exposures. J Cell Biochem Suppl 1996;25:15–22.
25. The Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94:606–16.
26. Freidlin B, Zheng G, Li Z, Gastwirth J. Trend tests for case-control studies of genetic markers: power, sample size and robustness. Hum Hered 2002;53:146–52.
27. Allred D, Brown P, Medina D. The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer. Breast Cancer Res 2004;6:240–45.
28. Anderson W, Chatterjee N, Ershler W, Brawley O. Estrogen receptor breast cancer phenotypes in the Surveillance, Epdiemiology and End Results database. Breast Cancer Res Treat 2002;76:27–36.
29. Petloketo H, Isomaa V, Ghosh D, Vihko P. Estrogen metabolism genes: HSD17B1 and HSD17B2. In: Henderson B, Ponder B, Ross R, editors. Hormones, genes, and cancer. New York: Oxford University Press; 2003. p. 181–98.
30. Colditz G, Rosner B, Chen W, Holmes M, Hankinson S. Risk factors for breast cancer according to estrogen and progesterone receptor status. J Natl Cancer Inst 2004;96:218–28.
31. Albain K. Adjuvant chemotherapy for lymph node-negative, estrogen receptor-negative breast cancer: a tale of three trials. J Natl Cancer Inst 2004;96:1801–4.
32. Althuis M, Fergenbaum J, Garcia-Closas M, Brinton L, Madigan M, Sherman M. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev 2004;13:1558–68.
33. Ursin G, Bernstein L, Lord S, et al. Reproductive factors and subtypes of breast cancer defined by hormone receptor and histology. Br J Cancer 2005;93:364–71.
34. Gruvberger S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 2001;61:5979–84.
35. Chang J, Hilsenbeck S, Fuqua S. The promise of microarrays in the management and treatment of breast cancer. Breast Cancer Res 2005;7:100–4.