

Genetic Variation at the *CYP19A1* Locus Predicts Circulating Estrogen Levels but not Breast Cancer Risk in Postmenopausal Women

Christopher A. Haiman,¹ Laure Dossus,² V. Wendy Setiawan,¹ Daniel O. Stram,¹ Alison M. Dunning,³ Gilles Thomas,⁴ Michael J. Thun,⁵ Demetrius Albanes,⁴ David Altshuler,⁶ Eva Ardanaz,⁷ Heiner Boeing,⁸ Julie Buring,⁹ Noël Burt, ⁶ Eugenia E. Calle,⁵ Stephen Chanock,⁴ Françoise Clavel-Chapelon,¹² Graham A. Colditz,¹¹ David G. Cox,¹¹ Heather Spencer Feigelson,⁵ Susan E. Hankinson,¹⁰ Richard B. Hayes,⁴ Brian E. Henderson,¹ Joel N. Hirschhorn,⁶ Robert Hoover,⁴ David J. Hunter,¹¹ Rudolf Kaaks,² Laurence N. Kolonel,¹³ Loïc Le Marchand,¹³ Per Lenner,¹⁴ Eiliv Lund,¹⁵ Salvatore Panico,¹⁶ Petra H. Peeters,¹⁷ Malcolm C. Pike,¹ Elio Riboli,¹⁸ Anne Tjønneland,¹⁹ Ruth Travis,²⁰ Dimitrios Trichopoulos,²¹ Sholom Wacholder,⁴ and Regina G. Ziegler⁴

¹Department of Preventive Medicine, University of Southern California, Los Angeles, California; ²Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; ³Cancer Research United Kingdom, Department of Oncology, Strangeways Research Laboratory, University of Cambridge, Cambridge, United Kingdom; ⁴Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; ⁵American Cancer Society, Atlanta, Georgia; ⁶Broad Institute of Harvard and MIT, Cambridge, Massachusetts; ⁷Public Health Institute of Navarra, Pamplona, Spain; ⁸Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany; ⁹Division of Preventive Medicine, ¹⁰Channing Laboratory, Brigham and Women's Hospital, Department of Medicine, Harvard Medical School and ¹¹Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts; ¹²Institut National de la Sante et de la Recherche Medicale, Institut Gustave Roussy, Villejuif, France; ¹³Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, Hawaii; ¹⁴Department of Radiation Sciences/Oncology, Umeå University, Umeå, Sweden; ¹⁵Institute of Community Medicine University of Tromsø, Tromsø, Norway; ¹⁶Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy; ¹⁷Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands; ¹⁸Imperial College, London, United Kingdom; ¹⁹Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark; ²⁰Cancer Research United Kingdom Epidemiology Unit, University of Oxford, Oxford, United Kingdom; and ²¹Department of Hygiene and Epidemiology, School of Medicine, University of Athens, Athens, Greece

Abstract

The *CYP19A1* gene encodes the enzyme aromatase, which is responsible for the final step in the biosynthesis of estrogens. In this study, we used a systematic two-step approach that included gene resequencing and a haplotype-based analysis to comprehensively survey common genetic variation across the *CYP19A1* locus in relation to circulating postmenopausal steroid hormone levels and breast cancer risk. This study was conducted among 5,356 invasive breast cancer cases and 7,129 controls comprised primarily of White women of European descent drawn from five large prospective cohorts within the National Cancer Institute Breast and Prostate Cancer Cohort Consortium. A high-density single-nucleotide polymorphism (SNP) map of 103 common SNPs ($\geq 5\%$ frequency) was used to identify the linkage disequilibrium and haplotype patterns across the *CYP19A1* locus, and 19 haplotype-tagging SNPs were selected to provide high predictability of the common haplotype patterns. We found haplotype-tagging SNPs and common haplotypes spanning the coding and proximal 5' region of *CYP19A1* to be significantly associated with a 10% to 20% increase in endogenous estrogen levels in postmenopausal

women [effect per copy of the two-SNP haplotype rs749292-rs727479 (A-A) versus noncarriers; $P = 4.4 \times 10^{-15}$]. No significant associations were observed, however, with these SNPs or common haplotypes and breast cancer risk. Thus, although genetic variation in *CYP19A1* produces measurable differences in estrogen levels among postmenopausal women, the magnitude of the change was insufficient to contribute detectably to breast cancer. [Cancer Res 2007;67(5):1893–7]

Introduction

A doubling of postmenopausal endogenous levels is estimated to increase breast cancer risk by $\sim 30\%$ (1). After menopause, estrogen biosynthesis takes place predominantly in adipose tissue and is catalyzed by the aromatase enzyme (encoded by the gene *CYP19A1*) which converts androgens to estrogens. *CYP19A1* spans ~ 123 kb at 15q21.1 and comprises nine coding exons (II–X) and several alternative, untranslated first exons that are expressed under the control of tissue-specific promoters (2). Studies have suggested that variation at this locus may be a marker of circulating estrogen levels in postmenopausal women and contribute to risk of breast cancer (3–8). Results, however, have been inconsistent, and most studies have been underpowered to detect the modest magnitude of effect anticipated for a common low-penetrance allele or have focused on a limited number of markers across the locus. A combination of genomic approaches is applied here; this includes resequencing of the coding exons and a linkage disequilibrium (LD)-based analysis to systematically search for markers of circulating steroid hormones and breast cancer risk at the *CYP19A1* locus.

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

Requests for reprints: Christopher A. Haiman, Department of Preventive Medicine, University of Southern California Keck School of Medicine, 1441 Eastlake Avenue, MS 44, Room 4441, Los Angeles, CA 90033-0800. Phone: 323-865-0429; Fax: 323-865-0127; E-mail: haiman@usc.edu.

©2007 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-06-4123

Materials and Methods

This study includes five breast cancer case-control studies nested within established prospective cohorts: American Cancer Society Cancer Prevention Study II (9), European Prospective Investigation into Cancer and Nutrition (EPIC) cohort (10), Harvard Nurses' Health Study (NHS; ref. 11); Women's Health Study (12); and Multiethnic Cohort Study (MEC; ref. 13) that represent the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (Supplementary Materials). The nested case-control studies comprised 5,356 invasive breast cancer cases and 7,129 controls. 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. Within each cohort, controls were individually or frequency matched to cases by ethnicity, age, and country of residence (EPIC). Eighty percent of cases were postmenopausal, and with the exception of the MEC (which also comprises African Americans, Latinos, Japanese Americans, and Native Hawaiians) over 93% of women in each cohort were of European descent (range, 93–100%). Thus, our study focused on genetic variation at the *CYP19A1* locus common among ancestral European Whites.

Levels of estradiol (E_2), estrone (E_1), testosterone (T) and androstenedione (A) had been measured in prospectively collected samples (E_2 , N 3,431; E_1 , N 3,146; T, N 3,435; A, N 3,224) as part of previous studies among postmenopausal women within the EPIC (14), NHS (15), and MEC (16) cohorts. Because the hormone fractions were measured in different laboratories by either direct or indirect RIA, results are presented as the percentage change in geometric mean levels by genotype and haplotype class.

We initially resequenced the coding exons of *CYP19A1* in a panel of 95 women in the MEC (19 each from African Americans, Latinos, Japanese Americans, Native Hawaiians and Whites) with invasive, nonlocalized breast cancer. This panel provided $\geq 85\%$ power to detect a SNP with an ethnic-specific minor allele frequency of $\geq 5\%$ or an overall minor allele frequency of $\geq 1\%$. In the haplotype analysis, LD structure of *CYP19A1* was determined by genotyping 103 common SNPs and two rare missense SNPs (R264C and T201M) among a panel of 349 women from the MEC (including 70 Whites). Seventy-four SNPs had been used to characterize LD patterns at this locus in a previous study (6). Markers were distributed over a 181.0-kb region, from 22.5 kb 5' of exon 1.1 through 29.4 kb 3' of the 3' untranslated region (UTR; average SNP spacing 1.76 kb; Supplementary Table S1). The *CYP19A1* locus has been shown to contain four blocks of LD (Fig. 1; ref. 6). Within

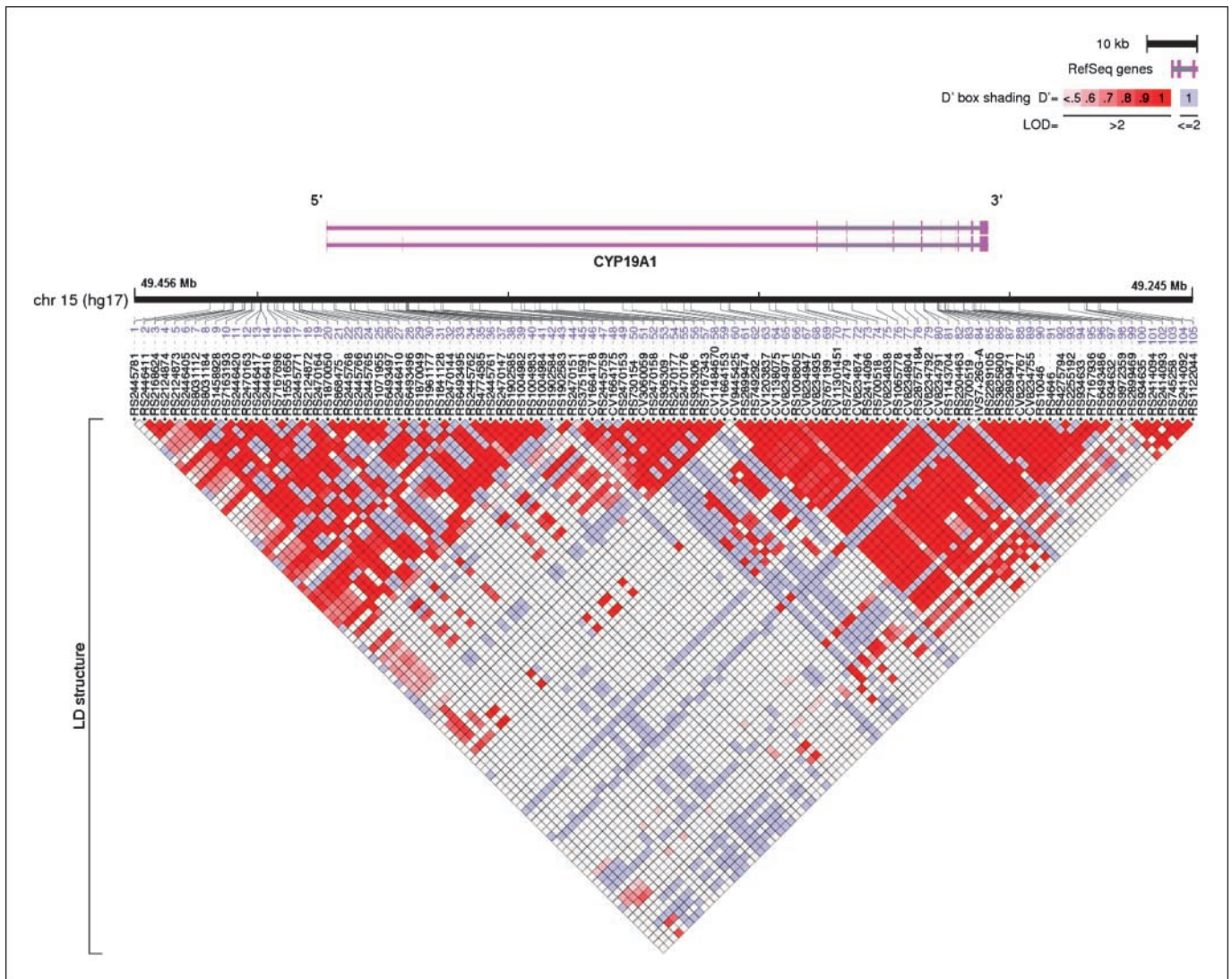


Figure 1. LD plot across the *CYP19A1* locus. Horizontal black line, region on chromosome 15 that contains the *CYP19A1* gene; purple, *CYP19A1* gene (RefSeq genes from the human genome assembly, May 2004); blue below the black line, 105 SNPs used for LD characterization. LD plot at the bottom is based on the measure D' . Diamonds, pairwise magnitudes of LD, strong LD (with red, $D' > 0.8$) and logarithm of odds score >2.0 . [Figure prepared with LocusView (<http://www.broad.mit.edu/mpg/locusview/>), Broad Institute, Cambridge, MA].

Downloaded from <http://aacrjournals.org/cancerres/article-pdf/67/5/1893/2867438/1893.pdf> by guest on 27 February 2024

Table 1. Associations of tagging SNPs in LD blocks 3 and 4 of *CYP19A1* with the percentage change in geometric mean E_2 and E_1 levels among postmenopausal women and breast cancer risk

SNP	$N = 3,431^*$	Increase E_2 versus ref. [†]	$N^* = 3,146$	Increase E_1 versus ref. [†]	Cases, $N = 5,356^\ddagger$	Controls, $N = 7,129^\ddagger$	Breast cancer OR (95% CI) [§]
Block 3							
rs749292 (49346023)							
GG	1103	ref.	1012	ref.	1,670	2,210	1.00 (ref.)
AG	1679	+4.9%	1548	+7.7%	2,530	3,396	1.00 (0.93–1.08)
AA	609	+14.4%	549	+12.1%	966	1,312	0.99 (0.89–1.08)
p^\P		1.3×10^{-10}		3.0×10^{-10}			0.79
CV8234971 (49337127)							
GG	1176	ref.	1085	ref.	1,842	2,417	1.00 (ref.)
AG	1653	+4.3%	1519	+5.4%	2,420	3,273	0.99 (0.92–1.06)
AA	522	+13.0%	474	+11.1%	837	1,093	1.01 (0.91–1.12)
p^\P		3.0×10^{-8}		7.0×10^{-8}			0.93
rs1008805 (49336891)							
GG	610	ref.	561	ref.	861	1,203	1.00 (ref.)
AG	1624	+4.2%	1497	+6.0%	2,440	3,220	1.05 (0.95–1.15)
AA	1151	+9.5%	1048	+10.8%	1,887	2,484	1.02 (0.93–1.13)
p^\P		4.8×10^{-6}		1.1×10^{-7}			0.81
Block 4							
rs727479 (49321839)							
CC	417	ref.	382	ref.	656	840	1.00 (ref.)
AC	1526	+9.2%	1409	+10.8%	2,209	3,006	0.95 (0.85–1.05)
AA	1411	+15.7%	1283	+17.2%	2,265	2,977	0.96 (0.86–1.07)
p^\P		6.4×10^{-11}		1.3×10^{-13}			0.72
rs2414096 (49317071)							
GG	942	ref.	873	ref.	1,583	2,070	1.00 (ref.)
AG	1702	+6.1%	1569	+6.7%	2,460	3,327	1.00 (0.92–1.08)
AA	742	+13.2%	660	+11.6%	1,147	1,523	1.00 (0.91–1.11)
p^\P		1.0×10^{-9}		6.7×10^{-9}			0.95
rs10046 (49290278)							
GG	835	ref.	768	ref.	1,378	1,817	1.00 (ref.)
AG	1691	+5.7%	1571	+5.4%	2,443	3,246	1.01 (0.93–1.09)
AA	799	+12.8%	710	+11.7%	1,251	1,671	1.00 (0.91–1.10)
p^\P		2.9×10^{-9}		1.1×10^{-8}			0.95
rs4646 (49290136)							
AA	278	ref.	258	ref.	426	585	1.00 (ref.)
AC	1347	+3.9%	1239	+5.4%	2,078	2,753	1.01 (0.90–1.15)
CC	1747	+7.5%	1594	+10.6%	2,656	3,526	1.01 (0.90–1.14)
p^\P		1.3×10^{-3}		1.7×10^{-6}			0.90

Abbreviations: OR, odds ratio.

* Total number of postmenopausal women analyzed for E_2 and E_1 .

† The results shown are adjusted for age at the time of blood draw, body mass index (BMI), laboratory batch, cohort study, country (EPIC), and ethnicity (MEC).

‡ The total number of cases and controls evaluated in the five cohort studies; numbers of cases and controls for each SNP do not add to the totals because of missing genotype data.

§ The odds ratios and 95% CI were estimated using conditional logistic regression.

|| SNP position is based on May 2004 human genome version 17. SNPs rs749292, CV8234971 (rs6493494) and rs2414096, rs10046 are strongly correlated $r^2 \geq 0.83$.

‡ P value for trend across genotype groups.

blocks, haplotype frequency estimates were constructed from genotype data for Whites using the expectation-maximization algorithm, and 19 tag SNPs were selected to predict ($r_{\text{H}}^2 \geq 0.7$) 15 common haplotypes (≥ 0.05 ; Supplementary Table S2; refs. 6, 17, 18).

Tag SNPs were genotyped using the TaqMan assay (Applied Biosystems, Foster City, CA) in three Breast and Prostate Cancer Cohort Consortium laboratories. Details of each assay can be found online.²² In each laboratory, assays were run on a designated set of 94 samples; concordance across laboratories was 99.6%. Genotype data quality from each laboratory was

assessed by typing 5% to 10% blinded samples; the concordance rates were $>99.7\%$. No deviation in Hardy-Weinberg equilibrium was observed (at the $P < 0.01$ level) across more than one study for any given SNP. The study protocol was approved by all institutional review boards.

Tag SNPs were genotyped in the HapMap CEU trios (19) to permit an assessment of coverage in relation to the HapMap database (phase II,

²² <http://www.uscnorris.com/MECGenetics/>

October 2005). These 19 tag SNPs were estimated to predict ($r^2 \geq 0.7$) 70% of all common SNPs genotyped in the HapMap CEU population across the four LD block regions.

In the analysis, we estimated haplotype frequencies and expected haplotype counts (0, 1, or 2) for each individual, conditional on that individual's genotype data (20). This was conducted within each study and by race (MEC) and country (EPIC). These expectations were included in conditional logistic regression models to estimate odds ratios and confidence intervals and, in linear models, to evaluate associations with circulating steroid hormone levels. To control for type I error over all the SNPs and haplotypes considered, we conducted global tests for the influence of any genomic variation in a given block by putting all common haplotypes or SNPs into a multidegree of freedom likelihood ratio or F test for risk/hormones, respectively.

Results and Discussion

We found highly significant relationships (Table 1) between tag SNPs and circulating estrogen levels among postmenopausal women with the strongest effects observed with SNPs in LD blocks 3 and 4 (P values for global effects by block for E_2/E_1 : block 1, 0.24/0.03; block 2, 0.16/ 1.4×10^{-5} ; block 3, 7.2×10^{-9} / 6.2×10^{-9} ; block 4, 5.9×10^{-11} / 1.3×10^{-12}). Some of the tagging SNPs in LD blocks 3 and 4 were highly correlated [$r^2 \geq 0.83$ for SNPs rs749292, CV8234971 (rs6493494) and rs2414096, rs10046]. The two most significantly associated SNPs, rs749292 (A allele) and rs727479 (A allele), were only modestly correlated with each other ($r^2 = 0.46$), and each remained significantly associated with circulating levels when modeled concurrently (E_2 , $P = 2.1 \times 10^{-3}$ and $P = 3.1 \times 10^{-4}$; E_1 , $P = 0.047$ and $P = 6.1 \times 10^{-8}$, respectively). A two-SNP haplotype ($A-A$) composed of these SNPs was found to be a more significant predictor of E_2 and E_1 levels ($P = 4.6 \times 10^{-13}$ and $P = 4.4 \times 10^{-15}$, respectively); however, it accounted for $\leq 2.0\%$ of the variation in estrogen levels. Haplotype $G-A$ (composed of these two SNPs) was also associated with levels when modeled together with

haplotype $A-A$ (E_2 , $P = 0.012$; E_1 , $P = 4.2 \times 10^{-5}$), indicating that neither of these multimarker haplotypes nor tagging SNPs fully capture the effects of the functional allele at this locus. These effects were observed in all three cohorts that measured hormone levels (Table 2). Blocks 3 and 4 span a >67-kb region encompassing the translated region, 3' UTR and promoters I.2, I.6, I.3, and PII.

Haplotypes in blocks 3 and 4 were also strongly associated with estrogen levels (P values for global haplotype effects by block for E_2/E_1 : block 1, 0.11/0.01; block 2, 0.07/ 4.3×10^{-5} ; block 3, 1.4×10^{-9} / 9.1×10^{-11} ; block 4, 1.1×10^{-10} / 7.5×10^{-14}). The magnitude of the associations was similar to the independent tagging SNPs in these blocks (Supplementary Tables S3 and S4). These SNPs and haplotypes were not significantly associated with androgen levels, whereas associations with androgen to estrogen conversion ratios (E_2/T and E_1/A), which were evaluated as a measure of enzyme activity, mirrored those of E_2 and E_1 (Supplementary Tables S5 and S6). Significant associations were observed with SNPs/haplotypes in blocks 1 and 2, although the effects were substantially smaller than those observed in blocks 3 and 4 and were most likely due to LD with the tagging SNPs in these blocks.

In the case-control analyses, none of the SNPs or haplotypes found to be related to postmenopausal E_2 and E_1 levels were significantly associated with breast cancer risk (Table 1). The odds ratio for the rs749292-rs727479 $A-A$ haplotype (versus $G-C$ haplotype) was 0.99 [95% confidence interval (95% CI), 0.94-1.05; $P = 0.74$]. And no SNPs or haplotypes in any other LD block were consistently associated with risk (Supplementary Tables S7 and S8). Global likelihood ratio tests to control for type I error over all SNPs or haplotypes in each block were not significant ($P \geq 0.32$). Associations were similar when limiting the analysis to cases with more advanced breast cancer ($n = 1,648$). Two correlated haplotypes ($r^2 = 0.56$) in blocks 2 and 3 (and associated tagging SNPs), although showing no overall association, seemed to display heterogeneous effects ($P < 0.01$) when viewed separately for each

Table 2. Percent change in E_2 and E_1 levels by haplotype $A-A$ (rs749292-rs727479) in *CYP19A1*

Study	Percentage increase in E_2 by number of copies of the $A-A$ haplotype			P^\dagger	Percentage increase in E_1 by number of copies of the $A-A$ haplotype			P^\dagger
	0	1	2		0	1	2	
Postmenopausal women*								
EPIC								
Cases	ref.	+3.4	+8.2	0.078	ref.	+2.3	+5.4	0.21
n^\ddagger	209	325	106		200	308	100	
Controls	ref.	+6.7	+12.4	9.8×10^{-5}	ref.	+8.9	+11.5	1.2×10^{-5}
n^\ddagger	458	624	211		443	602	210	
NHS								
Cases	ref.	+2.7	+19.3	8.6×10^{-3}	ref.	+14.6	+22.7	2.3×10^{-3}
n^\ddagger	102	151	78		83	128	55	
Controls	ref.	+7.0	+25.2	8.0×10^{-6}	ref.	+4.7	+13.3	0.012
n^\ddagger	201	331	115		151	265	87	
MEC								
Controls	ref.	+9.4	+23.9	5.0×10^{-4}	ref.	+12.3	+25.4	8.8×10^{-5}
n^\ddagger	193	244	77		193	241	76	

*The results shown are adjusted for age at the time of blood draw, BMI, laboratory batch and country (EPIC), or ethnicity (MEC).

$^\dagger P$ value for trend across genotype groups.

‡ The estimated number of cases and controls carrying 0, 1, or 2 copies of each haplotype.

cohort (Supplementary Table S9). Aside from the MEC, which did not contribute to the observed heterogeneous effects, these cohorts largely represent populations of European ancestry. Thus, heterogeneity of the results does not represent population differences and is more likely to reflect random variation than true biological differences. This highlights the importance of large-scale studies to resolve inconsistent findings from single small studies.

No novel nonsynonymous SNPs were identified through resequencing efforts. SNPs R264C (rs700519) and T201M (rs28757184) were not associated with breast cancer risk or circulating hormone levels in any single cohort or in the pooled analysis (Supplementary Tables S3 and S10).

In conclusion, we provide compelling statistical evidence that the *CYP19A1* locus harbors genetic variation that is associated with circulating estrogen levels among postmenopausal women. In a study among 1,747 postmenopausal women that examined only two polymorphisms at this locus, Dunning et al. (7) found SNP rs10046 to be a modest predictor of E_2 levels ($\sim 6\%$ mean increase per allele, $P = 2 \times 10^{-4}$). In our more comprehensive analysis of common variation spanning the entire *CYP19A1* locus, we not only corroborated the observation with this particular SNP but also identified even stronger markers of circulating postmenopausal hormone levels at this locus. Whereas through resequencing we have ruled out nonsynonymous variants, the identity of the functional variant at this locus remains unknown. Screening for variation in noncoding and regulatory regions spanning LD blocks 3 and 4 among individuals with the multimarker *A-A* haplotype

(rs749292-rs727479) should be done to enumerate all possible functional alleles at this locus.

Based on previous epidemiologic studies (1), one can estimate that a functional variant associated with a 10% to 20% increase in postmenopausal estrogen levels for heterozygous or homozygous carriers would generate, respectively, approximately a 1.07-fold or 1.15-fold increase in risk of breast cancer (assuming log linearity between levels and risk) and, therefore, be potentially detectable in a study of this size. Our combined analysis of five cohort studies, although extremely large by traditional standards, is still not sufficiently powered to consistently detect effects at <1.10 per allele even for SNPs or haplotypes with very high frequency. If variants in several genes affect estrogen exposure in a similar fashion, then the combination might have a more measurable effect on breast cancer risk. Whether aromatase inhibitors used as hormonal therapy or as chemoprevention may be more effective among women with modestly elevated estrogen profiles as determined by *CYP19A1* genotype should be examined.

Acknowledgments

Received 11/8/2006; revised 1/6/2007; accepted 1/18/2007.

Grant support: National Cancer Institute cooperative agreements UO1-CA98233, UO1-CA98710, UO1-CA98216, and UO1-CA98758 and Intramural Research Program of 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 cohort participants.

References

- Key T, Appleby P, Barnes I, Reeves G. 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.
- Harada N, Utsumi T, Takagi Y. Tissue-specific expression of the human aromatase cytochrome P-450 gene by alternative use of multiple exons 1 and promoters, and switching of tissue-specific exons 1 in carcinogenesis. *Proc Natl Acad Sci U S A* 1993;90:11312–6.
- Kristensen VN, Harada N, Yoshimura N, et al. Genetic variants of *CYP19A1* (aromatase) and breast cancer risk. *Oncogene* 2000;19:1329–33.
- Healey CS, Dunning AM, Durocher F, et al. Polymorphisms in the human aromatase cytochrome P450 gene (*CYP19A1*) and breast cancer risk. *Carcinogenesis* 2000;21:189–93.
- Haiman CA, Hankinson SE, Spiegelman D, et al. A tetranucleotide repeat polymorphism in *CYP19A1* and breast cancer risk. *Int J Cancer* 2000;87:204–10.
- Haiman CA, Stram DO, Pike MC, et al. A comprehensive haplotype analysis of *CYP19A1* and breast cancer: the Multiethnic Cohort. *Hum Mol Genet* 2003;12:2679–92.
- Dunning AM, Dowsett M, Healey CS, et al. Polymorphisms associated with circulating sex hormones in postmenopausal women. *J Natl Cancer Inst* 2004;96:936–45.
- Travis RC, Churchman M, Edwards SA, et al. No association of polymorphisms in *CYP17*, *CYP19A1*, and *HSD17B1* with plasma estradiol concentrations in 1,090 British women. *Cancer Epidemiol Biomark Prev* 2004;13:2282–4.
- 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.
- 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.
- Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;5:388–96.
- 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.
- Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;151:436–57.
- 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.
- 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.
- Setiawan VW, Haiman CA, Stanczyk F, Le Marchand L, Henderson BE. Racial/Ethnic differences in postmenopausal endogenous hormones: the Multiethnic cohort study. *Cancer Epidemiol Biomark Prev* 2006;15:1849–55.
- 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.
- Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–9.
- International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–320.
- Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered* 2002;53:79–91.