Assessment of moderate coffee consumption and risk of epithelial ovarian cancer: a Mendelian randomization study

Jue-Sheng Ong, Liang-Dar Hwang, Gabriel Cuellar-Partida, *(more OCAC authors here)*, Nicholas G. Martin, Georgia Chenevix-Trench, Michael C. J. Quinn, Marilyn C. Cornelis, Puya Gharahkhani, Penelope M. Webb, Stuart MacGregor

[Please refer to the attached author list document for the complete list of authors and affiliations.]

Number of Words (main text): 3 336 words

Abstract:

Background:

Coffee consumption has been shown to be associated with various health outcomes in observational studies. However, evidence for its association with epithelial ovarian cancer (EOC) is inconsistent and it is unclear whether these associations are causal.

Methods:

We used SNPs associated with (i) coffee and (ii) caffeine consumption to perform Mendelian randomisation on EOC risk. We conducted a two-sample MR using genetic data on 44,062 individuals of European ancestry from the Ovarian Cancer Association Consortium (OCAC) and combined instrumental variable estimates using a Wald-type ratio estimator.

Results:

For all EOC cases the causal odds ratio (COR) for genetically predicted consumption of one additional cup of coffee per day was 0.92 (95% confidence interval: 0.79, 1.06). The COR was 0.90 (95% CI: 0.73, 1.10) for high-grade serous EOC. The COR for genetically predicted consumption of an additional 80 mg caffeine was 1.01 (95% CI: 0.92, 1.11) for all EOC cases and 0.90 (95% CI: 0.73, 1.10) for high-grade serous.

Conclusion:

We found no evidence indicative of a strong association between EOC risk and genetically predicted coffee or caffeine levels. However, our estimates were not statistically inconsistent with earlier observational studies and we were unable to rule out small protective associations.

Key Message

- Evidence for association between coffee and ovarian cancer is inconsistent and it is unclear whether the relationship is causal
- Results from this study indicate no evidence for a strong causal association between coffee intake and ovarian cancer susceptibility.
- A subsequent analysis on caffeine intake also found no causal link between caffeine intake and ovarian cancer.
- The Mendelian randomization estimates were consistent to observational finding of non-causality, but are unable to rule out small protective effects.

Introduction:

Coffee is one of the most consumed beverages globally. A conventional cup of coffee can contain up to 1,000 types of bioactive compounds including various kinds of antioxidants, aromatic compounds and most importantly, caffeine. Caffeine has been found to suppress tumour growth in various animal models(1, 2), making it a potentially relevant therapeutic agent in cancer studies. Other compounds present in coffee are also found to have anti-inflammatory and anti-carcinogenic effects such as the induction of enzymes responsible for carcinogen detoxification, inhibition of carcinogen activation activities and stimulating intracellular antioxidant defence (1-3). Observational studies have investigated coffee and caffeine intake in relation to type 2 diabetes (4, 5), depression (6), insomnia (7) as well as various cancers (8, 9), but the directions of association have been inconsistent across diseases (10).

There are growing concerns regarding coffee consumption in relation to women's health. Epithelial ovarian cancer (EOC) is a gynaecological malignancy with a high fatality rate. Approximately 151 900 women worldwide die of the disease annually (11). The high-grade serous histology is the most

common EOC subtype (12). Whilst many individual studies have found conflicting directions of association with coffee consumption and EOC risk, subsequent meta-analysis studies found no evidence for an association (13-18). A more recent Danish study (19) suggested that moderate increase in daily caffeine intake (by one cup of coffee per day) might be protective against invasive EOC. Inconsistencies observed in the literature may be due to the lack of compatibility of categorical definitions (size of cup, content, caffeine intensity, method of brewing) and differences in definitions for baseline groups (i.e. non-drinkers). Some studies further combined consumption of tea and coffee to investigate caffeine intake specifically. However, more importantly, all studies to date examining the link between coffee/caffeine and EOC risk are observational studies where bias due to confounding may make it difficult to draw reliable conclusions (20). For example, we can hypothesize that women diagnosed with EOC may have temporal nutritional awareness and develop aversion to caffeinated beverages (such as coffee and cola), which may distort the true underlying association in case-control studies. Since randomized trials examining coffee consumption in relation to ovarian risk have not been conducted, to work around these potential biases, we can apply an instrumental variable technique, Mendelian randomization (MR) (21) to draw causal inferences on coffee consumption.

Twin studies have shown that coffee consumption has a substantial genetic component, with an estimated heritability ranging from 0.37 − 0.77 (22-24). This suggests that coffee consumption may be a suitable trait for MR studies. In this study we aim to refine the relationship between coffee and EOC susceptibility. We hypothesize that genetic predisposition towards higher coffee intake is inversely associated with i) overall EOC susceptibility and ii) high-grade serous EOC susceptibility, and draw inference on causality via MR.

Methods:

Data source

Participants for this study were drawn from the Ovarian Cancer Association Consortium (OCAC). Genotyping was performed using the customised Infinium OncoArray-500K array (Illumina) (25) consisting of ~322 000 variants. OncoArray data were available for 59 115 samples across 71 study cohorts worldwide, of which 56 479 passed initial quality control protocols. Each individual was assigned values to indicate the proportion of European, African or Asian ancestry they inherited based on genetic makeup, using principal component analysis. These values sum up to 1 and are used to categorise the subjects into one of the intercontinental ancestry groups. Following that, imputation into the 1000 Genomes Project reference panel was carried out with pre-phasing using SHAPEIT and IMPUTE2 (26, 27). First-degree related individuals and duplicated samples (n=1 732) were removed. DNA samples from women of non-European ancestry were excluded for this study. The total sample size used in this study was 44 062 women of European ancestry (Table 1 shows a breakdown of the sample size by EOC histology). Baseline characteristics of our study samples from OCAC according to weight, age, smoking status and other potential confounders are summarised in Supplementary Table 1.

Genetic variants for the MR analyses were identified through an extensive review of published GWAS findings for coffee, tea and/or caffeine consumption (28-33). SNPs associated with coffee consumption (measured as cups/day) that were considered for use were rs1481012 in the *ABCG2* gene, rs6968554 in the *AHR* gene, rs2470893 in the *CYP1A2* gene, rs17685 in the *POR* gene and rs6265 in the *BDNF* gene. In our subsequent analysis, we investigated whether the association with coffee intake (in cups per day) onto ovarian cancer was driven mainly by genetic predisposition for altered caffeine intake. SNPs reported to show association with caffeine and considered for use here were rs6968865 from the *AHR* gene and rs2472297 from the *CYP1A2* gene. All of the SNPs investigated were either directly genotyped or imputed with high quality (info-score > 0.9). Although

these variants are different SNPs in AHR and CYP1A2, they are in high linkage-disequillibrium (r^2=0.8), see discussion for more detail. In order to ensure that our SNPs of interest are strong instruments, we examined the statistical evidence in the literature for their association with coffee and with caffeine consumption respectively. The variance on coffee consumption explained by a particular SNP can be derived using $r^2_{SNP} = 2p(1-p)\beta^2/\sigma^2$ where r^2_{SNP} refers to the variance explained by the SNP, p refers to the MAF of the SNP, β is the measured magnitude of association per effect allele and $\sigma^2$is the coffee trait variance. The variance explained by our SNP instruments can hence be obtained by linearly summing up r^2_{SNP} across each independent SNP instrument. We subsequently tested each SNP against several potential confounders. For each of age at menarche, measures of glycaemia, education attainment, BMI, waist-hip ratio, body fat and smoking behaviour, we extracted previously published results from publicly available GWAS datasets (full details plus references in supplementary table 3).

Causal Effect estimation

To perform MR we utilised a two sample statistical model to estimate the magnitude of association between coffee consumption and ovarian cancer using summary statistics (34). We fitted an additive model in SNPTEST (35) to test for association between each SNP and ovarian cancer status. Within-ancestry principal components (PC1-PC9) were fitted to remove potential bias arising from intra-ethnic population difference. Additional covariates that might be confounders such as BMI, smoking status and alcohol consumption were not available for all the genotyped OCAC participants and hence were not included as covariates (although subject to the assumptions of MR, not including these potential confounders as covariates will not bias our results) to maximize sample size. The genomic control lambda value was computed using 483 972 SNPs genome-wide to assess the possibility of population stratification biasing the association between allele frequencies and phenotype.

For both coffee and caffeine consumption we used the Wald-type ratio estimator (36) to combine the SNP-estimates which uses the SNP-risk factor and SNP-cancer magnitude of association estimates to calculate the aggregated causal effect. We estimated a causal OR (COR) for all ovarian cancer and for the high grade serous subtype. High-grade serous was the only histological subtype with sufficient numbers for sub-set analysis.

Results:

SNP Selection

We shortlisted a total of 4 independent SNPs (rs1481012, rs6968554, rs2470893, rs17685) as proxies for genetically determined coffee consumption behaviour (31). For the analysis on caffeine, we used 2 SNPs (rs6968865,rs2472297) (33) as genetic proxies for total caffeine consumption per day (in mg). Each of these SNPs is robustly associated with p-values less than $p<5 \times 10^{-8}$ for coffee consumption in the original coffee GWAS. Due to the smaller sample size in the published analysis for caffeine consumption, the published p-values for the effects of rs6968865 and rs2472297 on caffeine consumption were not as strong as those for the SNP-coffee associations but both of the SNPs combined associate with caffeine consumption with a p-value=3.74×10^{-14} (33), with its direction of association verified in an Australian sample (Supplementary A1). Each of the SNPs thus satisfies the strong MR instrument criterion (F>>10).

In our pleiotropy assessment, the SNP rs6265 in the *BDNF* gene was found to have pleiotropic effects on other traits of relevance to ovarian cancer (BMI and age of menarche, supplementary material) so it was excluded from our analyses. After removing BDNF, the 4 coffee SNPs combined explain about ~1.2% of the variation in coffee intake (31), whereas the 2 SNPs combined for our MR caffeine study explain about ~1.3% of the variation in caffeine intake (33). We also tested the association

between established ovarian cancer risk factors (oral contraceptive use, estrogen use, parity) and our SNPs of interest. The results of our pleiotropy assessment are available in Supplementary Table 3(publicly available GWAS) and Supplementary Table 4 (OCAC dataset). In brief, no associations were found above chance level and we conclude that the assumptions of no-pleiotropy is not violated. In particular, coffee consumption and cigarette consumption are correlated in some populations but our chosen SNPs are not associated with smoking (Supplementary Table 3).

Instrumental variable analysis

The SNP-cancer association results for each genetic instrument used are available in Supplementary Table 2. We estimated the causal odds ratio associated with a genetically predicted one cup per day change in coffee consumption. For all EOC cases the COR for consuming one additional cup of coffee per day was 0.92 (95% confidence interval, CI: 0.79, 1.06). For high-grade serous EOC, the COR was 0.90 (95% CI: 0.73, 1.10). We also performed an additional analysis to investigate caffeine consumption, with the COR scaled in terms of an 80mg increase (the approximate caffeine content in a conventional cup of coffee). The COR for consuming an additional 80mg of caffeine was 1.01 (CI: 0.92, 1.11) for all EOC cases and 0.90 (CI: 0.73, 1.10) for high-grade serous. The CORs derived from individual SNP instruments are shown in Figure 1 for coffee consumption; and Figure 2 for caffeine intake.

Population Stratification and confounding

Due to the missing covariate data on some OCAC participants (see Supplementary Table 1), the analyses were performed by only fitting the first 9 genetic (ancestral) principal components as covariates. In a sensitivity analysis using participants with confounder data available (n~11 400), adjustment for potential confounders (age of menarche, education level, number of pregnancies,

oral contraceptive use, estrogen use, smoking and BMI) did not change the magnitude of the SNP-disease associations (See Supplementary Table 6). The genomic control lambda was 1.076 ($\lambda_{1000} = 1.007$, LD-score intercept=1.032) demonstrating that there is little evidence for inflation of the genome-wide association statistics due to population stratification. Plots of the ancestral principal components (PC1 against PC2) between cases and controls indicate that the cases and controls are homogeneous (See Supplementary Figure 1 and 2).

Discussion:

In our study sample of 44 062 European participants from OCAC, we found no evidence suggestive of a large causal association between (genetically predicted) coffee consumption and overall EOC risk nor on high-grade serous EOC. Similarly, our findings consistently suggest no causal link between caffeine intake and EOC susceptibility.

Research in context

Most epidemiological studies in the past investigated the association of EOC with coffee consumption by assessing the difference in EOC risk among non-coffee drinkers and strong coffee drinkers. Consumption of > 3 cups of coffee per day was used as a benchmark to indicate strong coffee drinking behaviour. To compare our results, we rescaled findings from these observational studies to reflect an averaged moderate change in daily coffee consumption (1 cup of coffee per day) using Equation 1 in Supplementary material. The resultant estimates from our study were broadly compatible with results of previous meta-analyses (Figure 3).

Although some individual observational studies have found associations between coffee consumption and risk of EOC, meta-analyses have found no evidence to show that coffee consumption protects against EOC (13). However, a common criticism of observational studies is inconsistency in the definition of categorised consumption (i.e. different studies adopt different

definitions of heavy drinkers) and the variability in types of coffee beverages, which may differ strongly in terms of nutritional content (most importantly, caffeine). These systematic differences can make the interpretation of meta-analysed findings difficult. Moreover, it is difficult to rule out the potential effects of selection bias in case-control studies and of unmeasured or uncontrolled confounding in observational studies in general. In contrast, here we use genetically predicted coffee intake to provide more uniform estimates of coffee consumption in a large sample size (coffee GWAS (31), n>80 000). Our 2-sample MR design allows us to investigate the underlying association without the issue of potential confounders such as education level, alcohol use and smoking behaviour, which was established by earlier studies to be strongly correlated to coffee consumption. In our pleiotropy assessment, the SNP instruments we employ are not associated with these potential confounders (Supplementary table 3).

Even though the MR analyses were performed separately for coffee consumption and caffeine intake with independent SNPs within each study, the inference we draw from these findings are not independent. This is due to the fact that, for each study the most important single SNPs (rs2470893 in *CYP1A2* which explains ~0.5% of the variance in coffee consumption (31) and rs2472297 in *CYP1A2* which explains ~0.8% of the variance in caffeine consumption (33)) are in high linkage disequilibrium (r^2=0.7 between the two SNPs). Hence, the effect of those SNPs (rs2470893, rs2472297) on coffee and caffeine consumption may not be separable (i.e. *CYP1A2* is involved in metabolizing common bioactive compounds in coffee). The same applies for SNP rs6968865 and rs6968554 in *AHR*.

Previous studies have highlighted a potential role of caffeine in inducing p53-dependent (tumour suppression gene) apoptosis (37). Since *TP53* mutations are found in almost all high-grade serous EOC (38), an analysis of high-grade serous alone was of particular interest. However in our study, coffee and caffeine intake did not appear to be associated with any risk of high-grade serous

carcinoma among Europeans.

Strengths and limitations

One of the strength of our study is that participants used in our analyses were all of European ancestry, limiting potential bias due to population heterogeneity. Furthermore, the use of ancestral principal components to define ethnicity also prevents heritage-reporting errors (i.e. ethnicity was determined based on SNP profiles, as summarised by ancestry principal components to avoid self-reporting biases). In our MR study, the use of GWAS findings to predict coffee/caffeine consumption rather than relying on self-reports of consumption should remove misclassification biases that can plague self-reported studies and contribute to statistical heterogeneity in meta-analyses of observational studies. Since coffee consumption generally stabilizes during adulthood, our 2-sample MR approach is protected by potential biases due to apparent age differences between the SNP-coffee samples and the OCAC samples. In other words, the estimated SNP-coffee association during adulthood remain a robust genetic predisposition to lifetime coffee intake behaviour.

For our MR to infer about causality, several MR assumptions have to be met. Firstly, the instruments (SNPs) used here were robustly associated (with F>>10) to coffee and caffeine intake respectively. Secondly, he SNPs used in this study showed no evidence for any pleiotropic effects that may confound the association with EOC susceptibility. The third MR assumption, that the genetic variants used in our study only influence EOC susceptibility through mediating coffee consumption, can be difficult to test directly. However, previous studies have examined the role of *CYP1A1*, *CYP1A2* and *AHR* in detail (28, 29, 32, 39). In each case, a SNP in or near the gene has been implicated by GWAS and we assume that the action of the SNP on coffee consumption is via the specified gene. Taking each in turn, *CYP1A2* encodes the primary enzyme that metabolises caffeine in the liver, while *CYP1A1* encodes protein that metabolises polycyclic aromatic hydrocarbons, which are more

commonly found in coffee beans. The *AHR* gene is known to induce both *CYP1A1* and *CYP1A2* via a DNA binding mechanism (29) and is also responsible in detection of toxic chemicals (39). Despite coffee intake being strongly correlated with smoking, our pleiotropy assessment indicated that none of the SNPs appear to be associated (Bonferroni corrected p-value > 0.05) with smoking behaviours. Moreover, the lack of a main effect of the SNPs on smoking makes a coffee-smoking interaction less likely - Thus, it seems very improbable that these SNPs directly influence ovarian cancer through other independent biological processes.

Although we found no evidence supportive of an association between the SNPs used and common risk factors for EOC (40, 41)(e.g. smoking, oral contraceptive use, parity, etc.), it is hard to rule out directly possibilities of residual pleiotropy. However, suppose that a SNP has a strong pleiotropic effect which biases our results - for us to observe the null causal odds ratio we find here, the other SNPs (or some combination of SNPs) must act pleiotropically in the opposite direction and with similar magnitude to the first SNP. Since this is unlikely, it is unlikely that pleiotropic effects have a considerable influence on our non-causality conclusion.

There are some limitations that should also be considered in our analyses. Firstly, our study was performed using only European ancestry women and our findings may not generalize to other populations. Even though our SNPs greatly exceed the traditional strong instruments criteria (F>10), our SNPs combined only account for a relatively small proportion of variation (~1.2%) in coffee consumption (cups per day), potentially leading to problems in power when applying MR. With a relatively small proportion of variance explained, we must extrapolate from small changes in predicted coffee consumption. If the sample size in our data linking genotype to ovarian cancer risk were small, the overall estimates of the causal odds ratio would be too large to be useful. However, as we have available a large dataset from an international consortium, the overall standard error in our causal odds ratio is relatively small, allowing us to make clear statements on the likely limits of

the causal effect of coffee consumption on ovarian cancer (e.g. for all histologies the causal OR is 0.92 with 95% confidence interval 0.79, 1.06).

The precision of our estimates is good for the most common subtype high grade serous (Causal OR 0.90 with 95% CI: 0.73, 1.10) but for the less common subtypes taken individually our power is low; we similarly have insufficient power to perform stratified analyses (e.g. based on groups with particular smoking or BMI status).

The difference in coffee consumption as quantified in our MR analysis can be hard to interpret. In our analysis, CORs are calculated based on one additional cup of coffee per day averaging across all possible quantities of coffee consumption among regular coffee drinkers (including non-drinkers). This made it difficult to compare our estimates reliably with those from studies that investigated extreme ends of the trait distribution (heavy coffee drinking (>5 cups) and/or coffee drinkers to non-drinkers). Here, it is difficult for our study to completely rule out previous findings that showed positive associations of EOC when comparing very heavy coffee-drinkers to other categories (13). That is, our findings only infer that moderate differences in coffee consumption (averaging over the entire trait distribution) do not influence risk of EOC as the MR framework assumes that modifiable exposures linearly affect the underlying risk factor; which might be violated if the outcome to exposure relationship is non-linear (follows a J-shaped curve).

An additional consideration is how to handle non-coffee drinkers. For caffeine this is not an issue because non-users are included in the SNP association studies. For coffee consumption, in our main analysis, we focus on “cups per day” coffee consumption. However, the GWASs to date on “cups per day” in coffee consumers also found (31) that the same SNPs were also strongly associated with drinking status (“high” versus “low/no” coffee consumption). Hence our findings in support of non-causality of “cups per day” probably extend to alternative definitions such as “high” versus “low/no”

status.

We found no evidence indicative of a strong association between EOC risk and genetically predicted coffee or caffeine levels. However, our estimates were not statistically inconsistent with earlier observational studies and we were unable to rule out small protective associations. Our MR based results are more readily interpretable than previous observational studies because they are unlikely to be adversely affected by confounding biases which can invalidate the conclusions from observational studies.

Funding

This work was supported by The COGS project which is funded through a European Commission' s Seventh Framework Programme Grant (agreement number 223175 HEALTH F2 2009-223175); the Genetic Associations and Mechanisms in Oncology (GAME- ON): A NCI Cancer Post-GWAS Initiative (U19-CA148112); the Ovarian Cancer Association Con-sortium is supported by a grant from the Ovarian Cancer Re-search Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PPD/RPCI.07). Funding of the constituent studies was provided by the California Cancer Research Program (00-01389V-20170, N01-CN25403, 2II0200); the Canadian Institutes of Health Research (MOP-86727); Cancer Australia; Cancer Council Victoria; Cancer Council Queensland; Cancer Council New South Wales; Cancer Council South Australia; Cancer Council Tasmania; Cancer Foundation of Western Australia; the Cancer Institute of New Jersey; Cancer Research UK (C490/A6187, C490/A10119, C490/A10124); the Danish Cancer Society (94-222-52); the ELAN Program of the University of Erlangen-Nuremberg; the Eve Appeal; the Helsinki University Central

Hospital Research Fund; Helse Vest; grants from the Swedish Cancer Society and European Research Concil Advanced Grant ERC-2011-294576; the Norwegian Cancer Society; the Norwegian Research Council; the Ovarian Cancer Research Fund; Nationaal Kankerplan of Belgium; the L & S Milken Foundation; the Polish Ministry of Science and Higher Education (4 PO5C 028 14, 2 PO5A 068 27); the Roswell Park Cancer Institute Alliance Foundation; Breast Cancer Now, the Institute of Cancer Research, NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre; the US National Cancer Institute (K07-CA095666, K07-CA80668, K07-CA143047, K22-CA138563, N01-CN55424, N01-PC67001, N01-PC067010, N01-PC035137, P01-CA017054, P01-CA087696, P30-CA072720, P30-CA15083, P30-CA008748, P50-CA159981, P50-CA105009, P50-CA136393, R01-CA149429, R01-CA014089, R01-CA016056, R01-CA017054, R01-CA049449, R01-CA050385, R01-CA054419, R01-CA058598, R01-CA058860, R01-CA061107, R01-CA061132, R01-CA063678, R01-CA063682, R01-CA067262, R01-CA071766, R01-CA074850, R01-CA080978, R01-CA083918, R01-CA087538, R01-CA092044, R01-CA095023, R01-CA122443, R01-CA112523, R01-CA114343, R01-CA126841, R01-CA136924, R03-CA113148, R03-CA115195, U01-CA069417, U01-CA071966, UM1-CA186107, UM1-CA176726 and Intramural research funds); the NIH/National Center for Research Resources/General Clinical Research Center (MO1-RR000056); the US Army Medical Research and Material Command (DAMD17-01-1-0729, DAMD17-02-1-0666, DAMD17-02-1-0669, W81XWH-07-0449, W81XWH-10-1-02802); the US Public Health Service (PSA-042205); the National Health and Medical Research Council of Australia (199600 and 400281); the German Federal Ministry of Education and Research of Germany Programme of Clinical Biomedical Research (01GB 9401); the State of Baden-Wurttemberg through Medical Faculty of the University of Ulm (P.685); the German Cancer Research

Center; the Minnesota Ovarian Cancer Alliance; the Mayo Foundation; the Fred C. and Katherine B. Andersen Foundation; the Lon V. Smith Foundation (LVS-39420); the Oak Foundation; Eve Appeal; the OHSU Foundation; the Mermaid I project; the Rudolf-Bartling Foundation; the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge, Imperial College London, University College Hospital 'Womens' Health Theme' and the Royal Marsden Hospital; and Work Safe BC 14.

Investigator-specific funding: JS and GC thank the University of Queensland and QIMR Berghofer Medical Research Institute for scholarship support. GCT and PMW are supported by the National Health and Medical Research Council (NHMRC). SM acknowledges funding support from an Australian Research Council Future Fellowship. This work is also supported by the NHMRC project grant 1123248.

Acknowledgements

We acknowledge with appreciation all the individuals who participated in the OCAC studies, and the many hospital directors and staff, gynaecologists, general practitioners and pathology services who provided assistance with confirmation of diagnoses, and the many research assistants and interviewers for assistance with the studies. For the OCAC studies, we thank: D.Bowtell, A. deFazio, D. Gertig, A. Green, P. Parsons, N. Hayward, and D. Whiteman (AUS); G. Peuteman, T. Van Brussel and D. Smeets (BEL); L. Gacucova (HMO); P. Schurmann, F. Kramer, W. Zheng, T. W. Park, Simon, K.

Beer-Grondke and D. Schmidt (HJO); S. Windebank, C. Hilker and J. Vollenweider (MAY); the state cancer registries of AL, AZ, AR, CA, CO, CT, DE, FL, GA, HI,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 WYL(NHS); L. Paddock, M. King, L. Rodriguez-Rodriguez, A. Samoila and Y. Bensman (NJO); M. Sherman, A. Hutchinson, N. Szeszenia-Dabrowska, B. Peplonska, W. Zatonski, A. Soni, P. Chao and M. Stagner (POL); C. Luccarini, P. Harrington the SEARCH team and ECRIC (SEA); I. Jacobs, M. Widschwendter, E. Wozniak, N. Balogun, A. Ryan, C. Karpinskyj and J. Ford (UKO); Carole Pye (UKR); A. Amin Al Olama, J. Dennis, E. Dicks, K. Michilaidou and K. Kuchenbaker (COGS).

Conflict of Interests

Mark T Goodman is a consultant for Johnson and Johnson Ltd. Usha Menon has stock ownership and research funding from Abcodia Ltd, a UCL spin-out company with interest in biomarkers and ovarian cancer screening. Peter Fasching conducts research with grants from Amgen and Novartis, and received honoraria from Amgen, Novartis, Roche, Celgene, Nanostring, Genomic Health and TEVA. James D. Brenton hold stock in Inivata, the makers of DNA assays, but declare no conflict of interest on this work. All remaining authors declare no competing interest.

Reference

1. Huber WW, Rossmanith W, Grusch M, Haslinger E, Prustomersky S, Peter-Vorosmarty B, et al. Effects of coffee and its chemopreventive components kahweol and cafestol on cytochrome P450 and sulfotransferase in rat liver. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association. 2008;46(4):1230-8.
2. Huber WW, Scharf G, Nagel G, Prustomersky S, Schulte-Hermann R, Kaina B. Coffee and its chemopreventive components Kahweol and Cafestol increase the activity of O6-methylguanine-DNA

methyltransferase in rat liver--comparison with phase II xenobiotic metabolism. Mutation research. 2003;522(1-2):57-68.
3. Boettler U, Volz N, Teller N, Haupt LM, Bakuradze T, Eisenbrand G, et al. Induction of antioxidative Nrf2 gene transcription by coffee in humans: depending on genotype? Molecular biology reports. 2012;39(6):7155-62.
4. Jiang X, Zhang D, Jiang W. Coffee and caffeine intake and incidence of type 2 diabetes mellitus: a meta-analysis of prospective studies. European journal of nutrition. 2014;53(1):25-38.
5. Lee JK, Kim K, Ahn Y, Yang M, Lee JE. Habitual coffee intake, genetic polymorphisms, and type 2 diabetes. European journal of endocrinology / European Federation of Endocrine Societies. 2015;172(5):595-601.
6. Lucas M, Mirzaei F, Pan A, Okereke OI, Willett WC, O'Reilly EJ, et al. Coffee, caffeine, and risk of depression among women. Archives of internal medicine. 2011;171(17):1571-8.
7. Skarupke C, Schlack R, Lange K, Goerke M, Dueck A, Thome J, et al. Insomnia complaints and substance use in German adolescents: did we underestimate the role of coffee consumption? Results of the KiGGS study. Journal of neural transmission. 2015.
8. Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. The New England journal of medicine. 2012;366(20):1891-904.
9. Crippa A, Discacciati A, Larsson SC, Wolk A, Orsini N. Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: a dose-response meta-analysis. American journal of epidemiology. 2014;180(8):763-75.
10. Schoenfeld JD, Ioannidis JP. Is everything we eat associated with cancer? A systematic cookbook review. Am J Clin Nutr. 2013;97(1):127-34.
11. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: a cancer journal for clinicians. 2015;65(2):87-108.
12. Koti M, Siu A, Clement I, Bidarimath M, Turashvili G, Edwards A, et al. A distinct pre-existing inflammatory tumour microenvironment is associated with chemotherapy resistance in high-grade serous epithelial ovarian cancer. British journal of cancer. 2015;112(7):1215-22.
13. Braem MG, Onland-Moret NC, Schouten LJ, Tjonneland A, Hansen L, Dahm CC, et al. Coffee and tea consumption and the risk of ovarian cancer: a prospective cohort study and updated meta-analysis. Am J Clin Nutr. 2012;95(5):1172-81.
14. Steevens J, Schouten LJ, Verhage BA, Goldbohm RA, van den Brandt PA. Tea and coffee drinking and ovarian cancer risk: results from the Netherlands Cohort Study and a meta-analysis. British journal of cancer. 2007;97(9):1291-4.
15. Song YJ, Kristal AR, Wicklund KG, Cushing-Haugen KL, Rossing MA. Coffee, tea, colas, and risk of epithelial ovarian cancer. Cancer Epidemiol Biomarkers Prev. 2008;17(3):712-6.
16. Lueth NA, Anderson KE, Harnack LJ, Fulkerson JA, Robien K. Coffee and caffeine intake and the risk of ovarian cancer: the Iowa Women's Health Study. Cancer causes & control : CCC. 2008;19(10):1365-72.
17. Silvera SA, Jain M, Howe GR, Miller AB, Rohan TE. Intake of coffee and tea and risk of ovarian cancer: a prospective cohort study. Nutrition and cancer. 2007;58(1):22-7.
18. Goodman MT, Tung KH, McDuffie K, Wilkens LR, Donlon TA. Association of caffeine intake and CYP1A2 genotype with ovarian cancer. Nutrition and cancer. 2003;46(1):23-9.
19. Gosvig CF, Kjaer SK, Blaakaer J, Hogdall E, Hogdall C, Jensen A. Coffee, tea, and caffeine consumption and risk of epithelial ovarian cancer and borderline ovarian tumors: Results from a Danish case-control study. Acta oncologica. 2015;54(8):1144-51.
20. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. Bmj. 1998;316(7125):140-4.
21. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63.

22. Vink JM, Staphorsius AS, Boomsma DI. A genetic analysis of coffee consumption in a sample of Dutch twins. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2009;12(2):127-31.
23. Laitala VS, Kaprio J, Silventoinen K. Genetics of coffee consumption and its stability. Addiction. 2008;103(12):2054-61.
24. Luciano M, Kirk KM, Heath AC, Martin NG. The genetics of tea and coffee drinking and preference for source of caffeine in a large community sample of Australian twins. Addiction. 2005;100(10):1510-7.
25. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nature genetics. 2017.
26. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature genetics. 2012;44(8):955-9.
27. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS genetics. 2014;10(4):e1004234.
28. Josse AR, Da Costa LA, Campos H, El-Sohemy A. Associations between polymorphisms in the AHR and CYP1A1-CYP1A2 gene regions and habitual caffeine consumption. Am J Clin Nutr. 2012;96(3):665-71.
29. Amin N, Byrne E, Johnson J, Chenevix-Trench G, Walter S, Nolte IM, et al. Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. Molecular psychiatry. 2012;17(11):1116-29.
30. Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, Bennett SN, et al. Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. PLoS genetics. 2011;7(4):e1002033.
31. Coffee, Caffeine Genetics C, Cornelis MC, Byrne EM, Esko T, Nalls MA, et al. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. Molecular psychiatry. 2015;20(5):647-56.
32. Sulem P, Gudbjartsson DF, Geller F, Prokopenko I, Feenstra B, Aben KK, et al. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. Human molecular genetics. 2011;20(10):2071-7.
33. McMahon G, Taylor AE, Davey Smith G, Munafo MR. Phenotype refinement strengthens the association of AHR and CYP1A1 genotype with caffeine consumption. PloS one. 2014;9(7):e103448.
34. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. European journal of epidemiology. 2015;30(7):543-52.
35. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nature genetics. 2007;39(7):906-13.
36. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic epidemiology. 2013;37(7):658-65.
37. He Z, Ma WY, Hashimoto T, Bode AM, Yang CS, Dong Z. Induction of apoptosis by caffeine is mediated by the p53, Bax, and caspase 3 pathways. Cancer research. 2003;63(15):4396-401.
38. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609-15.
39. Lensu S, Tuomisto JT, Tuomisto J, Viluksela M, Niittynen M, Pohjanvirta R. Immediate and highly sensitive aversion response to a novel food item linked to AH receptor stimulation. Toxicology letters. 2011;203(3):252-7.
40. Kurman RJ, Shih Ie M. Pathogenesis of ovarian cancer: lessons from morphology and molecular biology and their clinical implications. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists. 2008;27(2):151-60.

41. Salehi F, Dunfield L, Phillips KP, Krewski D, Vanderhyden BC. Risk factors for ovarian cancer: an overview with emphasis on hormonal factors. J Toxicol Environ Health B Crit Rev. 2008;11(3-4):301-21.

Supplementary Material

Equation 1

Let N be the number of ordered consumption categories, and O_i be the Odds ratio of the i-th category to the lowest category.

We have that $O_1 = 1$ (trivial). Then the averaged change in risk (Odds ratio) per additional cup of coffee per day, O_{Ave} is given by

$$O_{Ave} = exp(\frac{1}{N-1}\sum_{i=2}^{N}\frac{\log(O_i)}{i-1})$$

A1 Exploring the role of SNP instruments in caffeine consumption

It is important to note that the sample size of the genetic association study for caffeine intake (1) were much smaller than those in the published coffee GWAS. Although we opted to characterize the effect of our chosen instrumental variable SNPs on genetically predicted caffeine consumption using published data, we also confirmed the role these SNPs play using data on directly measured (self-reported) caffeine intake (through tea, cola, chocolate and coffee) from one of the studies participating in OCAC - 2,347 participants from the Australian Ovarian Cancer Study (see Supplementary Table 5). The results for the caffeine SNPs (rs6968865 and rs2472297) were consistent with the published findings (1). This serves as a validation of instrument strength, providing reassurance that the pattern of association of these SNPs is consistent across studies and that the results from our two-sample MR approach are robust (i.e. for the scenario where the SNP-caffeine associations come from a different sample than SNP-disease associations).

Table 1: Baseline Characteristics of OCAC Participants

	Eur Controls	Eur Cases
Participants	23,379	20,683
Age	56(47,64)	58(49,66)
Height	1.63(1.60,1.68)	1.63(1.60,1.68)
Weight(1yr ago)	68.2(60.3,80.0)	69.0(60.3,81.6)
Age at menarche	13(12,14)	13(12,14)
Missing	12,077	7,790
Highest level of education		
Less than high school	1,659	1,516
High school or more	9,278	9,678
Missing	12,442	9,489
Pregnancy		
Ever pregnant	12,276	11,117

Never pregnant	1,422	2,345
Median number of pregnancy	2(2,4)	2(1,3)
Missing	9,681	7,221
Median number of fullbirths	2(1,3)	2(1,3)
Smoking		
Current smoker	1,242	1,669
Former smoker	3,324	3,509
Never smoked	5,327	5,774
Missing	13,486	9,731
Oral contraceptive (OC)		
Ever used OC	9,790	8,199
Median total months of OC use	36(0,102)	12(0,72)
Never used OC	3,720	5,198

Missing	9,869	7,286
Estrogen		
Ever used estrogen	1,438	1,323
Never used	7,296	7,160
Missing	14,645	12,200

Table 2: Association of SNPs to risk of all EOC and High-grade serous EOC (HS).

Study	Gene	SNP ID	Risk Allele	Other Allele	Exposure	S.E.	Unit	Effect on EOC	S.E.	Effect on HS	S.E.
Coffee Consumption								Europeans (n=44,062)		Europeans (n=30,867)	
Cornelis et al.	ABCG2	rs1481012	A	G	0.06	0.01	cups/ day	0.019	0.022	0.041	0.031
Cornelis et al.	AHR	rs6968554	G	A	0.13	0.01	cups/day	-0.020	0.014	-0.032	0.02
Cornelis et al.	CYP1A1	rs2470893	T	C	0.12	0.01	cups/ day	0.011	0.015	0.017	0.021
Cornelis et al.	POR	rs17685	A	G	0.07	0.01	cups/ day	-0.033	0.015	-0.043	0.021
McMahon et al.	AHR	rs6968865	T	A	14.6	3.1	caffeine per day (mg)	-0.018	0.014	-0.032	0.02
McMahon et al.	CYP1A2	rs2472297	T	C	21.4	3.4	caffeine per day (mg)	0.012	0.017	0.025	0.023

Table 3: Pleiotropy assessment – Association of coffee/caffeine genetic variants with potential confounders through publicly available GWAS datasets

Trait	GWAS Consortia - variable	PubMed ID	Sample size	SNP ID	Risk Allele	Other Allele	Magnitude of Association	P-Value
Menarche (2)	ReproGen – Age at menarche	21102462	87,802	rs2470893	T	C	-0.004	5.30E-01
				rs2472297	T	C	-0.007	3.30E-01
				rs6968554	A	G	0.001	8.40E-01
				rs6968865	A	T	0.003	6.40E-01
				rs17685	A	G	0.012	1.10E-01
				rs6265	T	C	0.043	**9.30E-09**
Measures of Glycaemia (3)	MAGIC - Fasting glucose	20081858	46,186	rs2470893	T	C	-0.006	0.09935
				rs2472297	T	C	-0.009	0.03103
				rs6968554	A	G	0.004	0.2863

				rs6968865	A	T	0.004	0.2591
				rs17685	A	G	-0.005	0.2505
				rs6265	T	C	0.004	0.302
	MAGIC - Fasting insulin	20081858	46,186	rs2470893	T	C	-0.003	0.3833
				rs2472297	T	C	-0.004	0.257
				rs6968554	A	G	0.000	0.9575
				rs6968865	A	T	0.000	0.8843
				rs17685	A	G	0.002	0.622
				rs6265	T	C	0.004	0.2265
Education level (4)	SSGAC – Education attainment	23722424	126,559	rs2470893	T	C	-0.002	0.714
				rs2472297	T	C	-0.001	0.8807
				rs6968554	A	G	0.000	0.9079
				rs6968865	A	T	-0.002	0.7121
				rs17685	A	G	0.006	0.1748
				rs6265	T	C	0.007	0.1625

	SSGAC – College education	23722424	126,559	rs2470893	T	C	-0.002	0.2872
				rs2472297	T	C	-0.001	0.599
				rs6968554	A	G	-0.001	0.5006
				rs6968865	A	T	-0.002	0.3829
				rs17685	A	G	0.005	0.03126
				rs6265	T	C	0.003	0.1815
Body measurement (5, 6)	GIANT - Height	20881960	183,727	rs2470893	T	C	0.001	0.83
				rs2472297	T	C	-0.002	0.7
				rs6968554	A	G	0.009	0.0039
				rs6968865	A	T	0.008	0.018
				rs17685	A	G	-0.006	0.13
				rs6265	T	C	0.007	0.063
	GIANT - Waist hip ratio	20935629	77,167	rs2470893	T	C	0.009	0.022
				rs2472297	T	C	0.007	0.08
				rs6968554	G	A	0.006	0.059

				rs17685	A	G	0.006	0.28
				rs6265	**C**	**T**	**0.020**	**2.30E-06**
Obesity (7)	GIANT – Body Mass Index	25673413	339,224	rs2470893	T	C	0.008	0.01849
				rs2472297	T	C	0.005	0.1683
				rs6968554	G	A	0.009	0.004364
				rs6968865	T	A	0.006	0.172
				rs17685	A	G	0.010	0.03026
				rs6265	**C**	**T**	**0.042**	**2.99E-27**
	GIANT – BMI>30	25673413	339,224	rs17685	G	A	0.003	9.50E-01
				rs2470893	T	C	0.006	8.80E-01
				rs2472297	C	T	0.029	5.40E-01
				rs6968554	G	A	0.026	4.80E-01
				rs6968865	T	A	0.031	4.10E-01
				rs6265	**C**	**T**	**0.100**	**3.50E-10**
Smoking behavior (8)	TAG - Cigarette per day	20418890	68,028	rs2470893	T	C	-0.213	0.0274

				rs2472297	T	C	-0.161	0.2285
				rs6968554	A	G	0.056	0.5346
				rs6968865	A	T	0.050	0.5848
				rs17685	A	G	0.139	0.3136
				rs6265	T	C	-0.047	0.6564
	TAG - Ever/never smoke	20418890	74,035	rs2470893	T	C	0.009	0.5164
				rs2472297	T	C	-0.007	0.7498
				rs6968554	A	G	0.012	0.3518
				rs6968865	A	T	0.013	0.3114
				rs17685	A	G	-0.007	0.75
				rs6265	T	C	-0.063	1.72E-05
Body Fat (9)	Global Lipid Consortium - HDL	24097068	187,167	rs2470893	T	C	0.005	0.3613
				rs2472297	T	C	0.004	0.3967

				rs6968554	A	G	0.018	2.81E-06
				rs6968865	A	T	0.015	0.009228
				rs6265	T	C	0.008	0.07254
	Global Lipid Consortium - LDL	24097068	173,082	rs2470893	T	C	0.008	0.09758
				rs2472297	T	C	0.005	0.3315
				rs6968554	G	A	0.006	0.1251
				rs6968865	T	A	0.006	0.3134
				rs6265	C	T	0.003	0.6009
	Global Lipid Consortium - Total Cholesterol	24097068	187,365	rs2470893	T	C	0.006	0.2177
				rs2472297	T	C	0.003	0.4886
				rs6968554	G	A	0.003	0.4247
				rs6968865	T	A	0.004	0.3928
				rs6265	C	T	0.005	0.2669
	Global Lipid Consortium -	24097068	177,861	rs2470893	C	T	0.009	0.1374

	Triglyceride							
				rs2472297	C	T	0.006	0.352
				rs6968554	G	A	0.020	1.17E-06
				rs6968865	T	A	0.011	0.139
				rs6265	C	T	0.015	0.001204

Table 4: Pleiotropy assessment – Association of coffee/caffeine genetic variants with confounding variables using OCAC participant data

Risk Factor	**Unit of Measurement**	**Sample size**	**SNP ID**	**Risk Allele**	**Other Allele**	**Magnitude of Association**	**P-Value**
Estrogen use	No. of months	16,337	rs1481012	G	A	-0.0144	0.4292
			rs6968554	G	A	0.0100	0.3826
			rs2470893	T	C	-0.0299	0.0130
			rs17685	A	G	0.0144	0.2437
			rs6968865	T	A	0.0113	0.3224
			rs2472297	T	C	-0.0258	0.0601
Oral Contraceptive use	No. of months	25,699	rs1481012	G	A	0.0162	0.2585
			rs6968554	G	A	-0.0044	0.6278

			rs2470893	T	C	-0.0072	0.4465
			rs17685	A	G	-0.0044	0.6498
			rs6968865	T	A	-0.0050	0.5847
			rs2472297	T	C	-0.0081	0.4528
Parity	No. of pregnancies (regardless of outcome)	25,720	rs1481012	G	A	0.0057	0.6945
			rs6968554	G	A	-0.0087	0.3382
			rs2470893	T	C	0.0040	0.6714
			rs17685	A	G	0.0094	0.3372
			rs6968865	T	C	-0.0054	0.6209
			rs2472297	T	A	-0.0099	0.2792

Table 5: Sensitivity analysis - SNP association on caffeine consumption among OCAC AOCS participants (N=2,347)

SNP	Chromosome	Trait	Risk Allele	Other Allele	Magnitude of association (mg)	S.E.	P-value
rs6968865	7	Caffeine	T	A	11.7800648	5.83759195	0.043718
rs2472297	15	Caffeine	T	C	9.823905502	6.71210422	0.143446

AOCS refers to the Australian Ovarian Cancer Study.

Table 6: Sensitivity analysis – Adjustment of confounding variables in SNP-association

6A. Set 1: OCAC Participants with information on confounders available (excluding BMI), n=11,366										
				Model adjusted for potential confounders (exclude BMI)				**Original model**		
Gene	**SNP**	**EA**	**NEA**	**Pvalue**	**Beta**	**S.E.**		**Pvalue**	**Beta**	**S.E.**
Coffee										
ABCG2	rs1481012	G	A	0.7354	-0.0152	0.0449		0.5689	-0.0251	0.0440
AHR	rs6968554	G	A	0.7291	0.0098	0.0284		0.7607	0.0085	0.0279
CYP1A1	rs2470893	T	C	0.2732	0.0326	0.0298		0.3004	0.0303	0.0292
POR	rs17685	A	G	0.0074	-0.0819	0.0306		0.0085	-0.0791	0.0301
Caffeine										
AHR	rs6968865	T	A	0.7546	0.0089	0.0283		0.7671	0.0082	0.0278
CYP1A2	rs2472297	T	C	0.5422	0.0205	0.0337		0.6052	0.0171	0.0331

6B. Set 2: OCAC Participants with information on confounders available including BMI, n=4,718										
				Model adjusted for covariates including BMI				**Original model**		
Gene	**SNP**	**EA**	**NEA**	**Pvalue**	**Beta**	**S.E.**		**Pvalue**	**Beta**	**S.E.**
Coffee										
ABCG2	rs1481012	G	A	0.3318	0.0674	0.0694		0.3379	0.0650	0.0679
AHR	rs6968554	G	A	0.6501	-0.0200	0.0440		0.8276	-0.0094	0.0430

CYP1A1	rs2470893	T	C	0.4851	0.0326	0.0467		0.5542	0.0270	0.0457
POR	rs17685	A	G	0.0120	-0.1191	0.0475		0.0107	-0.1182	0.0464
Caffeine										
AHR	rs6968865	T	A	0.7155	-0.0160	0.0440		0.8935	-0.0057	0.0429
CYP1A2	rs2472297	T	C	0.3480	0.0500	0.0533		0.3399	0.0497	0.0521

EA refers to the effect allele, i.e. allele associated with increased coffee consumption; NEA refers to the non-effect allele.
The adjusted model is a logistic regression model on ovarian cancer status adjusted for 9 genetic principal components and covariates: education attainment, age at menarche, number of pregnancies, smoking, oral contraceptive use, estrogen use (and BMI in Set 2). The original model is a logistic model adjusted for only the 9 genetic principal components. The analysis on BMI is separated (reported in 7B) due to high number of missing values on BMI from the participants.

Table 7: Distribution of OCAC European participants

OCAC Acronym	**Study Name**	**Alt. Acronym**	**Country**	**Controls**	**Distribution on major EOC Histology/types**								
					All EOC	Invasive	All Serous	High-grade Serous	Low-grade Serous	Endometrioid	Mucinous	Clear-Cell	LMP*
AAS	African American Cancer Epidemiology Study	AACES	USA	0	0	0	0	0	0	0	0	0	0
AOCS/ ACS	Australia Ovarian Cancer Study & Australia Cancer Study	AOCS/ ACS	Australia	1139	1409	1133	813	733	40	118	39	68	262

	(Ovarian Cancer)												
AUS	merged with AOCS/ACS	AUS	Australia	0	109	88	63	56	4	9	5	2	21
BAV	Bavarian Ovarian Cancer Cases and Controls	BOCC	Germany	286	290	266	184	47	12	27	18	13	23
BEL	Belgium Ovarian Cancer Study	BOCS	Belgium	1287	792	601	474	362	16	45	40	25	124
BGS	Breakthrough Generations Study	BGS	UK	0	228	186	66	0	0	24	21	7	32
BVU	The BioVU DNA Repository	BioVU	USA	391	135	135	83	0	0	15	3	11	0
CAM	Cancer Research UK, Cambridge Research Institute	(none)	UK	0	233	228	155	0	0	10	0	17	0
CHA	Tianjin China Ovarian Cancer Study	(none)	China	0	0	0	0	0	0	0	0	0	0
CHN	Hebei Medical University	CHN	China	0	0	0	0	0	0	0	0	0	0
CNI	CNIO Ovarian Cancer Study	(none)	Spain	178	81	76	49	26	5	11	2	7	4
DKE	Duke University Clinic	(none)	USA	0	80	78	52	46	3	7	1	6	2
DOV	Diseases of the Ovary and their Evaluation	DOVE	USA	1459	1245	911	595	507	15	147	26	67	315

EPC	European Prospective Investigation into Nutrition and Cancer	EPIC	Europe	870	431	426	234	0	0	38	29	14	3
GER	Germany Ovarian Cancer Study	GOCS	Germany	376	202	180	117	90	14	18	19	6	19
GRC	Demokritos	DEM	Greece	0	0	0	0	0	0	0	0	0	0
GRR	Gilda Radner Familial Ovarian Cancer Registry	GRFOCR	Global	0	22	22	18	0	0	1	1	2	0
HAW	Hawaii Ovarian Cancer Study	(none)	USA	171	105	83	54	52	2	14	3	5	21
HJO	Hannover-Jena Ovarian Cancer Study	HJOCS	Germany	0	214	200	126	106	4	26	7	5	12
HMO	Hannover-Minsk Ovarian Cancer Study	HMOCS	Germany	285	65	65	35	7	0	5	3	1	0
HOC	Helsinki Ovarian Cancer Study	HOCS	Finland	280	264	256	140	0	0	35	50	16	7
HOP	Hormones and Ovarian Cancer Prediction	HOPE	USA	1189	525	470	268	248	14	71	25	36	37
HSK	Dr. Horst Schmidt Kliniken	(none)	Germany	0	122	118	101	98	3	12	1	0	4
HUO	Hannover-Ufa Ovarian Cancer Study	HUOCS	Germany	124	49	47	17	11	2	0	2	1	0

ICN	(merged to UK studies)	UK	UK	0	390	9	9	9	0	0	0	0	282
JPN	Hospital-based Epidemiologic Research Program at Aichi Cancer Center	HERPACC	Japan	0	0	0	0	0	0	0	0	0	0
KRA	Korean Epithelial Ovarian Cancer Study	Ko-EVE	Korea	0	0	0	0	0	0	0	0	0	0
LAX	Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute	WCP	USA	0	378	304	255	240	4	19	12	10	73
LUN	Departments of Cancer Epidemiology and Oncology, University Hospital, Lund	MISS	Sweden	1574	41	22	14	0	0	3	3	0	13
MAC	Mayo Clinic Case-Only Ovarian Cancer Study	(none)	USA	0	205	205	167	164	1	17	3	12	0
MAL	Danish Malignant Ovarian Tumor Study	MALOVA	Denmark	644	384	384	249	190	45	52	37	29	0
MAS	Malaysia Ovarian	MyOvCa	Malaysia	0	0	0	0	0	0	0	0	0	0

	Cancer Genetic Study												
MAY	Mayo Clinic Ovarian Cancer Case Control Study	(none)	USA	1130	1143	1036	771	755	11	126	30	58	93
MCC	Melbourne Collaborative Cohort Study	MCCS	Australia	142	134	109	62	20	3	11	15	7	24
MDA	MD Anderson Ovarian Cancer Study	(none)	USA	297	307	292	188	157	19	15	7	14	13
MEC	Multiethnic Cohort Study	MEC	USA	6	6	5	2	0	0	1	0	0	0
MOF	Moffitt Cancer Center Ovarian Cancer Study	MOF	USA	413	371	341	238	0	0	28	12	15	24
MSK	Memorial Sloan Kettering Cancer Center	MSKCC	USA	205	202	202	168	150	3	9	0	6	0
NCO	North Carolina Ovarian Cancer Study	NCOCS	USA	732	836	666	457	415	32	96	32	70	166
NEC	New England Case-Control Study	NECC	USA	566	502	424	239	217	12	97	28	31	70
NHS	Nurses' Health Study I and II	NHS	USA	314	336	261	130	0	0	49	16	16	59
NOR	University of Bergen, Haukeland University Hospital,	(none)	Norway	344	182	174	123	85	9	20	13	7	5

	Norway												
NTH	Nijmegen Ovarian Cancer Study	POLYGENE	Netherlands	584	254	252	126	74	22	63	32	20	2
OPL	Ovarian Cancer Prognosis and Lifestyle Study	OPAL	Australia	0	484	482	354	319	17	29	24	29	2
ORE	Oregon Ovarian Cancer Registry	OHSU-OOCR	USA	0	83	76	58	51	3	10	0	1	5
OVA	Ovarian Cancer in Alberta and British Columbia	OVAL-BC	Canada	722	660	499	284	0	0	81	24	45	137
PLC	The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	PLCO	USA	1117	263	233	130	0	0	19	5	8	22
POC	Polish Ovarian Cancer Study	IHCC	Poland	0	169	169	83	0	0	18	10	7	0
POL	Polish Ovarian cancer Case Control Study (NCI)	POCS	Poland	0	272	245	114	73	5	35	15	9	19
PVD	Danish Pelvic Mass Study	(none)	Denmark	0	194	194	152	141	9	15	11	9	0
RBH	Royal Brisbane Hospital	RBH	Australia	0	139	139	90	74	2	18	10	11	0

RMH	Royal Marsden Hospital Ovarian Cancer Study	(none)	UK	0	168	152	62	0	0	23	18	14	1
RPC	Roswell Park Cancer Institute Ovarian Cancer Cohort	(none)	USA	0	99	95	70	0	0	7	6	3	1
SEA	UK Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) Ovarian Cancer Study	SEARCH	UK	1823	2148	1945	890	431	459	255	247	181	198
SIS	The Sister Study	(none)	USA	1295	119	112	46	44	0	6	0	3	5
SMC	Swedish Mammography Cohort	(none)	Sweden	93	83	83	53	0	0	10	3	2	0
SOC	Southampton Ovarian Cancer Study	(none)	UK	0	298	272	118	87	9	63	33	10	19
SRO	Scottish Randomised Trial in Ovarian Cancer	SCOTROC	UK	0	3	3	3	1	2	0	0	0	0
STA	Family Registry for Ovarian Cancer AND Genetic	FROC & GEOCS	USA	310	282	203	128	113	11	27	16	14	73

	Epidemiology of Ovarian Cancer												
SWH	Shanghai Women's Health Study	SWHS	China	0	0	0	0	0	0	0	0	0	0
SZB	(merged to POC)			176	0	0	0	0	0	0	0	0	0
TBO	Tampa Bay Ovarian Cancer Study	TBOCS	USA	139	176	176	123	108	2	25	8	7	0
TOR	Familial Ovarian Tumor Study	FOTS	Canada	451	444	375	239	0	0	60	31	16	67
UCI	UC Irvine Ovarian Cancer Study	(none)	USA	292	258	145	91	80	5	35	7	10	112
UHN	Princess Margaret Cancer Centre	(none)	Canada	0	177	175	130	117	3	22	4	12	2
UKO	UK Ovarian Cancer Population Study	UKOPS	UK	985	729	729	387	313	24	112	71	71	0
UKR	UK Familial Ovarian Cancer Registry	UKFOCR	UK	0	42	41	23	16	1	3	3	1	0
USC	Los Angeles County Case-Control Studies of Ovarian Cancer	LAC-CCOC	USA	785	604	487	344	273	20	51	33	26	116
VAN	OVCARE Gynecologic Tissue Bank	(none)	Canada	0	172	154	139	136	0	3	3	6	18

	and Outcomes Unit												
WMH	Westmead Institute for Cancer Research - Westmead Hospital	(none)	Australia	0	145	142	118	105	12	13	0	5	3
WOC	Warsaw Ovarian Cancer Study	(none)	Poland	205	200	198	142	141	1	20	8	17	2
			Total	**23379**	**20683**	**17779**	**11213**	**7488**	**880**	**2199**	**1125**	**1121**	**2512**

*LMP refers to Low-malignant Potential.

Figure 1a: PCA plot of OCAC participants with EOC

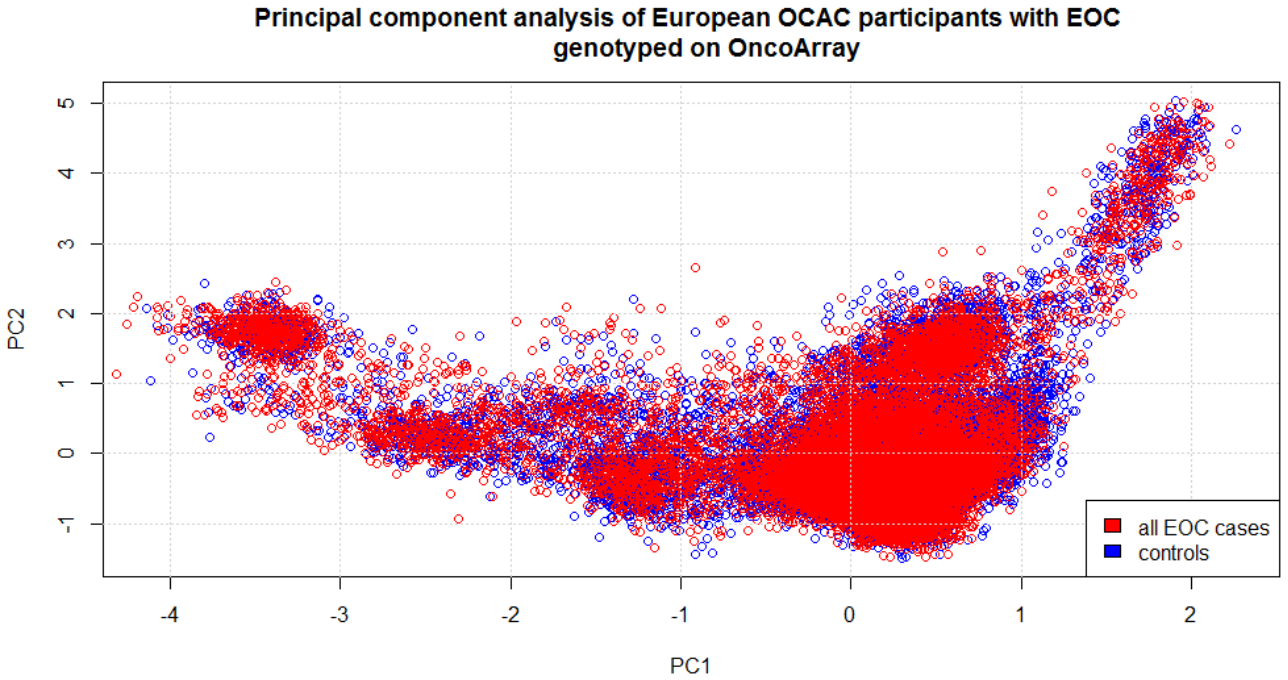
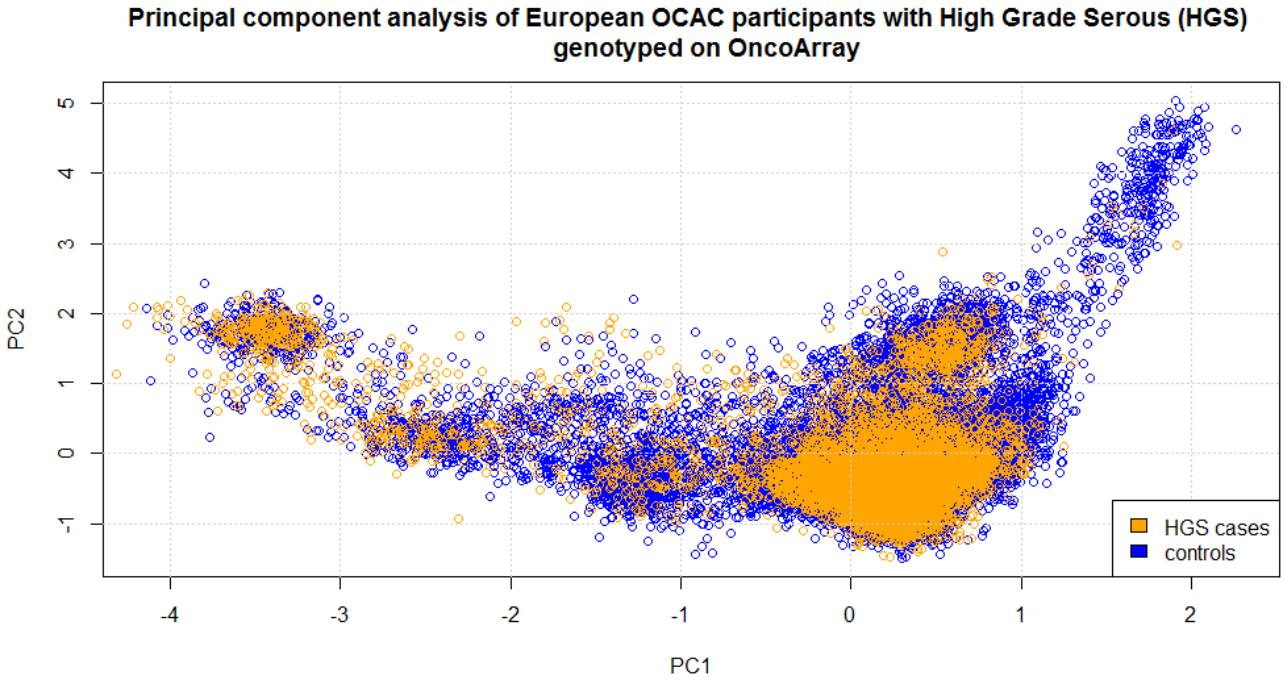


Figure 1b: PCA plot of OCAC participants with high grade serous



Reference

1. McMahon G, Taylor AE, Davey Smith G, Munafo MR. Phenotype refinement strengthens the association of AHR and CYP1A1 genotype with caffeine consumption. PloS one. 2014;9(7):e103448.
2. Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nature genetics. 2010;42(12):1077-85.
3. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics. 2010;42(2):105-16.
4. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013;340(6139):1467-71.
5. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467(7317):832-8.
6. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nature genetics. 2010;42(11):949-60.
7. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.
8. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature genetics. 2010;42(5):441-7.
9. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013;45(11):1274-83.

Table 1 : Distribution of EOC cases among European participants in OCAC.

Nature/Subtype	European Cases
Invasive	17,779
All serous¥	11,213
Endometrioid	2,199
Clearcell	1,121
Mucinous	1,125
All mucinous¥	2,023
High-grade serous	7,488
Low-grade serous	880
All EOC cases¥	20,683

¥Including unclassified and unknown serous/mucinous ovarian tumours.

Note: A complete breakdown of the EOC cases by each participating study is provided in Supplementary Material.