The American Journal of Human Genetics, Volume *109*

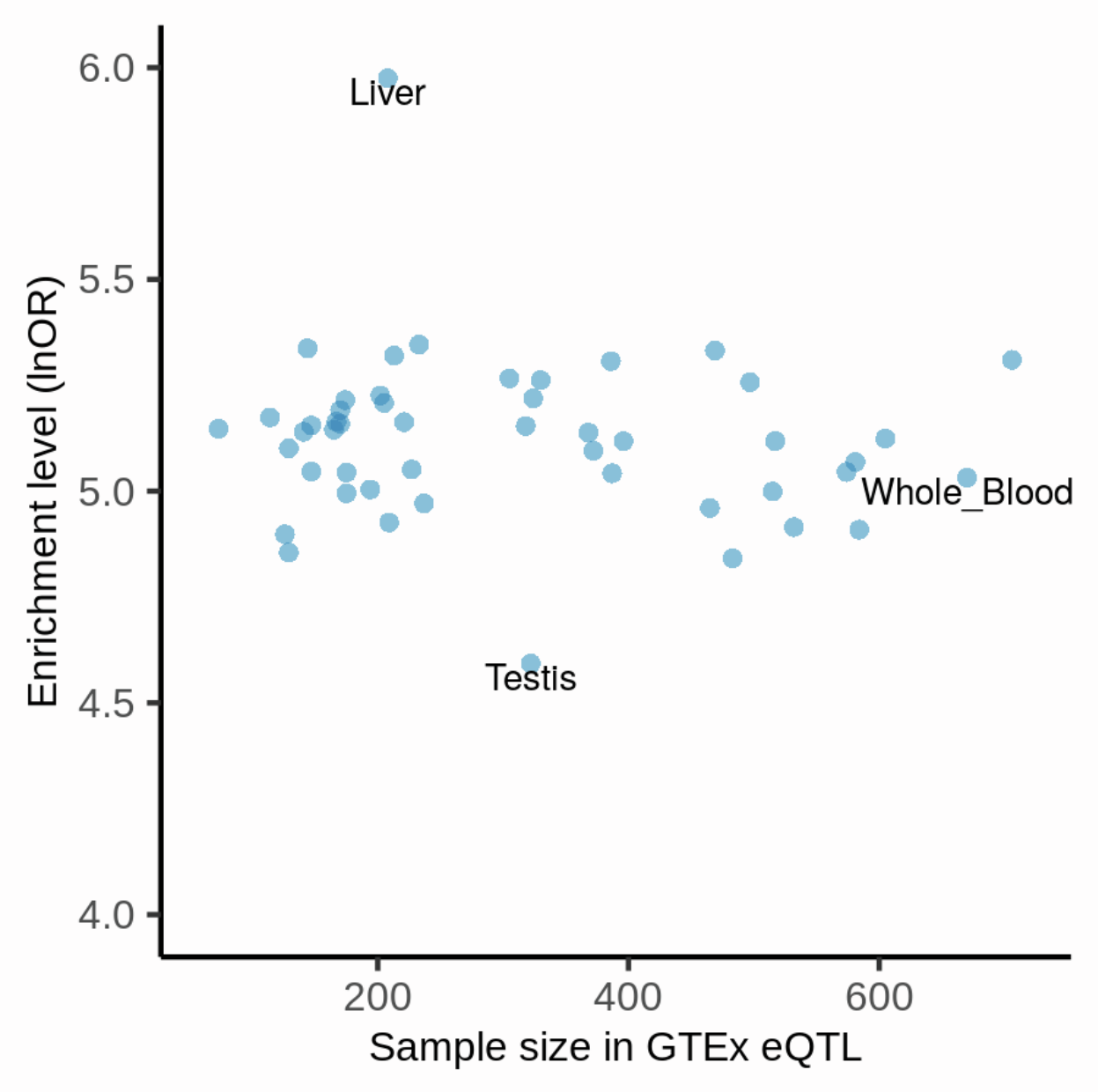
Supplemental information

Integrating transcriptomics, metabolomics, and GWAS helps reveal molecular mechanisms for metabolite levels and disease risk

Xianyong Yin, Debraj Bose, Annie Kwon, Sarah C. Hanks, Anne U. Jackson, Heather M. Stringham, Ryan Welch, Anniina Oravilahti, Lilian Fernandes Silva, FinnGen, Adam E. Locke, Christian Fuchsberger, Susan K. Service, Michael R. Erdos, Lori L. Bonnycastle, Johanna Kuusisto, Nathan O. Stitziel, Ira M. Hall, Jean Morrison, Samuli Ripatti, Aarno Palotie, Nelson B. Freimer, Francis S. Collins, Karen L. Mohlke, Laura J. Scott, Eric B. Fauman, Charles Burant, Michael Boehnke, Markku Laakso, and Xiaoquan Wen

Supplemental Figures

Figure S1. Relationship of GTEx eQTL enrichment levels in METSIM metabQTL with the sample sizes in the GTEx eQTL study.



lnOR: natural logarithm of odds ratio.

Figure S2. Number of associated protein-coding genes per metabolite identified with FDR<0.05 in PTWAS.

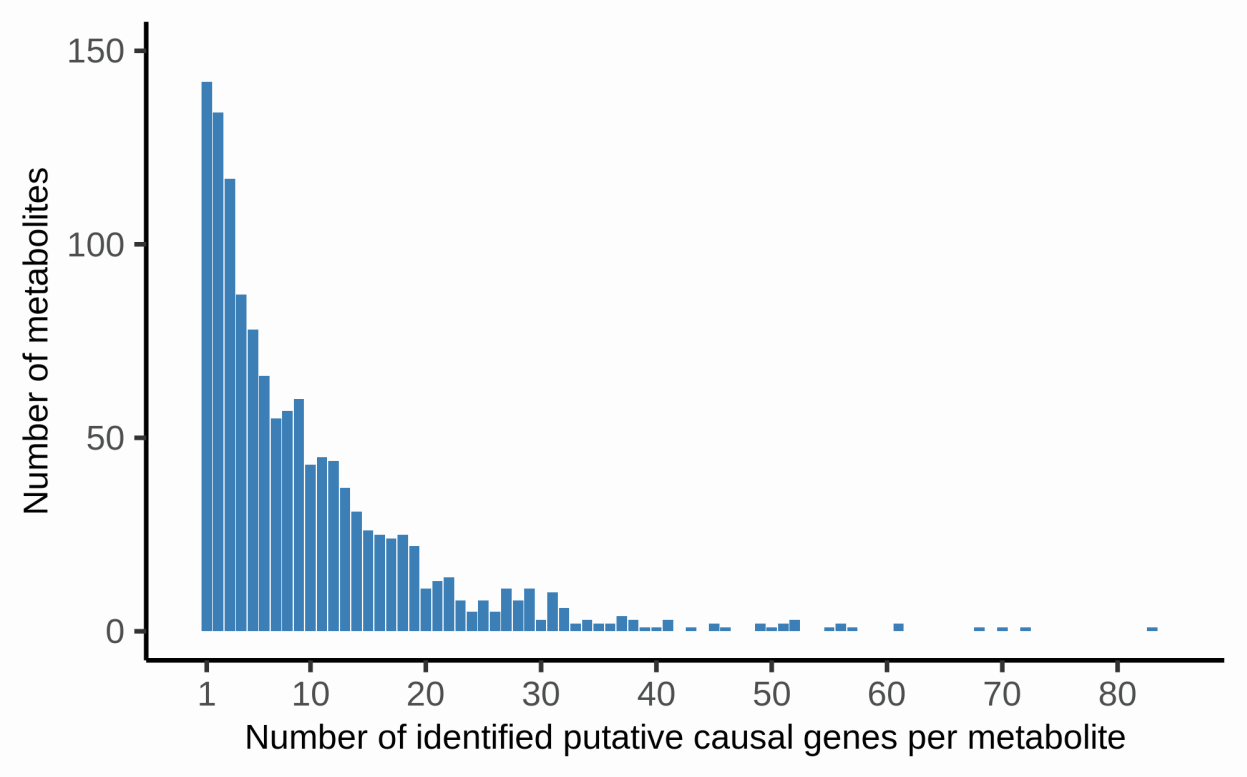
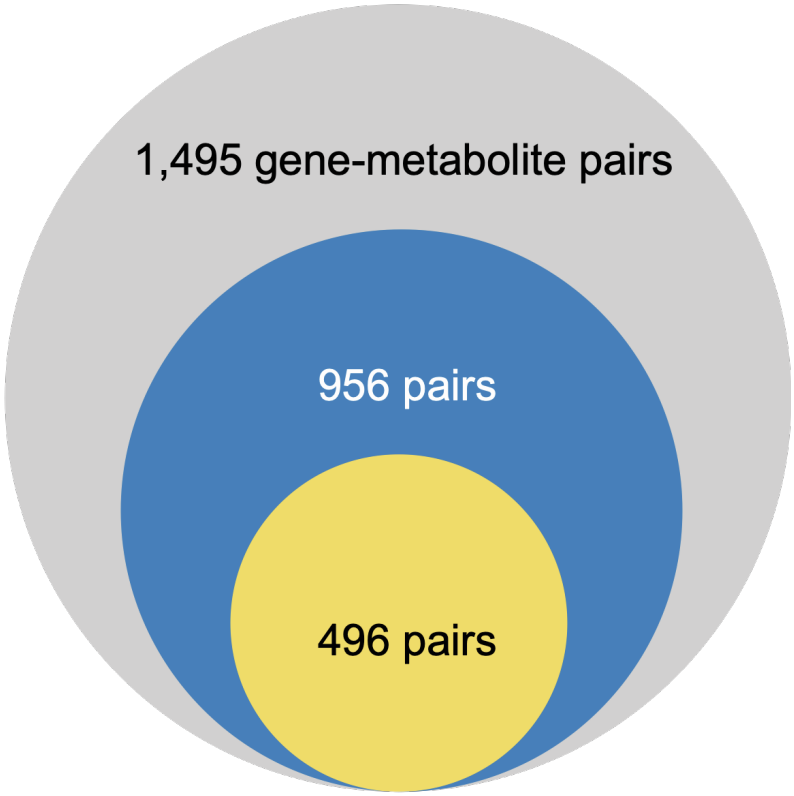
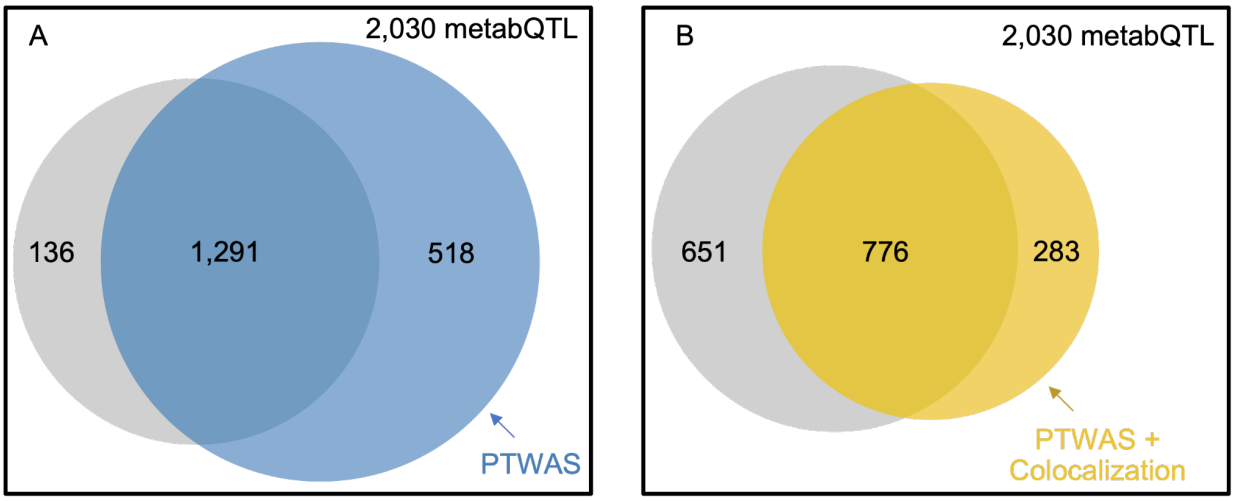


Figure S3. Venn diagram of gene-metabolite pairs identified in Yin et al. 2022[1] and in PTWAS and colocalization analysis.



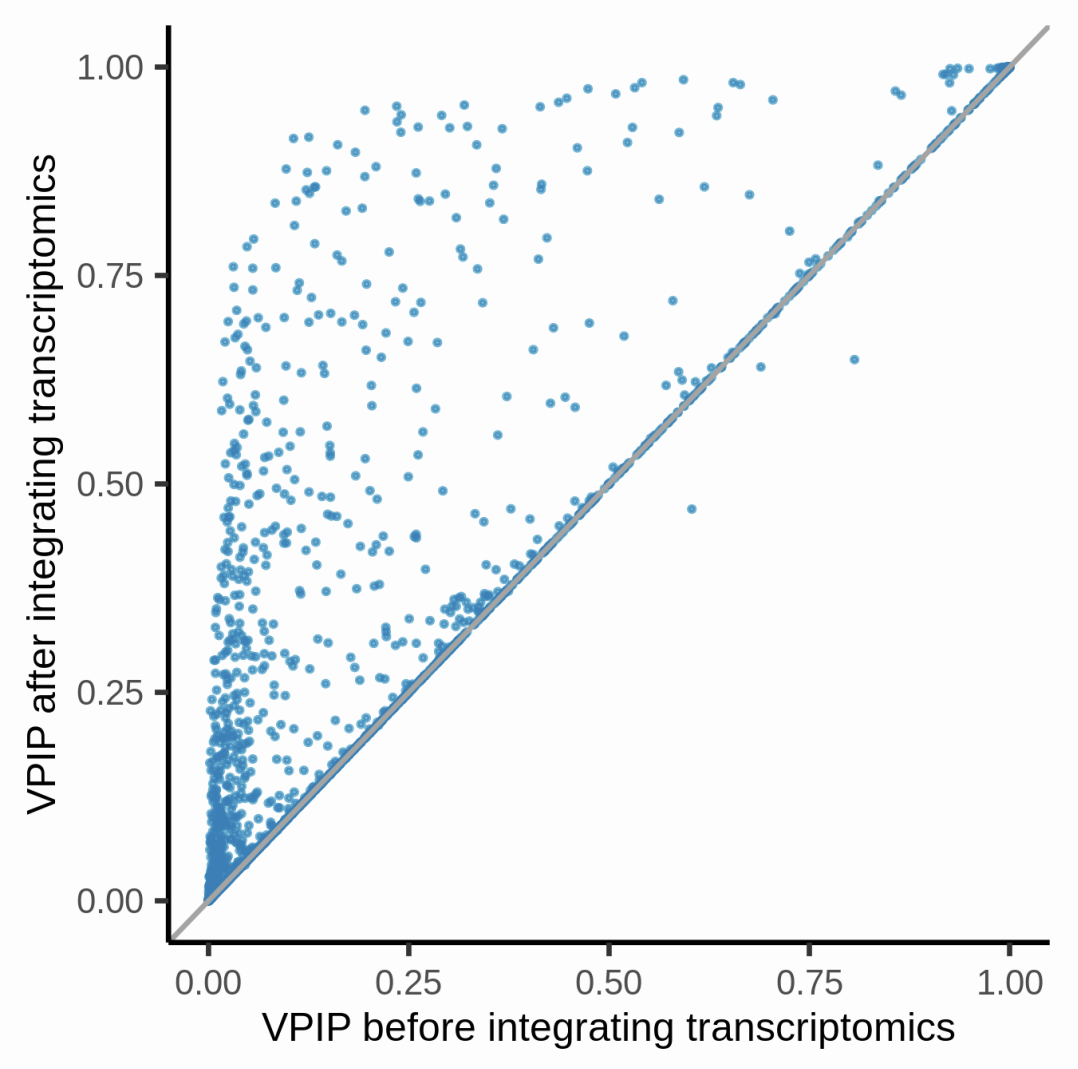
Yin et al. 2022[1] identified 1,495 gene-metabolite pairs by matching metabolite biochemical activities and nearby gene functions of metabQTL, which we used as ground truths (gray). 956 of them were identified with significance in PTWAS (blue). 496 of the 956 pairs further showed colocalizations (yellow).

Figure S4. Venn diagram of metabQTL with genes nominated in Yin et al. 2022[1] and in PTWAS and colocalization analysis. (A) Intersection of metabQTL with genes nominated in Yin et al. 2022[1] and in PTWAS only. (B) Intersection of metabQTL with genes nominated in Yin et al. 2022[1] and in both PTWAS and colocalization analysis.



The gray circle represents the number of metabQTL with gene nominations in Yin et al. 2022[1].

Figure S5. Colocalizing metabQTL with GTEx eQTL increases metabQTL fine-mapping resolution.



VPIP: variant posterior inclusion probability estimated in DAP-g fine-mapping.

Figure S6. Number of associated protein-coding genes per disease trait (FDR<0.05) in PTWAS.

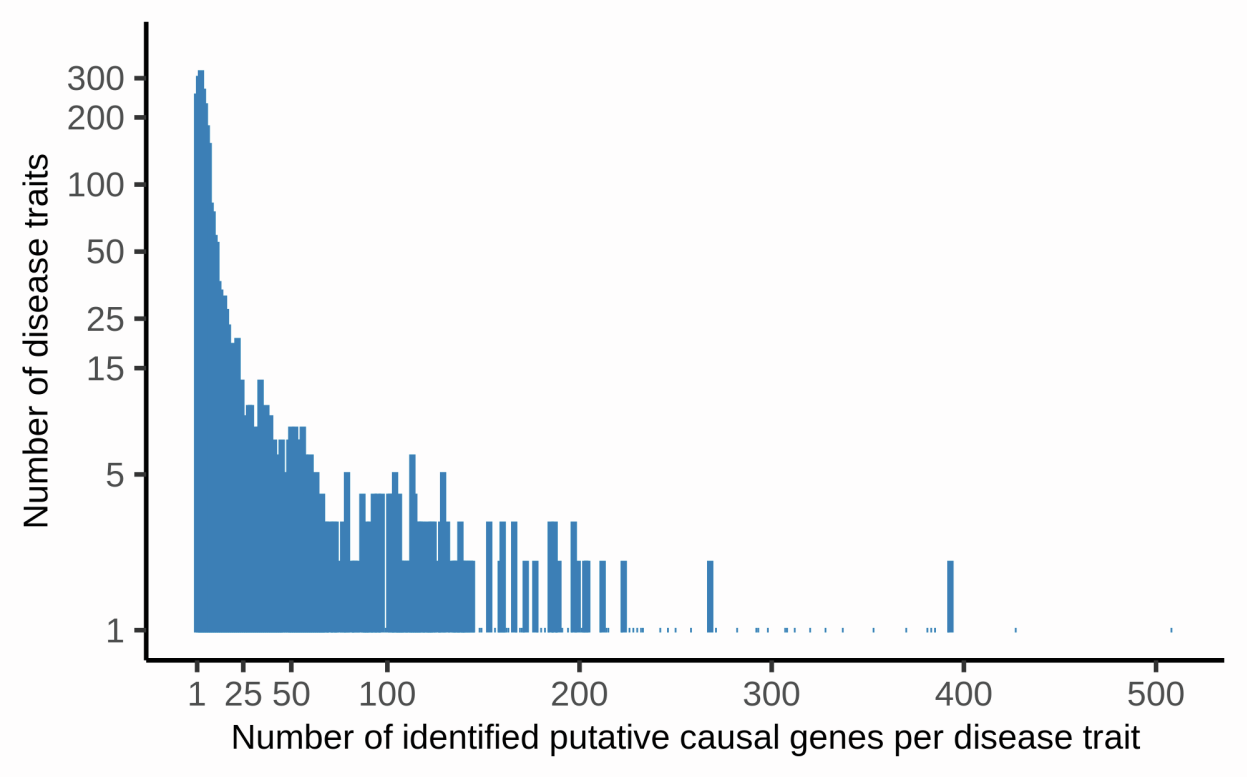
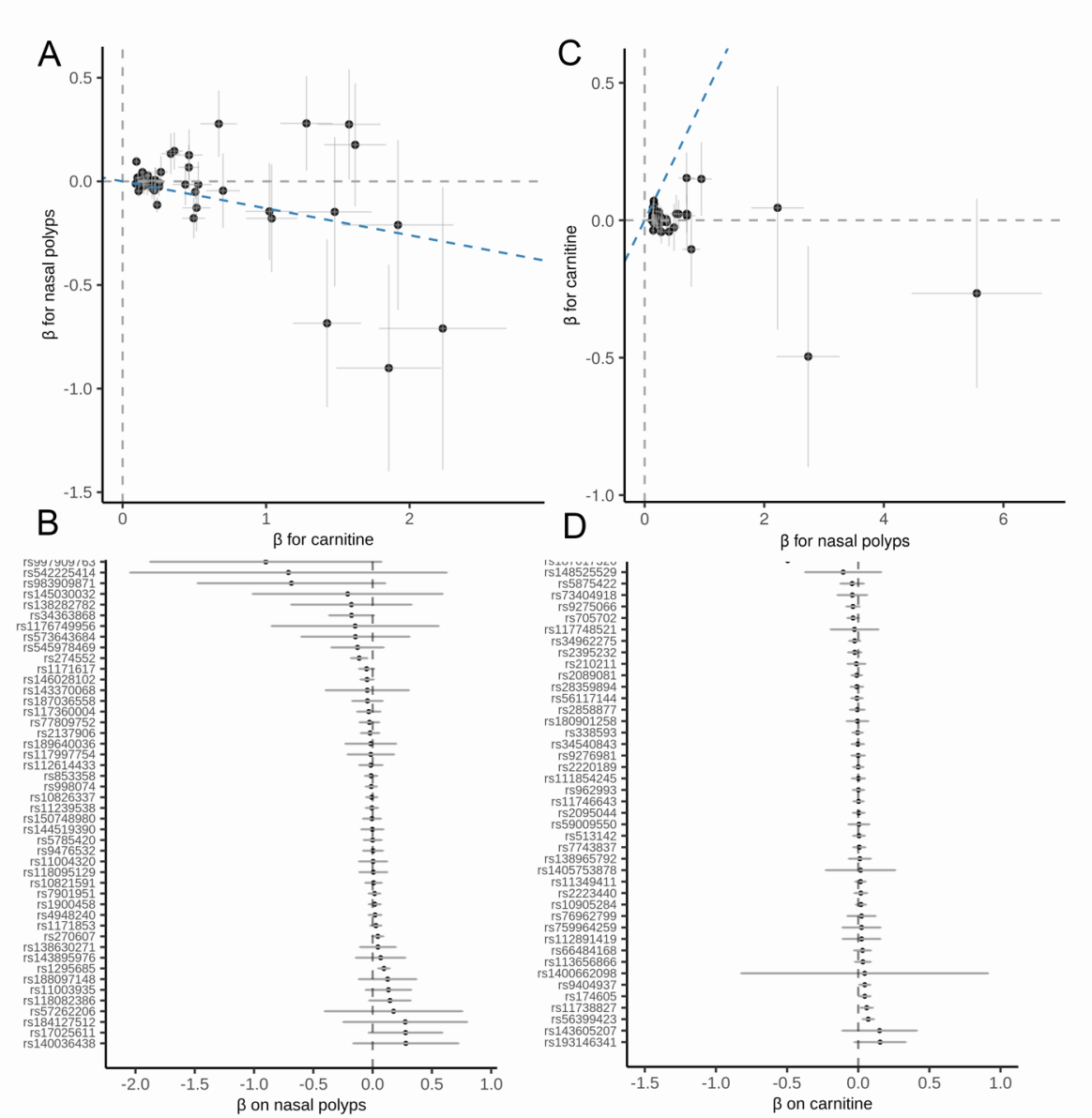


Figure S7. Mendelian randomization identified no causal relationship between plasma carnitine level and the risk of nasal polyps. (A, C) scatter plots of association effect sizes (β) with exposure and outcome for the instrument variables that are used in the Mendelian randomization analysis from plasma carnitine level to nasal polyps risk and vice versa. (B, D) forest plots of effect sizes (β) on nasal polyps and plasma carnitine level for instrument variables.



Each point in (A-D) represents an instrument variable. The vertical and horizontal dashed lines depict β=0 on exposure or outcome. Error bars at each point in (A, C) indicate ± standard error of the association effect of the instrument variable on the exposure and outcome. The slope of the blue dashed line in (A, C) depicts the estimated putative causal effect of exposure on outcome in the Mendelian randomization analysis. The error bar at each point in (B, D) represents the 95% confidence interval of the association effect on the outcome (nasal polyps or carnitine levels).

Figure S8. Phenotype correlations among the ten metabolites with protective effects on the risk of age-related macular degeneration.

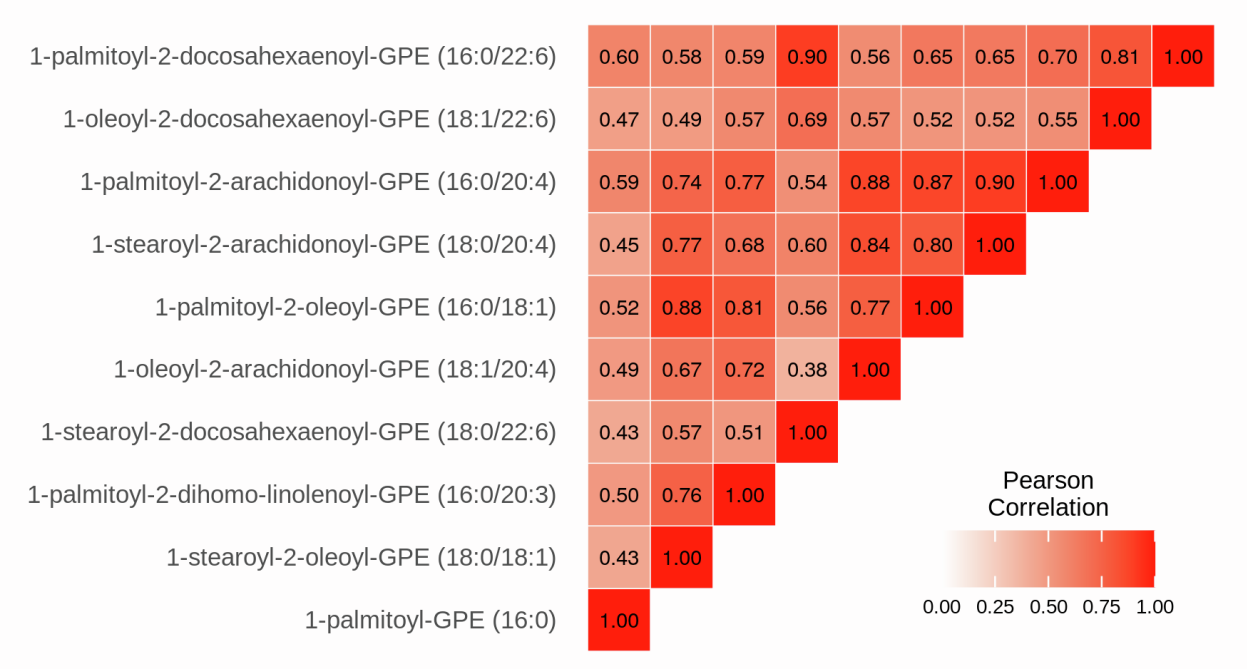
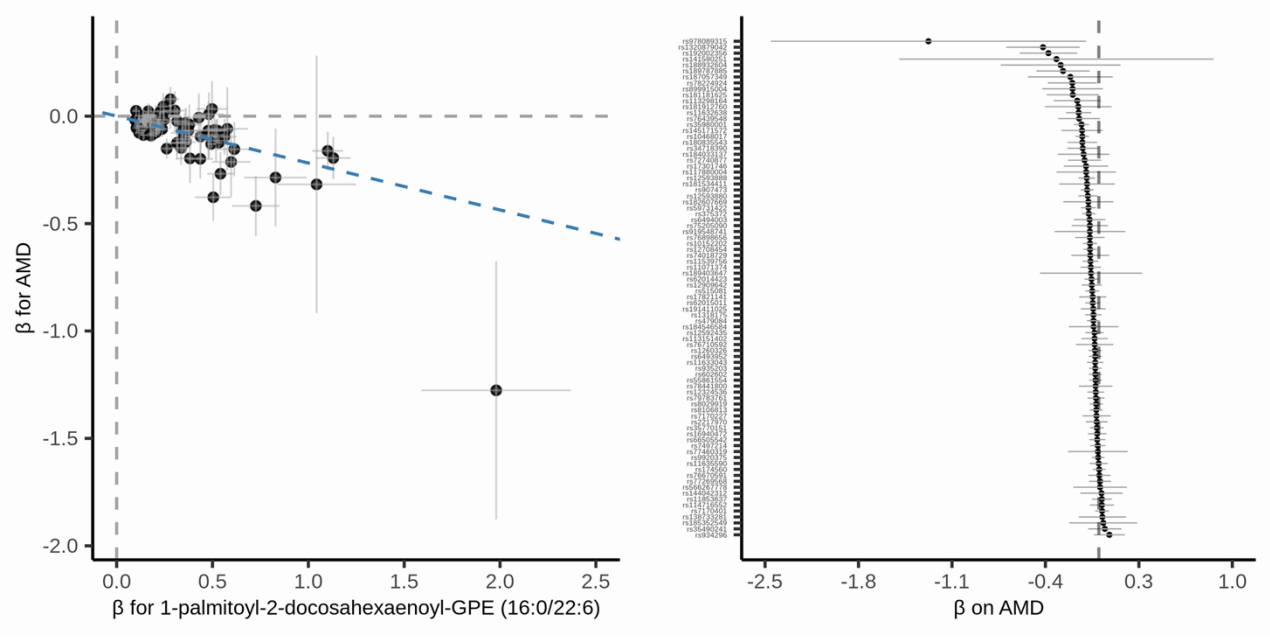
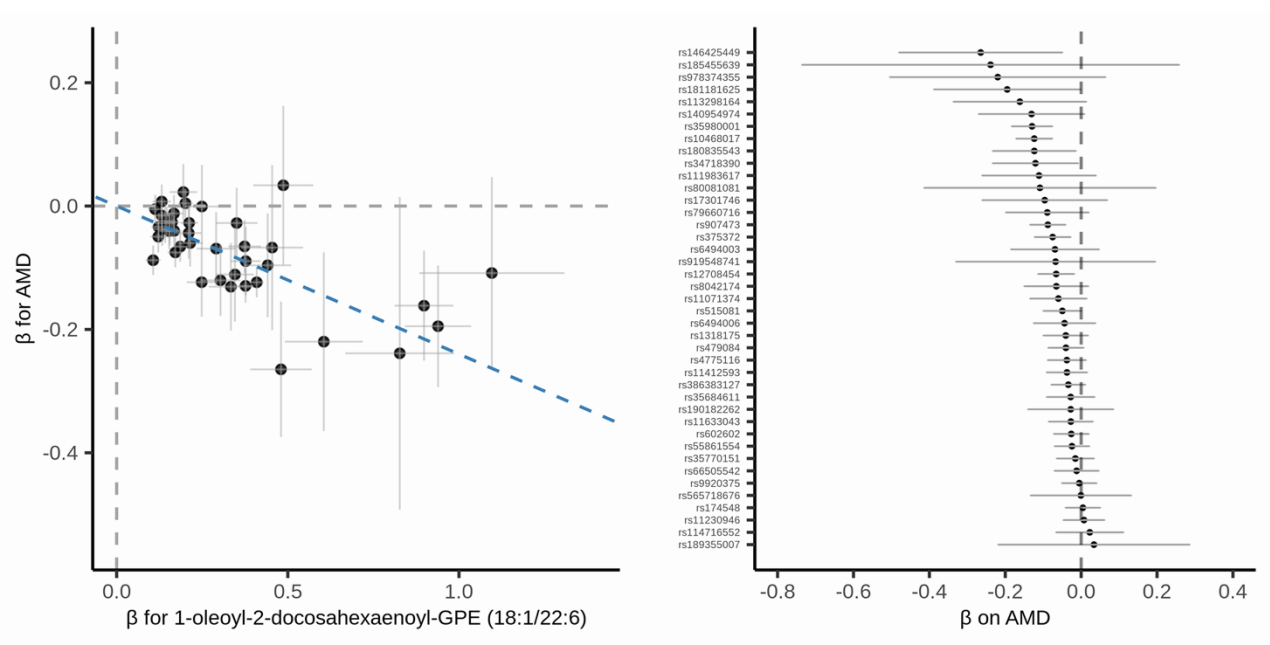


Figure S9. Mendelian randomization infers causal effects of ten plasma metabolites on the risk of age-related macular degeneration (AMD).

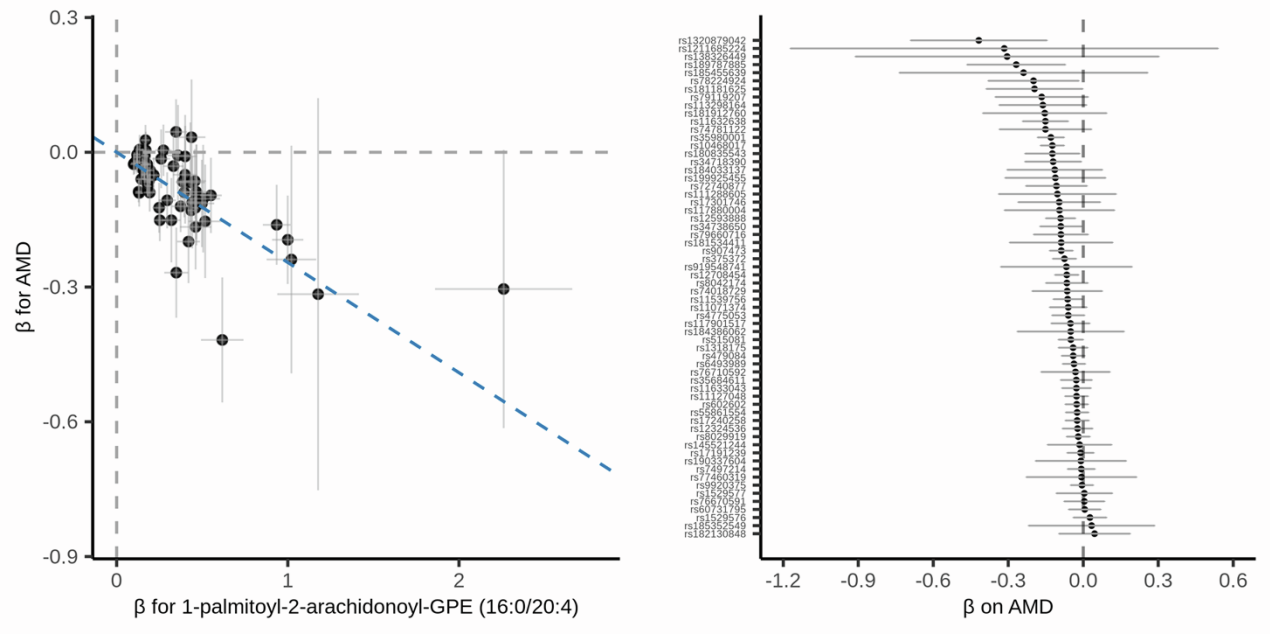
(A) 1-palmitoyl-2-docosahexaenoyl-GPE (16:0/22:6)



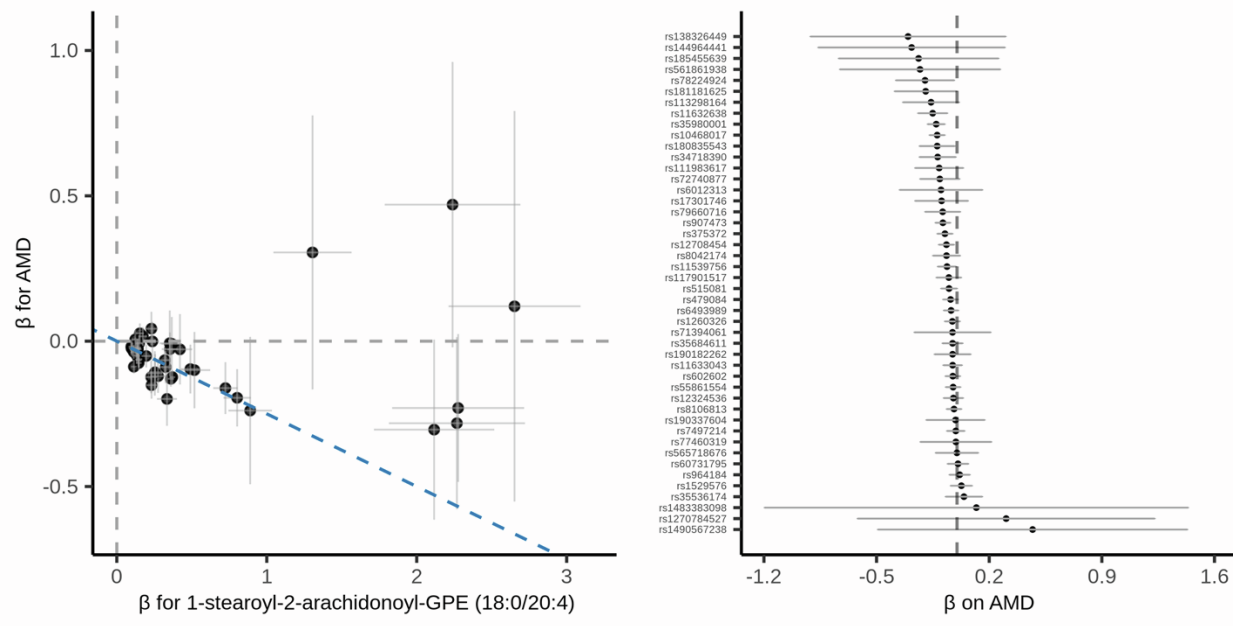
(B) 1-oleoyl-2-docosahexaenoyl-GPE (18:1/22:6)



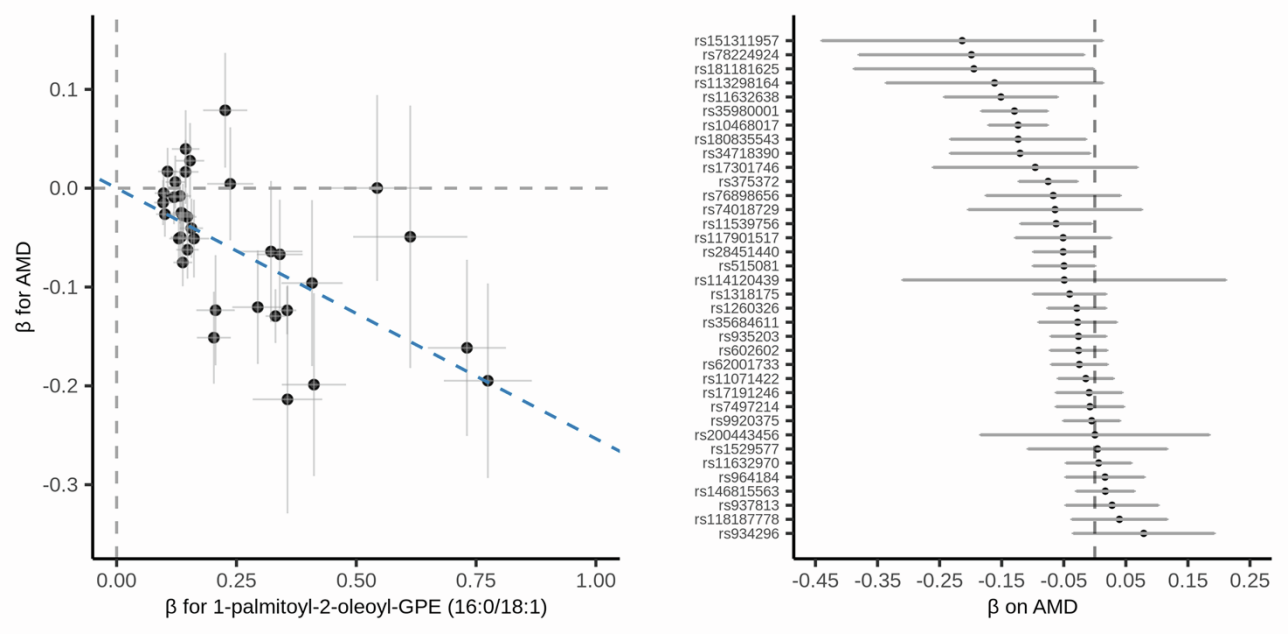
(C) 1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)



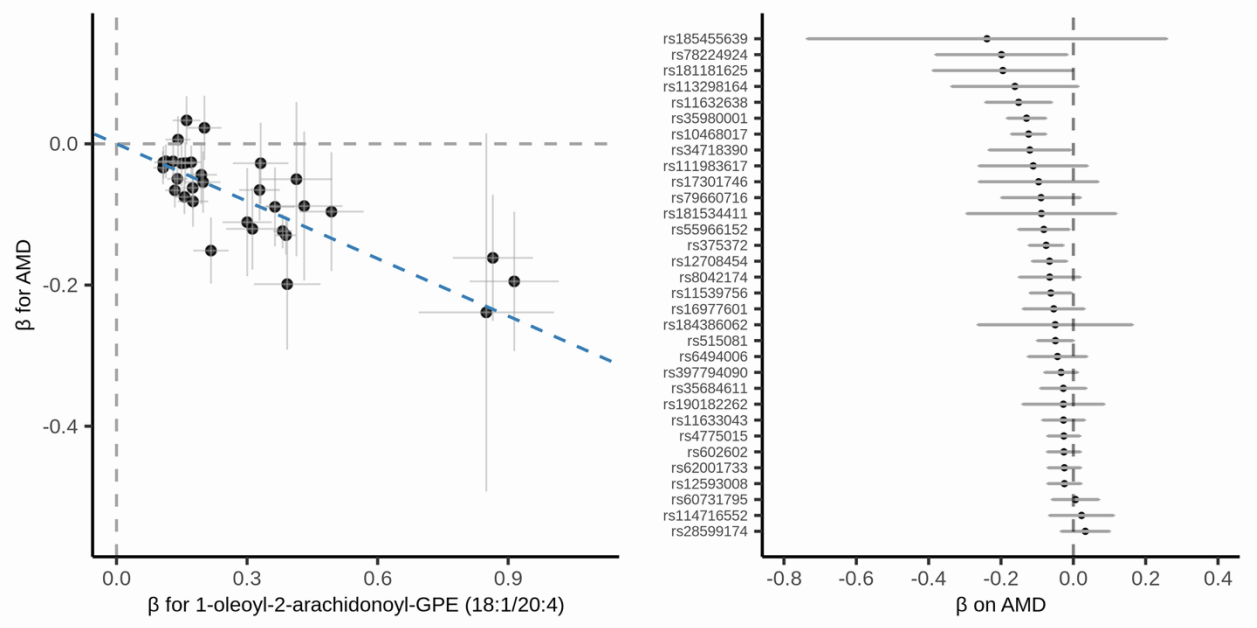
(D) 1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)



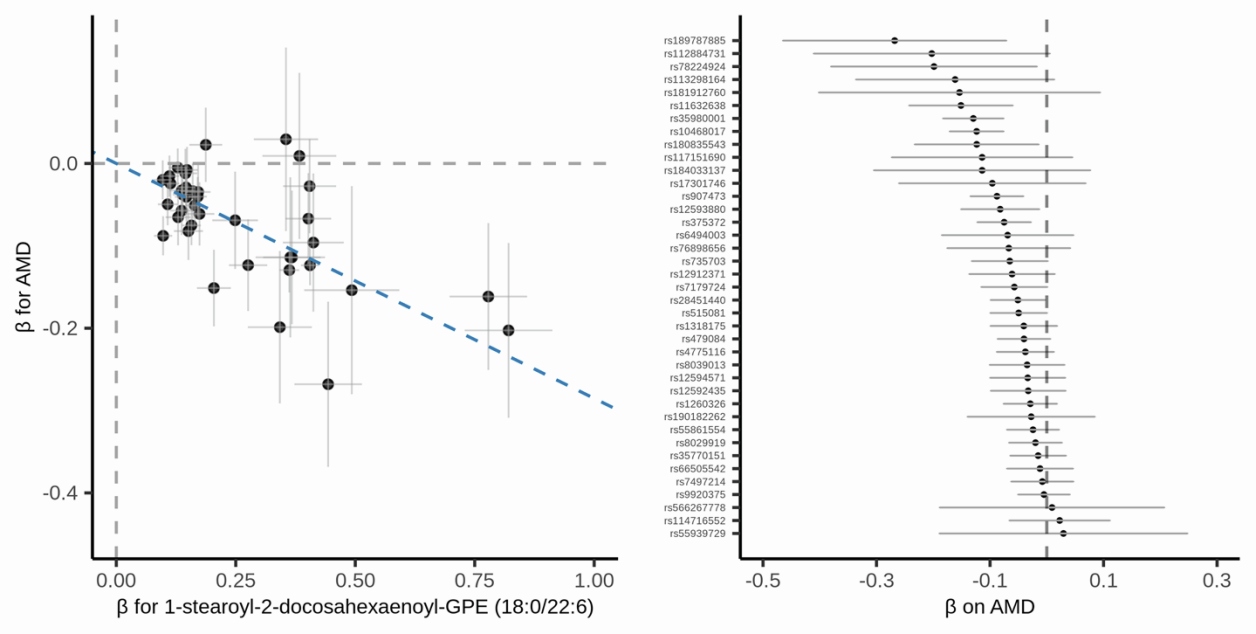
(E) 1-palmitoyl-2-oleoyl-GPE (16:0/18:1)



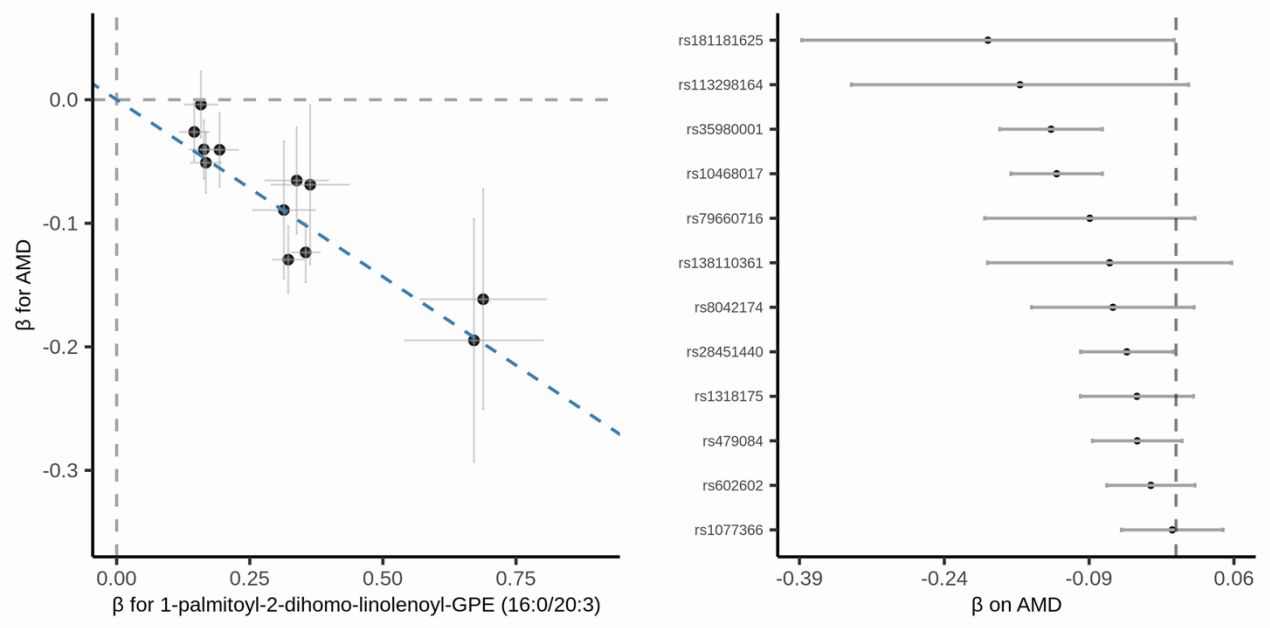
(F) 1-oleoyl-2-arachidonoyl-GPE (18:1/20:4)



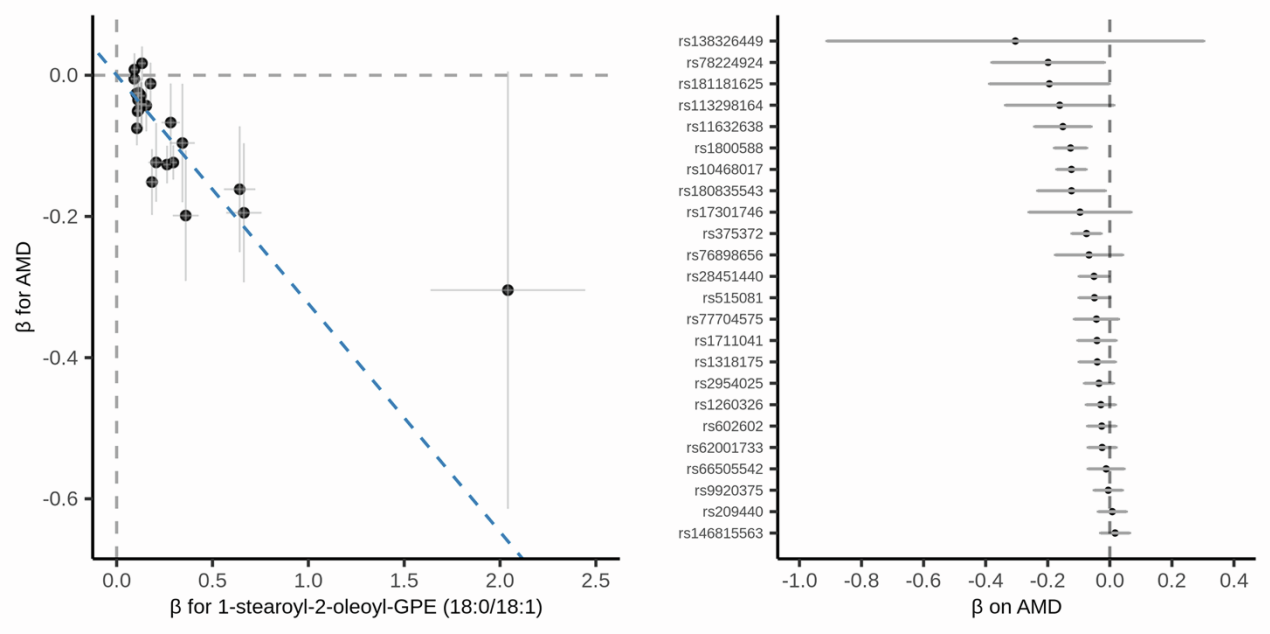
(G) 1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)



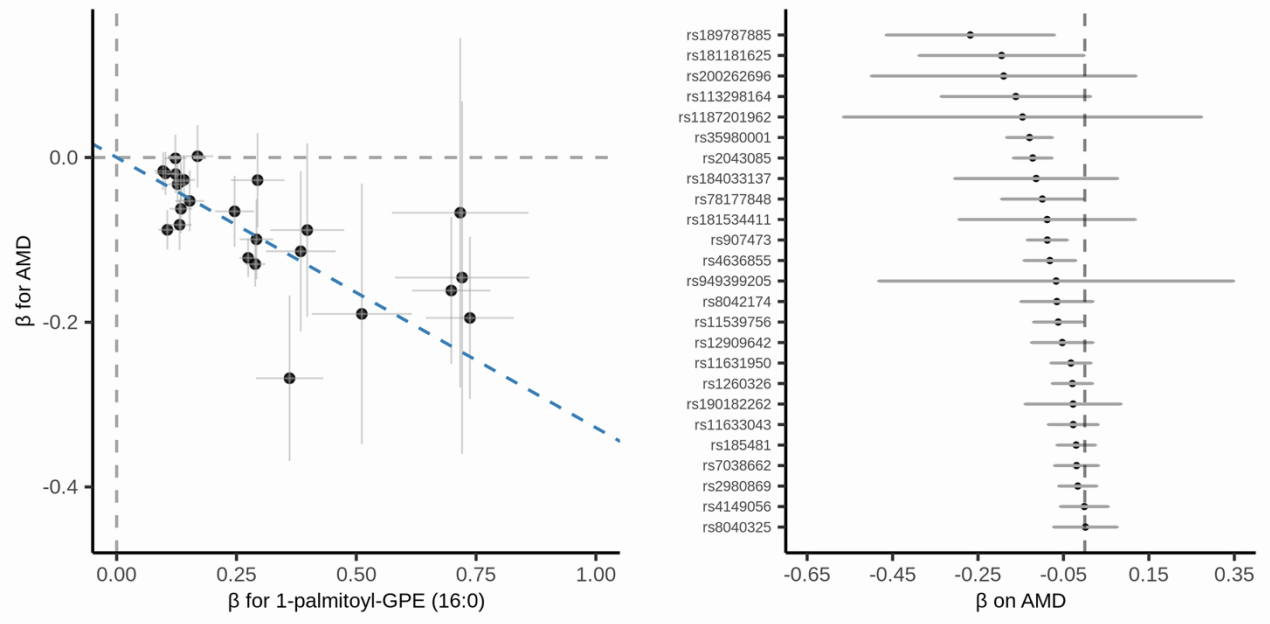
(H) 1-palmitoyl-2-dihomo-linolenoyl-GPE (16:0/20:3)



(I) 1-stearoyl-2-oleoyl-GPE (18:0/18:1)



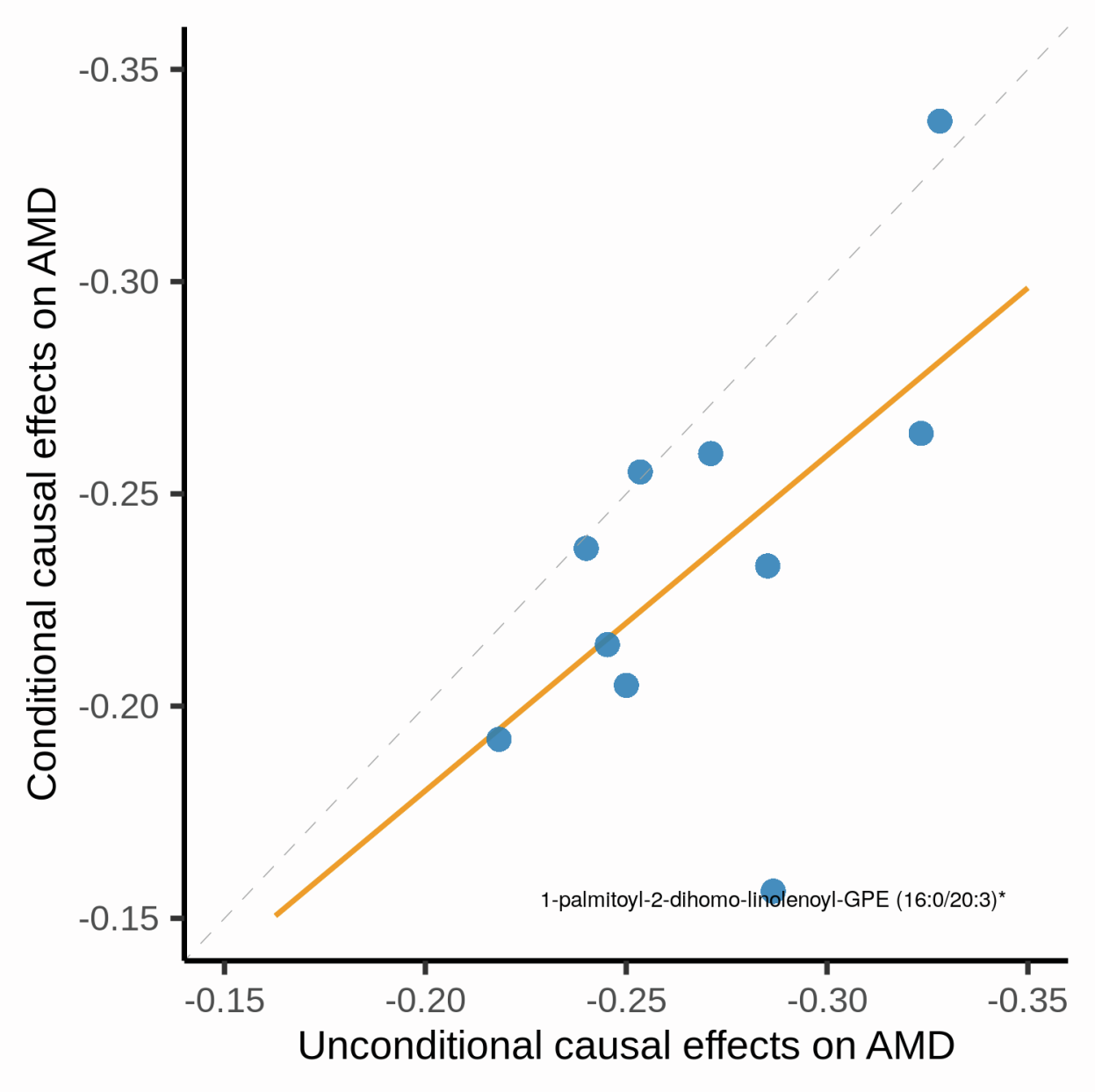
(J) 1-palmitoyl-GPE (16:0)



Each point in the figures represents an instrument variable. The vertical and horizontal dashed lines depict β=0 on exposure or outcome. Error bars at each point in the scatter plots indicate ±standard error of the association effect of the instrument variable on the exposure and outcome. The slope of the blue dashed line in the scatter plots depicts the estimated putative causal effect of corresponding metabolite on AMD in the Mendelian

randomization analysis. The error bar at each point in the forest plots represents the 95% confidence interval of the association effect on AMD.

Figure S10. Causal effects of ten metabolites on the risk of age-related macular degeneration (AMD) before and after controlling for the effects of HDL, LDL, and triglyceride.



Each point depicts a metabolite. The yellow line is the regression line between the two sets of causal effects across the ten metabolites (slope=0.79; R^2=0.33). The dashed one in gray color represents the line of slope one through the origin.

Supplemental References

1. Yin, X., Chan, L.S., Bose, D., Jackson, A.U., VandeHaar, P., Locke, A.E., Fuchsberger, C., Stringham, H.M., Welch, R., Yu, K., et al. (2022). Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. Nat. Commun. *13*, 1644.