Assessing a causal relationship between circulating lipids and breast cancer risk: Mendelian randomization study

Kelsey E. Johnson[1*], Katherine M. Siewert[2*], Derek Klarin [3,4,5], Scott M. Damrauer [6,7], the VA Million Veteran Program, Kyong-Mi Chang[6,8], Philip S. Tsao[9,10], Themistocles L. Assimes[9,10], Kara N. Maxwell[8,11], Benjamin F. Voight[6,11,12,13]

*Equal contribution

Affiliations:

1. Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
2. Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
3. Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
4. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
5. Boston VA Healthcare System, Boston, MA, USA
6. Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, USA
7. Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
8. Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
9. Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
10. VA Palo Alto Health Care System, Palo Alto, CA, USA
11. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
12. Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
13. Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Correspondence to:
Benjamin F. Voight, PhD
Associate Professor of Systems Pharmacology and Translational Therapeutics
Associate Professor of Genetics
University of Pennsylvania - Perelman School of Medicine
3400 Civic Center Boulevard
10-126 Smilow Center for Translational Research
Philadelphia, PA 19104
Email: bvoight@pennmedicine.upenn.edu

Abstract

Objective
To assess a potential causal relationship between genetic variants associated with plasma lipid traits (high-density lipoprotein cholesterol, HDL; low-density lipoprotein cholesterol, LDL; triglycerides, TG) with risk for breast cancer.

Design
Mendelian randomization (MR) study.

Setting and Participants
Data from genome-wide association studies in 210,967 subjects from the Million Veterans Project were used to construct genetic instruments for plasma lipid traits. The effect of these instruments on breast cancer risk was evaluated using genetic data from the BCAC consortium based on 137,045 breast cancer cases and 119,078 controls.

Exposures
Genetically modified plasma levels of LDL, HDL, or TG.

Main Outcomes and Measures
Odds ratio (OR) for breast cancer risk per standard-deviation increase in HDL, LDL, or TG.

Results
We observed that a 1-SD genetically determined increase in HDL levels is associated with an increased risk for all breast cancers (HDL: OR=1.08, 95% CI=1.04-1.13, $P=7.4x10^{-5}$). Multivariable MR analysis, which adjusted for the effects of LDL, TG, body mass index, and age at menarche, corroborated this observation for HDL (OR=1.06, 95% CI=1.03-1.10, $P=4.9x10^{-4}$) and also a relationship between LDL and breast cancer risk (OR=1.03, 95% CI=1.01-1.07, P=0.02). We did not observe a difference in these relationships when stratified by breast tumor estrogen receptor status. We repeated this analysis using genetic variants independent of the leading association at core HDL pathway genes and found that these variants were also associated with risk for breast cancers (OR=1.11, 95% CI=1.06–1.16, $P=1.5x10^{-6}$), including gene-specific associations at *ABCA1*, *APOE-APOC1-APOC4-APOC2* and *CETP*. In addition, we find evidence that genetic variation at the ABO locus affects both lipid levels and breast cancer.

Conclusions
Genetically elevated plasma HDL levels appear to increase breast cancer risk. Future studies are required to understand the mechanism underlying this putative causal relationship, with the goal to develop potential therapeutic strategies aimed at altering the HDL-mediated effect on breast cancer risk.

Introduction

Breast cancer is the second leading cause of death for women, motivating the need for a better understanding of its etiology and more effective treatments.[1] Cholesterol is a known risk factor for multiple diseases that have reported associations with breast cancer, including obesity, heart disease and diabetes.[2] However, it is unknown whether cholesterol plays a causal role in breast cancer susceptibility.

The body of epidemiological and clinical trial studies to date has yet to clearly determine if there is a causal relationship between cholesterol and breast cancer. Observational epidemiological studies have reported positive, negative, or no relationship between lipid levels and breast cancer risk, however these studies can suffer from confounding.[3–5] While recent evidence suggests that statin use may reduce breast cancer risk,[6] cumulative meta-analyses are inconclusive.[7] Recently, cholesterol-lowering medications have been associated with improved outcomes in breast cancer patients on hormonal therapy, suggesting an interaction of circulating cholesterol levels with estrogen-sensitive breast tissues.[8] These mixed findings motivate the need for a high-powered causal inference analysis of lipids on breast cancer.

To try to resolve these discrepancies, recent studies have applied the framework of Mendelian randomization (MR) to determine if genetically elevated lipid levels associate with breast cancer risk. In a small sample of 1,187 breast cancer cases, Orho-Melander *et al*. used multivariable MR to find suggestive evidence of a relationship between triglycerides and breast cancer, but no association between LDL-cholesterol or HDL-cholesterol and cancer.[9] In a second study, Nowak *et al*.[10] performed an MR analysis with genetic association data from large GWAS for lipids and breast cancer.[11,12] They reported nominal positive associations between LDL-cholesterol levels and all breast cancers, and between HDL-cholesterol levels and ER-positive breast cancers. While compelling, this study also had limitations. First, they used relatively few variants in their genetic instrument due to the need for addressing instrument heterogeneity and genetic correlation between lipid traits, resulting in a conservative analysis. Second, they analyzed each lipid trait separately rather than take advantage of multivariable methods to consider lipid traits together along with additional, potentially confounding causal risk factors. Third, the authors did not quantitatively assess heterogeneity to determine if the observed lipid associations were statistically different across breast cancer subtypes.

These studies motivate a Mendelian randomization study that considers multiple lipid traits concurrently to delineate the independent effect of each lipid trait on breast cancer susceptibility. Such an approach that includes the effects of all biomarkers plus confounding factors obviates the need to remove pleiotropic variants and the loss of statistical power which results from this removal. An ideal approach would consider the effects of body mass index (BMI) and age at menarche, known risk factors for breast cancer that are correlated with lipids.[13–19]

In what follows, we apply the causal inference framework of MR to determine if genetically elevated lipid traits modify breast cancer susceptibility, independent of one another and of BMI and age at menarche. We take advantage of a recent GWAS for lipid levels performed in 210,967 individuals of European ancestry[20] to provide power for our causal inference analyses. Using a joint, multivariable MR, we find that genetic elevation of HDL trait increases risk for breast cancer, even after accounting for the effects of our genetic instruments on BMI and age at menarche. In addition, we perform a gene-specific Mendelian randomization, which focuses on variants in genes most central to the metabolism of cholesterol. This gene-specific approach supports HDL as a causal risk factor for breast cancer. Finally, we perform genetic correlation analyses to look for both genome-wide and locus-based correlation in effect sizes between lipids and breast cancer. Our results suggest that genetically elevated levels of HDL, and genetically lowered levels of TG, raise breast cancer risk.

Methods

Study Populations

Lipids GWAS summary statistics were obtained from the Million Veteran Program (MVP) (210,967 European individuals)[20] and the Global Lipids Genetics Consortium (GLGC) (up to 188,577 genotyped individuals).[12] As additional exposures in multivariable MR analyses, we used BMI summary statistics from a meta-analysis of GWAS in up to 795,640 individuals, and age at menarche summary statistics from a meta-analysis of GWAS in up to 329,345 women of European ancestry.[16,21] Genome-wide association study (GWAS) summary statistics from 137,045 breast cancer cases and 119,078 controls were obtained from the Breast Cancer

Association Consortium (BCAC).[11] More details on these cohorts are in the **Supplementary Methods**.

Lipid meta-analysis

We performed a fixed-effects meta-analysis between each lipid trait (Total cholesterol (TC), LDL, HDL, and triglycerides) in GLGC and the corresponding lipid trait in the MVP cohort[12,20] using the default settings in plink.[22] These meta-analysis results appear well calibrated (**Supplementary Figure 1**).

Mendelian randomization analyses

Mendelian randomization analyses were performed using the TwoSampleMR R package (https://github.com/MRCIEU/TwoSampleMR).[23] For all analyses, we used the two-sample MR framework, which means that the lipid, BMI, age at menarche and breast cancer genetic associations were estimated in different cohorts. Unless otherwise noted, MR results reported in this manuscript used inverse-variance weighting assuming a random effects model. SNPs associated with each lipid trait were filtered for genome-wide significance ($P < 5x10^{-8}$) from the MVP lipid study[12], and then removed SNPs in linkage disequilibrium ($r^2 < 0.001$ in UK10K consortium).[24] Each of these independent, genome-wide significant SNPs was termed a genetic instrument. To reduce heterogeneity in our genetic instruments for single trait MR, we employed a pruning procedure (**Supplementary Methods**). Genetic instruments used in single trait MR are listed in **Supplementary Table 1**. For multivariable MR experiments, we generated genetic instruments by first filtering the genotyped variants for those present across all datasets. For each trait and data set combination (Yengo *et al*. for BMI; Day *et al*. for age at menarche, MVP and GLGC for HDL, LDL, and TG) we then filtered for genome-wide significance ($P < 5x10^{-8}$), and for linkage disequilibrium ($r^2 < 0.001$ in UK10K consortium).[24] We then tested for instrument strength and validity[25], and removed instruments driving heterogeneity (**Supplementary Methods**). Genetic instruments used in multivariable MR are listed in **Supplementary Table 2**. As the MR methods and tests we employed are highly correlated, we did not apply a multiple testing correction to the reported P-values.

Core HDL and LDL pathway genetic instrument development

We defined sets of core genes for HDL or LDL that met the following criteria: (1) their protein products are known to play a key role in HDL or LDL biology (plus *HMGCR* and *NPC1L1*, two targets of LDL lowering drugs, in the LDL gene set), and (2) there were conditionally independent lipid trait-associated variants within 100kb upstream or downstream of the RefSeq coordinates for the gene (or locus, in the case of *APOE-APOC1-APOC4-APOC2* and *APOA4-APOC3-APOA1*).[20] We then used the conditional HDL or LDL association statistics from Klarin et al. for those genes in gene-specific MR analyses.[20] The loci included in each set and the genetic instruments used in each locus-specific MR are listed in **Supplementary Table 3**. We performed a separate fixed effects inverse-variance weighted MR with the conditionally independent genetic instruments at each gene, and performed fixed effects inverse variance weighted meta-analysis of the results across HDL or LDL genes using the R package meta.[26]

Genetic correlation analyses

We performed cross-trait LD Score Regression using the LDSC toolkit with default parameters, with the BCAC association statistics for breast cancer and our meta-analysis of GLGC and MVP for lipid associations. We used the ρ-Hess software for our local genetic correlation analysis,[27] using the UK10K reference panel, and the LD-independent loci published in Berisa *et al*. to partition the genome.[28] We used a Bonferroni significance threshold based on the number of these independent loci (1,704 loci). There was minor cohort overlap between the GLGC and Breast Cancer GWAS due to the EPIC cohort.[10] We included this overlap when performing ρ-Hess, using the cross-trait LD score intercept to estimate phenotypic correlation.

Results

Single trait Mendelian randomization (MR) in breast cancer

We first performed single trait MR analyses using summary statistics from MVP[20] for each of four lipid traits (*i.e.*, total cholesterol - TC, high density lipoprotein cholesterol - HDL, low density lipoprotein cholesterol - LDL, and triglycerides - TG) as the intermediate biomarkers and risk for all breast cancers as the outcome (**Supplementary Figure 2**). We observed a significant relationship between genetically elevated HDL and breast cancer risk (OR=1.10 per standard deviation of lipid level increase, 95% CI=1.04-1.17, $P = 2.1 \times 10^{-3}$) and genetically decreased

triglyceride levels and breast cancer risk (OR=0.93, 95% CI=0.88-0.99, P=0.015; **Supplementary Table 4**). Sensitivity analyses identified heterogeneity (**Methods, Supplementary Table 5**), but there was no evidence of bias from directional pleiotropy (**Methods, Supplementary Table 6**). To mitigate concerns of instrument heterogeneity, we removed genetic variants from our genetic instrument for each lipid trait that were responsible for instrument heterogeneity (**Supplementary Methods**), and again observed a relationship with HDL cholesterol (OR=1.08, 95% CI=1.04-1.13, $P=7.4x10^{-5}$) and triglycerides (OR=0.94, 95% CI=0.90-0.98, $P=2.6x10^{-3}$) (**Figure 1, Supplementary Figures 3 and 4, Supplementary Table 7**). As HDL and TG are inversely correlated[14,29], the opposing relationship between these two lipid traits and breast cancer could be expected in single-trait analyses.

We also tested the relationship between lipid traits and breast cancer using a meta-analysis of the two major lipid GWAS from MVP and GLGC, and GLGC alone. Overall, single trait MR analyses with the meta-analysis and GLGC lipid associations produced consistent results to those with MVP alone (**Supplementary Figure 5).** In a reciprocal single trait MR testing the effect of genetically-determined breast cancer risk on each lipid trait, we observed no relationship with HDL or LDL cholesterol (**Supplementary Table 8**), but did see a relationship with TG. However, a Steiger test for directionality confirmed that breast cancer as outcome was the correct causal direction for all lipid traits (**Supplementary Table 7**).[30] We also performed genetic instrument pruning in the same manner as Nowak *et al*.: removing genetic instruments for LDL, HDL, and TG that were associated with at least one of the two other lipid traits (P<0.001).[10] After this pruning, we did not find a significant relationship with LDL, HDL, or TG, and noted that this pruning procedure increased the size of the confidence intervals considerably (**Supplementary Table 9**).

Multivariable Mendelian randomization with age at menarche and body mass index as exposures

It has been previously observed that body mass index (BMI) and age at menarche are both genetically correlated and epidemiologically associated with both breast cancer[19,31,32] and lipid traits.[14,29] To incorporate these potential confounders into our causal inference framework, we performed multivariable MR analyses using all three lipid traits (genetic effect estimates from MVP), age at menarche, and BMI as exposures, and breast cancer risk as the outcome (**Figure 1**). We observed relationships between genetically-influenced HDL, LDL, BMI, and age at

menarche with breast cancer (HDL: OR=1.06, 95% CI=1.03-1.10, P=4.93x10^{-4}; LDL: OR=1.04, 95% CI=1.01-1.07, P=0.02; BMI: OR=0.90, 95% CI=0.87-0.94, P=1.15x10^{-6}; age at menarche: OR=0.96, 95% CI=0.93-0.99, P=2.44x10^{-3}), but not TG (OR=0.98, 95% CI=0.95-1.00, P=0.10) (**Figure 1, Supplementary Table 10**). Our results were consistent before and after pruning for genetic instrument heterogeneity (**Supplementary Table 10)**, and when using summary statistics from three independent subsets of the breast cancer dataset (**Supplementary Figure 6**, **Supplementary Table 10**). We also performed multivariable MR with pairs of lipid traits with genetic effect estimates from different datasets (GLGC or MVP), with and without BMI, and saw consistent results (**Supplementary Table 11**, **Supplementary Figure 7**). Considering the genetic correlation between HDL and TG, the significant association of HDL compared to TG with breast cancer in multivariable analysis, and the consistent relationship between HDL and breast cancer across breast cancer datasets, we focused our further MR analyses on the relationship between HDL cholesterol and breast cancer, in addition to the previously reported association between LDL and breast cancer.[10]

MR Method	Odds Ratio (95% CI)	P-value
HDL		
IVW	1.08 (1.04-1.13)	7.42×10^{-5}
Weighted median	1.06 (1.00-1.13)	0.05
MR-Egger	1.09 (1.02-1.17)	0.01
MVMR	1.06 (1.03-1.10)	4.93×10^{-4}
LDL		
IVW	1.04 (1.00-1.09)	0.07
Weighted median	1.06 (1.00-1.12)	0.05
MR-Egger	1.06 (1.00-1.13)	0.05
MVMR	1.04 (1.01-1.07)	0.02
TG		
IVW	0.94 (0.90-0.98)	2.60×10^{-3}
Weighted median	0.92 (0.88-0.97)	3.47×10^{-3}
MR-Egger	0.91 (0.86-0.97)	3.98×10^{-3}
MVMR	0.98 (0.95-1.00)	0.10

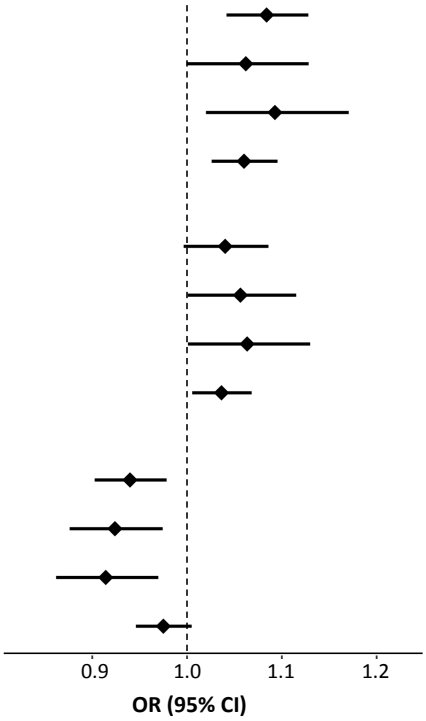


Figure 1. Results of MR analyses of the effects of HDL, LDL, or TG on breast cancer risk. IVW= inverse-variance weighted MR; MVMR = multivariable MR with HDL, LDL, TG, BMI, and age at menarche as exposures. Results plotted are after pruning for instrument heterogeneity.

Mendelian randomization with outcome stratified by estrogen receptor status

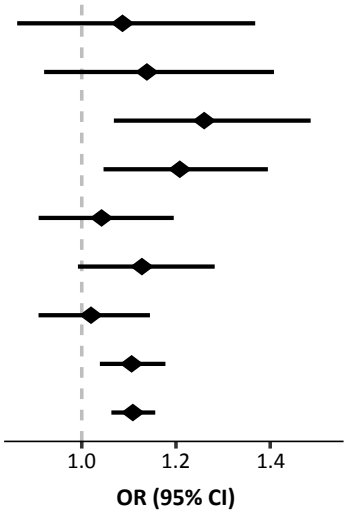
We next performed a MR analysis of the relationship between genetically-influenced lipids and breast cancer risk stratified by estrogen receptor positive (ER+) or negative (ER-) status. We observed similar effect size estimates of the four lipid traits on the breast cancer subtypes as on breast cancer not stratified by subtype (**Supplementary Figure 8**). A formal test for heterogeneity found no evidence to reject the null hypothesis of homogeneity between the cancer subtypes (e.g. HDL: Cochran's Q = 6.6×10^{-5}, P = 0.99; **Supplementary Table 12**).[33]. Thus, we observed no substantive difference in the relationship from any lipid trait to ER+ or ER- breast cancers, consistent with the strong genetic correlation between these two breast cancer subtypes (cross-trait LD Score regression genetic correlation estimate = 0.62, P = 2.9×10^{-83}).

HDL and LDL pathway gene-specific Mendelian randomization

To further clarify causal association, we next examined associations for breast cancer risk at genetic variants near core HDL and LDL genes. We identified conditionally independent associations at core genes, which we define as genes or loci previously annotated with a core role in the metabolism of each lipid trait, or an established drug target (HDL: *ABCA1, APOA4-APOC3-APOA1, APOE-APOC1-APOC4-APOC2, CETP, LCAT, LIPC, LIPG, PLTP, SCARB1*; LDL: *APOB, HMGCR, LDLR, LPA, MYLIP, NPC1L1, PCSK9*) (**Methods, Supplementary Table 3**). For each gene or locus with at least two conditionally independent genetic instruments (all except *LCAT* and *MYLIP*), we performed inverse variance weighted MR (fixed effects model) with conditional HDL or LDL effect size estimates as the exposure and breast cancer risk as the outcome (**Supplementary Figures 9 and 10**). We observed a positive relationship between HDL and breast cancer risk at three loci (*ABCA1, APOE-APOC1-APOC4-APOC2, CETP*; **Figure 2**), and between LDL and breast cancer risk at one locus (*HMGCR,* **Supplementary Figure 11**). Combining the effect estimates across core genes in a meta-analysis, we observed a positive relationship for HDL (OR = 1.11, 95% CI = 1.06-1.16, P = 1.53×10^{-6}; **Figure 2**) and LDL (OR = 1.07, 95% CI = 1.01-1.14, P = 0.02; **Supplementary Figure 11**). There was no evidence of heterogeneity across loci in either meta-analysis (HDL: Q=6.63, P=0.47; LDL: Q=5.53, P=0.35).

Figure 2. MR results for HDL gene-specific instruments (see **Supplementary Table 3**) and meta-analysis of effect estimates across genes.

Locus	N SNPs	Odds Ratio (95% CI)	P-value
SCARB1	5	1.09 (0.86-1.37)	0.48
PLTP	2	1.14 (0.92-1.41)	0.23
APOE-APOC1-APOC4-APOC2	6	1.26 (1.07-1.49)	6.09×10^{-3}
ABCA1	8	1.21 (1.05-1.40)	0.01
LIPG	6	1.04 (0.91-1.20)	0.56
APOA4-APOC3-APOA1	3	1.13 (0.99-1.28)	0.07
LIPC	16	1.02 (0.91 -1.14)	0.74
CETP	14	1.11 (1.04-1.18)	1.71×10^{-3}
Meta-analysis		1.11 (1.06-1.16)	1.53×10^{-6}



Genome-wide and local genetic correlation

If cholesterol levels were a causal risk factor for breast cancer, we might expect a correlation between the strength of genetic association with these two traits at genetic variants across the entire genome, in addition to those at genome-wide significant loci. To answer this question, we utilized two approaches to estimate genetic correlation between breast cancer and lipid traits. Consistent with our MR results, we found positive genetic correlation estimates for TC, LDL, and HDL; and a negative estimate for TG using cross-trait linkage disequilibrium score regression (**Supplementary Figure 12**).[14] However, most P-values were not significant; the only nominal association ($P < 0.05$) was with total cholesterol and ER-negative breast cancer ($P=0.04$). In contrast, the ρ-Hess method, which uses a different model of genetic correlation, found significant ($P<0.05$) correlations between all four lipid traits and breast cancer, with directions consistent with our MR results (**Supplementary Table 13**).

To discover specific new loci that may be specifically correlated, we used the ρ-Hess method to search for genomic regions with genetic correlation between this same lipid meta-analysis and breast cancer.[27] ρ-Hess identified one region that surpassed Bonferroni test correction, with a positive correlation between both LDL and TC and breast cancer (**Supplementary Table 14**). In this region there are two SNPs in high LD (rs532436 and rs635634, $r^2=0.99$) that are genome-

wide significantly associated with LDL (rs532436: P=9.98x 10^{-112}), TC (P=1.68 x 10^{-98}) and breast cancer (P=2.9 x 10^{-8}). These SNPs lie within an intron of the *ABO* gene and are associated with change in gene expression of *ABO* in multiple tissues, suggesting ABO is a possible causal gene for these trait associations.[34]

Discussion

Causal inference is one of the most challenging problems in biology and medicine, requiring a strong consensus of evidence from multiple sources. Excitingly, large-scale human genetics data provides the opportunity to bring an additional line of evidence to this important problem. To clarify how circulating levels of serum lipids levels might influence risk of breast cancer, we utilized the framework of Mendelian randomization. We provide evidence that genetically elevated HDL and LDL levels increase the risk for breast cancer in support of a causal hypothesis.

While substantial effort has been spent developing HDL raising therapies for cardiovascular disease prevention, independent studies have proposed an increase in all-cause mortality in individuals with high HDL levels.[35] Our results suggest that therapies that aim to reduce cardiovascular risk by raising HDL levels might have an unintended consequence of elevated breast cancer risk. Specifically, our gene-based score using HDL-raising variation at the *CETP* locus predicted that *CETP*-based inhibition would elevate breast cancer risk (OR=1.11, 95% CI=1.04-1.18, P=1.71x10^{-3}). Additionally, two recent Mendelian randomization studies reported causal evidence between elevated HDL and risk for age-related macular degeneration.[36,37] These potential disease-increasing consequences may not have been possible to identify in safety trials, given the limited window of study to monitor progression or incidence of disease, the putative causal effect estimates, and the demographics of the study population (*i.e.*, higher proportion of male participants). Our result supports the use of human genetics data as both a novel strategy for therapeutic targeting and for the discovery of potential drug complications to direct long term post-clinical trial followup.[38]

Although Nowak *et al*. previously used Mendelian randomization to discover associations between lipids and breast cancer[10,39], our report presents a reconsideration of these effects. Even

after conditioning on the effects of HDL, BMI, and age at menarche, our MR analysis suggests a potential causal relationship between LDL and BC. Nowak *et al*. only found a relationship between HDL cholesterol and ER+ breast cancer, while we found a relationship between HDL cholesterol and risk for all breast cancers. We also find a previously unreported association with triglycerides and breast cancer, though our multivariable analysis suggests this may be explained by correlation between triglycerides and HDL, and not an independent triglyceride effect. In their analyses, Nowak *et al*. used a strict pruning procedure in an attempt to isolate the effects of each lipid trait. However, this approach reduces power due to the high genetic correlation of these traits. The multivariable approach taken here is an alternative way to estimate the effect of an exposure while accounting for correlated exposures.

We note several caveats to our analyses. The first is that MR makes a number of assumptions that must be met for accurate causal inference.[40,41] Although we used statistical methods that try to detect and correct for violations of these assumptions, these methods are not guaranteed to correct for all types of confounding, and alternative causal inference frameworks outside of MR are warranted. We note that our effect estimates may be attenuated due to association of lipid IVS with the use of lipid-lowering medication. In addition, it is perhaps surprising that we did not find a significant genetic correlation between breast cancer and lipids using cross-trait LD score regression, however, our result corroborates a previous study which performed this analysis using smaller GWAS.[15] Our lack of significant results could be caused by limited polygenicity of either trait, which decreases the power of this method.[14] Lastly, we cannot be certain that the true underlying causal exposure is lipid levels, and not some phenotype for which lipids is a proxy. However, we are not aware of any process for which lipids is a proxy through which breast cancer would be affected.

The analyses presented here do not bring evidence on a specific mechanism for tumorigenesis, but they do bring renewed attention to potential mechanisms requiring future functional study. Cholesterol and its oxysterol metabolites, either in the circulatory system or in the local mammary microenvironment, may have direct effects on mammary tissue growth induction of breast tumorigenesis.[42,43]

Our study supports a causal relationship from increased HDL cholesterol to increased breast cancer risk, and this hypothesis warrants further exploration. Statins are widely prescribed to decrease LDL levels; however, statins also increase HDL levels. If further research substantiates the relationship between higher HDL levels and increased breast cancer risk, the consensus that HDL is "good cholesterol", or of benign effect, may require re-evaluation.

Acknowledgements

This work was supported by the US National Institutes of Health (R01 DK101478 to B.F.V., T32 GM008216 for K.E.J., T32 HG000046 for K.M.S.), and a Linda Pechenik Montague Investigator award (to B.F.V.). This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration and was supported by award #MVP000. This publication does not represent the views of the Department of Veterans Affairs or the United States Government. This research was also supported by two additional Department of Veterans Affairs awards (I01-BX003362 [Tsao/Chang], IK2-CX001780 [Damrauer]). This study makes use of data generated by the UK10K Consortium, derived from samples from the Avon Longitudinal Study of Parents and Children (ALSPAC) and the Department of Twin Research and Genetic Epidemiology (DTR). A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award WT091310.

Citations

1. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends--An Update. *Cancer Epidemiol Biomarkers Prev*. 2016;25(1):16-27. doi:10.1158/1055-9965.EPI-15-0578

2. Rose D, Gracheck P, Vona-Davis L, Rose DP, Gracheck PJ, Vona-Davis L. The Interactions of Obesity, Inflammation and Insulin Resistance in Breast Cancer. *Cancers (Basel)*. 2015;7(4):2147-2168. doi:10.3390/cancers7040883

3. Kuzu OF, Noory MA, Robertson GP. The role of cholesterol in cancer. *Cancer Res*. 2016;76(8):2063-2070. doi:10.1158/0008-5472.CAN-15-2613

4. Touvier M, Fassier P, His M, et al. Cholesterol and breast cancer risk: a systematic review and meta-analysis of prospective studies. *Br J Nutr*. 2015;114(03):347-357. doi:10.1017/S000711451500183X

5. Ni H, Liu H, Gao R. Serum Lipids and Breast Cancer Risk: A Meta-Analysis of Prospective Cohort Studies. Singh S, ed. *PLoS One*. 2015;10(11):e0142669. doi:10.1371/journal.pone.0142669

6. Statin use and mortality in cancer patients: Systematic review and meta-analysis of observational studies. *Cancer Treat Rev*. 2015;41(6):554-567. doi:10.1016/J.CTRV.2015.04.005

7. Ni H, Liu H, Gao R. Serum lipids and breast cancer risk: A meta-Analysis of prospective cohort studies. *PLoS One*. 2015;10(11). doi:10.1371/journal.pone.0142669

8. Borgquist S, Giobbie-Hurder A, Ahern TP, et al. Cholesterol, Cholesterol-Lowering Medication Use, and Breast Cancer Outcome in the BIG 1-98 Study. *J Clin Oncol*. 2017;35(11):1179-1188. doi:10.1200/JCO.2016.70.3116

9. Orho-Melander M, Hindy G, Borgquist S, et al. Blood lipid genetic scores, the HMGCR gene and cancer risk: a Mendelian randomization study. *Int J Epidemiol*. November 2017. doi:10.1093/ije/dyx237

10. Nowak C, Ärnlöv J. A Mendelian randomization study of the effects of blood lipids on breast cancer risk. *Nat Commun*. 2018. doi:10.1038/s41467-018-06467-9

11. Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551(7678):92-94. doi:10.1038/nature24284

12. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and Refinement of Loci Associated with Lipid Levels. *Nat Genet*. 2013;45(11):1274-1283. doi:10.1038/ng.2797.Discovery

13. Neuhouser ML, Aragaki AK, Prentice RL, et al. Overweight, Obesity, and Postmenopausal Invasive Breast Cancer Risk. *JAMA Oncol*. 2015;1(5):611.

doi:10.1001/jamaoncol.2015.1546

14. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406

15. Lindström S, Finucane H, Bulik-Sullivan B, et al. Quantifying the genetic correlation between multiple cancer types. *Cancer Epidemiol Biomarkers Prev*. 2017;26(9):1427-1435. doi:10.1158/1055-9965.EPI-17-0211

16. Day FR, Thompson DJ, Helgason H, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;49(6):834-841. doi:10.1038/ng.3841

17. Burgess S, Thompson DJ, Rees JMB, Day FR, Perry JR, Ong KK. Dissecting causal pathways using mendelian randomization with summarized genetic data: Application to age at menarche and risk of breast cancer. *Genetics*. 2017;207(2):481-487. doi:10.1534/genetics.117.300191

18. Feng Y, Hong X, Wilker E, et al. Effects of age at menarche, reproductive years, and menopause on metabolic risk factors for cardiovascular diseases. *Atherosclerosis*. 2008;196(2):590-597. doi:10.1016/j.atherosclerosis.2007.06.016

19. Hamajima N, Hirose K, Tajima K, et al. Menarche, menopause, and breast cancer risk: Individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol*. 2012;13(11):1141-1151. doi:10.1016/S1470-2045(12)70425-4

20. Klarin D, Damrauer SM, Cho K, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet*. 2018;50(11):1514-1523. doi:10.1038/s41588-018-0222-9

21. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641-3649. doi:10.1093/hmg/ddy271

22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795

23. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. doi:10.7554/eLife.34408

24. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526(7571):82-90. doi:10.1038/nature14962

25. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol*. December 2018. doi:10.1093/ije/dyy262

26. Schwarzer G. meta: An R package for meta-analysis. *R News*. 2007;7(3):40-45.

27. Shi H, Mancuso N, Spendlove S, Pasaniuc B. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *Am J Hum Genet*. 2017;101(5):737-751. doi:10.1016/J.AJHG.2017.09.022

28. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*. 2016;32(2):283-285. doi:10.1093/bioinformatics/btv546

29. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48(7):709-717. doi:10.1038/ng.3570

30. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. Li J, ed. *PLOS Genet*. 2017;13(11):e1007081. doi:10.1371/journal.pgen.1007081

31. Guo Y, Warren Andersen S, Shu XO, et al. Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization Analyses of Data from 145,000 Women of European Descent. *PLoS Med*. 2016;13(8). doi:10.1371/journal.pmed.1002105

32. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008;371(November):569-578. doi:10.1016/S0140-6736(08)60269-X

33. Cochran WG. The Combination of Estimates from Different Experiments. *Biometrics*. 1954;10(1):101. doi:10.2307/3001666

34. Consortium Gte. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213. doi:10.1038/nature24277

35. Madsen CM, Varbo A, Nordestgaard BG. Extreme high high-density lipoprotein cholesterol is paradoxically associated with high mortality in men and women: two prospective cohort studies. *Eur Heart J*. 2017;38(32):2478-2486. doi:10.1093/eurheartj/ehx163

36. Burgess S, Davey Smith G. Mendelian Randomization Implicates High-Density Lipoprotein Cholesterol–Associated Mechanisms in Etiology of Age-Related Macular Degeneration. *Ophthalmology*. 2017;124(8):1165-1174. doi:10.1016/j.ophtha.2017.03.042

37. Fan Q, Maranville JC, Fritsche L, et al. HDL-cholesterol levels and risk of age-related macular degeneration: a multiethnic genetic study using Mendelian randomization. *Int J Epidemiol*. 2017;46(6):1891-1902. doi:10.1093/ije/dyx189

38. Llewellyn-Bennett R, Edwards D, Roberts N, Hainsworth AH, Bulbulia R, Bowman L. Post-trial follow-up methodology in large randomised controlled trials: a systematic review. *Trials*. 2018;19(1):298. doi:10.1186/s13063-018-2653-0

39. Orho-Melander M, Hindy G, Borgquist S, et al. Blood lipid genetic scores, the HMGCR gene and cancer risk: a Mendelian randomization study. *Int J Epidemiol*. November 2017. doi:10.1093/ije/dyx237

40. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology*. 2014;25(3):427-435. doi:10.1097/EDE.0000000000000081

41. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet*. 2018;50(12):1728-1734. doi:10.1038/s41588-018-0255-0

42. Silvente-Poirot S, Poirot M. Cholesterol and cancer, in the balance. *Science (80-)*. 2014;343(6178):1445-1446. doi:10.1126/science.1252787

43. Nelson ER. The significance of cholesterol and its metabolite, 27-hydroxycholesterol in breast cancer. *Mol Cell Endocrinol*. 2018;466:73-80. doi:10.1016/J.MCE.2017.09.021