IJC
International Journal of Cancer

A comprehensive gene–environment interaction analysis in Ovarian Cancer using genome-wide significant common variants

Sehee Kim[1], Miao Wang[1], Jonathan P. Tyrer[2], Allan Jensen[3], Ashley Wiensch[4], Gang Liu[5], Alice W. Lee[6], Roberta B. Ness[7], Maxwell Salvatore[1,4], Shelley S. Tworoger[8,9], Alice S. Whittemore[10,11], Hoda Anton-Culver[12], Weiva Sieh[13], Sara H. Olson[14], Andrew Berchuck[15], Ellen L. Goode[16], Marc T. Goodman[17,18], Jennifer Anne Doherty[19], Georgia Chenevix-Trench[20], Mary Anne Rossing[21,22], Penelope M. Webb[23], Graham G. Giles[24,25,26], Kathryn L. Terry[9,27], Argyrios Ziogas[12], Renée T. Fortner[28], Usha Menon[29], Simon A. Gayther[30,31,32], Anna H. Wu[30], Honglin Song[2], Angela Brooks-Wilson[33,34], Elisa V. Bandera[35], Linda S. Cook[36,37], Daniel W. Cramer[9,27], Roger L. Milne[24,25,38], Stacey J. Winham[16], Susanne K. Kjaer[3,39], Francesmary Modugno[40,41,42], Pamela J. Thompson[17], Jenny Chang-Claude[28,43], Holly R. Harris[21], Joellen M. Schildkraut[44], Nhu D. Le[45], Nico Wentzensen[46], Britton Trabert[46], Estrid Høgdall[3,47], David Huntsman[48,49], Malcolm C. Pike[14,50], Paul D.P. Pharoah[2,51], Celeste Leigh Pearce[4,31] and Bhramar Mukherjee[1]

[1]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA
[2]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom
[3]Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark
[4]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA

Key words: ovarian cancer, genetics, additive interaction, G × E

Abbreviations: AR: absolute risk; BMI: body mass index; BSO: bilateral salpingo-oophorectomy; CI: confidence interval; df: degrees of freedom; G × E: gene–environment interaction; GWAS: genome-wide association study; LRT: likelihood ratio test; OCAC: Ovarian Cancer Association Consortium; OCP: oral contraceptive pill; OR: odds ratio; RD: risk difference; SNP: single nucleotide polymorphism

Additional Supporting Information may be found in the online version of this article.

S.K. and M. W. contributed equally to this work

Grant sponsor: California Cancer Center Research Program; **Grant numbers:** 00-01389V-20170, 2II0200; **Grant sponsor:** Canadian Institutes of Health Research; **Grant number:** MOP-86727; **Grant sponsor:** Cancer Councils of New South Wales, Victoria, Queensland, South Australia and Tasmania and The Cancer Foundation of Western Australia; **Grant numbers:** Multi-State Applications 182, 191, 211; **Grant sponsor:** Cancer Institute of New Jersey; **Grant sponsor:** Ovarian Cancer Research Fund; **Grant number:** PPD/RPCI.074; **Grant sponsor:** Deutsches Krebsforschungszentrum; **Grant sponsor:** Fred C. and Katherine B. Andersen Foundation; **Grant sponsor:** German Federal Ministry of Education and Research; **Grant number:** 01 GB 9401; **Grant sponsor:** Intramural Research Program of the National Cancer Institute; **Grant sponsor:** Kræftens Bekæmpelse; **Grant number:** 94 222 52; **Grant sponsor:** Lon V Smith Foundation; **Grant number:** LVS-39420; **Grant sponsor:** Mayo Foundation; **Grant sponsor:** Medical Research and Materiel Command; **Grant number:** DAMD17-01-1-0729; **Grant sponsor:** Mermaid I project; **Grant sponsor:** Minnesota Ovarian Cancer Alliance; **Grant sponsor:** National Cancer Institute; **Grant numbers:** 1 R01CA160669-01A1, K07-CA080668, NIH-K07 CA095666, NIH-K22-CA138563, P30 CA04, P30 CA046592, P30-CA008748, P30-CA072720, P50-CA159981, R01 CA076016, R01-CA61107, R01-CA83918, R01-CA95023, R21-CA222867; **Grant sponsor:** National Cancer Institute's Genetic Associations and Mechanisms in Oncology; **Grant number:** U19-CA148112; **Grant sponsor:** National Center for Research Resources; **Grant number:** M01-RR000056; **Grant sponsor:** National Health and Medical Research Council; **Grant numbers:** 1043134, 199600, ID400281, ID400413, 209057, 396414, 1074383; **Grant sponsor:** National Health Research and Development Program, Health Canada; **Grant number:** 6613-1415-53; **Grant sponsor:** National Institute for Health Research University College London Hospitals Biomedical Research Centre; **Grant sponsor:** National Institute of Environmental Health Sciences; **Grant number:** NIH ES 20811 BM; **Grant sponsor:** National Institutes of Health; **Grant numbers:** K22 CA193860, N01-CN-55424, N01-PC-67001, N01CN025403, N01PC67010, P01 CA87969, P01CA17054, P30-CA15083, P30CA14089, P50-CA105009, P50-CA136393, R01 CA49449, R01-CA0568860, R01-CA063678, R01 -CA074850, R01-CA080742, R01-CA112523, R01-CA122443, R01-CA54419, R01-CA58598, R01-CA67262, R01-CA76016, R01-CA87538, R01 CA61132, R03CA113148, R03CA115195, T32 ES013678, U01 CA69417, U01 CA71966, UM1 CA176726, UM1 CA186107; **Grant sponsor:** National Science Foundation; **Grant number:** NSF DMS 1406712 BM; **Grant sponsor:** Ovarian Cancer Australia; **Grant sponsor:** Peter MacCallum Foundation; **Grant sponsor:** Seventh Framework Programme; **Grant number:** 223175 HEALTH F2 2009-223175; **Grant sponsor:** The Eve Appeal (The Oak Foundation); **Grant sponsor:** U.S. Department of Defense; **Grant numbers:** DAMD17-02-1-0666DAMD17-02-1 -0669W81XWH-10-1-02802; **Grant sponsor:** Connecticut Department of Public Health; **Grant sponsor:** State of Connecticut Department of Public Health

DOI: 10.1002/ijc.32029

History: Received 15 Aug 2018; Accepted 24 Oct 2018; Online 29 Nov 2018

Correspondence to: Bhramar Mukherjee, SPH Tower, 1415 Washington Heights, Ann Arbor MI 48109, USA, Tel.: 734-764-6544; Fax: 734-764-3192, E-mail: bhramar@umich.edu

[5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[6]Department of Public Health, California State University, Fullerton, Fullerton, CA, USA
[7]University of Texas School of Public Health, Houston, TX, USA
[8]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA
[9]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[10]Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford, CA, USA
[11]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA
[12]Department of Epidemiology, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA
[13]Department of Genetics and Genomic Sciences, Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[14]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
[15]Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA
[16]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
[17]Cancer Prevention and Control, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[18]Department of Biomedical Sciences, Cedars-Sinai Medical Center, Community and Population Health Research Institute, Los Angeles, CA, USA
[19]Department of Population Health Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA
[20]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
[21]Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[22]Department of Epidemiology, University of Washington, Seattle, WA, USA
[23]Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
[24]Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, VIC, Australia
[25]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
[26]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia
[27]Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital, Boston, MA, USA
[28]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[29]Gynaecological Cancer Research Centre, Women's Cancer, Institute for Women's Health, University College London, London, United Kingdom
[30]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[31]Center for Cancer Prevention and Translational Genomics, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[32]Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[33]Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada
[34]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada
[35]Cancer Prevention and Control Program, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA
[36]University of New Mexico Health Sciences Center, University of New Mexico, Albuquerque, NM, USA
[37]Division of Cancer Care, Department of Population Health Research, Alberta Health Services, Calgary, AB, Canada
[38]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia
[39]Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
[40]Ovarian Cancer Center of Excellence, Womens Cancer Research Program, Magee-Womens Research Institute and Hillman Cancer Center, Pittsburgh, PA, USA
[41]Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA
[42]Division of Gynecologic Oncology, Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[43]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[44]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
[45]Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada
[46]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[47]Molecular Unit, Department of Pathology, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark
[48]British Columbia's Ovarian Cancer Research (OVCARE) program, Vancouver General Hospital, BC Cancer and University of British Columbia
[49]Department of Pathology and Laboratory Medicine, and Obstetrics and Gynecology, University of British Columbia, Vancouver, BC, Canada
[50]Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA
[51]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

As a follow-up to genome-wide association analysis of common variants associated with ovarian carcinoma (cancer), our study considers seven well-known ovarian cancer risk factors and their interactions with 28 genome-wide significant common genetic variants. The interaction analyses were based on data from 9971 ovarian cancer cases and 15,566 controls from 17 case–

control studies. Likelihood ratio and Wald tests for multiplicative interaction and for relative excess risk due to additive interaction were used. The top multiplicative interaction was noted between oral contraceptive pill (OCP) use (ever *vs.* never) and rs13255292 (p value = 3.48×10^{-4}). Among women with the TT genotype for this variant, the odds ratio for OCP use was 0.53 (95% CI = 0.46–0.60) compared to 0.71 (95%CI = 0.66–0.77) for women with the CC genotype. When stratified by duration of OCP use, women with 1–5 years of OCP use exhibited differential protective benefit across genotypes. However, no interaction on either the multiplicative or additive scale was found to be statistically significant after multiple testing correction. The results suggest that OCP use may offer increased benefit for women who are carriers of the T allele in rs13255292. On the other hand, for women carrying the C allele in this variant, longer (5+ years) use of OCP may reduce the impact of carrying the risk allele of this SNP. Replication of this finding is needed. The study presents a comprehensive analytic framework for conducting gene–environment analysis in ovarian cancer.

What's new?

Genetic and environmental risk factors for ovarian cancer have been identified separately but interactions between both remain largely unexplored. The authors identified a new gene x environment interaction between oral contraceptive pill (OCP) use and a single nucleotide polymorphism in the PVT1 gene, a long-noncoding RNA located on chromosome 8. The data suggest that the protective benefit of OCP use may be strongest in women with the T allele of PVT1 underscoring the need to tailor prevention strategies to individual genotypic profiles.

Introduction

Ovarian carcinoma (cancer) is a disease with high mortality; most women are diagnosed with advanced stage disease where five-year survival is less than 50%.[1] Effective screening modalities have been elusive,[2] and therefore primary prevention strategies remain the most promising avenue to minimize the incidence and mortality of ovarian cancer.

Several factors consistently associated with reduced or increased risk have been identified for ovarian cancer, including some that represent opportunities for chemoprevention or surgical intervention. Factors associated with reduced risk include oral contraceptive pill (OCP)[3] use, aspirin use,[4] tubal ligation,[5] parity,[3] salpingectomy[6–9] and bilateral salpingo-oophorectomy (BSO). Common germline genetic variation,[10–20] first-degree family history of ovarian cancer,[21,22] menopausal hormone therapy use,[23–25] greater body mass index (BMI)[26] and endometriosis[27] are risk factors for the disease. OCPs and aspirin use represent feasible chemoprevention strategies whereas salpingectomy is now recommended by many gynecologic societies as an ovarian cancer prevention approach for women seeking tubal sterilization, having a hysterectomy, or having other pelvic surgery.

Average lifetime risk of ovarian cancer diagnosis for women in the U.S. is 1.3%,[28] but this number varies greatly depending on the composite exposure history of risk factors.[29] Pearce *et al.* estimated the lifetime risk for women in the general population ranges from 0.35% (95%CI = 0.29% to 0.42%) to 8.8% (95% CI = 7.1% to 10.9%) depending on exposure history for six factors: OCP use, parity, tubal ligation, endometriosis, first degree family history of ovarian cancer and genetic risk score quintile.[29]

However, these lifetime risk estimates were limited to six risk factors and did not consider their interaction with individual genetic variants identified through genome-wide association studies (GWAS).[28] The multiplicative scale is commonly used for gene–environment interaction (G × E) analysis. Additive interaction analysis has been suggested for case–control studies in many recent papers for a more mechanistic interpretation.[30–34] Validity of a truly multiplicative model implies existence of additive interaction when the two factors under consideration have non-null main effects.[35] Thus, failure to detect G × E interaction on multiplicative scale may imply there exists interaction on additive scale, but the ability to detect it depends on the sample size and the main and interaction effect sizes.[35] We present here our efforts to evaluate both multiplicative and additive gene–environment interactions in ovarian cancer using data from the international Ovarian Cancer Association Consortium (OCAC) comprising 17 case–control studies.

We have included 28 common genetic variants previously associated with risk of ovarian cancer in genome-wide association analyses for our G × E analyses.[36] Environmental factors included in our analysis are OCP use, parity, tubal ligation, breastfeeding, menopausal hormone therapy, usual adult BMI, and endometriosis. A small number of studies in OCAC had data available on aspirin use and thus we have not included this risk factor in our analysis here. Among our list of environmental factors, BMI, OCP use, tubal ligation, breastfeeding, and menopausal hormone therapy are of special interest because they are modifiable targets for prevention.

Methods

Study population

The OCAC is an international multidisciplinary consortium formed in 2005 (http://apps.ccge.medschl.cam.ac.uk/consortia/

ocac/) with a goal of sharing data from worldwide ovarian cancer studies to establish reliable estimation of association between environmental and genetic factors related to risk of ovarian cancer.[23,37] Cases were defined as women with ovarian carcinoma (i.e., invasive epithelial ovarian cancers), fallopian tube cancer and primary peritoneal cancer. Controls were women without ovarian cancer and who had at least one ovary. For both cases and controls, individuals with prior cancers except non-melanoma skin cancers were excluded.

Genetic association analysis

In total, 28 single nucleotide polymorphisms (SNPs) previously identified through GWAS were included from 75 OCAC sites (Table 1). The first 26 SNPs were found to be significantly associated with either ovarian cancer overall or one or more histotypes.[36] In addition, rs13255292 and rs10962643 were included because they were in the same region as two other significant SNPs but showed a strong independent association with ovarian cancer risk. The SNP at locus 15q26 (rs8037137), which was found to be genome-wide significant,[13] was not included because not enough non-carriers were present in our analytic dataset for examining interactions. The genetic data included both genotyped and imputed variants (imputation being carried out using phase 2 Hapmap reference panel). More details regarding genotyping and imputation of the genetic data have been previously described.[12,17,18,20] The methods for analyzing the SNP data in the OCAC have also been described previously.[12,17,18,20] Briefly, logistic regression models were fit to examine the association between ovarian cancer and each genetic variant under an additive model (using risk allele dosage). The models were adjusted for ethnicity, genotyping panel and the leading principal components for each ethnicity. The summary results are shown in Table 1 and are also available through the OCAC website (http:/apps.ccge.medschl.cam.ac.uk/consortia/ocac/).

Environmental association analysis

Environmental variables (E). A total of seven established environmental risk factors for ovarian cancer were of primary interest (Table 2), including four associated with decreased risk and three with increased risk for ovarian cancer or one specific histotype. These included: OCP use, measured as both ever/never and duration of OCP use (never users including <1 one year of use, 1- < 5, 5+ years), tubal ligation (yes/no), breastfeeding (ever/never), parity (0, 1–2, 3+ full-term births lasting ≥6 months), type of menopausal hormone therapy use for more than 1 year after age 50 (never user, menopausal estrogen therapy only, any use of menopausal estrogen + progestin therapy), BMI (<25, 25- <30, 30+), and a history of endometriosis (yes/no).

Four other environmental variables were included in our analysis, as covariates: baseline age (<50, 50- <55, 55- <60, 60- <65, 65–70, 70+ years), Ethnic group (non-Hispanic white, Hispanic White, Black, Other), education (less than high school, high school graduate, some college, college graduate) and first-degree family history of ovarian cancer (yes/no). In addition to these four covariates, study site, OCP use, tubal ligation, parity, BMI and endometriosis were also included in all models for the environmental association analysis and gene by environment interaction analysis.

Harmonization and imputation of environmental data. A brief description of environmental data harmonization across OCAC study sites is provided in eMethod 1 in the Supporting Information. To optimize power and enhance the chance for discovery, we carried out multiple imputation of the environmental data. The maximal amount of data was used for imputation (see eMethod 1 and eFig. 1 in the Supporting Information for details). A total of 19 studies comprising 13,722 cases and 22,975 controls with partially missing data were included for imputation. Of these 19 studies, 12 were from the US, 4 from Europe, 2 from Canada and 1 from Australia (see Supporting Information, eTable 1 for a description of study sites). Further details for these 19 studies have been previously described (see Supporting Information). The environmental variables included in our analysis were multiply imputed by chained equations (MICE) to produce 10 imputed datasets. See details of imputation model in eMethod 2.1 in the Supporting Information.

All analyses were performed on each of the 10 imputed datasets, and coefficients/test statistics were properly combined to account for uncertainty due to imputation, after the recommended combination rule for multiply imputed datasets[38] (see details in eMethod 2.3 in the Supporting Information). Our marginal environmental association analysis was based on combined inference from the 10 imputed versions of this harmonized E data. Logistic regression models were used for evaluating marginal associations between the environmental risk factors with ovarian cancer after adjusting for covariate. The estimated ORs, their 95% CIs, as well as two-sided Wald tests after accounting for imputation uncertainty are presented in Table 2 along with summary statistics of complete cases before imputation. Full results of the complete cases analysis using logistic regression models are presented in Supporting Information, eTable 2.

Gene by environment interaction analysis

After marginal analysis of the genetic and environmental risk factors, we considered gene by environment (G × E) interaction analysis both on the multiplicative (odds ratio/relative risk) and the additive (relative excess risk due to interaction/absolute risk) scale.[39] From the 19 studies with imputed environmental data, a subset of 17 case–control studies with 9971 cases and 15,566 controls had available genetic data, thus G × E analyses were carried out on these 17 studies. Each imputed environmental dataset was merged with the genetic data for subsequent G × E analyses. Interaction analyses were then

Table 1. Odds ratios for marginal associations of 28 genetic susceptibility variants with ovarian cancer.

SNP	Previously published best hit[1]	Chr	Position	Risk Allele	Baseline Allele	RAF	OR[2]	p Value[2]
rs12023270	rs58722170[15]	1	38,086,578	T	C	0.264	1.08 (1.05,1.10)	2.65×10^{-8}
chr2:111818658	rs2165109[18]	2	111,818,658	C	A	0.277	1.06 (1.04,1.09)	2.03×10^{-6}
rs874898	rs752590[14]	2	113,974,196	C	G	0.262	1.00 (0.98,1.03)	7.36×10^{-1}*
rs1562314	rs711830[14]	2	177,045,560	T	A	0.638	1.10 (1.07,1.13)	2.84×10^{-14}
rs112071820[18]		3	138,849,110	allele 1	G	0.270	1.03 (1.00,1.06)	5.17×10^{-2}*
chr3:156397692	rs62274041[17]	3	156,397,692	T	C	0.048	1.47 (1.39,1.55)	7.73×10^{-47}*
rs9870207[18]		3	190,525,516	A	G	0.666	1.05 (1.03,1.08)	2.95×10^{-5}
rs7705526	rs10069690[10]	5	1,285,974	A	C	0.343	1.10 (1.07,1.12)	5.52×10^{-14}
chr5:66121089	rs555025179[18]	5	66,121,089	allele2	G	0.526	1.03 (1.00,1.05)	2.61×10^{-2}*
chr8:82653644	8:82668818[17]	8	82,653,644	G	A	0.064	1.18 (1.12,1.23)	3.25×10^{-12}*
rs9886651[18]		8	128,817,883	G	A	0.435	1.06 (1.03,1.08)	2.89×10^{-6}*
rs13255292[18]	NA	8	129,076,573	C	T	0.700	1.07 (1.05,1.10)	3.57×10^{-8}*
rs10103314	rs1400482[12]	8	129,560,744	A	C	0.883	1.15 (1.11,1.20)	5.76×10^{-15}*
chr9:16915105	rs10962692[20]	9	16,915,105	C	G	0.834	1.24 (1.20,1.28)	4.54×10^{-41}*
rs10962643	NA	9	16,857,403	C	A	0.699	1.17 (1.14,1.20)	1.13×10^{-35}*
rs320203[18]		9	104,943,226	C	A	0.842	1.03 (1.00,1.06)	5.21×10^{-2}
chr9:136138765[15]		9	136,138,765	G	allele 3	0.176	1.12 (1.08,1.15)	1.49×10^{-12}*
rs7084454	rs144962376[17]	10	21,821,274	A	G	0.301	1.07 (1.05,1.10)	3.32×10^{-8}*
rs7902587[18]		10	105,694,301	T	C	0.091	1.08 (1.03,1.12)	4.54×10^{-4}*
chr12:121403724	rs7953249[18]	12	121,403,724	A	G	0.570	1.05 (1.03,1.07)	2.58×10^{-5}*
chr15:91531995	rs8037137[13]	15	91,531,995	C	T	0.829	1.08 (1.05,1.12)	1.18×10^{-6}*
rs11658063	rs7405776[19]	17	36,103,872	G	C	0.614	1.04 (1.02,1.07)	2.98×10^{-4}*
chr17:43552537	rs1879586[17]	17	43,552,537	A	G	0.164	1.12 (1.08,1.15)	2.22×10^{-12}*
rs7217120	rs7207826[16]	17	46,484,755	C	T	0.275	1.10 (1.07,1.13)	8.69×10^{-13}*
rs8098244[18]		18	21,405,553	G	A	0.741	1.04 (1.01,1.07)	4.23×10^{-3}*
rs4808075[11]		19	17,390,291	C	T	0.268	1.13 (1.10,1.16)	1.49×10^{-20}*
rs74597329	rs688187[14]	19	39,739,155	G	T	0.301	1.02 (0.99,1.04)	2.63×10^{-1}
rs6005807[18]		22	28,934,313	T	C	0.095	1.09 (1.04,1.13)	3.35×10^{-5}*

Analysis used data with 26,864 cases and 48,034 controls from 75 study sites. Abbreviations: SNP, single-nucleotide polymorphism; RAF, risk allele frequency; Chr, chromosome; OR, odds ratio; allele1, GCCAGATTCAGAAT; allele2, GACACACAC; allele3, GCGCCCACCACTA.
[1] If not specified, the previously published best hit is the same as the current best hit.
[2] Logistic regression for ovarian cancer overall (regardless of histology), adjusted for ethnicity, study panel and leading principal components for each ethnicity (using a total of 47 principal components).
* p Value >0.01.

carried out separately on the 10 imputed G × E datasets, and then all tests and coefficients reported were combined using appropriate multiple imputation combination rules.[38]

For both multiplicative and additive interaction analysis, we started with global likelihood ratio tests (LRTs) for each G × E pair as several environmental factors had multiple categories resulting in tests for interactions with multiple degrees of freedom (df). These global joint tests, serving as a screening step for G × E interactions, were carried out for a total of 196 (7 × 28 = 196) G × E pairs. After the global tests, we then followed up on the suggestive interactions (with global test p value <0.2) and carried out a two-sided Wald test for interactions involving each separate category of an environmental risk factor.

For the k-th SNP G_k (k = 1, …, 28), coded as a continuous allelic dosage, the j-th environmental risk factor E_j (j = 1, …, 7), and a set of confounders/covariates $\{C_q\}$ (q = 1, …, Q), the basic fitted model for the probability of ovarian cancer of the i-th subject, namely, π_i, is of the after form:

$$\begin{aligned} &logit\left(\pi_i \mid G_{ki}, E_{ji}, C_{1i}, \ldots, C_{Qi}\right) \\ &\quad = \beta_0 + \beta_G G_{ki} + \sum_{l=1}^{L} \beta_{El} I\left(E_{ji} = l\right) \\ &\quad + \sum_{l=1}^{L} \beta_{GEl} I\left(E_{ji} = l\right) G_{ki} + \sum_{q=1}^{Q} \sum_{m=1}^{M_q} \beta_{C_q m} I\left(C_{qi} = m\right), \end{aligned} \tag{M1}$$

where L = (levels of E_j) − 1, M_q = (levels of C_q) − 1, and Q is the number of adjusted covariates.

Multiplicative interaction tests. For testing the multiplicative interaction between G_k and E_j, we first used the global LRT with L degrees of freedom to test for the joint null hypothesis

Table 2. Odds ratios for marginal associations of seven environmental risk factors with ovarian cancer risk with 13,722 cases and 22,975 controls from 19 study sites

	Before Imputation[1]		After Imputation[2]			
Environmental risk factor	**Control**	**Case**	**Control**	**Case**	**OR[3]**	***p* Value[3]**
OCP use						
Never	0.347	0.444	0.351	0.452	Ref	
Ever	0.645	0.536	0.649	0.548	0.62 (0.59,0.66)	5.24×10^{-73}
(missing)	0.008	0.020				
Duration of OCP use						
Never users (including <1 year)	0.425	0.542	0.430	0.554	Ref	
1- <5 year	0.229	0.208	0.232	0.215	0.70 (0.66,0.74)	8.23×10^{-32}
5+ year	0.332	0.222	0.338	0.231	0.48 (0.45,0.51)	2.20×10^{-133}
(missing)	0.014	0.028				
Tubal ligation						
No	0.693	0.777	0.762	0.824	Ref	
Yes	0.208	0.160	0.238	0.176	0.73 (0.69,0.78)	1.81×10^{-23}
(missing)	0.098	0.063				
Breastfeeding						
No	0.239	0.294	0.380	0.515	Ref	
Yes	0.532	0.410	0.620	0.485	0.76 (0.71,0.80)	4.80×10^{-21}
(missing)	0.229	0.296				
Parity (number of full-term births)						
0	0.148	0.241	0.149	0.243	Ref	
1–2	0.487	0.434	0.489	0.438	0.59 (0.55,0.63)	1.94×10^{-65}
3+	0.359	0.315	0.362	0.319	0.50 (0.46,0.53)	4.91×10^{-90}
(missing)	0.006	0.011				
Type of HT using more than 1 year after age 50						
Never use	0.687	0.647	0.789	0.782	Ref	
ET only	0.060	0.075	0.066	0.084	1.22 (1.12,1.34)	2.65×10^{-5}
Any EPT	0.131	0.118	0.145	0.134	0.97 (0.90,1.04)	3.55×10^{-1}
(missing)	0.121	0.160				
BMI						
<25	0.392	0.370	0.516	0.485	Ref	
25- <30	0.209	0.213	0.284	0.286	1.03 (0.98,1.09)	2.55×10^{-1}
30+	0.144	0.174	0.200	0.229	1.15 (1.08,1.22)	6.11×10^{-6}
(missing)	0.255	0.243				
Endometriosis						
No	0.703	0.695	0.937	0.902	Ref	
Yes	0.047	0.076	0.063	0.098	1.60 (1.46,1.75)	3.41×10^{-23}
(missing)	0.250	0.230				

Abbreviations: OR, odds ratio; OCP, oral contraceptive pills; BMI, body mass index; HT, menopausal hormone therapy; ET, menopausal estrogen therapy; EPT, menopausal estrogen + progestin therapy; Ref, reference group.
[1] Harmonized environmental data before imputation. Results of the complete cases analysis are provided in Supporting Information eTable 2.
[2] Based on 10 imputed E datasets.
[3] Logistic regression model adjusted for reference age, Ethnic group, education, family history, OCP use, tubal ligation, parity, BMI, endometriosis and study site.

H_0: $\beta_{GE1} = \beta_{GE2} = \ldots = \beta_{GEL} = 0$. If the global test p value <0.2, we further assessed the multiplicative interaction at each level of E_j by using a Wald test with one degree of freedom for the null hypothesis $H_0 : \beta_{GEl} = 0$ for the l-th level.

Additive interaction tests. Due to limitations of existing software (CGEN)[40] for testing additive interactions with continuous dosage data, we used the maximal probable genotype for imputed SNPs. We further conducted the LRTs with binary collapsing of SNPs assuming a dominant genetic susceptibility model (given the constraints in software).[31] For a given SNP G_k and an environmental risk factor E_j with L categories, a global LRT with L df was used for the after joint null hypothesis

$$H_0: \frac{\{\exp(\beta_{E1}) + \exp(\beta_G) - 1\}}{\exp(\beta_{E1} + \beta_G)} = \exp(\beta_{GE1}),$$
$$\ldots, \frac{\{\exp(\beta_{EL}) + \exp(\beta_G) - 1\}}{\exp(\beta_{EL} + \beta_G)} = \exp(\beta_{GEL}),$$

where the regression coefficients (β) are log odds ratio parameters described in model [M1]. This null hypothesis is based on a rare disease assumption,[41] which is tenable for our study (lifetime risk of ovarian cancer in the US is approximately 1.3%).[42] If the global LRT p value <0.2, we further assessed the additive interaction at each level of E_j through the relative excess risk due to interaction (RERI).[41] At the l-h level of E_j, a Wald test with one degree of freedom[35] was used to test for the null hypothesis: H_0: $RERI_{GEl} = 0$, where $RERI_{GEl} = \exp(\beta_{El} + \beta_{GEl} + \beta_G) - \exp(\beta_{El}) - \exp(\beta_G) + 1$.

After the screening step, we further explored the structure of the most promising interactions (defined as global test p value <0.01). This was accomplished by exploring odds ratios corresponding to E in sub-groups defined by G (for the multiplicative interaction) or absolute risks for ovarian cancer in each configuration of the values of (G, E) (for the additive interaction). To better understand these two different scales of interaction, we also compared the observed joint ORs with the corresponding expected ORs under the multiplicative and the additive nulls.

To estimate sub-group specific absolute risk (AR) for each stratum defined by a given SNP G_k and environmental risk factor, we need the relative risk and the joint distribution of G_k and E_j. The former was estimated from the fitted model [M1], and the latter was empirically estimated from the observed joint frequency of E_j and G_k in the control population (details in eMethod3 from the Supporting Information). Table 4 presents the bootstrap confidence intervals for the estimated ARs and the risk differences (RDs) (see details in eMethod4 in the Supporting Information). The results for G × E analysis are presented in Table 3 (multiplicative interaction), Table 4 (additive interaction) and Supporting Information, eTable 5 (observed and expected joint OR under the two different nulls). All calculations were performed in the statistical software R.[30,40]

Results

The marginal G analysis was carried out on 26,864 cases and 48,034 controls and the results are shown in Table 1. These results are available through the OCAC website (http://apps.ccge.medschl.cam.ac.uk/consortia/ocac/). A total of 36,697 women with 13,722 ovarian cancer cases from 19 sites were included in the marginal E analysis using the imputed datasets. All seven environmental risk factors were associated with ovarian cancer in the expected direction (Table 2). OCP use for five or more years was associated with a 52% decrease in risk of ovarian cancer compared to never users (OR = 0.48, 95%CI = 0.45 to 0.51). Tubal ligation (OR = 0.73, 95%CI = 0.69 to 0.78) and breastfeeding (OR = 0.76, 95%CI = 0.71 to 0.80) showed similar magnitudes of decreased risk. Also, having more than 3 children (*vs.* none) was associated with a 50% (OR = 0.5, 95%CI = 0.46 to 0.53) reduction in risk of ovarian cancer. Using menopausal estrogen therapy only for more than one year (OR = 1.22, 95%CI = 1.12 to 1.34), being obese (OR = 1.15, 95%CI = 1.08 to 1.22), and history of endometriosis (OR = 1.60, 95%CI = 1.46 to 1.75) were all associated with increased risk of ovarian cancer. The inference remained robust before and after imputation (Supporting Information, eTable 2.).

Gene by environment interaction results

Global likelihood ratio tests. The global LRT essentially serves as a screening approach to identify a list of potentially interesting interactions. All interactions with global LRT p value <0.2 (40 on multiplicative scale and 41 on additive scale) are listed in Supporting Information, eTable 3, while more detailed analysis of the top interactions, which showed the strongest significance (p value < 0.01; 4 on multiplicative and 2 on additive scale), are shown in Tables 3 and 4, respectively.

According to Global LRT results, the top interaction on the multiplicative scale was identified with the SNP rs13255292 and OCP use (ever and never use: p value = 3.48×10^{-4}; duration of use [<1 year, 1–5 years, 5+ years]: p value = 7.26×10^{-3}) (Table 3). None of the observed interactions were significant based on a Bonferroni threshold of $0.05/(28 \times 7) = 2.55 \times 10^{-4}$.

Wald tests for multiplicative interactions. For the most promising multiplicative interactions reported in Table 3 we carried out an in-depth analysis to better understand the structure of interactions by estimating the ORs (with accompanying Wald CIs and tests) corresponding to E in strata defined by G. For example, the OR for OCP use among women with the TT genotype for rs13255292 is estimated to be 0.53 (95% CI = 0.46 to 0.60), whereas for the CC genotype the estimated OR is 0.71 (95%CI = 0.66 to 0.77) suggesting a stronger protective effect of OCP use among TT genotypes (Table 3, Fig. 1*a*).

When OCP use was further stratified by duration, we observed an interesting pattern in its interaction with rs13255292. The estimated OR corresponding to 1–5 year of OCP use *vs.* <1 year use in the TT genotype group was 0.58 (95%CI = 0.50 to 0.69) compared to an OR of 0.79 (95% CI = 0.72 to 0.87) among women with CC genotype, showing effect modification by the risk allele (C) of rs13255292 (Table 3, Fig. 1*b*). This is akin to the result with ever/never user. However, the OR corresponding to 5+ years of OCP use *vs.* <1 year of use for the TT genotype group was 0.43 (95% CI = 0.37 to 0.50) and for the CC genotype was 0.53 (95% CI = 0.49 to 0.58) (Table 3, Fig. 1*c*). With overlapping confidence intervals, there is no significant difference in the odds ratios for long-term OCP users across genotype sub-groups. Table 3 shows that the p value of the Wald test for interaction

Table 3. Results from Multiplicative Interaction Analysis: Odds ratios corresponding to environmental risk factors, stratified by genotype (for G-E pairs with global likelihood ratio test p value <0.01

SNP	Environmental risk factor		N (cases/controls)[1]			Estimated OR[2] for E stratified by G (95%CI)			Global[3] LRT	Wald[4] Test
Risk/Baseline allele	Variable	Category	Genotype			Genotype			(df)	(df)
rs13255292 C/T	OCP use		TT	TC	CC	TT	TC	CC		
		Never	396/503	1758/2175	2077/2570	Ref			Ref	Ref
		Ever	446/1069	2286/4336	2768/4750	0.53 (0.46,0.60)	0.61 (0.57,0.66)	0.71 (0.66,0.77)	3.48×10^{-4} (1)	3.47×10^{-4} (1)
		Missing	24/15	96/56	120/96					
rs13255292 C/T	Duration of OCP use		TT	TC	CC	TT	TC	CC		
		<1 yr	451/636	2213/2670	2546/3145	Ref			Ref	Ref
		1- <5 yr	171/362	854/1522	1082/1662	0.58 (0.50,0.69)	0.68 (0.63,0.74)	0.79 (0.72,0.87)	7.26×10^{-3} (2)	4.74×10^{-3} (1)
		5+ yr	209/568	945/2269	1178/2470	0.43 (0.37,0.5)	0.48 (0.44,0.52)	0.53 (0.49,0.58)		2.43×10^{-2} (1)
		Missing	35/21	128/106	159/135					
rs10962643 C/A	Parity (full term birth)		AA	AC	CC	AA	AC	CC		
		0	230/220	940/940	1194/1080	Ref			Ref	Ref
		1–2	398/835	1741/3184	2202/3536	0.52 (0.44,0.61)	0.56 (0.51,0.6)	0.60 (0.54,0.66)	7.52×10^{-3} (2)	1.99×10^{-1} (1)
		3+	243/579	1242/2459	1664/2614	0.38 (0.32,0.46)	0.46 (0.42,0.5)	0.55 (0.49,0.61)		2.86×10^{-3} (1)
		Missing	11/15	47/58	59/46					
chr9:16915105 C/G	Parity (full term birth)		GG	GC	CC	GG	GC	CC		
		0	73/72	624/649	1667/1519	Ref			Ref	Ref
		1–2	111/300	1129/2285	3101/4970	0.46 (0.36,0.58)	0.52 (0.47,0.59)	0.60 (0.55,0.65)	5.25×10^{-3} (2)	5.10×10^{-2} (1)
		3+	70/220	749/1679	2330/3753	0.33 (0.26,0.43)	0.42 (0.37,0.48)	0.53 (0.48,0.58)		1.25×10^{-3} (1)
		Missing	2/7	37/36	78/76					

Analysis used the G × E data with 9971 cases and 15,566 controls from 17 study sites)

Abbreviation: SNP, single-nucleotide polymorphism; OR, odds ratio; OCP, oral contraceptive pills; yr, year; Ref, reference group; df, degree of freedom, LRT, likelihood ratio test.

[1] Number of cases and controls were estimated from the original merged G × E data (before imputation) with 9971 cases and 15,566 controls from 17 study sites, using maximal probable genotypes for imputed SNPs.

[2] ORs were estimated from the logistic regression model with SNP, E variable, SNP × E variable.

[3] LRT was performed for jointly testing multiplicative interactions.

[4] Wald test for individual multiplicative interaction.

All models were estimated from the logistic regression model with SNP, E variable, SNP × E variable, assuming log-additive model, using dosage data for imputed SNPs, adjusted for reference age, ethnic group, education, family history, OCP use, tubal ligation, parity, BMI, endometriosis and study site and were performed on imputed datasets of G-E (9971 cases, 15,566 controls) with proper pooling.

Table 4. Absolute risks and risk differences stratified by levels of environmental risk factor and levels of genotype (for G-E pairs with global likelihood ratio test p value <0.01 on additive scale

SNPs	Environmental risk factor		N (cases/controls)[1]			Estimated ARs or RDs for E stratified by SNPs (95%CI)[2]			Global LRT[3]	Wald Test[4]
risk/baseline allele	variable	category	Genotype			Genotype			(df)	(df)
rs11658063 G/C	Type of HT		CC	CG	GG	CC	CG	GG		
		Neither	589/1142	2609/4518	3310/4956	1.27% (1.23%,1.32%)	1.30% (1.28%,1.33%)	1.33% (1.26%,1.40%)	Ref	Ref
		ET only	66/98	281/409	416/454	1.36% (1.15%,1.57%)	1.63% (1.46%,1.79%)	1.96% (1.59%,2.33%)		
		RD[5]				0.09% (−0.14%,0.31%)	0.33% (0.15%,0.50%)	0.63% (0.24%,1.02%)	3.29×10^{-3} (2)	3.01×10^{-2} (1)
		Any EPT	105/207	498/952	606/1046	1.16% (1.04%,1.28%)	1.21% (1.12%,1.30%)	1.27% (1.09%,1.44%)		
		RD				−0.12% (−0.26%,0.03%)	−0.09% (−0.20%,0.01%)	−0.06% (−0.26%,0.13%)		7.04×10^{-1} (1)
		missing	122/202	582/762	787/820					
rs9886651 G/A	OCP use		AA	AG	GG	AA	AG	GG		
		Never	1278/1718	2053/2502	900/1028	1.52% (1.42%,1.62%)	1.70% (1.64%,1.76%)	1.91% (1.77%,2.04%)	Ref	Ref
		Ever	1666/3105	2640/4978	1194/2072	1.07% (1.02%,1.12%)	1.10% (1.07%,1.13%)	1.14% (1.07%,1.21%)		
		RD				−0.45% (−0.57%, −0.33%)	−0.60% (−0.69%, −0.51%)	−0.77% (−0.93%, −0.60%)	5.32×10^{-3} (2)	9.90×10^{-3} (1)
		missing	70/47	113/79	57/37					

Analysis used the G × E data with 9971 cases and 15,566 controls from 17 study sites)

Abbreviation: SNP, single-nucleotide polymorphism; AR, absolute risk; RD, risk difference; OCP, oral contraceptive pills; HT, menopausal hormone therapy; ET, menopausal estrogen therapy; EPT, menopausal estrogen + progestin therapy; Ref, reference group; df, degree of freedom.

All models were estimated from logistic regression model with SNP, E variable, SNP × E variable, assuming log-additive model (except for additive LRT which assumes dominant effect), using maximal probable genotypes for imputed SNPs, adjusted for reference age, Ethnic group, education, family history, OCP use, tubal ligation, parity, BMI, endometriosis and study site and were performed on imputed datasets of G-E (9,971 cases, 15,566 controls) with proper pooling.

[1] Number of cases and controls were estimated from the original merged G × E data (before imputation) with 9971 cases and 15,566 controls from 17 study sites, using maximal probable genotypes for imputed SNPs.

[2] ARs were estimated from logistic regression model by empirically estimated distribution of E and SNPs, while fixing all other covariates at their mode (determined from the original data).

[3] LRT was performed for jointly testing additive interactions, assuming dominant effect model of SNPs (due to limitation of software).

[4] 1-df Wald test corresponds to the test individual RERI term (SNP = 2 *vs.* SNP = 0, E = k *vs.* E = reference group) is zero or not.

[5] The risk difference corresponds to given category compared to the reference group, stratified by SNP.

All models were estimated from logistic regression model with SNP, E variable, SNP × E variable, assuming log-additive model (except for additive LRT which assumes dominant effect), using maximal probable genotypes for imputed SNPs, adjusted for reference age, ethnic group, education, family history, OCP use, tubal ligation, parity, BMI, endometriosis and study site and were performed on imputed datasets of G-E (9,971 cases, 15,566 controls) with proper pooling.

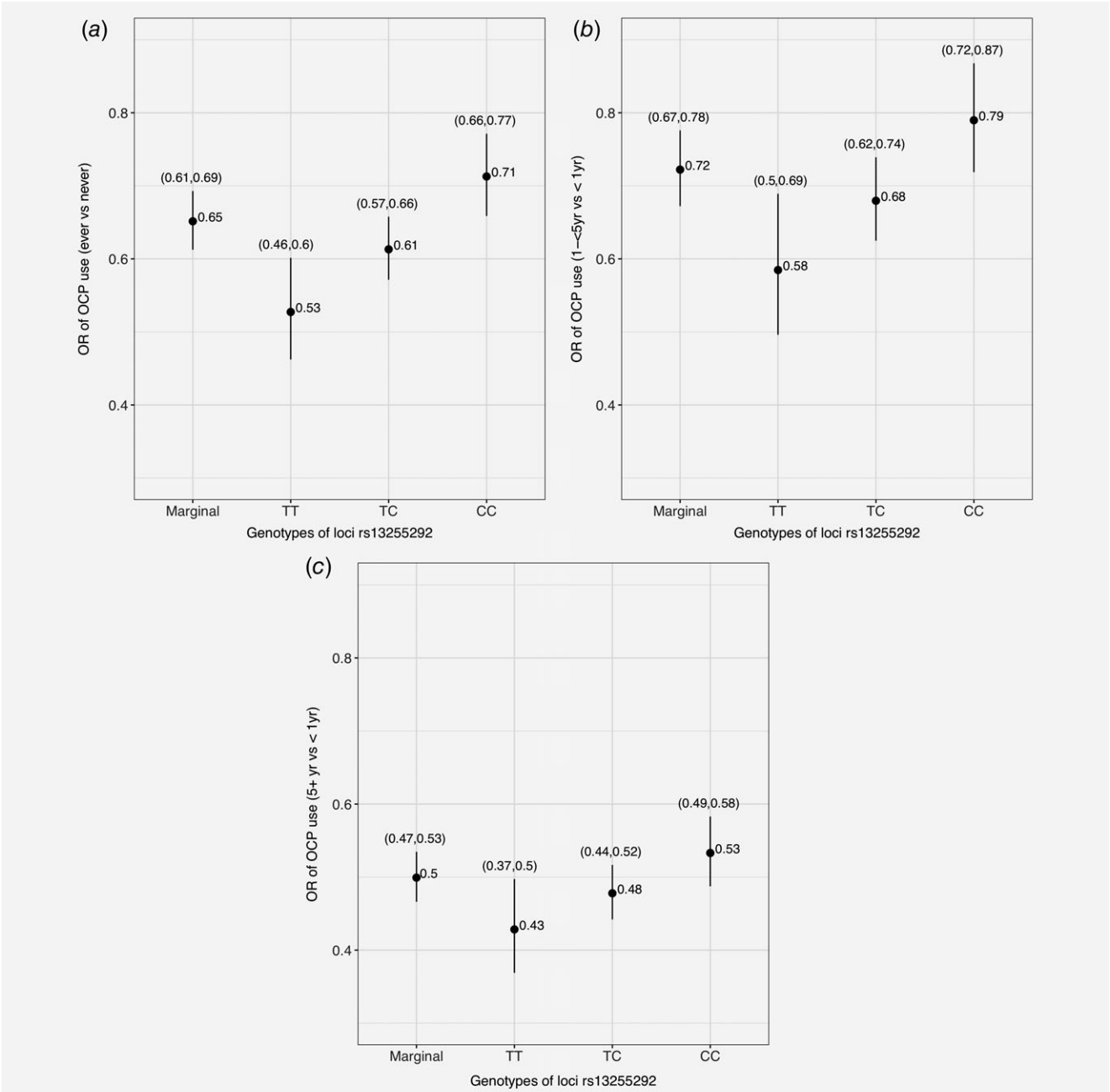

Figure 1. (*a–c*) ORs of oral contraceptive (OCP) use, marginally, or stratified by number of risk allele of rs13255292. The ORs were calculated from a logistic regression model assuming log-additive effect of SNPs. (*a*) OR of OCP (ever *vs*. never) (*b*) OR of 1 to 5 years of OCP use (*vs*. ‹1 year) (*c*) OR of more than 5 years of OCP use (*vs*. ‹1 year).

of rs13255292 and 1–5 years of OCP use (*vs*. <1 year) was lower (p value = 4.74×10^{-3}), when compared to the p value for interaction of the same variant with 5+ years of OCP use (*vs*. <1 year) (p value = 2.43×10^{-2}).

Wald test for additive interaction/RERI. For the most statistically significant additive interactions in Table 4, we estimated the sub-group specific absolute risks (ARs) and risk differences (RDs) in each E by G stratum. For example, for the strongest additive interaction based on the global likelihood ratio tests in Table 4, there was suggestive evidence that rs11658063 modified the effect of menopausal estrogen therapy use, compared to never use of menopausal hormone therapy (p value = 3.01×10^{-2}). Among women with the GG genotype, never users of menopausal hormone therapy had an estimated AR of 1.33% (95%CI = 1.26% to 1.40%) while

women who used menopausal estrogen therapy had an estimated AR of 1.96% (95%CI = 1.59% to 2.33%), leading to an absolute risk increase of 0.63% (95%CI = 0.24% to 1.02%) (Table 4, eFig. 2, Supporting Information).

For women with the CC genotype, the estimated AR was 1.27% (95%CI = 1.23% to 1.32%) for never receiving menopausal hormone therapy and 1.36% (95%CI = 1.15% to 1.57%) for receiving menopausal estrogen only therapy. This implies virtually no increased risk from taking menopausal estrogen only therapy among women with the CC genotype (95%CI = −0.14% to 0.31%; Table 4, eFig. 2, Supporting Information). The results on the additive interactions were in general weaker in terms of the strength of p values.

Discussion

We have conducted a comprehensive multiplicative and additive interaction analysis of previously identified common genetic variants and environmental factors unequivocally associated with ovarian cancer risk. We observed six suggestive interactions (with p value <0.01), four on the multiplicative scale and two on the additive scale. The lack of statistical significance of interactions after multiple testing correction from a large collection of data and well-curated studies enable us to conclude that it is unlikely that there are substantive interactions with single variants and environmental factors regardless of the choice of scale. This is consistent with what has been observed for other cancers. One may argue that the Bonferroni threshold for multiple comparisons is likely to be conservative for this set of correlated environmental factors, but the general pattern of findings remains consistent with smaller magnitude of interaction effect sizes. However, there are several interesting findings from this analysis that may be worthwhile to follow-up in future G × E studies of ovarian cancer.

Mechanistic insight. In addition to guiding targeted prevention strategies, G × E analysis has the potential to provide mechanistic insight into the complex multifactorial structure of the underlying biological pathway. One issue complicating observed gene–environment interactions of even confirmed susceptibility loci is that the true casual alleles and the biological impact of the variants are unknown. Our top interaction is between OCP use and rs13255292. This variant lies in the 8q24 region which harbors several risk loci for ovarian cancer[18] and other cancers.[43,44] The SNP is in the *PVT1* gene which interacts with the oncogene *MYC*.[45] *MYC* has long been reported to be at least in part under hormonal control[46,47] thus an interaction with OCP use is plausible. Conversely, our top additive interaction is between menopausal estrogen use and rs11658063 which falls in *HNF1B*. To our knowledge there is no relationship between *HNF1B* and hormones thus underscoring the difficulty of understanding these gene–environment interactions given our limited understanding of the function of the variants and even more broadly the biological role of the genes.

Exposure pathways and potential for targeted prevention. The strongest interactions are observed with OCP use or menopausal estrogen use which are modifiable exposures. Our most promising finding is the potential interaction between SNP rs13255292 and OCP use. This finding, if replicated could potentially lead to improved understanding of exposure pathways.

Analytic architecture and the choice of scale for measuring interaction. We present a comprehensive analytical framework to carry out post-GWAS G × E analysis on both multiplicative and additive scale. Our framework starting with data harmonization and imputation followed by global likelihood ratio tests and single df Wald tests provides a principled analytic architecture for such analysis. Our analysis reiterates the well-known fact that testing the additive and multiplicative nulls are very similar when the marginal associations are weak but could depart when both marginal associations are large in magnitude and the sample size is finite. In Supporting Information, eTable 5, we present observed joint odds ratios for strata defined by G and E along with the expected odds ratios under the multiplicative null and the additive null. We use our top hit rs13255292 and OCP use (ever *vs.* never) and length of OCP use (<1 year, 1- <5 years, 5+ years) as an illustration. One can note that the expected ORs are fairly close under both models. However, their estimated departure from the observed joint OR is more pronounced for the 1- <5 years sub-group when compared to 5+ years, explaining the suggestive evidence for rejecting the null.

We discussed the multiplicative interaction results for rs13255292 and OCP use in the previous section. We now explore the structure of additive interaction for this G × E result (Fig. 2*a,c*). Marginally, without including any genetic information, from a pure environmental association analysis we observed a relationship between duration of OCP use and risk reduction for ovarian cancer. For 1–5 years of OCP use (*vs.* <1 year) the estimated absolute risk difference was 0.47% (95%CI = 0.37% to 0.56%), while the estimated absolute risk difference for long-term use of OCPs (5+ year *vs.* <1 year) was 0.84% (95%CI = 0.77% to 0.92%) (Fig. 2*b,c,* Supporting Information, eTable 4), in agreement with previous findings that longer duration of OCP use is associated with larger risk reduction in ovarian cancer.[3] However, when stratified by rs13255292 genotype, we observed an interesting pattern. Among individuals with TT genotype, the corresponding absolute risk difference estimate for 1–5 year of OCP use (*vs.* <1 year) was 0.69% (95%CI = 0.49% to 0.88%), whereas among individuals with CC genotypes the corresponding risk reduction estimate was 0.36% (95%CI = 0.22% to 0.50%), implying potential effect modification by the C allele at locus rs13255292 (p value = 1.12×10^{-2}) (Fig. 2*b*, Supporting Information, eTable 4). In contrast, the absolute risk difference is

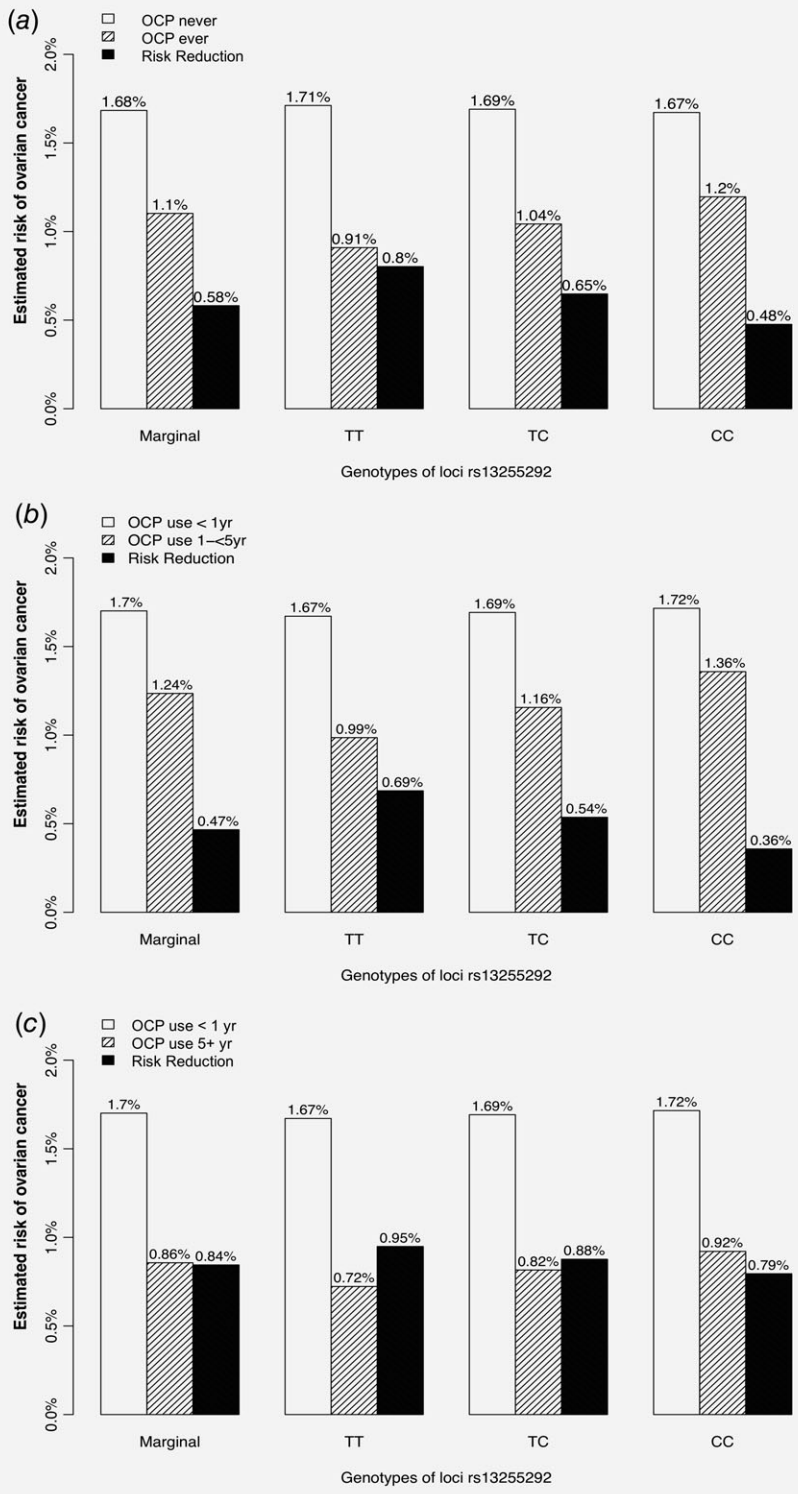

Figure 2. Estimated absolute risk (AR) of ovarian cancer given OCP use and number of copies of C allele, among non-Hispanic white college graduates aged below 50 with no family history of ovarian cancer, BMI below 25, no tubal ligation, no endometriosis, with one child. The ARs were calculated from a logistic regression model assuming log-additive effect of SNPs while all covariates fixed at their most frequent level as described above. (*a*) ARs stratified by OCP (ever *vs*. never) and genotype (*b*) ARs stratified by 1 to 5 years of OCP use (*vs*. <1 year) and genotype (F) ARs stratified by more than 5 years of OCP use (*vs*. <1 year) and genotype. Risk differences were also reported as the solid black bar.

estimated at 0.95% (95%CI = 0.78% to 1.12%) for women with TT genotype and at 0.79% (95%CI = 0.69% to 0.90%) in women with CC genotype. This indicates that longer OC use is associated with greater risk reduction overall and the risk reduction might be even greater for women with the TT genotype than those with the CC genotype. From Fig. 2b,c, we observe the interplay between "nature vs. nurture" with risk due to germline genetic mutations offset by long-term use of a modifiable protective factor. This analysis also highlights the benefit of measuring duration of exposure as opposed to a coarse indicator of ever/never use.

Prior work in G × E for ovarian cancer has focused solely on multiplicative interactions. We previously reported no departures from a multiplicative model with the first six risk loci identified through GWAS with a reduced set of exposures.[3] Follow-up work identified an interaction with menopausal estrogen therapy use and rs10069690 in the *TERT* gene,[48] but that finding was not replicated in the present analysis which included a larger set of studies. Fridley and colleagues have reported on G × E taking a candidate gene approach with several promising findings.[49] There are several studies in other cancers examining G × E on the multiplicative scale with limited success in identifying interactions, but to our knowledge, only prostate cancer and bladder cancer have been studied on the additive scale. In prostate cancer, suggestive additive interactions between vitamin D, confirmed genetic variants and risk have been identified.[50] In bladder cancer, additive interaction has been explored between confirmed genetic loci and smoking with risk of disease.[31] In this work the authors were able to demonstrate that the absolute risk of bladder cancer for current smokers varied from 2.9% to 9.9% based on the polygenic risk score quartile. These results are similar to our findings on the additive scale with absolute risk differing based on genetics and hormone therapy use; an interesting next step for our work is to consider the polygenic risk score for all of these confirmed ovarian cancer susceptibility alleles.

There are several limitations of the current analysis. Though we considered both multiplicative and additive interactions, the logistic model in (M1) is linear in covariates and exposures. We ignored potential non-linearity and exposure × exposure as well as exposure × covariate interactions. Similarly, we ignored any higher order interactions. A completely non-parametric machine learning approach, based on a recursive partition of the predictor space may avoid misspecification of the model, but would lack interpretability from an epidemiologic and public health perspective. We also acknowledge that this exploration of interaction is purely statistical, a more causal interpretation in a biological sense will require functional validation. One may also want to explore G × E interaction with loci that are not significant at genome-wide threshold but are significant at a less stringent threshold or even conduct genome-wide G × E scans.

The associations between ovarian cancer risk and some of the variants included here were limited to specific histotypes of ovarian cancer, however we have only presented results for all epithelial ovarian cancers combined. Developing histotype-specific risk stratification approaches is not feasible because for any given histotype the absolute risk is unlikely to ever reach an actionable threshold on a population level. In addition, risk reducing strategies are the same across histotypes and thus there is little benefit to considering histotype specific results from a precision prevention perspective. Heterogeneous associations between environmental risk factors and ovarian cancer risk by histology has previously been well characterized.[3,23,27] There is value in understanding histotype associations for disease etiology and mechanisms and this will be the focus of future work.

The analyses presented here offer insight into potential biological mechanisms, opportunities for ovarian cancer risk stratification, and approaches to studying gene–environment interactions. Ideally, replication for the six promising findings would be undertaken, but this is challenging with ovarian cancer given that most studies with the relevant data are included here. Functional studies for the regions harboring our most promising findings are underway and it is possible that the association described here may help inform those investigations [51]. Also, gene–environment interaction analyses can also be used to identify novel genetic associations[51] and thus a deeper evaluation of variants that are still borderline significant, but do not exactly achieve a genome-wide threshold is warranted for subsequent G × E analysis. Of particular interest will be to conduct risk stratification and risk prediction analysis using a summative polygenic risk score and to conduct an agnostic genome-wide search for G × E interaction. Despite the limitations the comprehensive framework of data harmonization, imputation, screening test followed by characterization of effect and risk estimates that has been used in this analysis can serve as a robust model for future gene–environment interaction analyses.

Acknowledgements

The AOCS also acknowledges the cooperation of the participating institutions in Australia and acknowledges the contribution of the study nurses, research assistants and all clinical and scientific collaborators to the study. The complete AOCS Study Group can be found at www.aocstudy.org. We would like to thank all of the women who participated in these research programs. The cooperation of the 32 Connecticut hospitals, including Stamford Hospital, in allowing patient access, is gratefully acknowledged. Our study was approved by the State of Connecticut Department of Public Health Human Investigation Committee. Certain data used in our study were obtained from the Connecticut Tumor Registry in the Connecticut Department of Public Health. The authors assume full responsibility for analyses and interpretation of these data. The German Ovarian Cancer Study (GER) thank Ursula Eilber for competent technical assistance. The NHS/NHSII studies thank the after state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY. We particularly thank I. Jacobs, M.Widschwendter, E. Wozniak, A. Ryan, J. Ford and N. Balogun for their contribution to the study. We also thank Harvey Risch for contributing data for this analysis. Usha Menon has shares in Abcodia Ltd., which has an interest in early cancer detection.

References

1. Society AC. Cancer Facts & Figures 2017. American Cancer Society 2017.
2. Jacobs IJ, Menon U, Ryan A, et al. Ovarian cancer screening and mortality in the UK collaborative trial of Ovarian Cancer screening (UKCTOCS): a randomised controlled trial. *Lancet (London, England)* 2016;387:945–56.
3. Pearce CL, Rossing MA, Lee AW, et al. Combined and interactive effects of environmental and GWAS-identified risk factors in ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2013;22:880–90.
4. Trabert B, Ness RB, Lo-Ciganic WH, et al. Aspirin, nonaspirin nonsteroidal anti-inflammatory drug, and acetaminophen use and risk of invasive epithelial ovarian cancer: a pooled analysis in the Ovarian Cancer association consortium. *J Natl Cancer Inst* 2014;106:djt431.
5. Sieh W, Salvador S, McGuire V, et al. Tubal ligation and risk of ovarian cancer subtypes: a pooled analysis of case-control studies. *Int J Epidemiol* 2013;42:579–89.
6. Falconer H, Yin L, Gronberg H, et al. Ovarian cancer risk after salpingectomy: a nationwide population-based study. *J Natl Cancer Inst* 2015; 107:dju410.
7. Lessard-Anderson CR, Handlogten KS, Molitor RJ, et al. Effect of tubal sterilization technique on risk of serous epithelial ovarian and primary peritoneal carcinoma. *Gynecol Oncol* 2014; 135:423–7.
8. Madsen C, Baandrup L, Dehlendorff C, et al. Tubal ligation and salpingectomy and the risk of epithelial ovarian cancer and borderline ovarian tumors: a nationwide case-control study. *Acta Obstet Gynecol Scand* 2015;94: 86–94.
9. Yoon SH, Kim SN, Shim SH, et al. Bilateral salpingectomy can reduce the risk of ovarian cancer in the general population: a meta-analysis. *Eur J Cancer* 2016;55:38–46.
10. Bojesen SE, Pooley KA, Johnatty SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* 2013;45: 371–84. 84e1-2.
11. Bolton KL, Tyrer J, Song H, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat Genet* 2010;42:880–4.
12. Goode EL, Chenevix-Trench G, Song H, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet* 2010;42:874–9.
13. Kar SP, Beesley J, Amin Al Olama A, et al. Genome-wide meta-analyses of breast, Ovarian, and prostate Cancer association studies identify multiple new susceptibility loci shared by at least two Cancer types. *Cancer Discov* 2016;6:1052–67.
14. Kelemen LE, Lawrenson K, Tyrer J, et al. Genome-wide significant risk associations for mucinous ovarian carcinoma. *Nat Genet* 2015;47:888–97.
15. Kuchenbaecker KB, Ramus SJ, Tyrer J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet* 2015;47:164–71.
16. Permuth-Wey J, Lawrenson K, Shen HC, et al. Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. *Nature communications* 2013;4: 1627.
17. Pharoah PD, Tsai YY, Ramus SJ, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet* 2013;45:362–70. 70e1-2.
18. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* 2017;49:680–91.
19. Shen H, Fridley BL, Song H, et al. Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. *Nat Commun* 2013;4:1628.
20. Song H, Ramus SJ, Tyrer J, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet* 2009;41: 996–1000.
21. Auranen A, Pukkala E, Makinen J, et al. Cancer incidence in the first-degree relatives of ovarian cancer patients. *Duodecim; laaketieteellinen aikakauskirja* 1997;113:46–50.
22. Stratton JF, Pharoah P, Smith SK, et al. A systematic review and meta-analysis of family history and risk of ovarian cancer. *Br J Obstet Gynaecol* 1998;105:493–9.
23. Lee AW, Ness RB, Roman LD, et al. Association between menopausal estrogen-only therapy and Ovarian carcinoma risk. *Obstet Gynecol* 2016;127: 828–36.
24. Pearce CL, Chung K, Pike MC, et al. Increased ovarian cancer risk associated with menopausal estrogen therapy is reduced by adding a progestin. *Cancer* 2009;115:531–9.
25. Collaborative Group On Epidemiological Studies Of Ovarian C, Beral V, Gaitskell K, et al. Menopausal hormone use and ovarian cancer risk: individual participant meta-analysis of 52 epidemiological studies. *Lancet (London, England)* 2015;385:1835–42.
26. Olsen CM, Nagle CM, Whiteman DC, et al. Obesity and risk of ovarian cancer subtypes: evidence from the Ovarian Cancer association consortium. *Endocr Relat Cancer* 2013;20:251–62.
27. Pearce CL, Templeman C, Rossing MA, et al. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol* 2012;13:385–94.
28. SEER Cancer Statistics Review 1975-2014, Based on November 2016 SEER data submission, posted to the SEER web site, April 2017.
29. Pearce CL, Stram DO, Ness RB, et al. Population distribution of lifetime risk of Ovarian Cancer in the United States. *Cancer Epidemiol Biomarkers Prev* 2015;24:671–6.
30. Liu G, Lee S, Lee AW, et al. Robust tests for additive gene-environment interaction in case-control studies using gene-environment Independence. *Am J Epidemiol* 2017;187:366–377.
31. Garcia-Closas M, Rothman N, Figueroa JD, et al. Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer Res* 2013;73:2211–20.
32. Knol MJ, Vanderweele TJ, Groenwold RH, et al. Estimating measures of interaction on an additive scale for preventive exposures. *Eur J Epidemiol* 2011;26:433–8.
33. Liu G, Mukherjee B, Lee S, et al. Robust tests for additive gene-environment interaction in case-control studies using gene-environment Independence. *Am J Epidemiol* 2018;187: 366–77.
34. Vanderweele TJ, Vansteelandt S. A weighting approach to causal effects and additive interaction in case-control studies: marginal structural linear odds models. *Am J Epidemiol* 2011;174:1197–203.
35. Vanderweele TJ. Sample size and power calculations for additive interactions. *Epidemiol Methods* 2012;1:159–88.
36. Bolton KL, Ganda C, Berchuck A, et al. Role of common genetic variants in ovarian cancer susceptibility and outcome: progress to date from the Ovarian Cancer association consortium (OCAC). *J Intern Med* 2012;271:366–78.
37. Little RJA, Rubin DB. *Chapter 10: Bayes and multiple imputation statistical analysis with missing data*, 2nd edn. NJ: John Wiley & Sons, 2002.
38. Nickels S, Truong T, Hein R, et al. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. *PLoS Genet* 2013;9: e1003284.
39. Bhattacharjee S Cn, Han S and Wheeler W. CGEN: An R package for analysis of case-control studies in genetic epidemiology, 2012.
40. Han SS, Rosenberg PS, Garcia-Closas M, et al. Likelihood ratio test for detecting gene (G)-environment (E) interactions under an additive risk model exploiting G-E independence for case-control data. *Am J Epidemiol* 2012;176:1060–7.
41. SEER, ed. *Cancer stat facts: Ovarian Cancer*. In: SEER, ed., vol. 2017. National Cancer Institute. Bethesda, MD.
42. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007;39:638–44.
43. Shi J, Zhang Y, Zheng W, et al. Fine-scale mapping of 8q24 locus identifies multiple independent risk variants for breast cancer. *Int J Cancer* 2016; 139:1303–17.
44. Tseng YY, Bagchi A. The PVT1-MYC duet in cancer. *Mol Cell Oncol* 2015;2:e974467.
45. Shang Y, Brown M. Molecular determinants for the tissue specificity of SERMs. *Science* 2002;295: 2465–8.
46. Wang C, Mayer JA, Mazumdar A, et al. Estrogen induces c-myc gene expression via an upstream enhancer activated by the estrogen receptor and the AP-1 transcription factor. *Mol Endocrinol* 2011;25:1527–38.
47. Lee AW, Bomkamp A, Bandera EV, et al. A splicing variant of TERT identified by GWAS interacts with menopausal estrogen therapy in risk of ovarian cancer. *Int J Cancer* 2016;139:2646–54.
48. Usset JL, Raghavan R, Tyrer JP, et al. Assessment of multifactor gene-environment interactions and Ovarian Cancer risk: candidate genes, obesity, and hormone-related risk factors. *Cancer Epidemiol Biomarkers Prev* 2016;25:780–90.
49. Dimitrakopoulou VI, Travis RC, Shui IM, et al. Interactions between genome-wide significant genetic variants and circulating concentrations of 25-Hydroxyvitamin D in relation to prostate Cancer risk in the National Cancer Institute BPC3. *Am J Epidemiol* 2017;185:452–64.
50. McAllister K, Mechanic LE, Amos C, et al. Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *Am J Epidemiol* 2017;186:753–61.