

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

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

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)
<u>Age at menarche</u>		
	13(12,14)	13(12,14)
Missing	12,077	7,790
<u>Highest level of education</u>		
Less than high school	1,659	1,516
High school or more	9,278	9,678
Missing	12,442	9,489
<u>Pregnancy</u>		
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)
<u>Smoking</u>		
Current smoker	1,242	1,669
Former smoker	3,324	3,509
Never smoked	5,327	5,774
Missing	13,486	9,731
<u>Oral contraceptive (OC)</u>		
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									
Gene	SNP	EA	NEA	<u>Model adjusted for potential confounders (exclude BMI)</u>			<u>Original model</u>		
				Pvalue	Beta	S.E.	Pvalue	Beta	S.E.
<i>Coffee</i>									
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
<i>Caffeine</i>									
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									
Gene	SNP	EA	NEA	<u>Model adjusted for covariates including BMI</u>			<u>Original model</u>		
				Pvalue	Beta	S.E.	Pvalue	Beta	S.E.
<i>Coffee</i>									
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
<i>Caffeine</i>										
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

	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	POLYGEN E	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	OV CARE 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

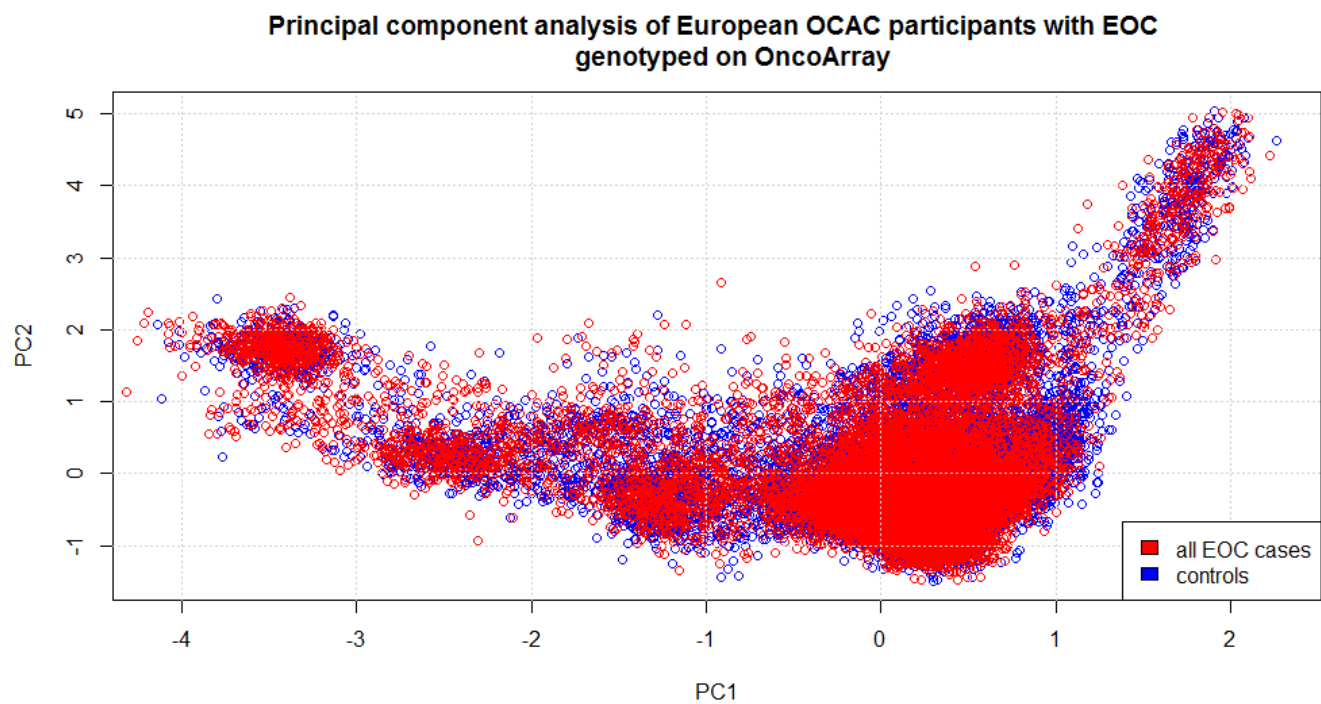
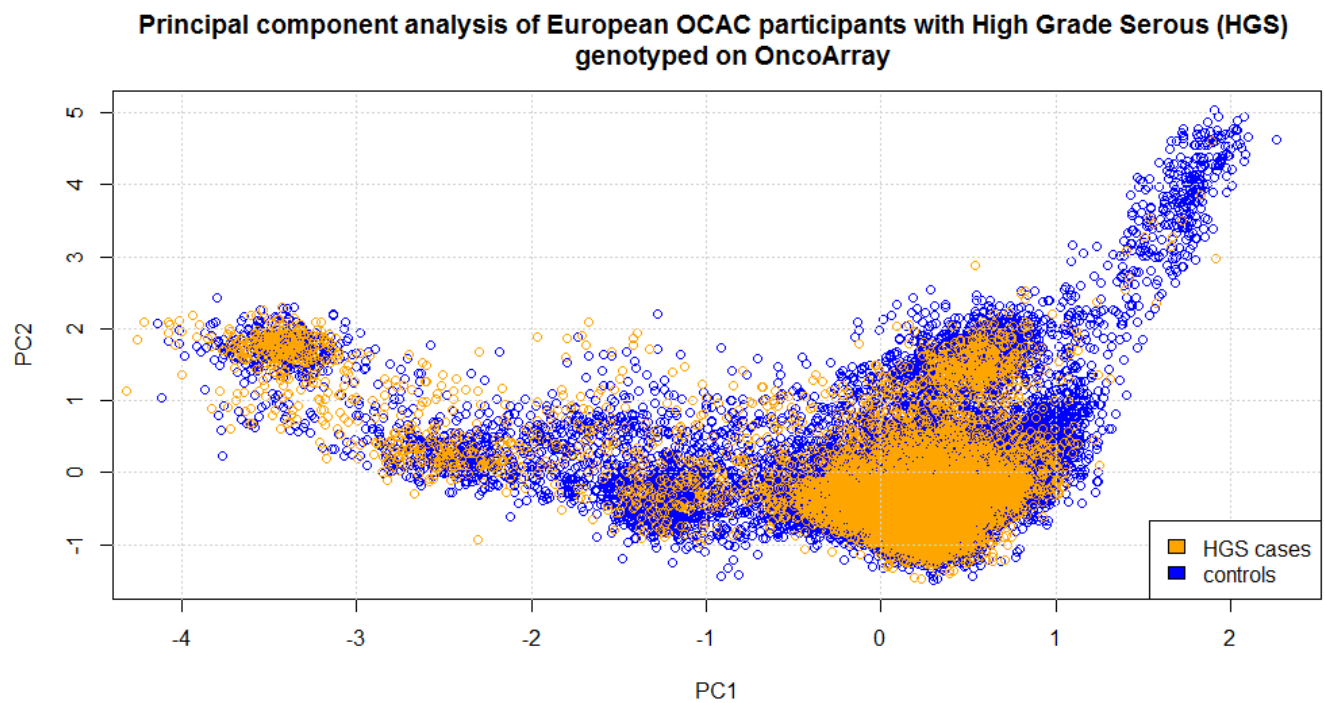


Figure 1b: PCA plot of OCAC participants with high grade serous



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, Steinhilber S, 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.