Common genetic variation in the sex hormone metabolic pathway and endometrial cancer risk: pathway-based evaluation of candidate genes

Hannah P. Yang1,2, *, Jesus Gonzalez Bosquet1, Qizhai Li1,7, Elizabeth A. Platz2, Louise A. Brinton1, Mark E. Sherman1, James V. Lacey Jr1, Mia M. Gaudet2, Laurie A. Burdette1,4, Jonine D. Figueroa1,8, Julia C. Ciampa1, Jolanta Lissowska5, Beata Peplonska6, Stephen J. Chanock1,4 and Montserrat Garcia-Closas1

1Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20852, USA, 2Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA, 3Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY 10461, USA, 4Core Genotype Facility at the National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA, 5Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland and 6Department of Occupational and Environmental Epidemiology, Nofer Institute of Occupational Medicine, Lodz, Poland

*Present address: Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing, China

†Present address: Division of Cancer Etiology, Department of Population Sciences, Beckman Research Institute, City of Hope, Duarte, CA 91010, USA

‡To whom correspondence should be addressed. National Cancer Institute, 6120 Executive Boulevard, E0S/5034a, Rockville, MD 20852, USA. Tel: +1 301 594 5640; Fax: +1 301 402 0916; Email: yanghan@mail.nih.gov

Background. Estrogen plays a major role in endometrial carcinogenesis, suggesting that common variants of genes in the sex hormone metabolic pathway may be related to endometrial cancer risk. In support of this view, variants in CYP19A1 [cytochrome P450 (CYP), family 19, subfamily A, polypeptide 1] have been associated with both circulating estrogen levels and endometrial cancer risk. Associations with variants in other genes have been suggested, but findings have been inconsistent. Methods: We examined 36 sex hormone-related genes using a tagging approach in a population-based case-control study of 417 endometrial cancer cases and 407 controls conducted in Poland. We evaluated common variation in these genes in relation to endometrial cancer risk using sequential haplotype scan, variable-sized sliding window and adaptive rank-truncated product (ARTP) methods. Results. In our case-control study, the strongest association with endometrial cancer risk was for AR (androgen receptor; ARTP P = 0.006). Multilocus analyses also identified boundaries for a region of interest in AR and in CYP19A1 around a previously identified susceptibility loci. We did not find evidence for consistent associations between previously reported candidate single-nucleotide polymorphisms in this pathway and endometrial cancer risk. Discussion. In summary, we identified regions in AR and CYP19A1 that are of interest for further evaluation in relation to endometrial cancer risk in future haplotype and subsequent fine mapping studies in larger study populations.

Introduction

Endometrial cancer is the fourth most common cancer among women and the most common gynecological cancer in the USA (1). Hormonal factors have been identified as major etiologic factors for endometrial cancer (2). Data from cohort and case-control studies have demonstrated consistent links between elevated circulating sex steroid hormones and endometrial cancer risk (3–5). Epidemiologic data show that many exposures that are associated with endometrial cancer may increase total estrogen exposure, including use of exogenous unopposed estrogen or tamoxifen, early ages at menarche, late ages at menopause, nulliparity and obesity (2). In addition, a 3-fold increased endometrial risk has been associated with a family history of endometrial cancer (6), suggesting the contribution of genetic predisposition of endometrial cancer.

These data suggest that interindividual variation in genes involved in sex hormone metabolic pathway could be related to endometrial cancer risk. Two variants in CYP19A1 [cytochrome P450 (CYP), family 19, subfamily A, polypeptide 1] (aromatase), which have been shown to be related to circulating estrogen levels (7), have been demonstrated to be associated with endometrial cancer risk in a recent pooled analysis of 10 studies including nearly 5000 endometrial cancer cases (8). A large number of independent studies have also assessed associations between other genes in the sex hormone metabolic pathway and endometrial cancer risk with inconsistent results. To date, single-nucleotide polymorphisms (SNPs) in AR (androgen receptor) (9,10), COMT (catechol-O-methyltransferase) (11–15), CYP11A1 (CYP, family 1, subfamily A, polypeptide 1) (11,14,16–23), CYP12A1 (CYP, family 1, subfamily A, polypeptide 2) (11,14,19,21), CYP3A4 (CYP, family 3, subfamily A, polypeptide 4) (14,24), CYP17A1 (CYP, family 17, subfamily A, polypeptide 1) (19,25–33), CYP19A1 (CYP, family 19, subfamily A, polypeptide 1) (6,8,21,26,31,32,34), CYP1B1 (CYP, family 1, subfamily B, polypeptide 1) (11–15,19,35,36), ESR1 (estrogen receptor 1) (37–41), ESR2 (estrogen receptor 2) (37,42), HSD17B1 [hydroxysteroid (17-beta) dehydrogenase 1] (43,44), PGR (progesterone receptor) (14,45–48), SHBG (sex hormone-binding globulin) (14,49,50), SULT1A1 (sulfotransferase family, cytosolic, 1A, phenol-prefering, member 1) (14,19,21), SULT1E1 (sulfotransferase family 1E, estrogen-prefering, member 1) (14,19), UGT1A1 (uridine diphosphate glucuronosyltransferase 1A1, family 1, subfamily A, polypeptide 1), UGT1A2 (uridine diphosphate glucuronosyltransferase 1A2, family 1, subfamily A, polypeptide 2), UGT2B1 (uridine diphosphate glucuronosyltransferase 2B7, family 1, subfamily B, polypeptide 7) (53) have been examined. Most of these studies, however, focused on candidate SNPs, rather than a more comprehensive assessment of common genetic variation on potentially important hormone-related genes in a pathway-based analysis.

In the present population-based case-control study, we evaluated 36 sex hormone-related genes using a tagging approach. These include SNPs in genes involved in synthesis and bioactivation of estrogen and other sex hormones [AKR1C1 (aldo-keto reductase family 1, member C1), AKR2C2 (AKR family 2, member C2), AR, CYP17, CYP19A1, CYP11A1 (CYP, family 11, subfamily A, polypeptide 1), CYP11B1 (CYP, family 11, subfamily A, polypeptide 1), CYP11B2 (CYP, family 11, subfamily B, polypeptide 2), CYP21A2 (CYP, family 21, subfamily A, polypeptide 2), ESR1, ESR2, HSD17B1, HSD17B2, HSD17B3, HSD17B4, HSD3B1 (hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1), HSD3B2, HA02 (hydroxycacid oxidase 2), PGR, SHBG, SRS5A5 (steroid-5-alpha-reductase, alpha polypeptide 1), SRS5A2, STAR (steroidogenic acute regulatory protein), SULT1A1, SULT1A2, SULT1B1 and SULT1E1, synthesis of catecholestrogen (CYP1A1, CYP1A2, CYP1B1 and CYP3A4) and inactivation of estrogen, catecholestrogen or its products (COMT, UGT1A8, UGT1A9, UGT1A10, UGT2B11 and UGT2B17). We examined individual SNPs and individual genes using sequential haplotype scan, variable-sized sliding window and adaptive rank-truncated product (ARTP) methods.
Study population

Data were derived from a population-based case–control study carried out in two cities in Poland (Warsaw and Łódź) during 2001–2003. The design and conduct of this study have been described in detail elsewhere (27,54). Briefly, the cases were women aged 20–74 years old, newly diagnosed with pathologically confirmed endometrial cancer and identified through participating hospitals and local cancer registries. Controls without a history of breast or endometrial cancer and with an intact uterus at time of enrollment were randomly selected from a database of all residents and frequency matched to cases. The cases were women aged 20–74 years old, newly diagnosed with pathological endometrial cancer and with an intact uterus at time of enrollment were randomly selected from a database of all residents and frequency matched to cases (1:2 case:control ratio) by study site (Warsaw and Łódź) and age in 5 year categories. For comparability, endometrial cancer cases included in the study did not have a history of breast cancer.

In-person interviews, including information on demographic characteristics and known or suspected endometrial cancer risk factors, were obtained from 551 cases (79% of the 695 eligible cases) and 1925 controls (68% of the 2843 eligible controls). Trained interviewers collected venous blood from 88% of participating cases and 94% of participating controls. The study protocol was reviewed and approved by local Polish and US National Cancer Institute institutional review boards. All participants provided written informed consent.

SNP selection

Candidate genes were identified from reviews of the sex hormone metabolic pathway (55–57) and cross-referenced with the Gene Ontology database (58,59) and the Kegg Pathway (60) to confirm pathway information. The 36 candidate genes that are included in this study and their roles in sex hormone metabolic pathway are shown in Figure 1. Tag SNPs were selected to capture the common genetic variation across the gene based on the Carlson method (61) and linkage disequilibrium information in the regions of interest in the International HapMap Project (Release 21 with National Center for Biotechnology Information build 36, www.hapmap.org/), using the Genotype Library and Utilities software (http://code.google.com/p/glu-genetics/). The following criteria were used to select SNPs in these genes: located in the respective gene or within the 10 kb upstream or 5 kb downstream of the gene, had a minor allele frequency ≥5% in various ethnic groups and not already represented by a current tag SNP at an linkage disequilibrium of r² ≥0.80. All selected SNPs were included in the design unless their design failed. Additional information for the SNPs in this study is provided in supplementary Table 1 (available at Carcinogenesis Online).

Genotyping

To approximate a 1:1 case–control ratio for genotyping, we selected all cases and randomly selected 5 year age categories and study site frequency-matched controls who had donated blood. Adequate amounts of DNA for genotyping were available for 447 cases and 439 controls after excluding samples with poor DNA quality or quantity and plating requirements. Genotyping was completed at the Core Genotyping Facility of the National Cancer Institute’s Division of Cancer Epidemiology and Genetics using the iSelect bead chip (Illumina Custom Infinium, http://www.illumina.com/pages.illum?ID=158) as part of a large-scale genotyping effort that included ~29 000 SNPs with coverage on >1300 genes for different tumor types. Description and methods for assays can be found at http://cgf.nci.nih.gov/operations/multiplex-genotyping.html.

Subject completion percentages for cases and controls were >95% for all assays, after removing 30 cases and 32 controls that failed lab validation. To determine assay reliability specifically for our study, we scattered 53 blinded duplicate samples from our study on the DNA plates. The concordance between quality control study samples was >98%. To monitor quality control procedure, we included 270 HapMap samples to calculate concordance and completion proportions. We excluded those SNPs with <90% completion and <95% concordance (78 SNPs for 36 genes). Thus, the analysis is based on 815 SNPs for 36 genes in 417 cases and 407 controls.

Statistical analyses

Gene-based analyses. We calculated a standardized summary for the association between a gene and endometrial cancer risk through adaptive combination of P-values using the ARTP method (62). For the ARTP statistic, we used a set of five truncated points at the SNP level and at only one truncation point for the gene-level summary. To obtain the empirical distribution of P-values under the null hypothesis of no association between the SNP and endometrial cancer, we performed 10 000 permutations of randomly generated populations. Multiple comparisons have been adjusted by permutation testing.

In addition, we performed haplotype analyses for genes showing evidence of association from our gene-based analyses as well as CYP19 that was recently reported in a recent pooled analysis to be related to endometrial cancer risk (8). We reconstructed haplotypes using fastPHASE (63) and identified significant haplotypes using the sequential haplotype scan (R package ‘seqhap’) (64) and variable-sized sliding window (R package ‘vsrww’) (65) methods. Both methods examine the association of a binary trait over a set of biallelic loci and optimize the choice of tag SNPs for haplotype analysis using different algorithms. In the sequential haplotype scan method, adjacent tag SNPs are added to form the haplotype if an additional marker contributes information for detecting haplotype associations, conditional on the initial locus. This process extends the haplotype while testing the contribution of the additional allele (64). In the variable-sized sliding window method, at each initial locus, the maximum window size for the haplotype analysis is determined and then

![Fig. 1. Description of the sex hormone metabolic pathway: genes involved in synthesis, bioactivation, metabolism, and inactivation of estrogen and other hormones.](http://example.com/fig1.png)
a regularized regression method is used for the joint analysis effect of all haplotypes of different lengths (65). Multiple comparisons have been adjusted by permutation testing and Bonferroni correction for the sequential scan and variable-sized window methods, respectively. For the significant haplotypes identified from these two methods, haplo.stats (66) was used to evaluate the overall difference in haplotype frequencies between cases and controls as well as the individual haplotype frequencies and to calculate odds ratios and 95% confidence intervals.

SNP-based analyses. The frequency of each SNP was compared between cases and controls using the χ² test. Odds ratios and 95% confidence intervals were calculated using unconditional logistic regression with adjustment for frequency-matching variables, age in 5 year categories and study site. Genotypes were coded using indicator variables with the homozygous genotype of the most common allele serving as the reference. We assumed a log additive mode of inheritance (i.e. risk increases or decreases with each additional minor allele) to calculate the mode of inheritance (i.e. risk increases or decreases with each additional minor allele) to calculate the expected protective association. We assumed a log additive mode of inheritance (i.e. risk increases or decreases with each additional minor allele) to calculate the expected protective association. We assumed a log additive mode of inheritance (i.e. risk increases or decreases with each additional minor allele) to calculate the expected protective association. We assumed a log additive mode of inheritance (i.e. risk increases or decreases with each additional minor allele) to calculate the expected protective association.

Results
The characteristics of the endometrial cancer risk cases and controls and reasons for non-participation have been previously reported (27,54). In brief, cases have more years of education, earlier ages at menarche, fewer pregnancies, younger age at first birth; are heavier and non-users of oral contraceptives and users of oral menopausal hormone than controls. The majority of cases and controls included in this analysis (77 and 85%, respectively) were post-menopausal.

Gene-based analyses
AR (ARTP P = 0.006) was statistically significantly associated with endometrial cancer risk; other genes were weakly or not associated with endometrial cancer risk (P > 0.01; Table I). In addition, we performed sequential haplotype scan and variable-sized sliding window analyses for AR, the top hit gene from our gene-based analyses, and CYP19, the endometrial cancer susceptibility gene reported in a recent pooled analysis (8). We observed that particular regions within AR, largely marked by rs12011793 (chromosome X: position 66834816) and rs12011518 (66849761), and CYP19A1, primarily marked by rs28757184 (chromosome 15: position 49301864), rs16964201 (49302643), rs9944225 (49315372), rs700518 (49315404) and rs17038833 (4917389), were associated with endometrial cancer risk (Figure 2). A summary of the odds ratios for the most significant haplotypes, identified from sequential scan and variable-sized sliding window methods, and endometrial cancer risk is provided in supplementary Table 2 (available at Carcinogenesis Online).

SNP-based analyses
Although the main analyses were gene-based, we also evaluated individual SNPs in relation to endometrial cancer risk since they can be more powerful if associations are driven by a single-tagged SNP or candidate SNP. These analyses also make comparisons easier with...
other studies using different tag SNPs or evaluating candidate SNPs. One hundred and forty SNPs had a minor allele frequency <0.006 (i.e., three or more cells with zero counts in a two by three genotype table for cases and controls) and were not considered for individual SNP analyses. The log quantile–quantile plot of the P-trend of these SNPs is provided in supplementary Figure 1 (available at Carcinogenesis Online). Comparison of the observed and expected distribution of the univariate P-trend showed that the observed values lie close to the line expected under the null hypothesis of no association, providing no evidence of inflation of the test statistic that would suggest systematic bias. There was, however, some deviation of the higher observed values from those expected, suggesting multiple weak associations.

Figure 3 summarizes the association between each SNP and endometrial cancer (P-trend per minor allele using unconditional logistic regression). After accounting for correlated SNPs within each gene ($r^2 > 0.70$), one SNP in AKR1C2 (rs11252887), three in AR (rs12011793, rs5919393 and rs4827545), one in CYP11B1 (rs9567021), one in HSD17B2 (rs2911422), two in CYP19A1 (rs700519 and rs3784308), one in CYP1A1/A2 (rs16972208), one in SHBG (rs6259) and five (rs8192120, rs169052, rs824811, rs1892236 and rs1687779) in SRD5A1 were statistically significantly associated with endometrial cancer ($P < 0.05$). A summary of the magnitude of association for each SNP and endometrial cancer is provided in supplementary Table 1 (available at Carcinogenesis Online).

**Discussion**

We evaluated 36 sex hormone-related genes using a tagging approach in a population-based case–control study. The strongest evidence for an association with endometrial cancer risk was for common variation in AR. Multilocus analyses identified boundaries for a region of interest within this gene, as well as within CYP19A1 around rs727479, previously identified as being associated with endometrial cancer risk (8).

Our previous results for six haplotype tagging SNPs in AR and endometrial cancer risk (10) conflicted with a previous report of the same SNPs (9). However, neither of these previous studies included one of our statistically significant tag SNP rs12011793 from our current study. This SNP is in a region of AR identified from our multilocus analysis as being associated with endometrial cancer risk. This association could be mediated through potential associations between the identified variant and elevated serum androgen levels that have been associated with endometrial cancer in a few studies (4,5), although not all (3). This observation is also consistent with the epidemiologic data on the association between endometrial cancer and polycystic ovarian syndrome that is hypothesized to be related to hyperandrogenism (67). Furthermore, experimental data suggest that estrogens are accumulated by conversion of androgens through the aromatase pathway with tumor–stromal interactions, although not in
normal or hyperplastic endometrium (68,69). Although the underlying mechanism is unresolved, androgens have been proposed to increase endometrial cancer risk, either indirectly through the aromatization of androgen into estrogens or directly by the increase in cellular growth and proliferative effects in the endometrium (70). Additionally, AR has been detected in normal human endometrium and endometrial carcinomas (71). Thus, although there is biological plausibility for the association between AR and endometrial cancer risk, previous reports may not have captured the area of interest. Additional studies are needed in this region marked by rs12011793.

CYP19A1 is involved in final and rate-limiting step of estrogen biosynthesis and associated common haplotypes spanning the coding and proximal 5’ region have been associated with circulating estrogen levels in post-menopausal women (7). Previous studies have evaluated the associations between various CYP19A1 polymorphisms and endometrial cancer risk with inconsistent results (6,8,21,26,31,32,34), except for rs749292 and rs727479 (8) that have been associated with estrogen levels (7). Our most statistically significant CYP19 SNP, rs700519, has been reported to have a multiplicative interaction with body mass index among post-menopausal women (P = 0.01) in study of 1033 cases and 1030 controls in Shanghai, China (34). In our current analysis, we also examined the densely tagged CYP19A1 using sequential haplotype scan and variable-sized sliding window methods and identified a genetic region primarily marked by rs28757184 (chromosome 15; position 49301864), rs16964201 (49302643), rs9944225 (49315372), rs700519 (49315404) and rs17703883 (4917389) as being associated with endometrial cancer risk. These SNPs, which are in the same region as rs727479 and rs700519, a region defined as block 4 (positions 49332746–49283122) by Haiman et al. (7), had stronger endometrial cancer associations than with rs727479 in our data. This region also contains another SNP (rs10046), which has been found to be associated with estrogen levels among post-menopausal women (7,72), suggesting that other markers of endometrial cancer susceptibility might exist. Additional studies are needed in this region that might be other markers related to estrogen levels and endometrial cancer risk.

To date, 12 candidate SNPs in sex hormone-related genes (CYP1A1, CYP1A2, CYP1B1, COMT, ESR1, ESR2, SHBG and SULT1E1) have been reported in at least three independent populations, including in our current study (11–20,23,27,35–38,40,43,44,49,50). We conducted a meta-analysis of these candidate SNPs and endometrial cancer risk to quantify the current evidence for an association (supplementary Figure 2 available at Carcinogenesis Online). Our meta-analytic approach did not identify consistent associations with endometrial cancer risk. For the candidate SNP for SHBG, rs6259, which was one of our top hits from the SNP-based analysis, we observed an increased rather than decreased risk of endometrial cancer associated with the variant allele, unlike previous studies (19,50). However, we did observe some possible exceptions although they were based on small numbers: variants of rs1256049 (ESR2) and rs37336599 (SULT1E1) were associated with significantly increased endometrial cancer risk based on the dominant inheritance model. Candidate SNPs should be evaluated in larger studies or consortia efforts to determine if they are indeed associated with endometrial cancer.

Strengths of our study include the detailed assessment of known common genetic variation in sex hormone-related genes by densely tagging 36 genes in this pathway. Our population-based study had high participation rates and all participants were of Caucasian descent, minimizing concerns of selection bias or population stratification, respectively. However, the power of our study to detect associations of weak magnitude, particularly for less common SNPs was limited, thus increasing the chances of false-positive and false-negative findings. To address this limitation, we used the robust gene-based ARTP method (62) to confirm associations between the sex hormone genes and endometrial cancer risk. Furthermore, interactions between SNPs in hormone metabolizing genes, as well as interactions with hormonally related risk factors for endometrial cancer are possible. However, since our study had very low power to uncover complex relationships, we limited our analyses to individual SNPs and genes.

In summary, we identified boundaries within AR and CYP19A1 that are associated with endometrial cancer risk. These regions are of interest for future haplotype and subsequent fine mapping studies aimed at identifying the best markers and eventually causal variants. Evaluation of common variation in consortial efforts will be needed to attain adequate power to establish or rule out putative association in the hormone pathway.

Supplementary material
Supplementary Tables 1 and 2 and Figures 1and 2 can be found at http://carcin.oxfordjournals.org/

Funding
Intramural program of the National Cancer Institute, National Institutes of Health.

Acknowledgements
The authors would like to thank Neonila Szeszenia-Dabrowska of the Nofer Institute of Occupational Medicine (Lodz, Poland) and Witold Zatonski of the M. Sklodowska-Curie Institute of Oncology and Cancer Center (Warsaw, Poland) for their contribution to the Polish Endometrial Cancer Study. The authors would also like to recognize Phil Taylor and Dennis Maeder of the National Cancer Institute for their contribution to the iSelect Rare Cancer Project and Pei Chao and Michael Stagner (IMS, Silver Spring, MD), who have been invaluable to the management of the study. This work would not be possible without the dedicated efforts of the physicians, nurses, interviewers and study participants.

Conflict of Interest Statement: None declared.

References


Received September 17, 2009; revised November 25, 2009; accepted December 15, 2009