Supplementary Online Content

Yarmolinsky J, Bull CJ, Vincent EE, et al. Association between genetically proxied inhibition of HMG-CoA reductase and epithelial ovarian cancer. *JAMA*. doi:10.1001/jama.2020.0150

eMethods

This supplementary material has been provided by the authors to give readers additional information about their work.

OCAC and CIMBA analyses

The OCAC dataset comprises 63 genotyping project/case-control sets with some studies contributing samples to more than one genotyping project and some case-control sets representing a combination of multiple individual studies. The CIMBA dataset comprises 63 studies where the majority of participants were sampled from cancer genetics clinics. In CIMBA, mutation carriers were followed until the age of ovarian cancer diagnosis, or risk-reducing salpingo-oophorectomy or age at study recruitment. In both datasets, genotype data were obtained by either direct genotyping using an Illumina Custom Infinium array (OncoArray) consisting of approximately 530,000 single-nucleotide polymorphisms (SNPs) or by imputation with reference to the 1000 Genomes Project Phase Three dataset. All SNPs with a call rate of <95%, evidence of violation of Hardy-Weinberg equilibrium ($P<10^{-7}$ in controls or unrelated samples in CIMBA and $P<10^{-12}$ in cases), or SNPs with a concordance <98% among 5,280 duplicate pairs were removed. In imputation, all SNPs with a minor allele frequency of <1% and a call rate of <98% and SNPs that could not be linked to the 1000 Genomes reference panel or differed significantly in frequency from the panel (European samples) in addition to 1,128 SNPs where the cluster plot was judged to be inadequate were removed. OCAC OncoArray analyses were pooled with those from a previous genotyping project (COGS) and five additional GWAS datasets using fixed-effects meta-analysis. The analyses were adjusted for study and for population substructure by including the eigenvectors of project-specific principal components as covariates in the model. CIMBA OncoArray analyses were pooled with the COGS samples using a fixed-effects meta-analysis.

Extended Methods

SNP selection procedure

A linkage disequilibrium clumping approach was used to construct instrument for each drug target and LDL cholesterol. This approach filters out SNPs with *P*-values larger than a specific threshold, clusters the remaining SNPs by linkage disequilibrium and physical distance between SNPs, and selects the top associated SNP (i.e., SNP with the lowest *P*-value) from each clump. In analyses of drug targets, all genome-wide significant ($P < 5 \times 10^{-8}$) SNPs were first selected within ±100kb windows from the gene encoding each respective target and the clumping procedure was applied using a linkage disequilibrium r^2 threshold of 0.20 and a physical distance threshold of 250kb. In analyses of LDL cholesterol, all genome-wide significant ($P < 5 \times 10^{-8}$) SNPs (independent to genomic position) were selected and the clumping procedure was applied using a linkage disequilibrium r^2 threshold of 0.001 and a physical distance threshold of 250kb.

Effect estimation in Mendelian randomization

Summary genetic association data (effect estimates, standard errors, effect alleles, non-effect alleles, effect allele frequencies) were obtained for all SNPs used to proxy risk factors (drug targets or LDL cholesterol) from both risk factor and ovarian cancer (general population or BRCA1/2 mutation carriers) GWAS datasets. For all risk factors, SNPs used to proxy the risk factor were matched to the ovarian cancer dataset by assigning them the same effect allele (for all analyses, this represented the effect allele that lowered LDL cholesterol). For each drug target and LDL cholesterol analysis, effect estimates were first generated per individual SNP using the Wald ratio (also termed the "Ratio method") and standard errors were approximated using the delta method. The Wald ratio represents the SNP-ovarian cancer effect estimate divided by the SNP-risk factor effect estimate and the delta method approximation represents the SNP-ovarian cancer standard error divided by SNP-risk factor

effect estimate. Dividing the SNP-ovarian cancer effect estimate by the SNP-risk factor (in mmol/L) effect estimate permits scaling of Mendelian randomization estimates to represent the equivalent of a 1 mmol/L (38.7 mg/dL) reduction in LDL cholesterol. The inverse-variance weighted model represents the mean of Wald ratios from two or more SNPs weighted by the inverse variance of each of their respective SNP-ovarian cancer associations.

Multivariable Mendelian randomization

In contrast to conventional Mendelian randomization which estimates the total effect of an exposure on an outcome, multivariable Mendelian randomization estimates direct effects of two (or more) traits adjusted for each other. This method requires that there are at least as many genetic instruments available as there are exposures to proxy and that variants are associated with the two (or more) exposures examined.

MR-Egger

MR-Egger relaxes the exclusion restriction criterion and thus can provide unbiased estimates of causal effects even when all single-nucleotide polymorphisms in an instrument are invalid through violation of this assumption. This approach performs a weighted generalized linear regression of the SNP-outcome effect estimates on the SNP-exposure effect estimates with an unconstrained intercept term (i.e., unconstrained to pass through zero). Provided that the InSIDE (Instrument Strength Independent of Direct Effect) assumption is met (that no association exists between the strength of SNP-risk factor associations and the strength of bias due to horizontal pleiotropy) and that measurement error in the genetic instrument is negligible ("No Measurement Error" or NOME assumption), the slope generated from MR-Egger regression can provide an estimate of the causal effect of a risk factor on a disease outcomes that is adjusted for directional pleiotropy (where the

horizontally pleiotropic effect across a genetic instrument do not average to zero) and the intercept term can provide a formal statistical test for directional pleiotropy.

Weighted median estimator

The weighted median estimator (WME) approach provides an estimate of the weighted median of a distribution in which individual SNP effect estimates in an instrument are ordered and weighted by the inverse of their variance. Unlike MR-Egger which can provide an unbiased causal effect even when all SNPs are invalid instruments, WME requires that at least 50% of the information in a multi-allelic instrument is coming from SNPs that are valid instrumental variables in order to provide an unbiased estimate of a causal effect in an MR analysis. However, the WME has two advantages over MR-Egger in that it provides improved precision as compared to the latter and does not rely on the InSIDE assumption.

Weighted mode estimator

The weighted mode-based estimator generates a causal estimate using the mode of a smoothed empirical density function of individual SNP effect estimates in a multi-allelic instrument, weighted by the inverse variance of the SNP-outcome association. This approach operates under the assumption that the most common effect estimate of individual SNPs in a multi-allelic instrument arises from valid instruments (called the Zero Modal Pleiotropy Assumption, or ZEMPA). If this assumption holds, the mode can provide a consistent causal estimate even if most of the (non-modal) SNPs are invalid. Mode-based approaches have less power to detect a causal effect than the weighted median estimator but greater power than MR-Egger regression under the condition of no invalid instruments. Similar to the weighted

median estimator, mode-based approaches are also (by default) less susceptible to bias from outlying variants in a risk score.

Colocalization analysis

Colocalization tests the probability of shared causal variants between two (or more) traits. The presence of shared causal variants– as opposed to distinct causal variants that are in linkage disequilibrium with each other - is necessary in order to infer potential causality between these traits (though it is not sufficient as it does not account for horizontal pleiotropy and cannot inform on direction of association between traits). In order to examine whether there was evidence of colocalization across SNPs proxying HMG-CoA reductase and invasive epithelial ovarian cancer in/near *HMGCR*, we used the eCAVIAR package to quantify the probability of shared causal variants across datasets. eCAVIAR offers the following advantages over other colocalization packages: 1) it can account for multiple causal variants within a given locus, 2) analyses can be performed exclusively using summary genetic association data. We generated z-score estimates and obtained pair-wise correlations of SNP-LDL cholesterol and SNP-ovarian cancer estimates for all SNPs within ±100 kb windows from *HMGCR*. The package generates a colocalization posterior probability (CLPP) to estimate the degree of colocalization across both datasets and a 95% credible set which represents the minimum number of variants having a cumulative posterior probability of greater than 0.95. A cut-off threshold of 0.01 for the CLPP was used to conclude that drug targets and ovarian cancer outcomes shared a causal variant.

eTables

eTable 1. Comparison of Effect Allele Frequencies across GWAS for measured LDL-C levels, invasive epithelial ovarian cancer in the general population (OCAC), and epithelial ovarian cancer among BRCA1/2 mutation carriers (CIMBA)

SNP	Willer (EAF)	OCAC (EAF)	CIMBA (EAF)
HMG-CoA reductase			
rs12916	0.57	0.58	0.58
rs10515198	0.90	0.89	0.89
rs12173076	0.88	0.88	0.88
rs3857388	0.87	0.88	0.88
rs7711235	0.73	0.77	0.77
NPC1L1			
rs2073547	0.81	0.80	0.80
rs217386	0.41	0.43	0.43
rs7791240	0.91	0.91	0.91
PCSK9			
rs11591147	0.02	0.02	0.02
rs11206510	0.15	0.18	0.18
rs2479409	0.67	0.66	0.66
rs585131	0.18	0.18	0.17
rs11206514	0.39	0.38	0.39
rs2495477	0.40	0.39	0.39
rs572512	0.65	0.63	0.63
rs2479394	0.72	0.71	0.71
rs12067569	0.97	0.96	0.96
rs10493176	0.11	0.10	0.09
rs11583974	0.97	0.96	0.96

SNP= Single-Nucleotide Polymorphism, EA= Effect Allele Frequency, OCAC= Ovarian Cancer Association Consortium, CIMBA= Consortium of Investigators of Modifiers of BRCA1/2.

eTable 2. Characteristics of LDL cholesterol lowering genetic variants

SNP	EA*/ NEA	EAF	Effect (SE)	*P*-value
rs10195252	C/T	0.42	-0.024 (0.004)	3.8×10^{-8}
rs10490626	A/G	0.08	-0.051 (0.007)	1.7×10^{-12}
rs10832962	C/T	0.28	-0.032 (0.004)	6.6×10^{-14}
rs10893499	G/A	0.86	-0.052 (0.005)	3.9×10^{-21}
rs10903129	A/G	0.46	-0.033 (0.004)	3.0×10^{-17}
rs10947332	G/A	0.87	-0.050 (0.006)	7.0×10^{-18}
rs112201728	C/T	0.94	-0.068 (0.010)	8.5×10^{-10}
rs11563251	C/T	0.87	-0.035 (0.006)	4.5×10^{-8}
rs11591147	T/G	0.02	-0.497 (0.018)	8.6×10^{-143}
rs12066643	T/C	0.12	-0.039 (0.006)	1.1×10^{-8}
rs1250229	T/C	0.21	-0.024 (0.004)	3.1×10^{-8}
rs12721109	A/G	0.02	-0.446 (0.018)	3.0×10^{-122}
rs12748152	C/T	0.93	-0.050 (0.007)	3.2×10^{-12}
rs12916	T/C	0.57	-0.073 (0.004)	7.8×10^{-78}
rs13206249	A/G	0.22	-0.038 (0.006)	4.5×10^{-8}
rs13277801	T/C	0.65	-0.034 (0.004)	4.0×10^{-17}
rs1367117	G/A	0.71	-0.119 (0.004)	9.5×10^{-183}
rs1408272	G/T	0.05	-0.052 (0.008)	3.7×10^{-9}
rs1564348	T/C	0.85	-0.048 (0.005)	2.8×10^{-21}
rs16831243	C/T	0.82	-0.038 (0.006)	9.1×10^{-12}
rs16891156	A/C	0.98	-0.097 (0.017)	8.2×10^{-9}
rs17404153	T/G	0.14	-0.034 (0.005)	1.8×10^{-9}
rs174583	T/C	0.37	-0.052 (0.004)	7.0×10^{-41}
rs1800961	T/C	0.03	-0.069 (0.011)	6.0×10^{-10}
rs1801689	A/C	0.96	-0.102 (0.014)	9.8×10^{-12}
rs1883025	T/C	0.24	-0.030 (0.004)	6.1×10^{-11}
rs2000999	G/A	0.82	-0.065 (0.005)	4.2×10^{-41}
rs2030746	C/T	0.60	-0.021 (0.004)	8.6×10^{-9}
rs2073547	A/G	0.81	-0.049 (0.005)	1.9×10^{-21}
rs2228603	T/C	0.07	-0.104 (0.007)	4.4×10^{-44}
rs2315065	C/A	0.91	-0.110 (0.016)	5.2×10^{-12}
rs2328223	A/C	0.75	-0.030 (0.005)	5.6×10^{-9}
rs2390536	G/A	0.63	-0.022 (0.004)	2.0×10^{-8}
rs2419604	G/A	0.68	-0.030 (0.004)	7.5×10^{-14}
rs247616	T/C	0.29	-0.055 (0.004)	2.6×10^{-37}
rs2495495	C/T	0.87	-0.034 (0.006)	3.5×10^{-8}
rs2587534	G/A	0.47	-0.039 (0.004)	8.1×10^{-25}
rs2642438	A/G	0.25	-0.035 (0.004)	7.3×10^{-16}
rs267733	G/A	0.14	-0.033 (0.005)	5.3×10^{-9}
rs2710642	G/A	0.38	-0.024 (0.004)	6.1×10^{-9}
rs2737252	A/G	0.26	-0.031 (0.004)	7.0×10^{-14}
rs2886232	C/T	0.88	-0.045 (0.006)	3.9×10^{-11}
rs2965157	C/T	0.02	-0.190 (0.011)	7.3×10^{-62}
rs314253	C/T	0.34	-0.024 (0.004)	3.4×10^{-10}
rs364585	A/G	0.37	-0.0250 (0.004)	4.3×10^{-10}
rs3757354	T/C	0.21	-0.038 (0.004)	2.1×10^{-17}
rs3780181	G/A	0.05	-0.045 (0.007)	1.8×10^{-9}
rs4253776	A/G	0.88	-0.031 (0.006)	3.4×10^{-8}
rs4530754	G/A	0.42	-0.028 (0.004)	3.6×10^{-12}

SNP	EA*/ NEA	EAF	Effect (SE)	*P*-value
rs4722551	T/C	0.83	-0.039 (0.005)	3.9×10^{-14}
rs4942486	C/T	0.54	-0.024 (0.004)	2.3×10^{-11}
rs4970712	A/C	0.19	-0.0340 (0.004)	2.5×10^{-13}
rs5763662	C/T	0.97	-0.077 (0.012)	1.2×10^{-8}
rs579459	T/C	0.79	-0.067 (0.005)	2.4×10^{-44}
rs6016373	G/A	0.37	-0.035 (0.004)	7.9×10^{-19}
rs6065311	T/C	0.54	-0.042 (0.004)	1.7×10^{-30}
rs646776	C/T	0.21	-0.160 (0.004)	1.6×10^{-272}
rs6504872	C/T	0.53	-0.027 (0.004)	3.5×10^{-13}
rs6511720	T/G	0.10	-0.221 (0.006)	3.9×10^{-262}
rs6544713	C/T	0.71	-0.081 (0.004)	4.8×10^{-83}
rs6709904	G/A	0.11	-0.055 (0.009)	4.6×10^{-10}
rs676388	T/C	0.54	-0.027 (0.004)	1.3×10^{-11}
rs6818397	G/T	0.59	-0.022 (0.004)	1.7×10^{-8}
rs6882076	T/C	0.33	-0.046 (0.004)	3.3×10^{-31}
rs6909746	T/C	0.39	-0.026 (0.004)	7.9×10^{-11}
rs7254892	A/G	0.03	-0.485 (0.012)	$0.0 \times 10^{+00}$
rs72902576	G/T	0.04	-0.093 (0.013)	9.6×10^{-12}
rs7534572	C/G	0.31	-0.041 (0.006)	1.3×10^{-11}
rs7551981	G/T	0.41	-0.047 (0.004)	1.4×10^{-33}
rs75687619	G/T	0.98	-0.174 (0.016)	8.1×10^{-24}
rs7640978	T/C	0.11	-0.039 (0.007)	9.8×10^{-9}
rs7832643	G/T	0.60	-0.034 (0.004)	2.7×10^{-17}
rs8017377	G/A	0.54	-0.030 (0.004)	2.5×10^{-15}
rs964184	C/G	0.84	-0.086 (0.008)	2.0×10^{-26}
rs9875338	A/G	0.39	-0.027 (0.004)	2.2×10^{-11}
rs9987289	A/G	0.08	-0.071 (0.007)	8.5×10^{-24}

*LDL Cholesterol lowering allele, EA= Effect Allele, NEA = Non-Effect Allele, EAF = Effect Allele Frequency. Effect represents the change in LDL cholesterol levels (mmol/L) per copy of the effect allele. To convert estimates in mmol/L to mg/dL, multiply mmol/L estimate by 38.7. EAF estimates were obtained from the 1000 Genomes Phase 3 panel (European samples).

eTable 3. F-statistic estimates for genetic instruments and statistical power (%) estimates for primary analyses

Outcome	**HMG-CoA reductase (F=128.2)**	**NPC1L1 (F=71.7)**	**PCSK9 (F=196.4)**	**LDL-C (F=173.7)**
Invasive epithelial ovarian cancer	98.6	67.5	100.0	100.0
High grade serous carcinoma	88.8	45.3	99.9	100.0
Low grade serous carcinoma	14.1	7.4	36.6	97.3
Mucinous carcinoma	18.1	8.8	48.2	99.6
Endometrioid carcinoma	31.4	13.5	76.9	99.9
Clear cell carcinoma	17.6	8.6	46.8	99.5
Epithelial ovarian cancer (*BRCA1/2* mutation carriers)	33.0	14.0	79.4	99.9

Power calculations represent statistical power to detect an odds ratio of OR 0.50 per 1 mmo/L (38.7 mg/dL) reduction in LDL cholesterol at a 5% false positive rate.

eTable 4. Association between genetically-proxied LDL cholesterol levels with risk of overall and histotype-specific invasive epithelial ovarian cancer in the Ovarian Cancer Association Consortium in sensitivity analyses examining horizontal pleiotropy

Outcome	**MR-Egger regression OR (95% CI) *P*-value**	**MR-Egger intercept OR (95% CI) *P*-value**	**Weighted median estimator OR (95% CI) *P*-value**	**Weighted mode estimator OR (95% CI) *P*-value**
Invasive epithelial ovarian cancer	1.01 (0.90-1.13) 0.87	1.00 (0.99-1.00) 0.43	1.00 (0.91-1.09) 0.99	1.00 (0.93-1.07) 0.98
High grade serous carcinoma	1.01 (0.89-1.15) 0.86	1.00 (0.99-1.01) 0.81	1.02 (0.92-1.14) 0.66	1.04 (0.95-1.13) 0.41
Low grade serous carcinoma	1.13 (0.83-1.54) 0.43	0.99 (0.97-1.01) 0.49	1.14 (0.81-1.60) 0.44	1.11 (0.81-1.53) 0.52
Mucinous carcinoma	0.79 (0.58-1.08) 0.14	1.00 (0.98-1.02) 0.97	0.76 (0.56-1.01) 0.06	0.78 (0.62-1.00) 0.05
Endometrioid carcinoma	1.07 (0.84-1.36) 0.61	0.99 (0.97-1.00) 0.12	0.97 (0.79-1.19) 0.77	0.99 (0.84-1.18) 0.95
Clear cell carcinoma	1.05 (0.81-1.36) 0.71	1.00 (0.98-1.01) 0.70	0.99 (0.75-1.31) 0.95	1.00 (0.78-1.28) 0.98

OR (95% CI) represents the exponential change in odds of overall and histotype-specific invasive epithelial ovarian cancer per genetically-proxied 1 mmol/L (38.7 mg/dL) decrease in LDL cholesterol.

eTable 5. Colocalization analysis of *HMGCR* variants and their association with LDL-C levels and risk of invasive epithelial ovarian cancer in the Ovarian Cancer Association Consortium

Causal variant*	**CLPP**	**Credible set posterior probability**	**LDL-C *P*-value**	**Epithelial ovarian cancer *P*-value**
rs7703051	0.014	0.71	1.4×10^{-77}	3.8×10^{-3}
rs11749783	0.004	0.19	4.4×10^{-76}	3.9×10^{-3}
rs3846663	0.002	0.10	1.13×10^{-75}	2.7×10^{-3}

CLPP = Colocalization posterior probability. *rs11749783 and rs384663 are in strong LD with rs7703051 variant (r^2=1.00 and 0.98, respectively), using 1000 Genomes Phase 3 reference panel (CEU Population)

eTable 6. Association between genetically-proxied inhibition of HMG-CoA reductase and previously reported risk factors for ovarian cancer

Proposed risk factor	**N**	**Effect estimate (95% CI)**	***P*-value**
Age at menarche (years)	182,416	-0.18 (-0.33, -0.03)	0.02
Age at natural menopause (years)	69,360	0.08 (-0.75,0.92)	0.84
Body mass index (kg/m^2)	339,224	0.20 (0.06, 0.34)	0.006
Self-reported endometriosis	2,999 cases; 191,154 controls	1.14 (0.78-1.67)	0.44
Ever oral contraceptive use	159,724 cases; 33,989 controls	0.96 (0.85-1.08)	0.54
Smoking initiation	41,969 cases; 32,066 controls	1.18 (0.86-1.63)	0.31

Effect estimate (95% CI) represents the change in previously reported risk factor per genetically-proxied inhibition equivalent to a 1 mmol/L (38.7 mg/dL) reduction in LDL cholesterol

eTable 7. Association between genetically-proxied inhibition of HMG-CoA reductase and invasive epithelial ovarian cancer in the Ovarian Cancer Association Consortium and risk of epithelial ovarian cancer in the Consortium of Investigators of Modifiers of BRCA1/2, adjusted for body mass index and age at menarche

Population	**Proposed risk factor**	**Effect estimate**	***P*-value**
OCAC		**OR (95% C)**	
	Body mass index	0.46 (0.24-0.89)	0.02
	Age at menarche	0.34 (0.17-0.74)	0.007
CIMBA		**HR (95% CI)**	
	Body mass index	0.66 (0.37-1.17)	0.15
	Age at menarche	0.51 (0.26-1.00)	0.05

Effect estimate represents the change in ovarian cancer outcome per genetically-proxied inhibition equivalent to a 1 mmol/L (38.7 mg/dL) reduction in LDL cholesterol, adjusted for each previously reported risk factor

eTable 8. Association between genetically-proxied inhibition of HMG-CoA reductase and invasive epithelial ovarian cancer in the Ovarian Cancer Association Consortium and risk of epithelial ovarian cancer in the Consortium of Investigators of Modifiers of BRCA1/2 in leave-one-out analysis

Sample	**SNP removed**	**Effect estimate (95% CI)**	***P*-value**
OCAC		**OR (95% CI)**	
	rs12916	0.59 (0.42-0.85)	0.004
	rs10515198	0.61 (0.43-0.86)	0.005
	rs12173076	0.57 (0.41-0.81)	0.002
	rs3857388	0.60 (0.43-0.84)	0.003
	rs7711235	0.60 (0.42-0.84)	0.003
CIMBA		**HR (95% CI)**	
	rs12916	0.69 (0.50-0.94)	0.02
	rs10515198	0.70 (0.51-0.94)	0.02
	rs12173076	0.67 (0.49-0.91)	0.01
	rs3857388	0.70 (0.52-0.94)	0.02
	rs7711235	0.69 (0.51-0.93)	0.02

OCAC = Ovarian Cancer Association Consortium, CIMBA = Consortium of Investigators of Modifiers of BRCA1/2. Effect estimate represents the change in ovarian cancer outcome per genetically-proxied inhibition equivalent to a 1 mmol/L (38.7 mg/dL) reduction in LDL cholesterol