



ORIGINAL ARTICLE

Candidate Genetic Polymorphisms and Haplotypes Associated with Endometrial Cancer Risk (United States)

Jane A McElroy ^{1,2*}, Robin L Kruse¹, J David Robertson^{2,3}, Helen Yampara-Iquise⁴, Elizabeth C Bryda⁵ and Jeremy F Taylor⁴

¹Family and Community Medicine Department, University of Missouri, Columbia, MO, USA

²Missouri Research Reactor, University of Missouri, Columbia, MO, USA

³Department of Chemistry, University of Missouri, Columbia, MO, USA

⁴Division of Animal Sciences, University of Missouri, Columbia, MO, USA

⁵Department of Veterinary Pathobiology, University of Missouri, Columbia, MO, USA

*Corresponding author: Jane A McElroy, Family and Community Medicine Department, University of Missouri, 7 Hospital Drive, MA306 Medical Science Building, DC32.00 Columbia, MO 65211, USA, Tel: 573-882-4993



Abstract

Background: Exploration of single nucleotide polymorphisms (SNPs) that alter the expression or function of a gene may enable the development of diagnostics for endometrial cancer susceptibility.

Methods: We evaluated eleven candidate SNPs that have previously been reported in the literature or that are associated with cadmium sequestering (i.e., metallothionein) for their effects on endometrial cancer risk. We also predicted haplotypes for SNPs within genes on chromosomes 6, 14 and X and tested haplotype effects for cancer risk.

Results: Our population-based study comprised 480 cases and 513 controls with 96% non-Hispanic White. Nine of the 11 SNPs were successfully assayed from the forward and reverse DNA strands. Concordance rate among genotype pairs was 99.73%. Our study found suggestive evidence of an increased endometrial cancer risk for two SNPs within the *ESR1* and *ESR2* genes (*ESR1:rs2234693*; *ESR2:rs944050*) but not for several other SNPs within those genes (*ESR1:rs9340799*; *rs3020314* or *ESR2:rs1255998*). Likewise, an increased risk for carriers of two rare haplotypes (*ESR1* haplotype:*rs2234693:rs9340799:rs3020314*) and (*ESR2* haplotype:*rs1255998:rs944050*) was found using conditional logistic regression multivariable models. We found no association between SNPs located in metallothionein (*rs2836603*, *rs3918249*), progesterone receptor (*rs1042638*), androgen receptor (*rs12011793*, *rs12011518*), aromatase enzyme (*rs4775936*), or chemokine c-x-c motif (*rs352038*) and endometrial cancer risk.

Conclusions: Our data suggest that the *ESR1* (*rs2234693*) and the *ESR2* (*rs944050*) polymorphisms are associated with an increased risk of developing endometrial cancer. Further elucidation of genetic influence on endometrial cancer risk adds evidence to can inform precision medicine approaches.

Keywords

Endometrial cancer, Cadmium, Single nucleotide polymorphisms, Estrogen receptors, Metallothionein

Introduction

Endometrial cancer is the fourth most common cancer of women in the United States and occurs primarily in postmenopausal women. The American Cancer Society (2018) [1] reported that 63.230 women were expected to be diagnosed annually with cancer of the body of the uterus (endometrial cancers and uterine sarcomas), with 11.350 deaths annually. The 5-year relative survival rate is about 84% for white women and 62% for black women post-diagnosis (2018). Endometrial cancer is divided into two types. Type 1 tumors comprise the majority of diagnoses, with most being classified as adenocarcinomas. These tumors are associated with endogenous and exogenous estrogen exposure [2]. Type II tumors are primarily serous carcinomas and are often char-

acterized as estrogen-independent. Type II tumors are believed to arise from atrophic endometrium and account for a disproportional number of deaths [2,3]. Cadmium, a toxic non-essential metal, mimics estrogen and may increase estrogen-receptor-mediated proliferation for the Type I cancer [4]. Other established risk factors for Type I cancer include obesity, nulliparity, and diabetes. Risk factors for Type II endometrial cancer are less well established [4,5].

According to Meyer, et al., [5] candidate single nucleotide polymorphisms (SNPs) may be involved in mechanisms that affect Type I and/or Type II endometrial cancer such as DNA damage repair, apoptosis, steroid receptor genetic pathways, carcinogen metabolism, steroid metabolism, and cell-cycle control. SNPs and/or haplotypes constructed from SNPs within candidate genes have been reported in a limited number of studies as being associated with endometrial cancer risk, though the findings are inconsistent [5,6].

In this study, we assessed the effects of genetic variants that had previously been identified as being significantly associated with endometrial cancer risk as well as SNPs associated with genes responsible for cadmium sequestration on risk of endometrial cancer. Of the 12 SNPs originally selected as candidates for endometrial cancer risk, *rs1042638* (a 3' UTR variant in *TPD52* with progesterone receptor gene *PGR*) was monomorphic and removed from further analysis. The remaining 11 SNPs were located within 7 genes: chromosome X: *rs12011793* and *rs12011518* (*AR*: androgen receptor); chromosome 6: *rs2234693*, *rs3020314*, *rs9340799* (*ESR1*: estrogen receptor 1); chromosome 14: *rs944050*, *rs1255998* (*ESR2*: estrogen receptor 2); chromosome 4: *rs352038* (*CXCL3*: chemokine c-x-c motif, ligand 3); chromosome 20: *rs3918249* (*MMP9*: matrix metalloproteinase 9); chromosome 15: *rs4775936* (*CYP19A1*: cytochrome P450); and chromosome 16: *rs28366003* (*MT2A*: metallothionein 2A). Individual SNP effects were investigated using single-marker analyses, and combined SNP effects were assessed by testing additive effects of haplotypes construction for the SNPs located within *ESR1*, *ESR2* and *AR* on chromosomes 6, 14 and X, respectively.

Materials and Methods

Study population

Genetic data were collected from participants in a population-based endometrial case-control study conducted during 2011-2012. The design and participation proportions were described in detail elsewhere [7]. Briefly, ascertainment of incident endometrial cancer cases came from three cancer registries: Arkansas Central Cancer Registry, Iowa Cancer Registry and Missouri Cancer Registry and Research

Center. Each registry contacted potential cases to gain written permission to pass their names to Health and Environmental Exposure (HEER) study staff. Using available data from Iowa and Missouri voter registration lists, controls were randomly selected within age strata. For each case, four controls with the same birth date (Iowa) or approximate age (Missouri) were sent invitation letters. Ineligible participants included those who contacted HEER staff to decline participation, those in which a telephone number or current address could not be found.

Trained interviewers called potential participants to obtain information on known and suspected risk factors. The age at, or year of, cancer diagnosis for cases was used as the reference age, or year, for the age-matched controls, respectively. At the conclusion of the interview, mailing addresses were confirmed and saliva and urine collection kits were mailed to the interviewed participants for home collection. These samples were returned to the study center for processing in pre-stamped addressed envelopes that were included with the collection kits. Participants who returned urine kits were sent a \$25 check as an indication of appreciation. This study was approved by the University of Missouri Health Sciences Institutional Review Board as well as the review boards of the respective state cancer registries. Each participant provided informed consent.

Cadmium analysis

Urine cadmium content was quantified by using inductively coupled plasma-dynamic reaction cell-mass spectrometry (ICP-DRC-MS) following the established protocol from the National Health and Nutritional Examination Study (NHANES) [8] with an additional correction for the effect of strontium on the tin interference. A comprehensive quality-control program incorporating numerous methods including bottle blanks, monitoring multiple cadmium isotopes, internal and external controls, stability samples, replicates, spikes, and routine inclusion of National Institute of Standards and Technology (NIST) standard reference materials was utilized to ensure high-quality data. The method detection limit of 0.0082 µg/L cadmium enabled quantification of metal levels in every urine sample. Urine creatinine level was measured using a colorimetric assay based on the Jaffé reaction to control for kidney function [9] See McElroy, et al. [7] for more details about cadmium quantification.

DNA extraction

Participants collected saliva samples using the Ora-gene DNA/saliva kit disk format (DNA Genotek, Ottawa, Ontario) and mailed the sample to the study center. For processing, the vials were placed in an air incubator at 50 °C overnight. Next, they were mixed with 500 µl of

Oragene* DNA Purifier and incubated on ice for 10 minutes. After incubation, they were centrifuged at $15.000 \times g$ for 5 minutes at room temperature. To precipitate the DNA, the supernatants were recovered, mixed with 500 μ l 100% ethanol and centrifuged at $15.000 \times g$ for 5 minutes at room temperature. The recovered pellets were then washed with 250 μ l 70% ethanol and centrifuged at $15.000 \times g$ for 5 minutes at room temperature. The final step included rehydrating the DNA pellets by adding 100 μ l of TE buffer (10 mM Tris-HCL, 1 mM EDTA, pH 8.0) and incubating the pellets at 50°C for 1 hour with occasional vortexing.

DNA concentration and quality was assessed using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific, Waltham, MA). Samples identified as having an $A_{260/280}$ optical density ratio of less than 1.8 were re-extracted using a standard phenol-chloroform isoamyl alcohol extraction. Aliquots of 500 ng of DNA were sent to GeneSeek (a Neogen Company, Lincoln, NE) for SNP genotyping.

Genotyping assays

Genotypes for the 11 variable SNPs located within 7 genes (1 *CXCL3*, 3 *ESR1*, 2 *ESR2*, 1 *CYP19A1*, 1 *MT2A*, 1 *MMP9*, 2 *AR*) were obtained using a Sequenom Mass ARRAY® Analyzer 4 at GeneSeek. Each SNP was assayed twice with detection probes designed from each DNA strand. For SNPs located within the same gene, haplotypes were phased and 146 missing values (2.1% of the data) were estimated using Beagle 4.0 (Browning and Browning, 2007) to produce complete data for the loci (*ESR1*hap *rs2234693:rs9340799:rs3020314*, *ESR2*hap *rs1255998:rs944050* and *AR*hap*rs12011793:rs12011518*). Allele frequencies were estimated by allele counting. Genotypes were tested for departures from Hardy-Weinberg equilibrium (HWE) using a chi-square test with Bonferroni correction for multiple testing.

Statistical analysis

We determined associations between SNPs and case-control status after controlling for other known risk factors for endometrial cancer. To determine which participant characteristics were associated with endometrial cancer, we used chi-square and t-test analyses to examine the association of categorical and continuous survey variables, respectively, with case-control status. We analyzed additional survey variables in categories (e.g., medical history, demographics, dietary habits) to determine which were the most strongly associated with case-control status. Variables with $P < 0.1$ that had the strongest association within categories were then selected as additional candidates for inclusion in the multivariable conditional logistic regression analysis. Model variables were selected by both forward stepwise inclusion and backwards elimination to determine if the same factors were chosen, using $P < 0.1$ for both

model entry and exit criteria. Using all available survey data, we developed a multivariable conditional logistic model for case-control status, stratifying on age at diagnosis. Interactions between variables in the model were tested; none were significant and were thus not retained.

Stepwise and backwards variable selection retained the same set of variables, yielding a final set of independent variables of: Age at diagnosis; non-Hispanic African-American race; marital status; body mass index (BMI); history of trying to lose weight; smoking status; pack-years of smoking; history of endometriosis, breast cancer, ovarian cancer, or uterine fibroids; endometrial cancer in a first degree relative; years of oral contraceptive use; years of unopposed estrogen use; menopause status; menopause after age 56; number of protein shakes consumed per week; and 5 or more servings per week of whole milk.

We used Fisher's exact test to compare genotype frequencies between the cases and controls. With log-transformed creatinine-adjusted cadmium concentration as the dependent variable, we tested the association of 11 SNPs and 3 loci with cadmium in simple regression models. Each SNP and locus was subsequently included as a covariate in a conditional logistic model for case-control status that included the selected covariates described above as well as log-transformed cadmium concentration. For those loci for which a statistically significant association with urine cadmium levels was found, cadmium was excluded from the multivariable model used in the case-control analysis.

Because not all women returned their home biospecimen kits, and because women who did and did not return these samples differed on several characteristics (for example, alcohol consumption, medical history, education, and income), we developed a multivariable logistic model for the probability of an eligible participant returning a kit. The probability of returning a sample predicted by this model was used to adjust for non-response bias in the final case-control models by including the inverse of the probability as a weight [10].

Genotype data were added to the conditional logistic regression models to determine if SNPs or the loci were associated with case-control status after controlling for other risk factors. We tested the additive effects of SNPs by fitting a model with genotypes treated as allele dosages (e.g., CC = 0, CT = 1, TT = 2) with increasing dosages representing increasing numbers of copies of the rare (minor) allele at each locus. We tested non-additive effects at each SNP using two variables (a and d) following the notation of Falconer (1960) [11]. Samples with 0 copies of the common allele were coded as a = -1, d = 0; those with 1 copy were coded as a = 0, d = 1; and those with 2 copies were coded as a = 1, d = 0. Three SNPs (*rs352038*,

rs944050, and *rs12011518*) had very rare minor alleles that prevented convergence of the logistic models, and were excluded from the non-additive effect analyses. Additive analyses were also performed for the haplotype effects by forming dummy variables for each haplotype in which the number of copies (0, 1, or 2) of each haplotype present in each diplo-type (genotype based on haplotypes) were scored. For some diplotypes and haplotypes that were rare, these were dropped from the analysis.

While missing phenotypic data were rare, data imputation is considered to result in less bias than only analyzing cases with no missing data [12]. We therefore imputed missing phenotypic values using multiple imputation in SAS to create 20 datasets using the Markov chain Monte Carlo method. All final multivariable models for case-control status were run using the 20 imputed datasets and the MIANALYZE procedure in SAS was

used to summarize results. Data were analyzed using SAS for Windows 9.4 (Cary, NC).

Results

Characteristics of participants

Sample derivation has previously been described [7] and is summarized in Figure 1. Briefly, of the 1519 women who completed the telephone survey, 517 failed to return analyzable saliva samples. We excluded nine controls that reported a history of endometrial cancer, leaving 480 cases and 513 controls in the study.

Endometrial cancer case participants and population-based controls were significantly different for all but 5 of the characteristics that we had selected *a priori* based on the literature (Table 1). These were

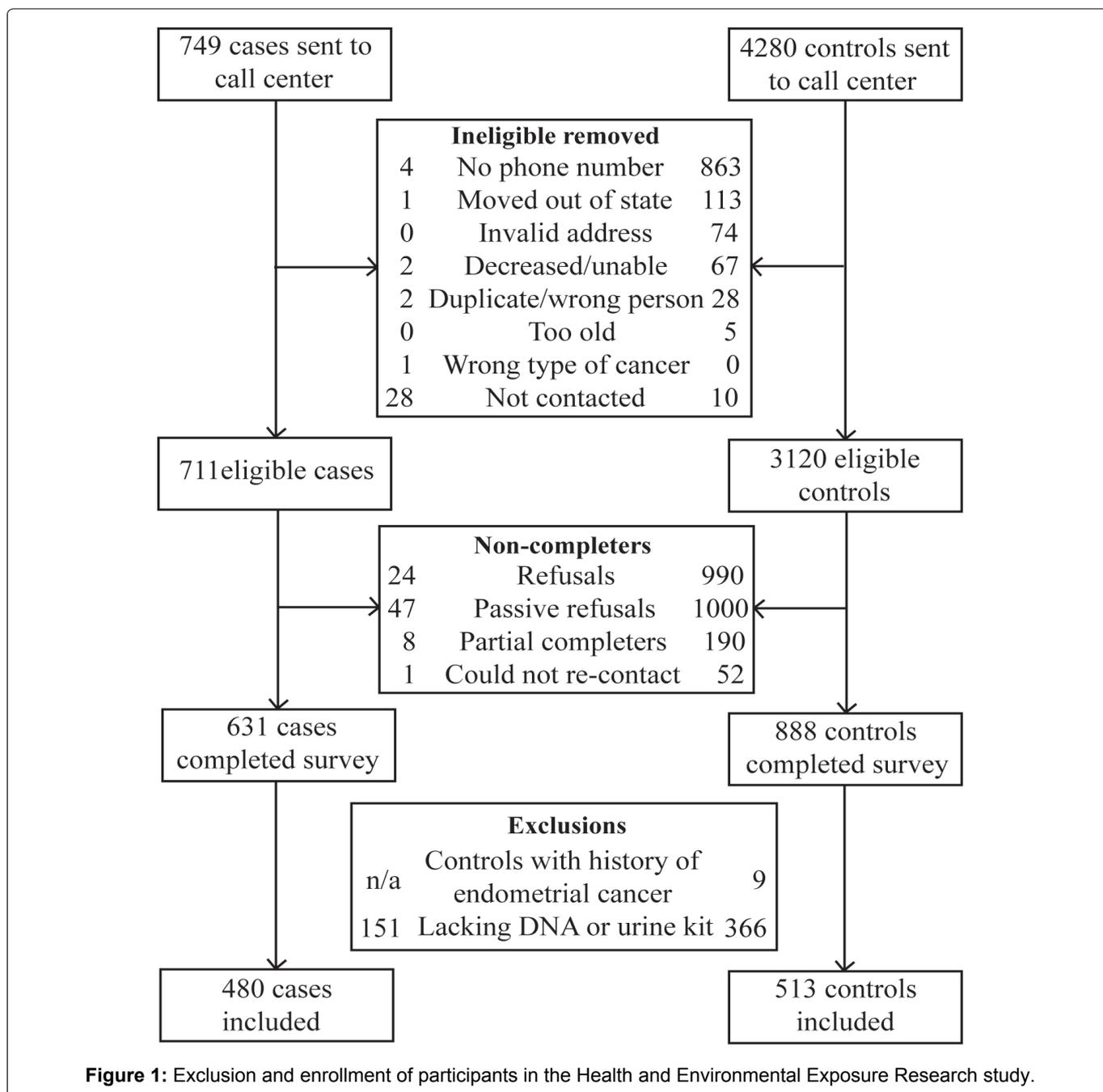


Figure 1: Exclusion and enrollment of participants in the Health and Environmental Exposure Research study.

Table 1: Characteristics of endometrial cancer cases and population-based controls who returned urine and saliva samples*.

Characteristic (N missing)	Number (Percent)	Cases (N = 480)	Controls (N = 513)	p-value†
Demographics				
Age at diagnosis				
< 40	14 (1.4)	12 (2.5)	2 (0.4)	< 0.0001
40-44	21 (2.1)	15 (3.1)	6 (1.2)	
45-49	37 (3.7)	23 (4.8)	14 (2.7)	
50-54	98 (9.9)	59 (12.3)	39 (7.6)	
55+	823 (82.9)	371 (77.3)	452 (88.1)	
Mean (SD)	61.8 (8.6)	60.5 (9.2)	63.1 (7.9)	< 0.0001
Education				
High school or less	371 (37.4)	176 (36.7)	195 (38.0)	0.83
1-3 years of college	261 (26.3)	125 (26.0)	136 (26.5)	
College or more	361 (36.4)	179 (37.3)	182 (35.5)	
Employment				
Not in labor force	472 (47.5)	217 (45.2)	255 (49.7)	0.16
Working	521 (52.5)	263 (54.8)	258 (50.3)	
Family income, annual (62)				
≤ \$30,000	245 (26.3)	138 (30.0)	107 (22.7)	0.006
\$30,001 - \$50,000	265 (28.5)	140 (30.4)	125 (26.5)	
\$50,001 - \$100,000	298 (32.0)	130 (28.3)	168 (35.7)	
> \$100,000	52 (11.3)	52 (11.3)	71 (15.1)	
Marital status				
Married/living with partner	714 (71.9)	318 (66.3)	396 (77.2)	< 0.0001
Divorced, separated, widowed	220 (22.2)	118 (24.6)	102 (19.9)	
Never married	59 (5.9)	44 (9.2)	15 (2.9)	
Race/ethnicity (4)				
Hispanic	6 (0.6)	4 (0.8)	2 (0.4)	0.004
Non-Hispanic Black	27 (2.7)	21 (4.4)	6 (1.2)	
Non-Hispanic White	946 (95.6)	452 (94.4)	494 (96.9)	
Other Non-Hispanic	10 (1.0)	2 (0.4)	8 (1.6)	
Work schedule (473)				
Regular daytime shift	439 (84.4)	221 (84.4)	218 (84.5)	0.96
Other shift	81 (15.6)	41 (15.6)	40 (15.5)	
Dietary habits				
Days/week alcoholic beverages consumed (10)				
0	440 (44.8)	224 (47.1)	216 (42.6)	0.041
< 1	269 (27.4)	139 (29.2)	130 (25.6)	
1-2.9	125 (12.7)	54 (11.3)	71 (14.0)	
3 or more	149 (15.2)	59 (12.4)	90 (17.8)	
Protein shake consumption (5)				
None	877 (88.8)	414 (86.4)	463 (91.0)	0.0051
Less than 2 per week	48 (4.9)	22 (4.6)	26 (5.1)	
2 per week or more	63 (6.4)	43 (9.0)	20 (3.9)	
Whole milk consumption				
5 or more times per week	50 (5.0)	32 (6.7)	18 (3.5)	0.023
Less than 5 times per week	943 (95.0)	448 (93.3)	495 (96.5)	
Physical activity and weight				
Body mass index, year before diagnosis (14)				
Mean (SD)	32.1 (9.2)	35.3 (10.1)	29.0 (6.9)	< 0.0001

METs-minutes/week [^] , total (1)				
Mean (SD)	36.4 (54.3)	31.5 (48.8)	41.0 (58.6)	0.006
Any weight gain since age 25 (14)				
Mean (SD)	20.1 (18.0)	24.1 (19.6)	16.4 (15.6)	< 0.0001
Smoking history				
Smoke cigarettes (3)				
Never	652 (65.9)	325 (68.0)	327 (63.9)	0.1
Former	274 (27.7)	130 (27.2)	144 (28.1)	
Current	64 (6.5)	23 (4.8)	41 (8.0)	
Smoking, total pack-years (6)				
Mean for ever-smokers (SD)	8.2 (17.8)	6.3 (15.1)	10.0 (19.9)	0.001
Medical history				
Breast cancer (2)				
Yes	50 (5.0)	13 (2.7)	37 (7.2)	0.0012
No	941 (95.0)	466 (97.3)	475 (92.8)	
Diabetes (3)				
Yes	183 (18.5)	118 (24.7)	65 (12.7)	< 0.0001
No	807 (81.5)	360 (75.3)	447 (87.3)	
Endometrial cancer in first degree relative (36)				
Yes	31 (3.2)	25 (5.4)	6 (1.2)	0.0002
No	926 (96.8)	435 (94.6)	491 (98.8)	
Endometriosis (20)				
Yes	164 (16.9)	84 (18.0)	80 (15.8)	0.35
No	809 (83.1)	382 (82.0)	427 (84.2)	
Hypertension (2)				
Yes	526 (53.1)	275 (57.5)	251 (48.9)	0.007
No	465 (46.9)	203 (42.5)	262 (51.1)	
Ovarian cancer (7)				
Yes	33 (3.4)	30 (6.3)	3 (0.6)	< 0.0001
No	953 (96.6)	444 (93.7)	509 (99.4)	
Medications and treatments				
Birth control, years used (3)				
Mean (SD) [‡]	6.0 (7.3)	5.0 (6.2)	7.0 (8.0)	< 0.0001
Hormone replacement therapy, years used (14)				
Mean (SD)	3.3 (7.1)	2.1 (5.9)	4.4 (8.0)	< 0.0001
Reproductive history				
Age of menarche 11 or earlier (2)				
Yes	224 (22.6)	123 (25.7)	101 (19.7)	0.025
No	767 (77.4)	356 (74.3)	411 (80.3)	
Menopause status, year before diagnosis				
Premenopausal	149 (15.0)	107 (22.3)	42 (8.2)	< 0.0001
Postmenopausal	840 (84.6)	372 (77.5)	468 (91.2)	
Unknown	4 (0.4)	1 (0.2)	3 (0.6)	
Nulliparous (4)	148 (15.0)	89 (18.6)	59 (11.5)	0.0018

[^]9 controls who reported a diagnosis of endometrial cancer were excluded. Variables refer to the year prior to diagnosis of endometrial cancer unless otherwise specified.

[‡]P-value for chi-square test except for means, which were compared with a t-test.

[‡]Birth control refers to methods that release hormones, such as birth control pills, injections, patches, progesterone implants (Norplant), progestin-releasing intrauterine devices, or vaginal rings.

[^]METs minutes is a measure of the volume of activity which is computed by weighting each type of activity by its energy requirements *defined* in METs to yield a *score* in *MET-minutes*.

Table 2: Association of alleles with creatinine-adjusted cadmium concentration.

Locus (N missing)	Gene and coordinates*	Alleles or haplotypes	P-value	
			Log ₂ -transformed	Not transformed
rs352038 (6)	CXCL3 4:74052981	C/T	0.58 [‡]	0.74 [‡]
rs2234693	ESR1 6:151842200	C/T	0.18 [‡]	0.44 [‡]
rs9340799	ESR1 6:151842246	A/G	0.99 [‡]	0.45 [‡]
rs3020314	ESR1 6:151949537	C/T	0.62 [‡]	0.64 [‡]
rs1255998	ESR2 14:64227153	C/G	0.76 [‡]	0.77 [‡]
rs944050	ESR2 14:64233327	A/G	0.47 [‡]	0.97 [‡]
rs4775936 (12)	CYP19A1 15:51243825	C/T	0.34 [‡]	0.05 [‡]
rs28366003 (6)	MT2A 16:56608579	A/G	0.15 [‡]	0.10 [‡]
rs3918249 (9)	MMP9 20:46009497	C/T	0.59 [‡]	0.29 [‡]
rs12011793	AR X:67698249	C/T	0.48 [‡]	0.62 [‡]
rs12011518	AR X:67713194	G/T	0.94 [‡]	0.88 [‡]
ESR1Hap [§]		CAC	0.14 [†]	0.02 [†]
		CAT	0.74 [†]	0.14 [†]
		CGC	0.83 [†]	0.65 [†]
		CGT	0.17 [†]	0.26 [†]
		TAC	0.44 [†]	0.51 [†]
		TAT	0.70 [†]	0.47 [†]
		TGC	0.94 [†]	0.85 [†]
		TGT	0.06 [†]	0.10 [†]
ESR2Hap ^d		CA	0.46 [†]	0.82 [†]
		GA	0.47 [†]	0.18 [†]
		GG	0.22 [†]	0.74 [†]
ARHap ^d		CG	0.67 [†]	0.54 [†]
		CT	0.87 [†]	0.58 [†]
		TG	0.35 [†]	0.73 [†]
		TT	-	-

*Chromosome number: Position in bp based on Genome Reference Consortium Human Build 38 patch release 7 [29].

[†]Independent group t-test; [‡]Analysis of variance; [§]Haplotype based on the SNPs located within the corresponding gene.

smoking status, history of endometriosis, educational attainment, employment status and shift work. Among the commonly cited risk factors for endometrial cancer [13], case participants had higher average BMI (35.3 versus 29.0 kg/m²); a higher likelihood of diabetes (24.7% versus 12.7%), hypertension (57.5% versus 48.9%), first degree relatives with endometrial cancer (5.4% versus 1.2%), or ovarian cancer (6.3% versus 0.6%); menarche at age 11 years or earlier (25.7% versus 19.7%); and were more likely to be nulliparous (18.6% versus 11.5%) relative to controls. Among the 16 histology types identified from the cancer registry data, over 78% were endometrioid carcinoma (ICD-03 code: 8380), 9% mixed cell adenocarcinoma (ICD-03 code: 8323), and 5% adenocarcinoma, NOS (ICD-03 code: 8140). The remaining histology types ranged from 1 to 12 cases, with 76 missing a histology type. Among the histology types with known histology 14% were Type 11 and 86% were Type 1. Since case ascertainment came directly from the cancer registries, the time of interview varied from 3.3 years since di-

agnosis to 4 months since diagnosis. Mean length of time from diagnosis was 1.6 years.

Cadmium analysis

Creatinine-adjusted urine cadmium levels varied from 0.006 to 0.649 (mean 0.041) µg/g in control subjects and from 0.005 to 0.417 (mean 0.037) µg/g in case participants. Urine cadmium concentration did not differ between cases and controls ($P = 0.11$). One SNP (*rs4775936*) and two of the *ESR1* Haplotypes (CAC, TGT) were associated or nearly so with variation in creatinine-adjusted urine cadmium concentration (Table 2). For these loci, cadmium concentration was excluded from the conditional logistic regression model for case-control status since cadmium concentration made very little difference on parameter estimates.

Genetic analysis

Eleven SNPs were successfully assayed from both DNA strands with a concordance rate among genotype pairs of 99.73%. Discordant genotypes were set to be

Table 3: Frequencies of participants' genotypes among endometrial cancer cases and population-based controls who returned urine and saliva samples.

SNP (N missing) Genotype	Number of individuals (percentage)	Cases (N = 480)	Controls (N = 513)	P-value*
<i>rs352038</i> (6) CC	1 (0.1)	1 (0.2)	0 (0.0)	0.39
CT	28 (2.8)	16 (3.3)	12 (2.4)	
TT	958 (97.1)	463 (96.5)	495 (97.6)	
<i>rs2234693</i> CC	127 (12.8)	68 (14.2)	59 (11.5)	0.29
CT	416 (41.9)	205 (42.7)	211 (41.1)	
TT	450 (45.3)	207 (43.1)	243 (47.4)	
<i>rs9340799</i> AA	400 (40.3)	194 (40.4)	206 (40.2)	0.9
AG	461 (46.4)	220 (45.8)	241 (47.0)	
GG	132 (13.3)	66 (13.8)	66 (12.9)	
<i>rs3020314</i> CC	106 (10.7)	50 (10.4)	56 (10.9)	0.66
CT	461 (46.4)	230 (47.9)	231 (45.0)	
TT	426 (42.9)	200 (41.7)	226 (44.1)	
<i>rs1255998</i> CC	795 (80.1)	380 (79.2)	415 (80.9)	0.18
CG	181 (18.2)	88 (18.3)	93 (18.1)	
GG	17 (1.7)	12 (2.5)	5 (1.0)	
<i>rs944050</i> AA	923 (93.0)	451 (94.0)	472 (92.0)	0.19
AG	69 (6.9)	28 (5.8)	41 (8.0)	
GG	1 (0.1)	1 (0.2)	0 (0.0)	
<i>rs4775936</i> (12) CC	301 (30.7)	139 (29.3)	162 (32.0)	0.48
CT	469 (47.8)	236 (49.8)	233 (46.0)	
TT	211 (21.5)	99 (20.9)	112 (22.1)	
<i>rs28366003</i> (6) AA	890 (90.2)	434 (91.0)	456 (89.4)	0.5
AG	94 (9.5)	41 (8.6)	53 (10.4)	
GG	3 (0.3)	2 (0.4)	1 (0.2)	
<i>rs3918249</i> (9) CC	123 (12.5)	68 (14.3)	55 (10.8)	0.16
CT	445 (45.2)	219 (45.9)	226 (44.6)	
TT	416 (42.3)	190 (39.8)	226 (44.6)	
<i>rs12011793</i> CC	7 (0.7)	5 (1.0)	2 (0.4)	0.19
CT	124 (12.5)	67 (14.0)	57 (11.1)	
TT	862 (86.8)	408 (85.0)	454 (88.5)	
<i>rs12011518</i> GG	870 (87.6)	414 (86.2)	456 (88.9)	0.23
GT	122 (12.3)	65 (13.5)	57 (11.1)	
TT	1 (0.1)	1 (0.2)	0 (0.0)	
<i>ESR1</i> Haplotype CAC	13 (1.3)	11 (2.3)	2 (0.4)	0.01
CAT	108 (10.9)	56 (11.7)	52 (10.1)	0.48
CGC	98 (9.9)	49 (10.2)	49 (9.6)	0.75
CGT	404 (40.7)	198 (41.2)	206 (40.2)	0.75
TAC	486 (48.9)	238 (49.6)	248 (48.3)	0.7
TAT	492 (49.5)	234 (48.8)	258 (50.3)	0.66
TGC	24 (2.4)	12 (2.5)	12 (2.3)	0.99
TGT	144 (14.5)	63 (13.1)	81 (15.8)	0.24
<i>ESR2</i> Haplotype CA	976 (98.3)	468 (97.5)	508 (99.0)	0.09
GA	133 (13.4)	73 (15.2)	60 (11.7)	0.11
GG	70 (7.1)	29 (6.0)	41 (8.0)	0.26
<i>AR</i> Haplotype CG	13 (1.3)	10 (2.1)	3 (0.6)	0.05
CT	122 (12.3)	65 (13.5)	57 (11.1)	0.25
TG	986 (99.3)	475 (99.0)	511 (99.6)	0.27
TT	1 (0.1)	1 (0.2)	0 (0.0)	0.48

ESR1 = estrogen receptor 1; ESR2 = estrogen receptor 2; AR = androgen receptor; *Fisher's Exact Test.

Table 4: Association of loci or haplotypes with case-control status in multivariable logistic regression using additive coding.

Allele and gene description	Parameter estimate (SE)	p-value*
<i>rs352038 (CXCL3)</i>	0.205 (0.088)	0.67
<i>rs2234693 (ESR1)</i>	0.220 (0.111)	0.048
<i>rs9340799 (ESR1)</i>	-0.115 (0.113)	0.31
<i>rs3020314 (ESR1)</i>	0.019 (0.119)	0.88
<i>rs1255998 (ESR2)</i>	0.136 (0.184)	0.46
<i>rs944050 (ESR2)</i>	0.587 (0.315)	0.062
<i>rs4775936 (CYP19A1)[†]</i>	-0.156 (0.102)	0.12
<i>rs28366003 (MT2A)</i>	-0.053 (0.227)	0.81
<i>rs3918249 (MMP9)</i>	0.052 (0.115)	0.65
<i>rs12011793 (AR X)</i>	0.154 (0.207)	0.46
<i>rs12011518 (AR X)</i>	-0.190 (0.217)	0.38
<i>ESR1Hap[†]-rs2234693:rs9340799:rs3020314</i>		
<i>CGC-CAT</i>	-0.175 (0.357)	0.62
<i>CGT-CAT</i>	-0.188 (0.289)	0.52
<i>TAC-CAT</i>	0.063 (0.285)	0.82
<i>TAT-CAT</i>	-0.011 (0.267)	0.97
<i>TGC-CAT</i>	-0.645 (0.543)	0.23
<i>TGT-CAT</i>	0.224 (0.314)	0.48
<i>CAC-CAT</i>	-1.906 (0.931)	0.041
<i>CGC-CAC</i>	1.731 (0.918)	0.059
<i>CGT-CAC</i>	1.718 (0.869)	0.048
<i>TAC-CAC</i>	1.969 (0.875)	0.024
<i>TAT-CAC</i>	1.896 (0.872)	0.03
<i>TGC-CAC</i>	1.261 (0.986)	0.2
<i>TGT-CAC</i>	2.130 (0.887)	0.016
<i>CGT-CGC</i>	-0.013 (0.281)	0.96
<i>TAC-CGC</i>	0.238 (0.283)	0.4
<i>TAT-CGC</i>	0.165 (0.259)	0.53
<i>TGC-CGC</i>	-0.470 (0.542)	0.39
<i>TGT-CGC</i>	0.399 (0.315)	0.21
<i>TAC-CGT</i>	0.251 (0.167)	0.13
<i>TAT-CGT</i>	0.177 (0.148)	0.23
<i>TGC-CGT</i>	-0.457 (0.495)	0.36
<i>TGT-CGT</i>	0.412 (0.224)	0.067
<i>TAT-TAC</i>	-0.074 (0.148)	0.62
<i>TGC-TAC</i>	-0.708 (0.500)	0.16
<i>TGT-TAC</i>	0.161 (0.227)	0.48
<i>TGC-TAT</i>	-0.635 (0.483)	0.19
<i>TGT-TAT</i>	0.234 (0.219)	0.28
<i>TGT-TGC</i>	0.869 (0.547)	0.11
<i>ESR2Hap -rs1255998:rs944050</i>		
<i>CA-GG</i>	-0.580 (0.316)	0.066
<i>GA-GG</i>	-0.646 (0.362)	0.074
<i>CA-GA</i>	0.067 (0.215)	0.76
<i>ARHap -rs12011793:rs12011518^{††}</i>		
<i>CG-TG</i>	0.126 (0.614)	0.84
<i>CT-TG</i>	-0.184 (0.218)	0.4
<i>CG-CT</i>	0.310 (0.655)	0.64

CXCL = chemokine (C-X-C motif) ligand 3; ESR1 = estrogen receptor 1; ESR2 = estrogen receptor 2; CYP19A1 = cytochrome P450, family 19, subfamily A, polypeptide 1; MT2A = metallothionein 2A; MMP9 = matrix metalloproteinase 9; AR = androgen receptor.

*Multivariable model included: age at diagnosis, non-Hispanic African-American race, marital status, BMI, ever tried to lose weight, smoking status, pack-years smoking, history of endometriosis, history of breast cancer, history of ovarian cancer, history of uterine fibroids, endometrial cancer in 1st degree relative, years of oral contraceptive use, years of unopposed estrogen use, menopause after age 56, menopause status, number of protein shakes per week, 5 or more servings per week whole milk; †Log2-transformed adjusted cadmium concentration was not included in models for rs4775936 and ESR1hap; ‡Only one copy of the TT allele was identified so not included in haplotype analysis.

Table 5: Association of alleles with case-control status in multivariable logistic regression using non-additive coding[†].

Allele [†]	a parameter		d parameter	
	Parameter estimate (SE)	p-value	Parameter estimate (SE)	p-value
rs2234693	0.25 (0.12)	0.039	0.11 (0.17)	0.52
rs9340799	-0.14 (0.12)	0.24	0.11 (0.16)	0.52
rs3020314	-0.068 (0.14)	0.62	-0.25 (0.18)	0.18
rs1255998	-0.24 (0.29)	0.41	0.48 (0.34)	0.15
rs4775936 [‡]	-0.14 (0.10)	0.17	-0.20 (0.16)	0.2
rs28366003	0.070 (0.39)	0.86	-0.14 (0.46)	0.76
rs3918249	0.063 (0.13)	0.62	0.035 (0.16)	0.83
rs12011793	-0.036 (0.38)	0.93	-0.22 (0.43)	0.6

missing values. Overall genotype call rate for the 11 SNPs was 98.35%. All SNPs were in Hardy-Weinberg Equilibrium (HWE; Bonferroni $P > 0.08$). In the single marker analyses, there were no genotype or allele frequency differences between the cases and controls (Table 3). Haplotype CAC at *ESR1* ($P = 0.01$) and CG at *AR* ($P = 0.05$) were at a slightly higher frequency in the cases than in the controls (Table 3).

Of the 11 SNPs, the C allele at rs2234693 in *ESR1* ($P = 0.048$) was associated with an increased risk of endometrial cancer in the conditional logistic regression multivariable models that fit the additive effect of SNP alleles on cancer risk (Table 4). Similarly, in the additive models for haplotype effects, carriers of the rare *ESR1*CAC haplotype (rs2234693:rs9340799:rs3020314) were associated with a significantly risk of endometrial cancer relative to carriers of all other haplotypes except TGC at *ESR1* (Table 4). Marginal significance (p-value = 0.07) for endometrial cancer was observed for *ESR2* GG haplotype (rs1255998:rs944050). In the non-additive model analyses, the dominance deviation d was not significantly different from 0 but the difference between alternate homozygotes was significant for rs2234693 in *ESR1* ($P = 0.039$) (Table 5), consistent with an additive effect of this locus on risk of endometrial cancer revealed in the additive model analyses (Table 4).

Discussion

Our study found suggestive evidence for associations between variants in *ESR1* and *ESR2* and risk for endometrial cancer. We found no evidence for associ-

ations between risk of endometrial cancer and variants in metallothionein 2A, a protein that sequesters cadmium, progesterone receptor, or androgen receptor. Our confidence in these finding is strengthened due to the extensive individual level data that were gathered and that allowed us to adjust the risk for known and suspected endometrial cancer risk factors.

Functionally, critical single nucleotide polymorphisms in many genes have been suggested as risk factors for the development of cancers such as endometrial cancer [5]. One consistently found and primary risk factor for endometrial cancer is estrogen exposure [14]. Previous studies that have evaluated the estrogen receptor 1 polymorphism rs2234693 in various populations (i.e., China, Sweden, Australia, Japan, Poland, USA, UK) with mixed results [6,15-26]. In a meta-analysis of thirteen studies, the estrogen receptor 1 polymorphism rs2234693 was significant for endometrial cancer risk (T vs. C allele: OR = 1.08, 95% CI, 1.00-1.17; and TT vs. CC allele: OR = 1.18, 95% CI, 1.00-1.38). In support of Zhou's finding, we also found an association between rs2234693 and endometrial cancer [27]. To our knowledge, only one other study has evaluated the estrogen receptor 2 polymorphism rs944050A/G allele for its effect on endometrial cancer risk and found an association [28]. In haplotype analyses of *ESR1* (rs2234693:rs9340799:rs3020314) our study found an increased risk of endometrial cancer in women carrying the CAC haplotype. Finally, our study evaluated haplotype of *ESR2* (rs1255998:rs944050) and found a marginally significant decreased risk of endometrial cancer in women carrying the GG haplotypes.

One limitation of our study was the very high proportion of white participants (> 95%), which prevented us from stratifying the analysis by race/ethnicity. Another potential source of ascertainment bias may have been caused by the fact that approximately one-third of the participants who contributed phenotypic data did not return saliva samples.

In conclusion, our study provides evidence that women carrying specific variants within estrogen receptor genes are at a slightly increased risk of endometrial cancer.

Acknowledgement

The authors acknowledge that the case data used in this report were provided by three cancer registries: The Arkansas Department of Health, Arkansas Central Cancer Registry (ACCR), Little Rock, AR; State

Health Registry of Iowa, Iowa Cancer Registry (ICR), Iowa City Iowa; and the Missouri Cancer Registry and Research Center (MCR-ARC), Columbia, MO. Michele West and Jason Brubaker were key Iowa personnel involved in data collection for this project. Telephone interviews were conducted by the University of South Carolina's Survey Research Laboratory.

Support

This work was supported by the American Cancer Society (RSG-10-197-01-CNE www.acs.org). Analytical instrumentation used in this study was acquired with support of the National Science Foundation, grant # BCS-0922374. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- (2018) Cancer facts & figures 2018. American Cancer Society.
- Setiawan VW, Yang HP, Pike MC, McCann SE, Yu H, et al. (2013) Type I and II endometrial cancers: have they different risk factors? *Journal of Clinical Oncology* 31: 2607-2618.
- Boruta DM, PA Gehrig, AN Fader, AB Olawaiye (2009) Management of women with uterine papillary serous cancer: a Society of Gynecologic Oncology (SGO) review. *Gynecol Oncol* 115: 142-153.
- Johnson MD, Kenney N, Stoica A, Hilakivi-Clarke L, Singh B, et al. (2003) Cadmium mimics the in vivo effects of estrogen in the uterus and mammary gland. *Nat Med* 9: 1081-1084.
- Meyer LA, Westin SN, Lu KH, Milam MR (2008) Genetic polymorphisms and endometrial cancer risk. *Expert Rev Anticancer Ther* 8: 1159-1167.
- Chiu HF, Tsai SS, Wu TN, CY Yang (2010) Effect modification of the association between trihalomethanes and pancreatic cancer by drinking water hardness: evidence from an ecological study. *Environmental Research* 110: 513-518.
- McElroy JA, Kruse RL, Guthrie J, Gangnon RE, Robertson JD (2017) Cadmium exposure and endometrial cancer risk: A large midwestern U.S. population-based case-control study. *PLoS One* 12: e0179360.
- Jarrett JM, G Xiao, Caldwell KL, Henahan D, Shakirova G, et al. (2008) Eliminating molybdenum oxide interference in urine cadmium biomonitoring using ICP-DRC-MS. *Journal of Analytical Atomic Spectrometry* 23: 962-967.
- Elinder CG (1985) Normal values for cadmium in human tissues, blood, and urine in different countries. In: L Friberg, CG Elinder, T Kjellstrom, GF Nordberg, Cadmium and health: a toxicological and epidemiological appraisal, volume I, exposure, dose, and metabolism. CRC Press, Boca Raton, FL, 81-102.
- Rosenbaum PR, DB Rubi (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41-55.
- Falconer Douglas Scott (1960) Introduction to quantitative genetics. Edinburgh: Oliver and Boyd.
- van der Heijden GJ, AR Donders, T Stijnen, KG Moons (2006) Imputation of missing values is superior to complete case analysis and the missing-indicator method in multi-variable diagnostic research: a clinical example. *Journal of Clinical Epidemiology* 59: 1102-1119.
- JL Davies, NB Rosenshein, CM Antunes, PD Stolley (1981) A review of the risk factors for endometrial carcinoma. *Obstet Gynecol Surv* 36: 107-116.
- Cook LS, Weiss NS, Doherty JA, C Chen (2006) Endometrial cancer. In: David Schottenfeld, JF Fraumeni, Cancer epidemiology and prevention. New York, Oxford University Press, 1027-1043.
- E Weiderpass, I Persson, H Melhus, S Wedrén, A Kindmark (2000) Estrogen receptor alpha gene polymorphisms and endometrial cancer risk. *Carcinogenesis* 21: 623-627.
- Sasaki M, Tanaka Y, Kaneuchi M, Sakuragi N, Dahiya R (2002) Polymorphisms of estrogen receptor alpha gene in endometrial cancer. *Biochemical & Biophysical Research Communications* 297: 558-564.
- Iwamoto I, Fujino T, Douchi T, Nagata Y (2003) Association of estrogen receptor alpha and beta3-adrenergic receptor polymorphisms with endometrial cancer. *Obstetrics & Gynecology* 102: 506-511.
- Shan D, JH Ben, YY Guo (2003) Study on the relationship between polymorphism of the estrogen receptor gene and endometrial carcinoma. *Chin J Clin Obstet Gynecol* 4: 201-205.
- Yang XS, Liu YH, Zhong X, Zhang, Y Wang (2007) The relationship between single nucleotide polymorphisms in estrogen metabolizing genes CYP1A1, CYP17, COMT and estrogen receptor alpha and the risk of endometrial adenocarcinoma among the Chinese women. *Prog Obstet Gynecol* 16: 76-80.
- Wedrén S, Lovmar L, Humphreys K, Magnusson C, Melhus H, et al. (2008) Estrogen receptor alpha gene polymorphism and endometrial cancer risk—a case-control study. *BMC Cancer* 8: 322.
- Sobczuk A, Smolarz B, Romanowicz-Makowska H, Perłyński T (2009) Estrogen receptor alpha (ER-alpha) gene polymorphism in patients from the Lodz region of Poland with sporadic endometrial cancer. *European Journal of Gynaecological Oncology* 30: 431-434.
- Ashton KA, Proietto A, Otton G, Symonds I, McEvoy M, et al. (2010) Polymorphisms in genes of the steroid hormone biosynthesis and metabolism pathways and endometrial cancer risk. *Cancer Epidemiology* 34: 328-337.
- Sliwinski T, Sitarek P, Stetkiewicz T, Sobczuk A, Blasiak J (2010) Polymorphism of the ERalpha and CYP1B1 genes in endometrial cancer in a Polish subpopulation. *Journal of Obstetrics & Gynaecology Research* 36: 311-317.
- Li G, Xiang YB, Courtney R, Cheng JR, Huang B, et al. (2011) Association of a single nucleotide polymorphism at 6q25.1,rs2046210, with endometrial cancer risk among Chinese women. *Chinese Journal of Cancer* 30: 138-143.
- Healey CS, Ahmed S, O'Mara TA, Ferguson K, Lambrechts D, et al. (2011) Breast cancer susceptibility polymorphisms and endometrial cancer risk: a Collaborative Endometrial Cancer Study. *Carcinogenesis* 32: 1862-1866.
- Lundin E, Wirgin I, Lukanova A, Afanasyeva Y, Krogh V, et al. (2012) Selected polymorphisms in sex hormone-related genes, circulating sex hormones and risk of endometrial cancer. *Cancer Epidemiology* 36: 445-452.
- Zhou X, Gu Y, Wang DN, Ni S, Yan J (2013) Eight functional

- polymorphisms in the estrogen receptor 1 gene and endometrial cancer risk: a meta-analysis. PLoS One 8: e60851.
28. Ashton KA, Proietto A, Otton G, Symonds I, McEvoy M, et al. (2009) Estrogen receptor polymorphisms and the risk of endometrial cancer. BJOG 116: 1053-1061.
29. GRCh38.p7 (2016) Genome Assembly NCBI, MD: National Center for Biotechnology Information, Bethesda, US. National Library of Medicine.