

**Genome-wide analysis of blood lipid metabolites in over 5000 South Asians  
reveals biological insights at cardiometabolic disease loci**

**SUPPLEMENTARY MATERIAL**

**Supplementary Tables**

- Supplementary Table 1. Lipids measured in PROMIS
- Supplementary Table 2. Lipids measured in INTERVAL
- Supplementary Table 3. Number of significant variants and loci associated with each lipid subclass
- Supplementary Table 4. Summary of significant associations between lipid metabolites and genetic variants from univariate genome-wide association study in PROMIS
- Supplementary Table 5. Summary of significant associations between lipid metabolites and genetic variants from conditional analyses of univariate genome-wide association study results in PROMIS
- Supplementary Table 6. Summary of significant associations between lipid metabolites and genetic variants from univariate genome-wide association study in INTERVAL
- Supplementary Table 7. Summary and classification of analysed ratios in PROMIS
- Supplementary Table 8. Summary of significant associations between ratios of lipid metabolites and genetic variants from univariate genome-wide association study in PROMIS
- Supplementary Table 9. Prediction of causal genes in PROMIS based on integration of information from bottom-up and top-down SNP annotation approaches
- Supplementary Table 10. Annotation of genome-wide significant associations in PROMIS from the level of the variant (bottom-up)
- Supplementary Table 11. Annotation of genome-wide significant associations in PROMIS by proximal ( $\pm 500$ -Kb from lead variant) gene function (top-down)

- Supplementary Table 12. Lead variants in PROMIS residing in exonic sequence
- Supplementary Table 13. Lead variants in PROMIS in high linkage disequilibrium ( $r^2 \geq 0.8$ ) with  $\geq 1$  non-synonymous SNP
- Supplementary Table 14. Lead SNP *cis*-eQTLs in PROMIS in lipid-relevant human tissues
- Supplementary Table 15. Enrichment analysis of cell-type specific enhancer overlap with our set of 90 lead SNPs from conditional analyses in PROMIS using HaploReg v4.1
- Supplementary Table 16. Prediction of causal genes in INTERVAL based on integration of information from bottom-up and top-down SNP annotation approaches
- Supplementary Table 17. Summary of associations between lipid metabolites in PROMIS and 175 major lipid loci
- Supplementary Table 18. Summary of significant associations between lipid metabolites and genetic variants from univariate genome-wide association study in PROMIS with adjustment for clinical lipid measures

### **Supplementary Figures**

- Supplementary Figure 1. Extended heat map showing associations of significant loci from conditional analyses in PROMIS with all lipid metabolites, major lipids, and lipid-related diseases/disorders
- Supplementary Figure 2. Genetic architecture of serum lipid levels in PROMIS
- Supplementary Figure 3. Comparison of associations in PROMIS and INTERVAL
- Supplementary Figure 4. Number of lipids in PROMIS associated with each variant
- Supplementary Figure 5. Increased *de novo* lipogenesis in lipodystrophy and NAFLD patients
- Supplementary Figure 6. Flow diagram outlining strategy for mediating gene prioritisation
- Supplementary Figure 7. Association of lipids in PROMIS with significantly associated loci from conditional analyses

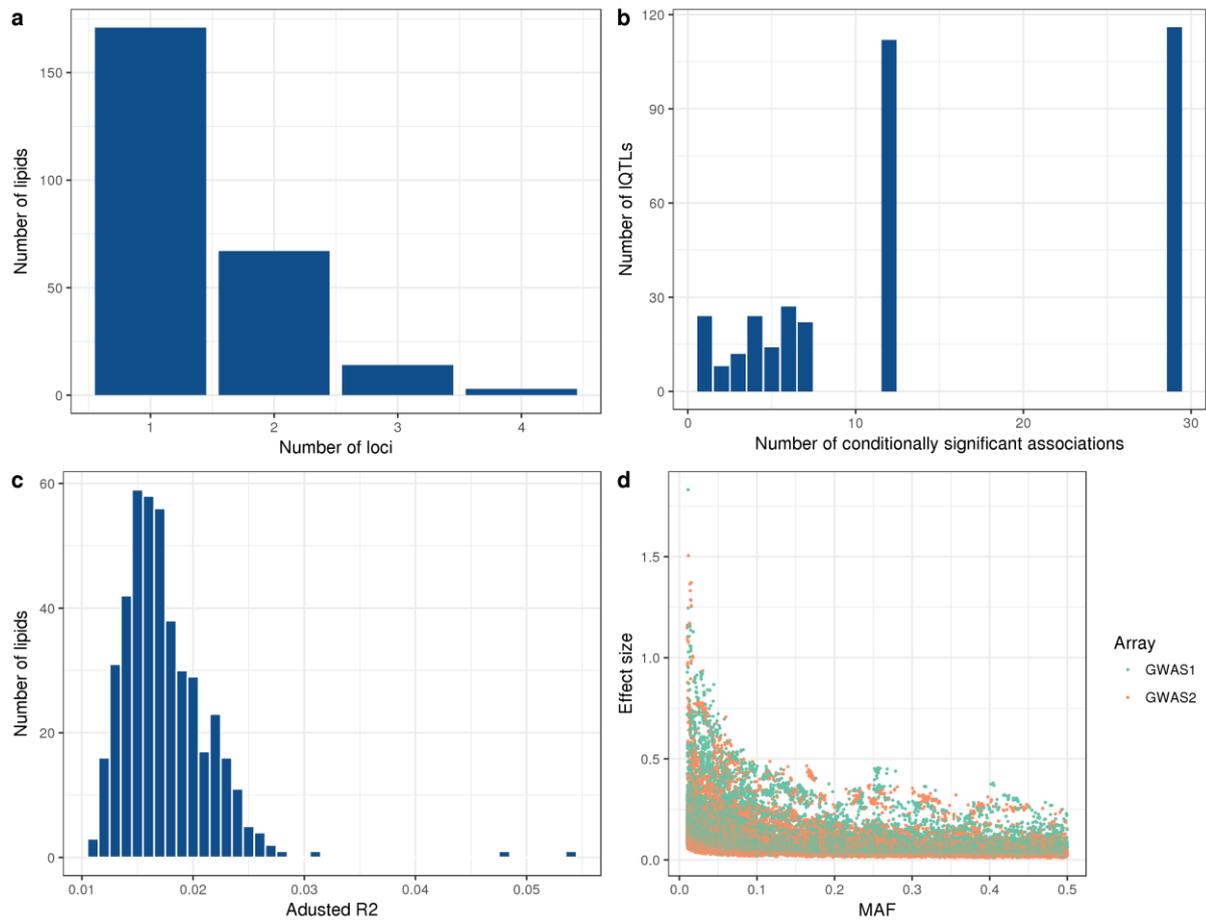
Supplementary Figure 8. Comparison of genetic associations with lipids pre- and post-adjustment for clinical lipid measures



**Caption for Supplementary Figure 1 (previous page):**

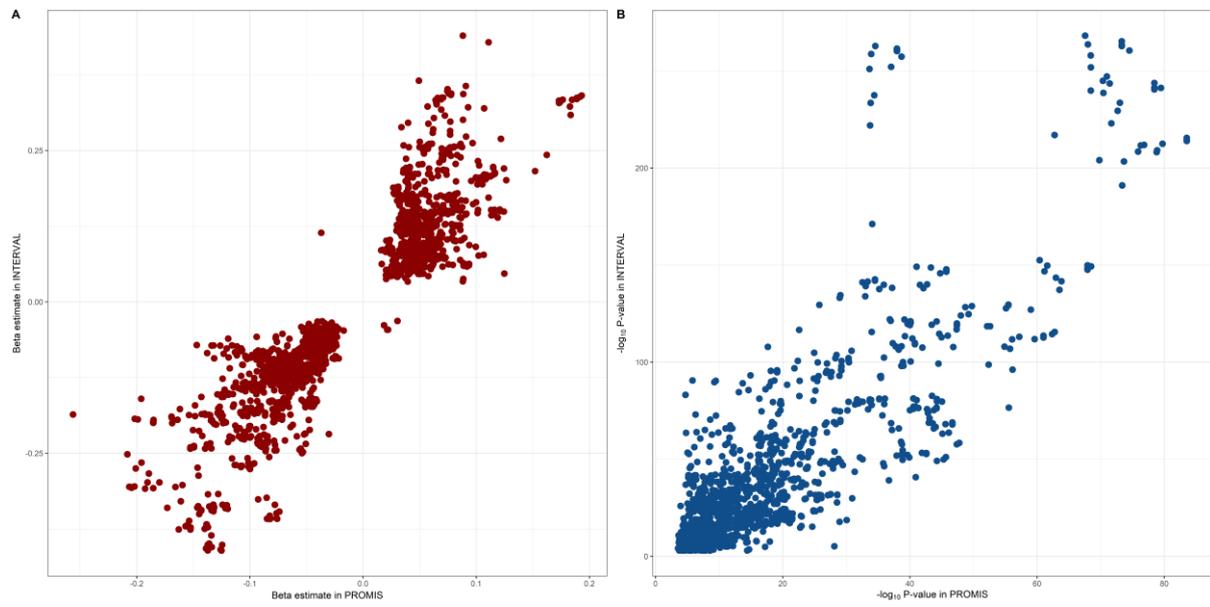
The effect estimates of the associations between significant variants with all lipid metabolites in PROMIS are plotted as a heat map. Results are shown for the association of all lipid metabolites (rows) with the most strongly associated genetic variant within each locus (columns). The associations with major lipids from the GLGC (total cholesterol, HDL-C, LDL-C, and triglycerides), DIAGRAM Consortium (type 2 diabetes), and CARDIoGRAMplusC4D Consortium (coronary artery disease) are also shown. The magnitude and direction of the effect estimates (standardised per 1-SD) are indicated by a colour scale, with blue indicating a negative association and red indicating a positive association with respect to the SNP effect on the trait. Asterisks indicates the degree of significance of the  $P$ -values of association. \* =  $P < 1 \times 10^{-4}$ ; \*\* =  $P < 5 \times 10^{-8}$ ; \*\*\* =  $P < 8.9 \times 10^{-10}$ . **Note:** A high-resolution version of this figure is available as Additional file 3.

## Supplementary Figure 2. Genetic architecture of serum lipid levels in PROMIS



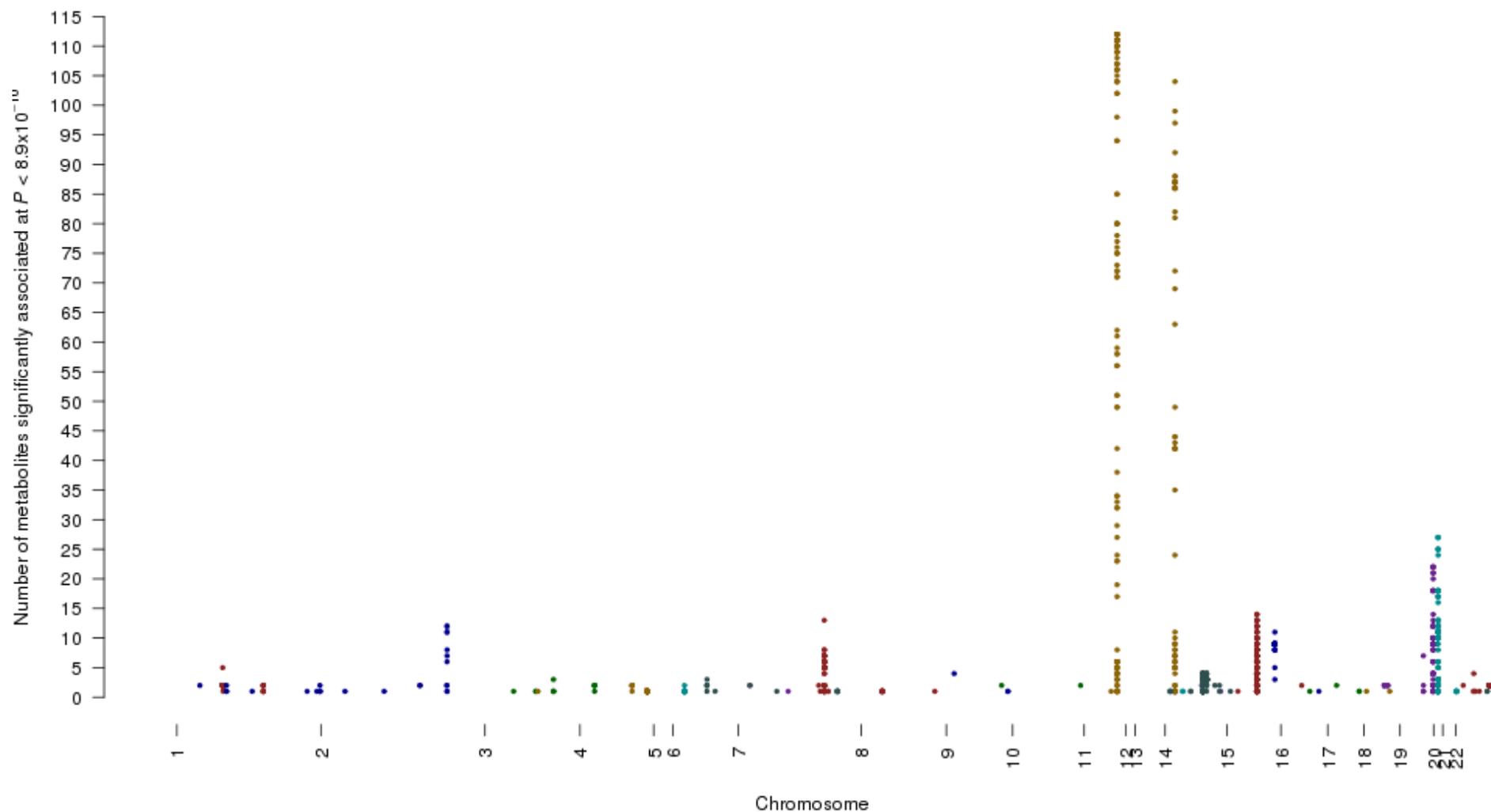
n = 5,662 participants from PROMIS. (a) Number of significantly associated loci per lipid. (b) Number of conditionally significant associations within each lipid QTL. (c) Histogram of variance explained by conditionally independent variants. (d) Effect size versus MAF.

**Supplementary Figure 3.** Comparison of associations in PROMIS and INTERVAL



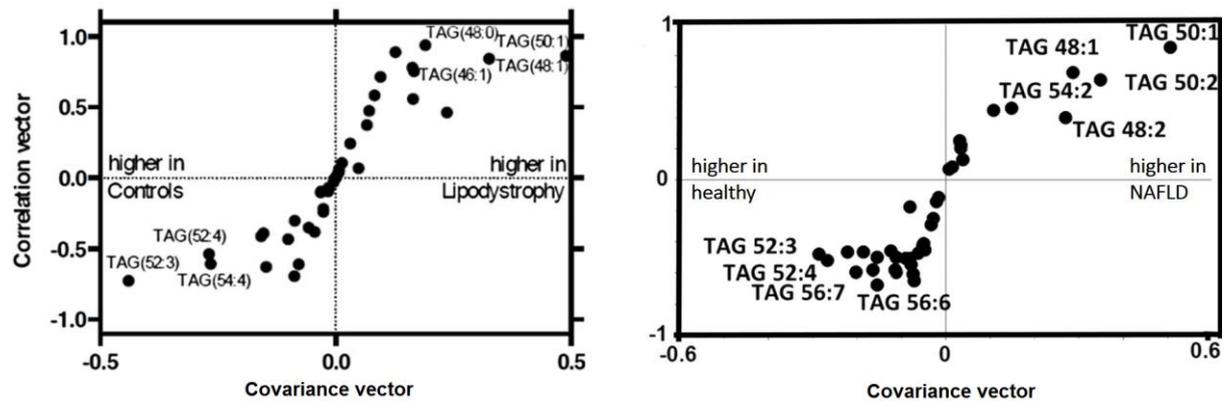
(a) Beta estimates in INTERVAL vs PROMIS. (b)  $-\log_{10}$   $P$ -values in INTERVAL vs PROMIS.

**Supplementary Figure 4.** Number of lipids in PROMIS associated with each variant



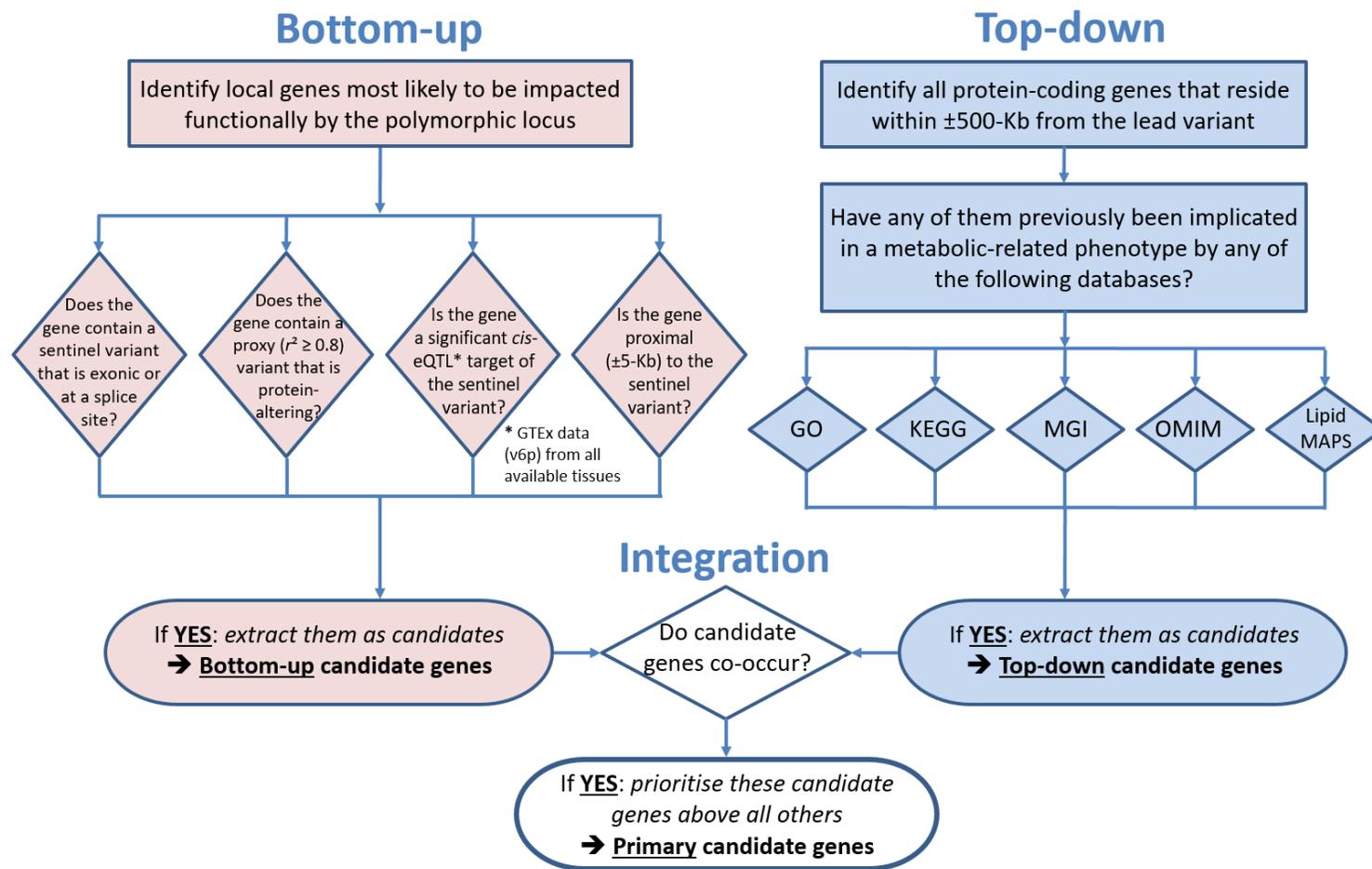
Results are shown for the number of lipids in PROMIS associated with each variant at genome-wide significance ( $P < 8.9 \times 10^{-10}$ ).

**Supplementary Figure 5.** Increased *de novo* lipogenesis in lipodystrophy and NAFLD patients



Increased *de novo* lipogenesis in lipodystrophy patients (left panel based on Eiden et al 2015) [49] and NAFLD patients (right panel based on Sanders et al 2018) [50] both show an increase in TAG(48:1) and TG(50:1) originating from the liver, leading to lower levels of TG(52:4) and other triglycerides associated with *APOA5-C3* variants.

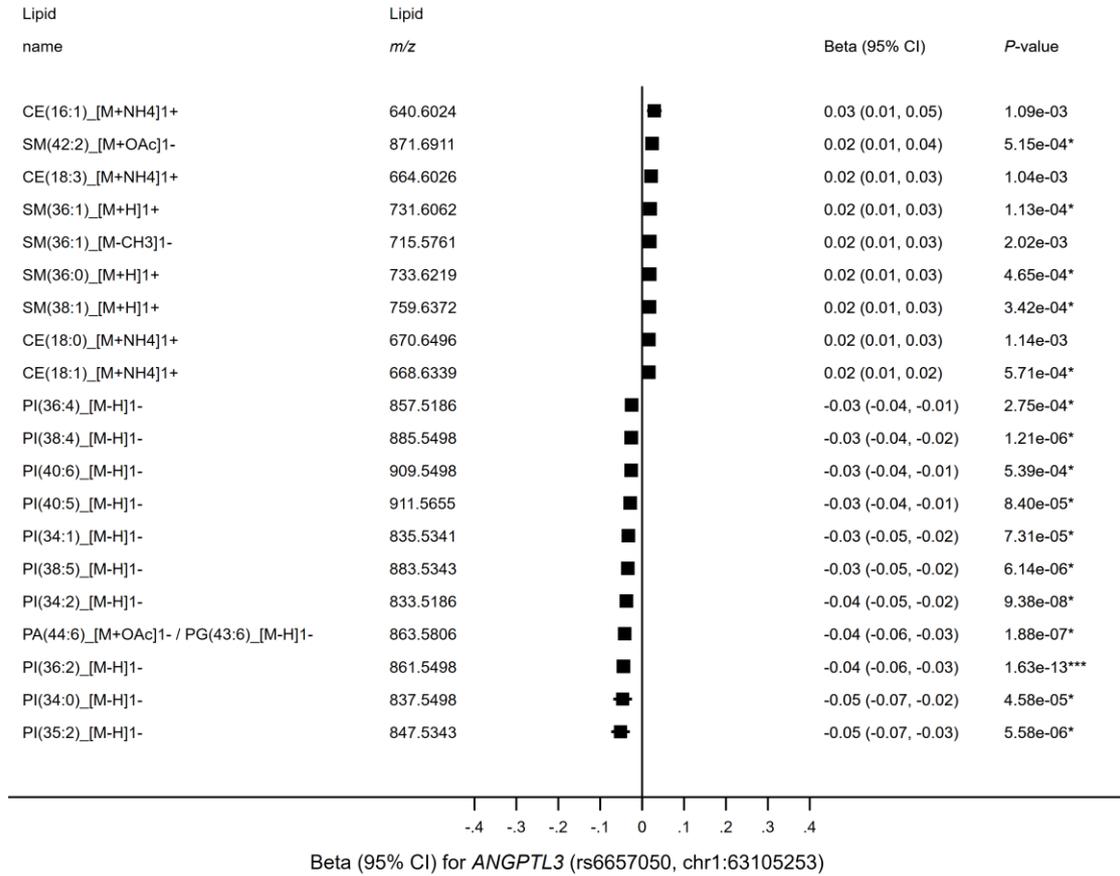
**Supplementary Figure 6.** Flow diagram outlining strategy for mediating gene prioritisation



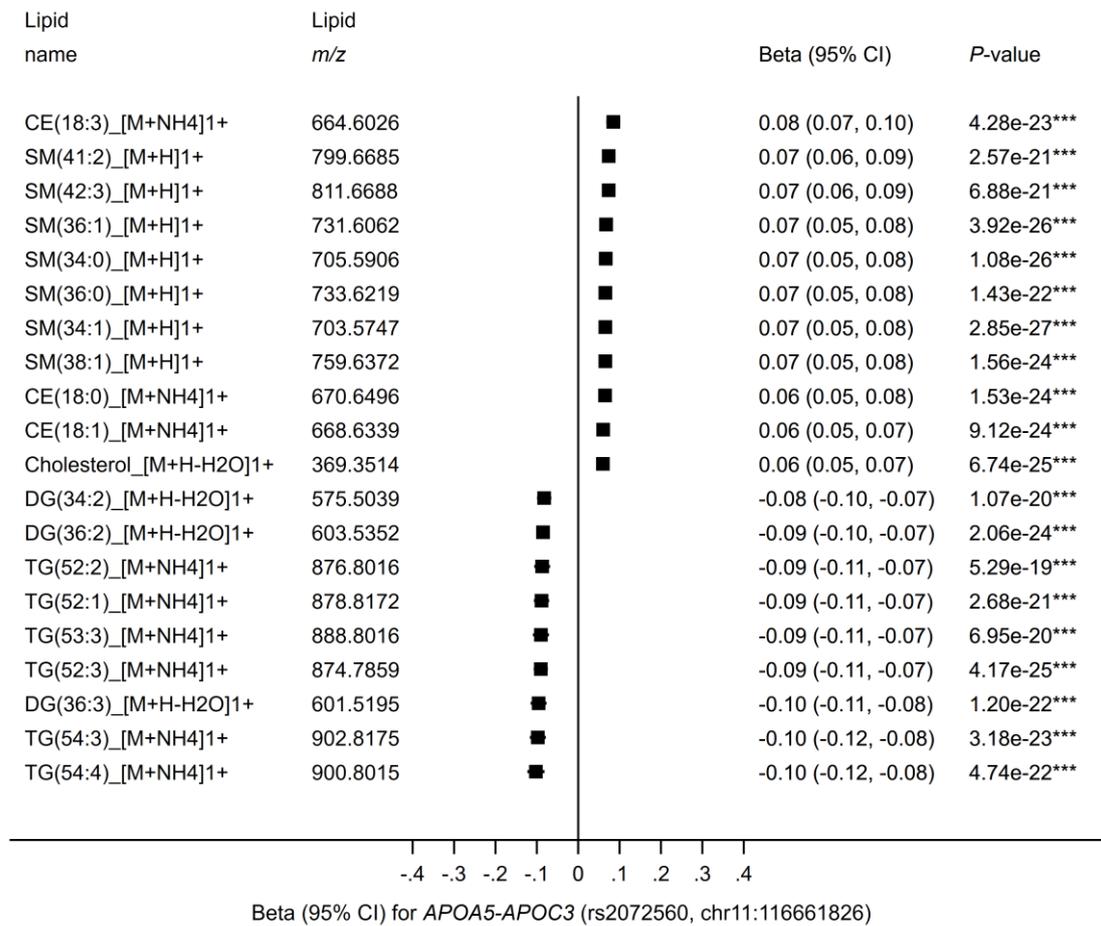
The flow diagram shows how the “bottom-up” and “top-down” approaches were used and then integrated to identify probable causal genes for each significantly associated variant. A proxy is defined as those variants with  $r^2 \geq 0.8$  with the lead (EUR population, 1000 Genomes). **Abbreviations:** eQTL = Expression Quantitative Trait Locus; **GO** = Gene Ontology; **GTEx** = Genotype-Tissue Expression; **KEGG** = Kyoto Encyclopedia of Genes and Genomes; **Lipid MAPS** = Lipid Metabolites and Pathways Strategy; **MGI** = Mouse Genome Informatics; **OMIM** = Online Inheritance in Man. Adapted from Stacey et al [14].

**Supplementary Figure 7.** Association of lipids in PROMIS with significantly associated loci from conditional analyses

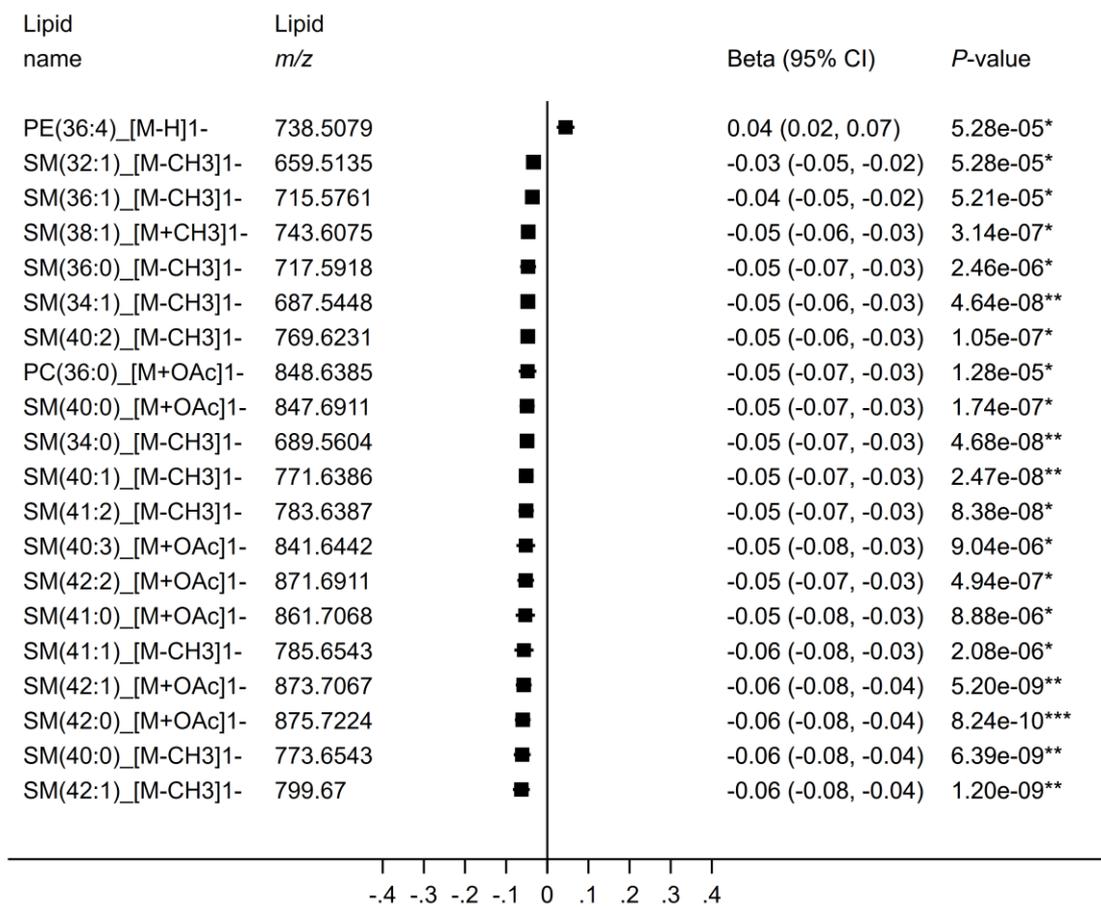
**(a) *ANGPTL3***



**(b) APOA5-APOC3**

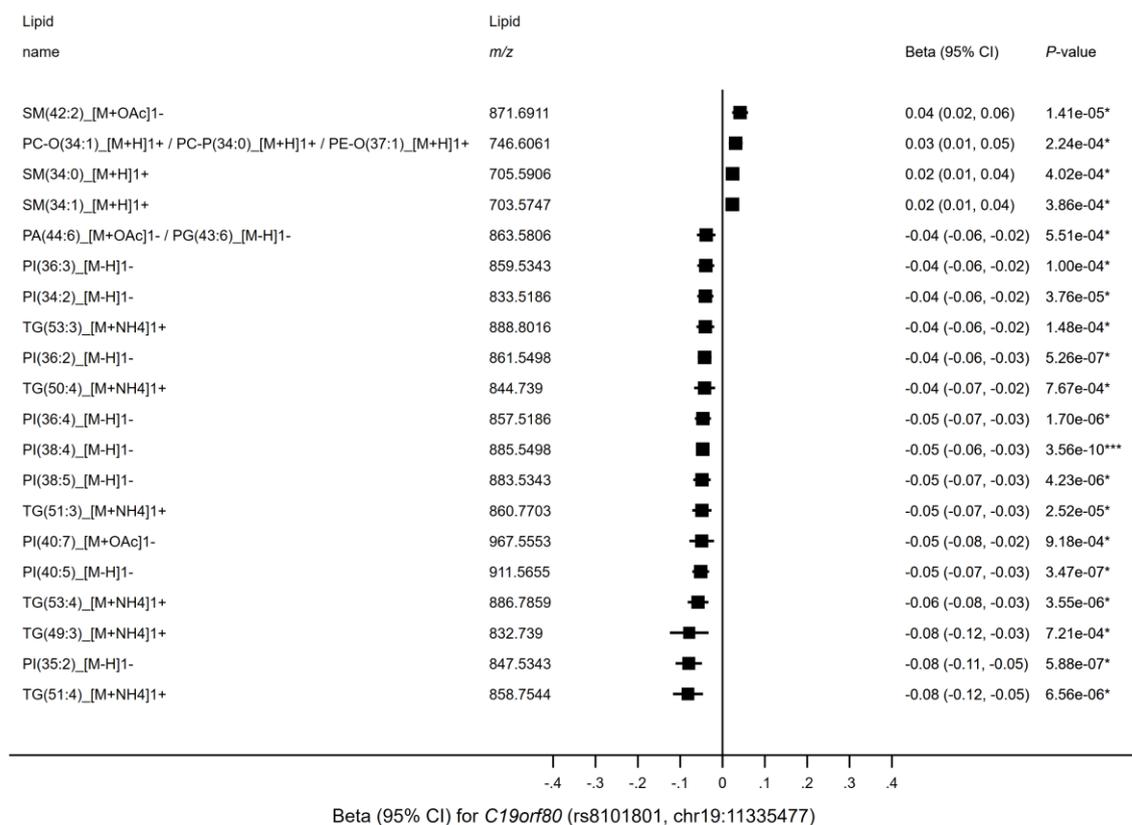


**(c) APOE-APOC1-APOC2-APOC4**

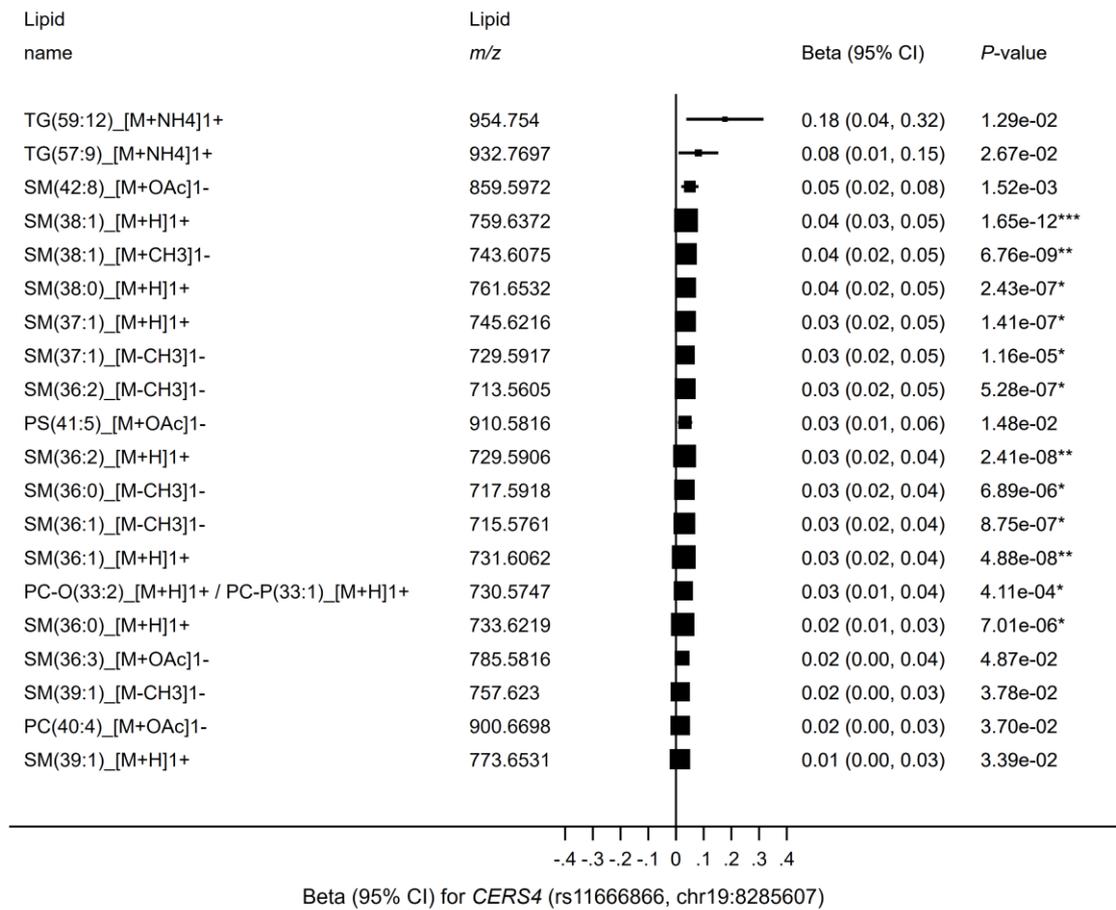


Beta (95% CI) for APOE-APOC1-APOC2-APOC4 (rs75627662, chr19:45413576)

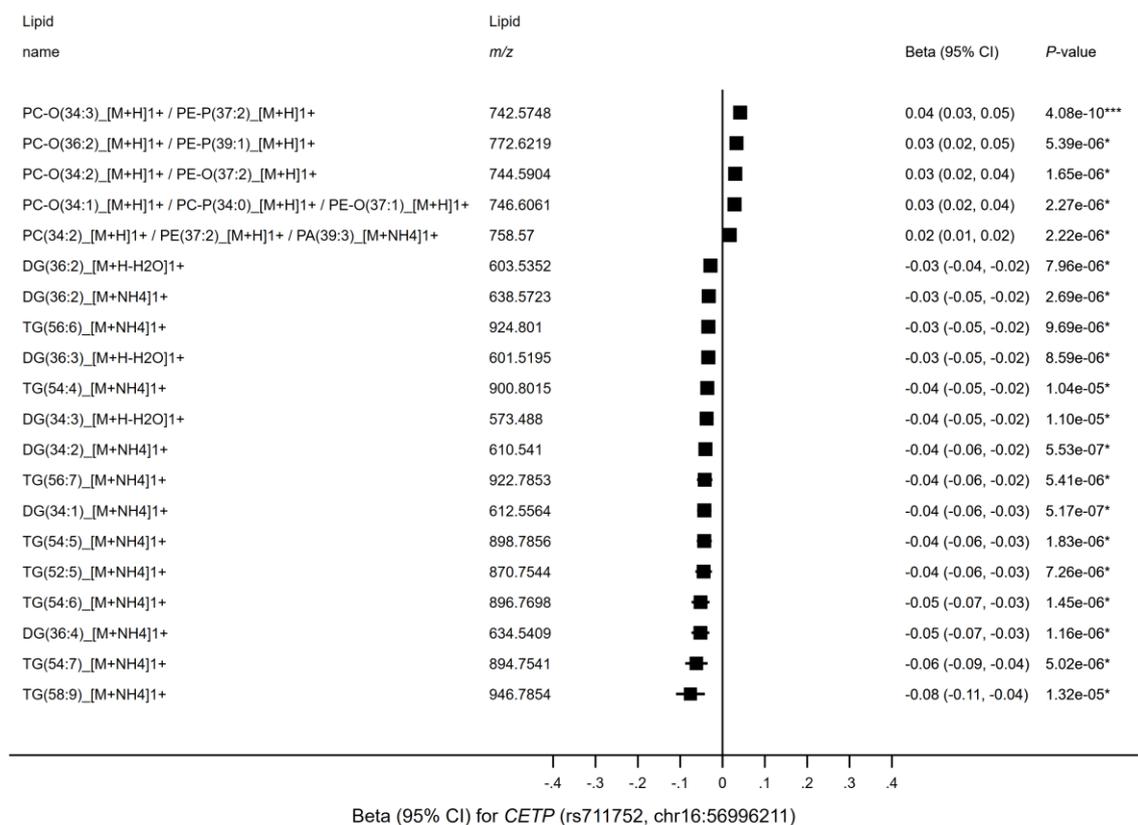
**(d) C19orf80**



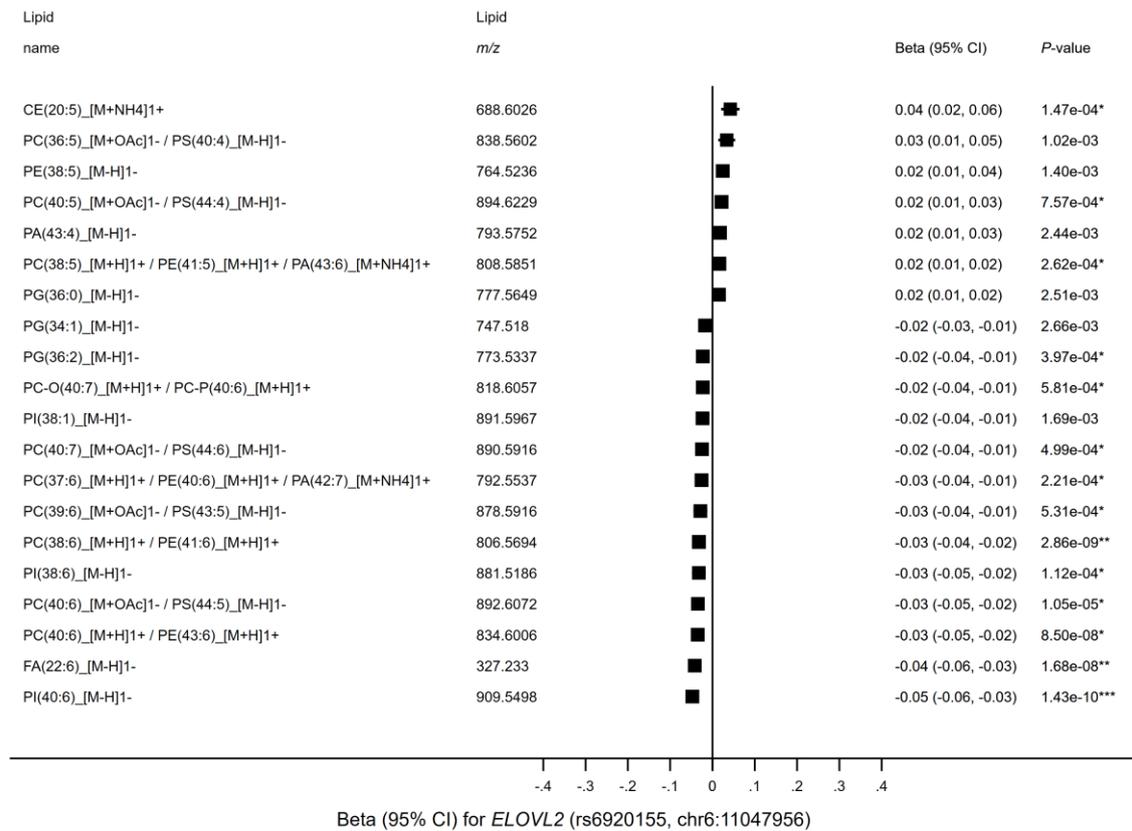
**(e) CERS4**



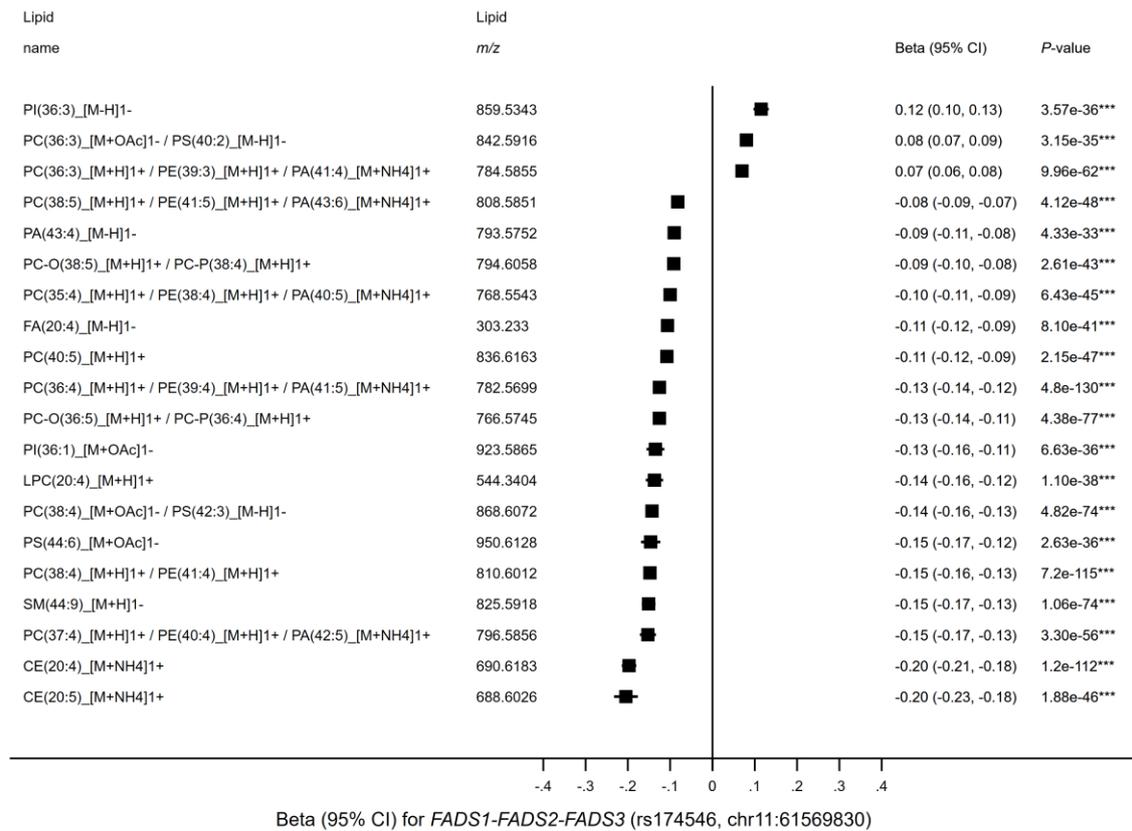
**(f) CETP**



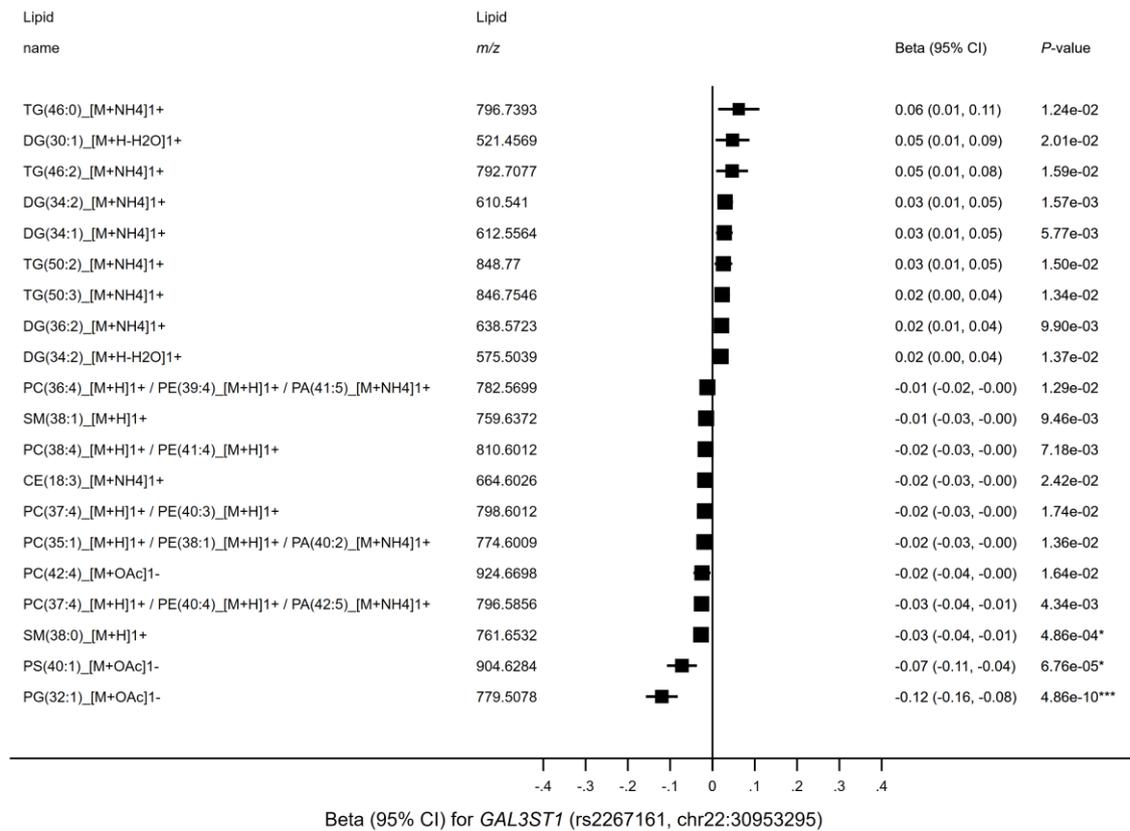
**(g) ELOVL2**



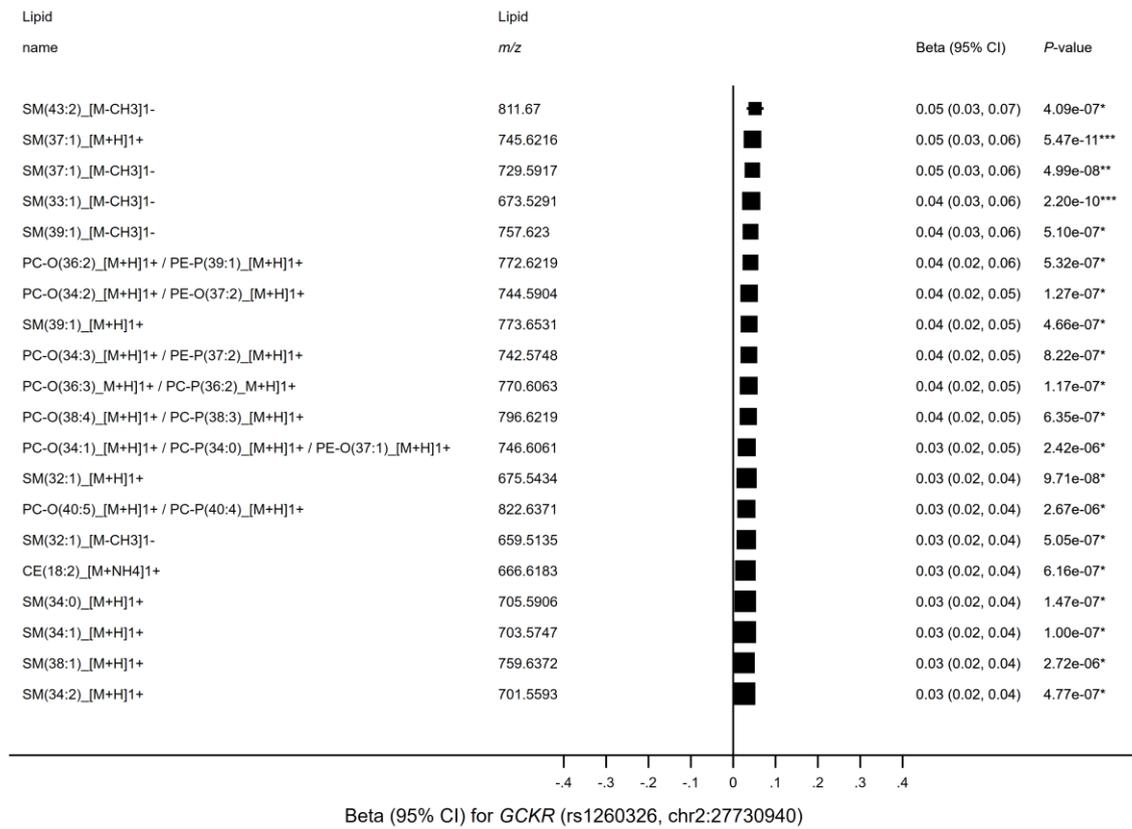
**(h) FADS1-FADS2-FADS3**



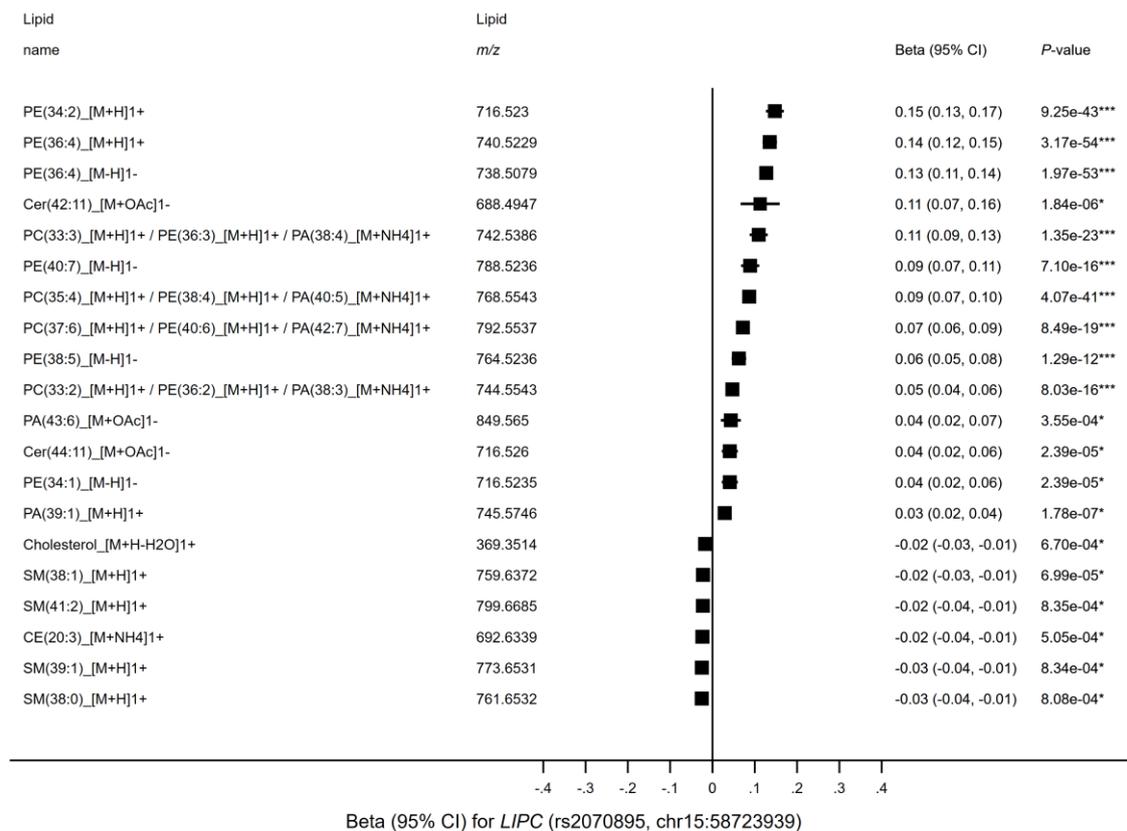
**(i) GAL3ST1**



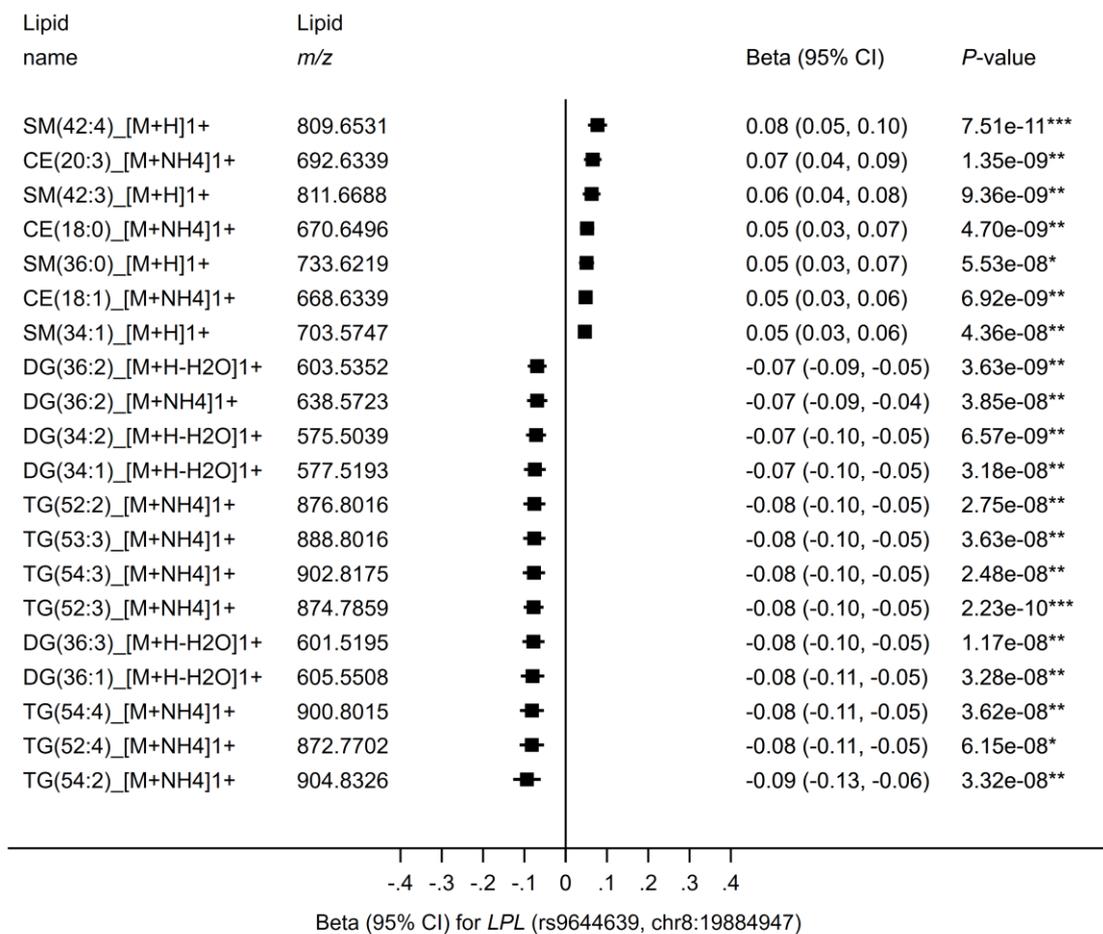
**(j) GCKR**



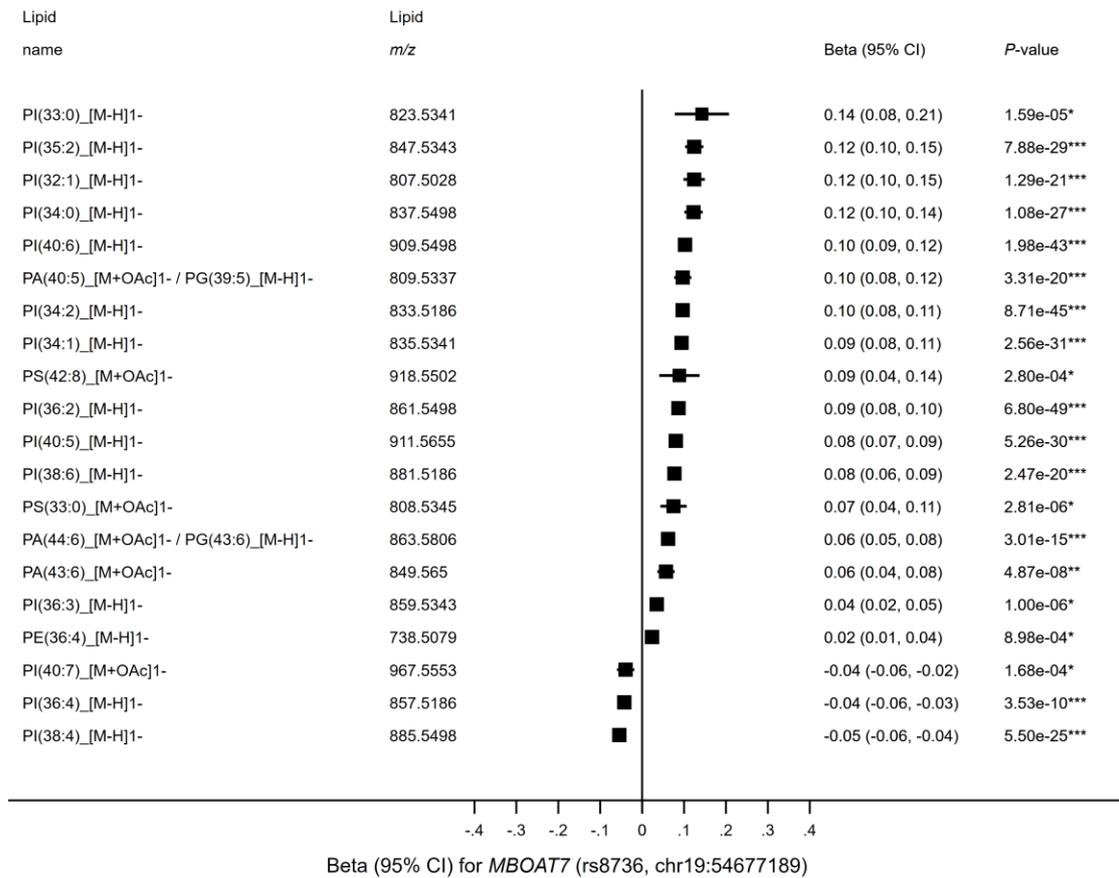
**(k) LIPC**



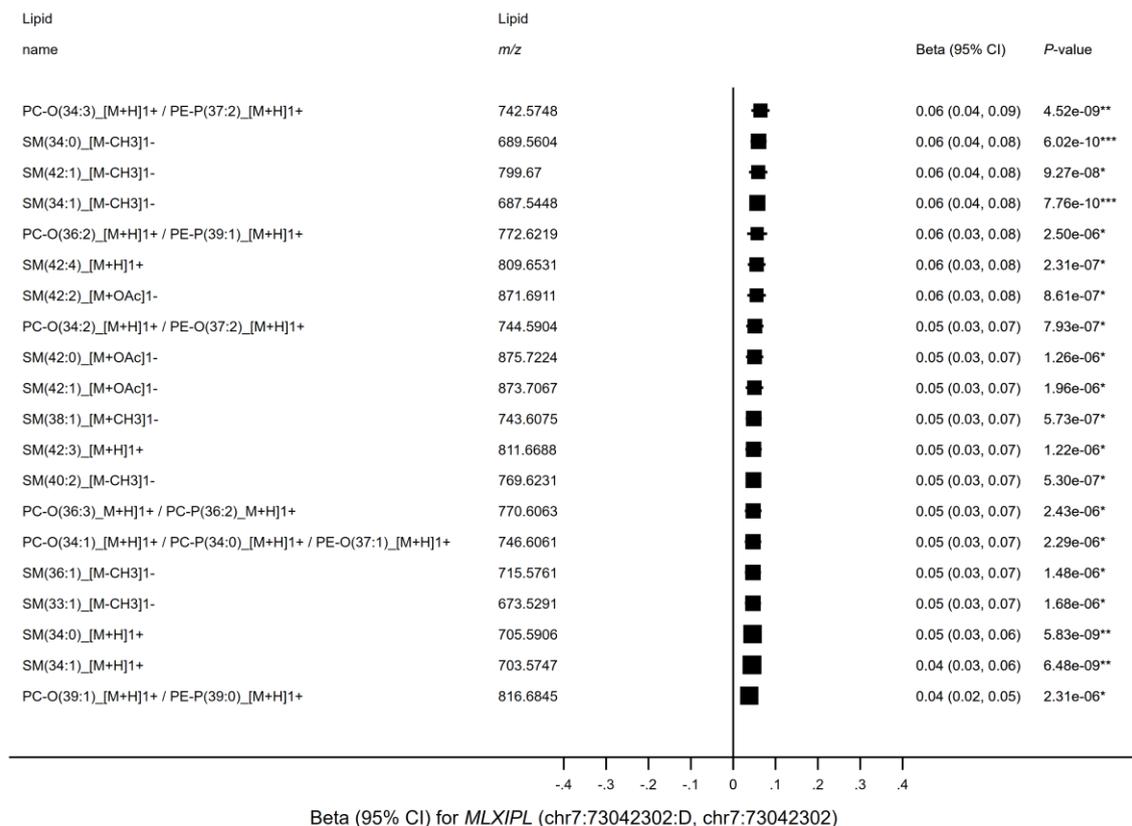
**(I) LPL**



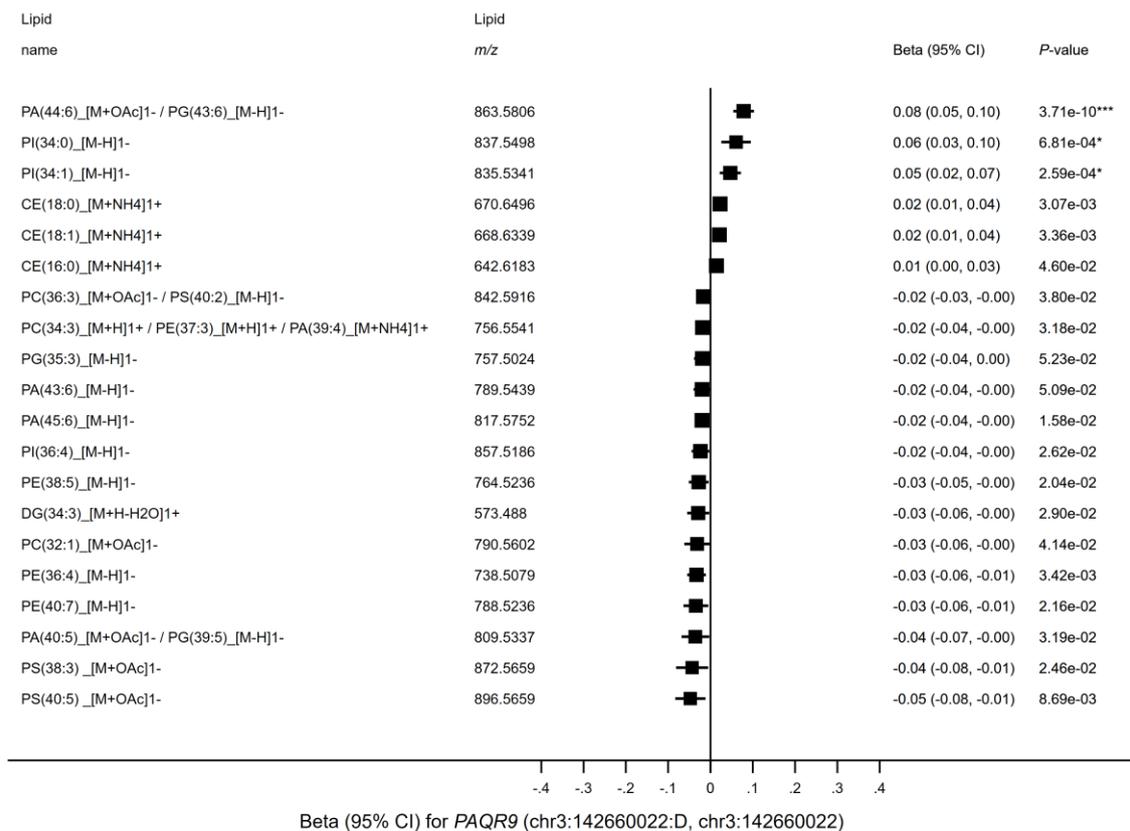
**(m) MBOAT7**



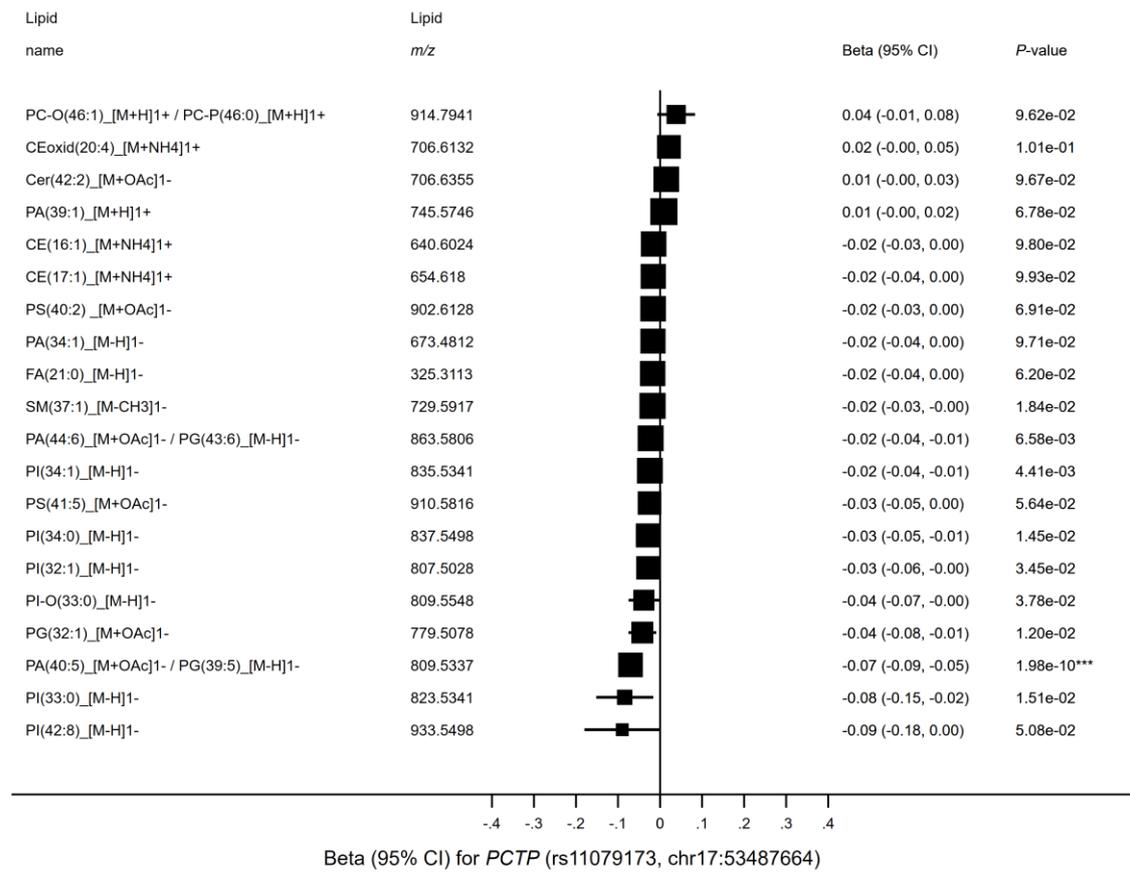
**(n) MLXIPL**



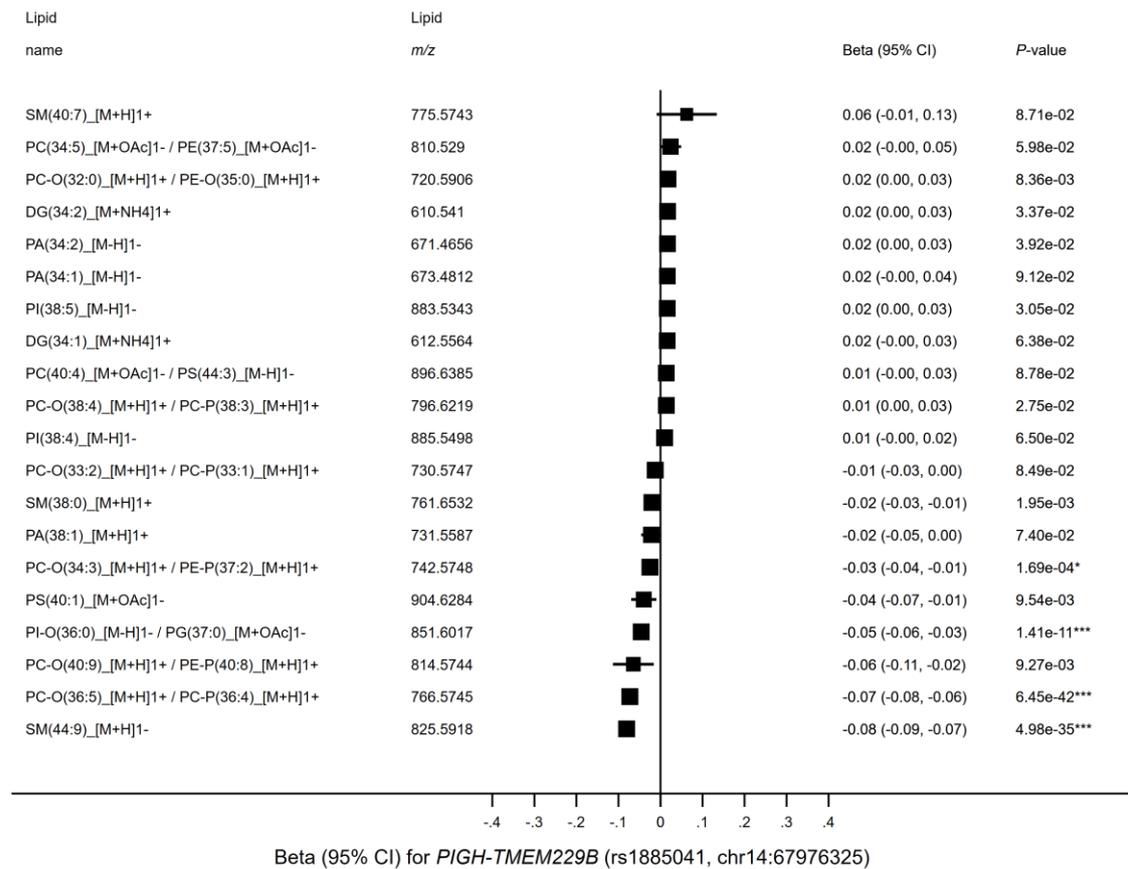
**(o) PAQR9**



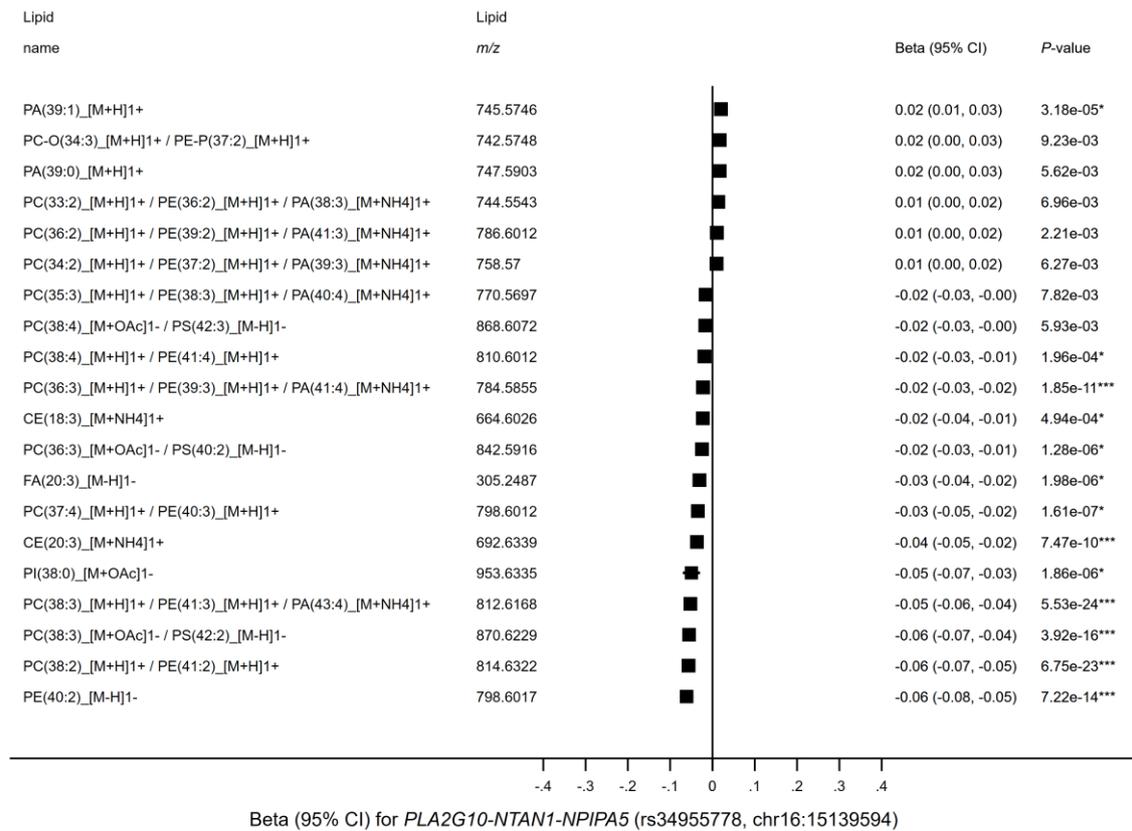
**(p) PCTP**



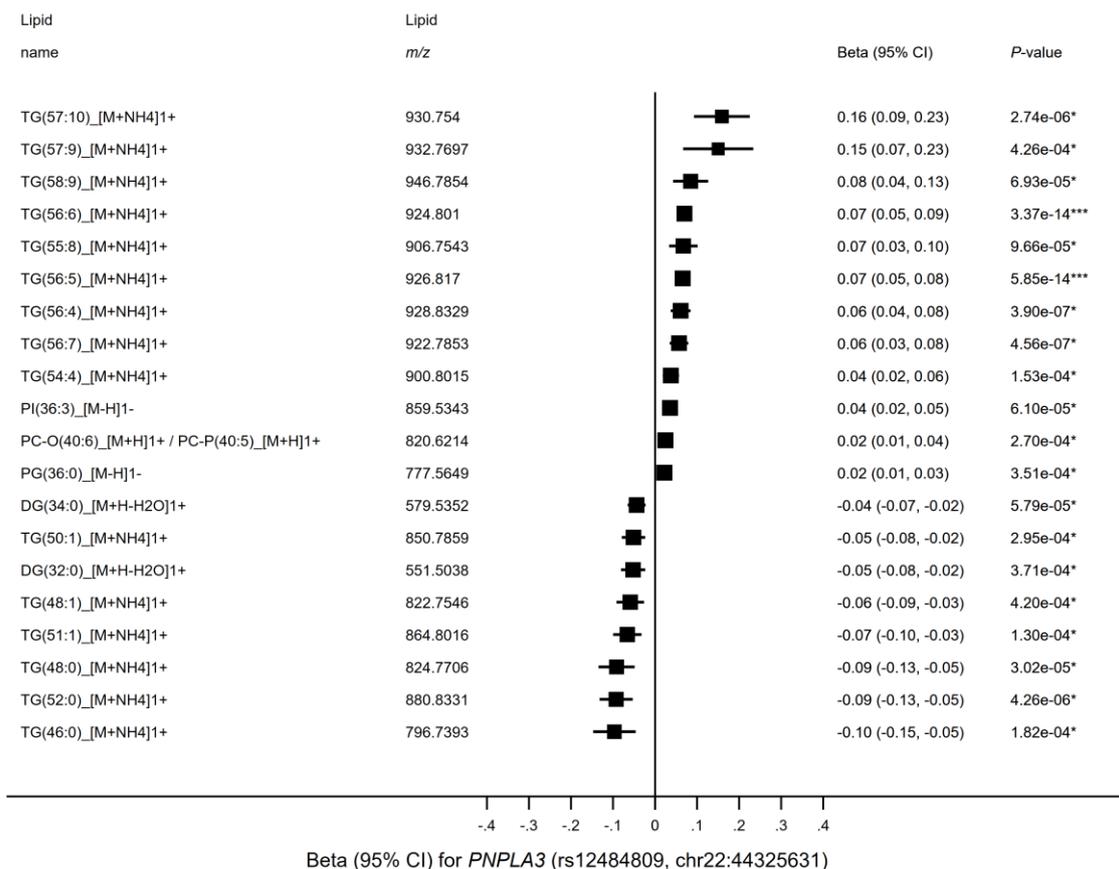
**(q) PIGH-TMEM229B**



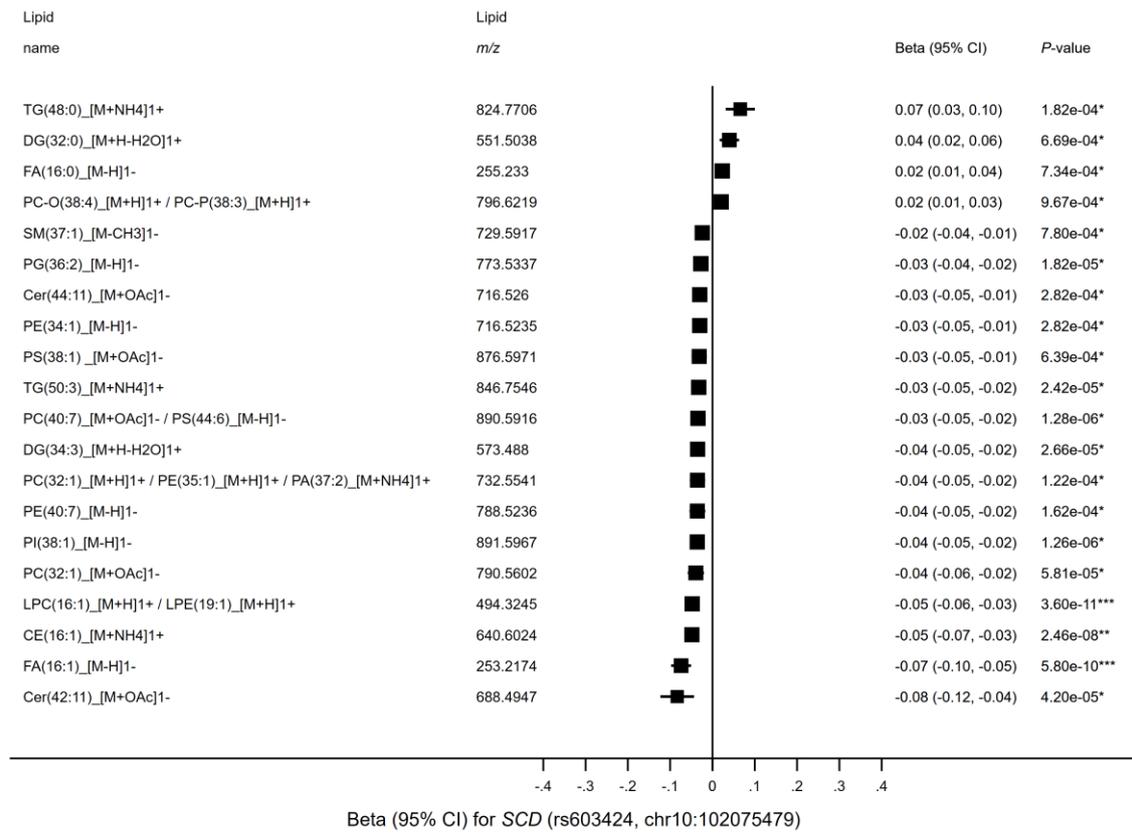
**(r) PLA2G10-NTAN1-NPIPA5**



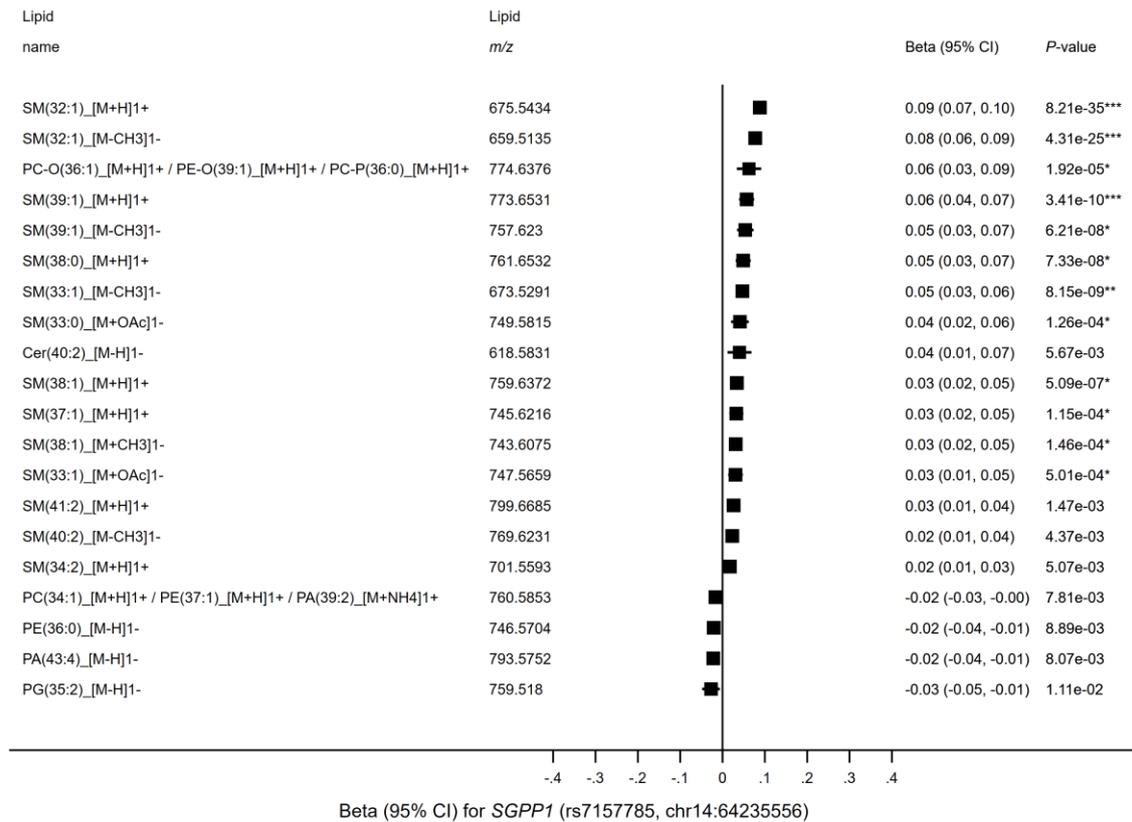
**(s) PNPLA3**



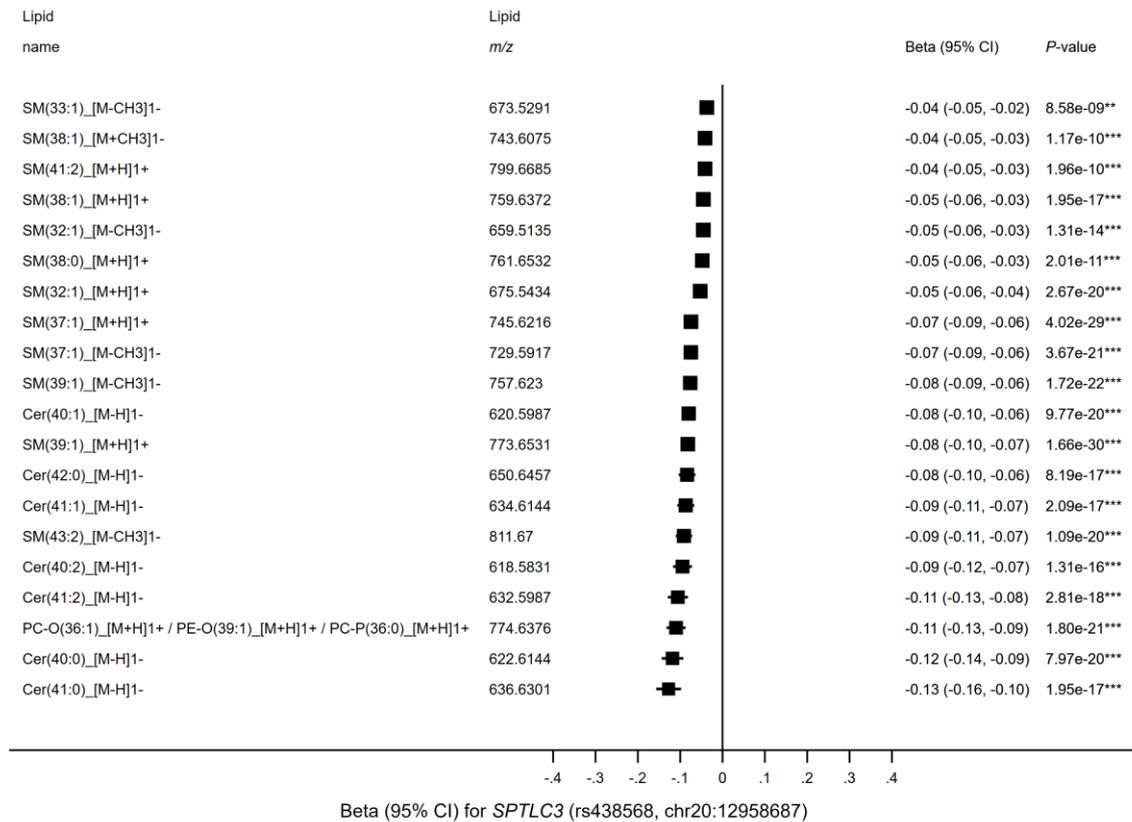
**(t) SCD**



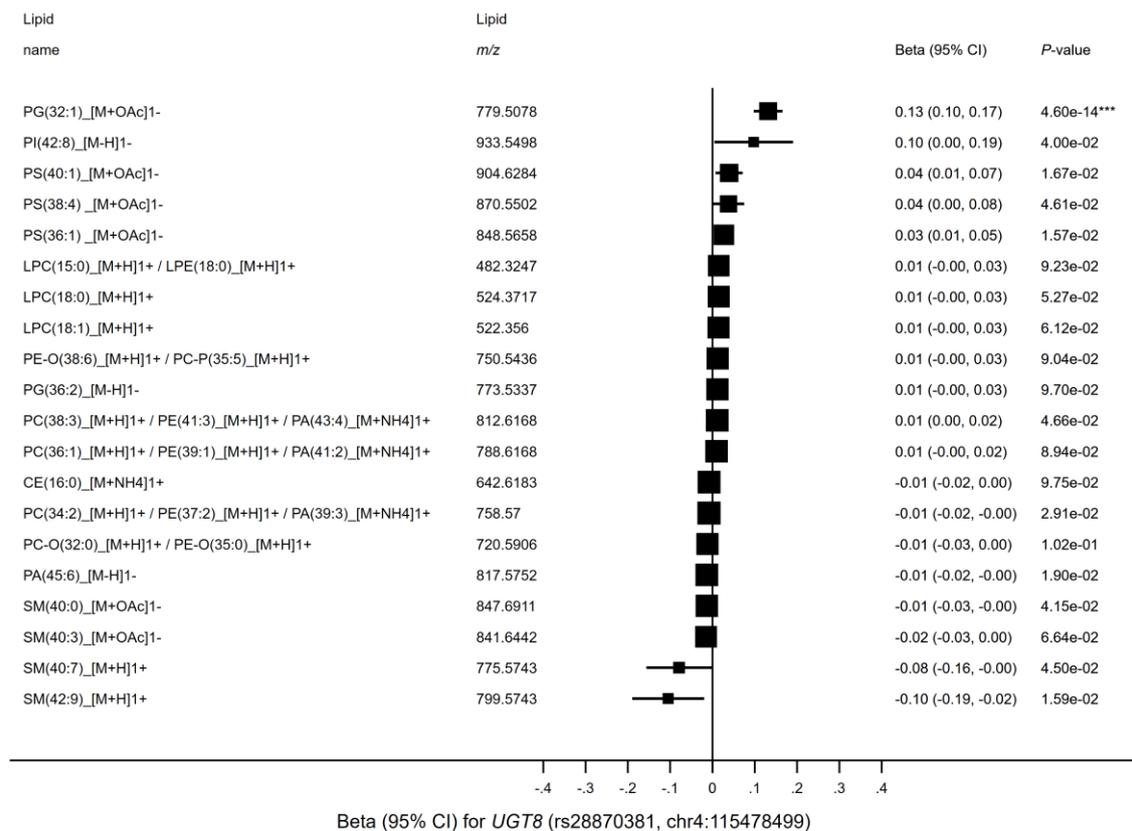
**(u) SGPP1**



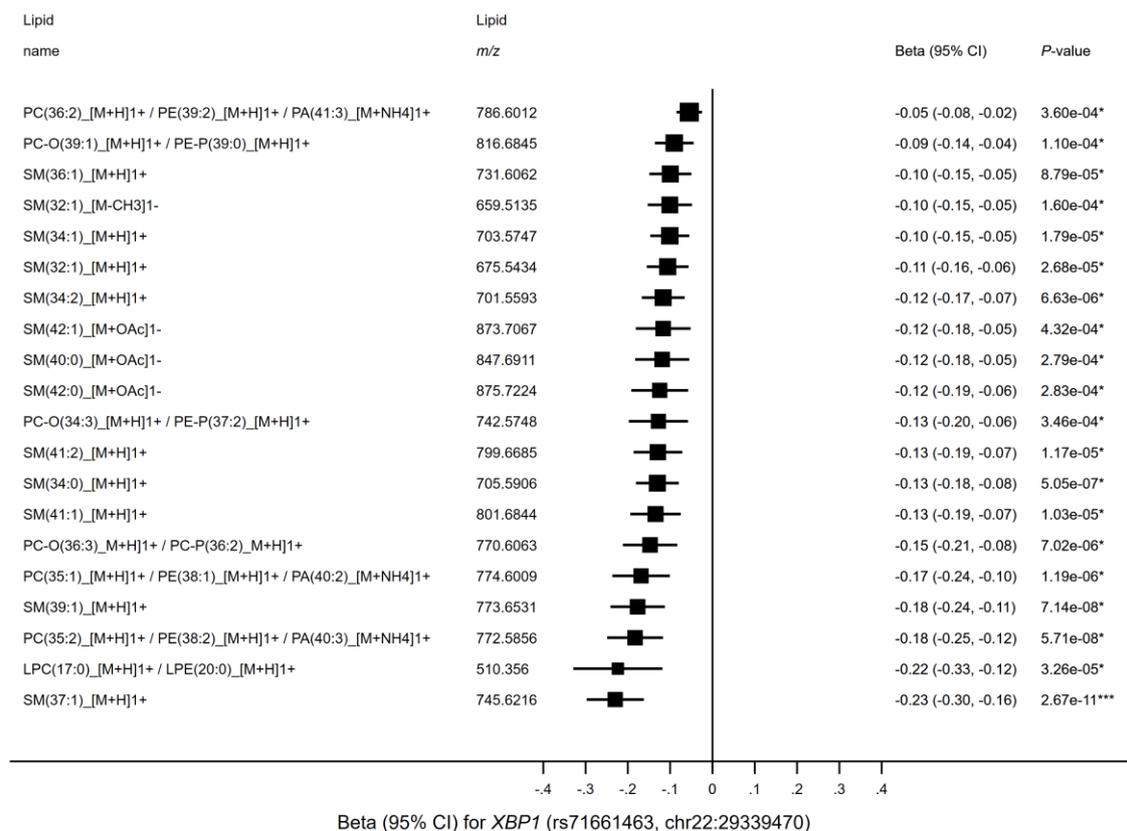
**(v) SPTLC3**



**(w) UGT8**

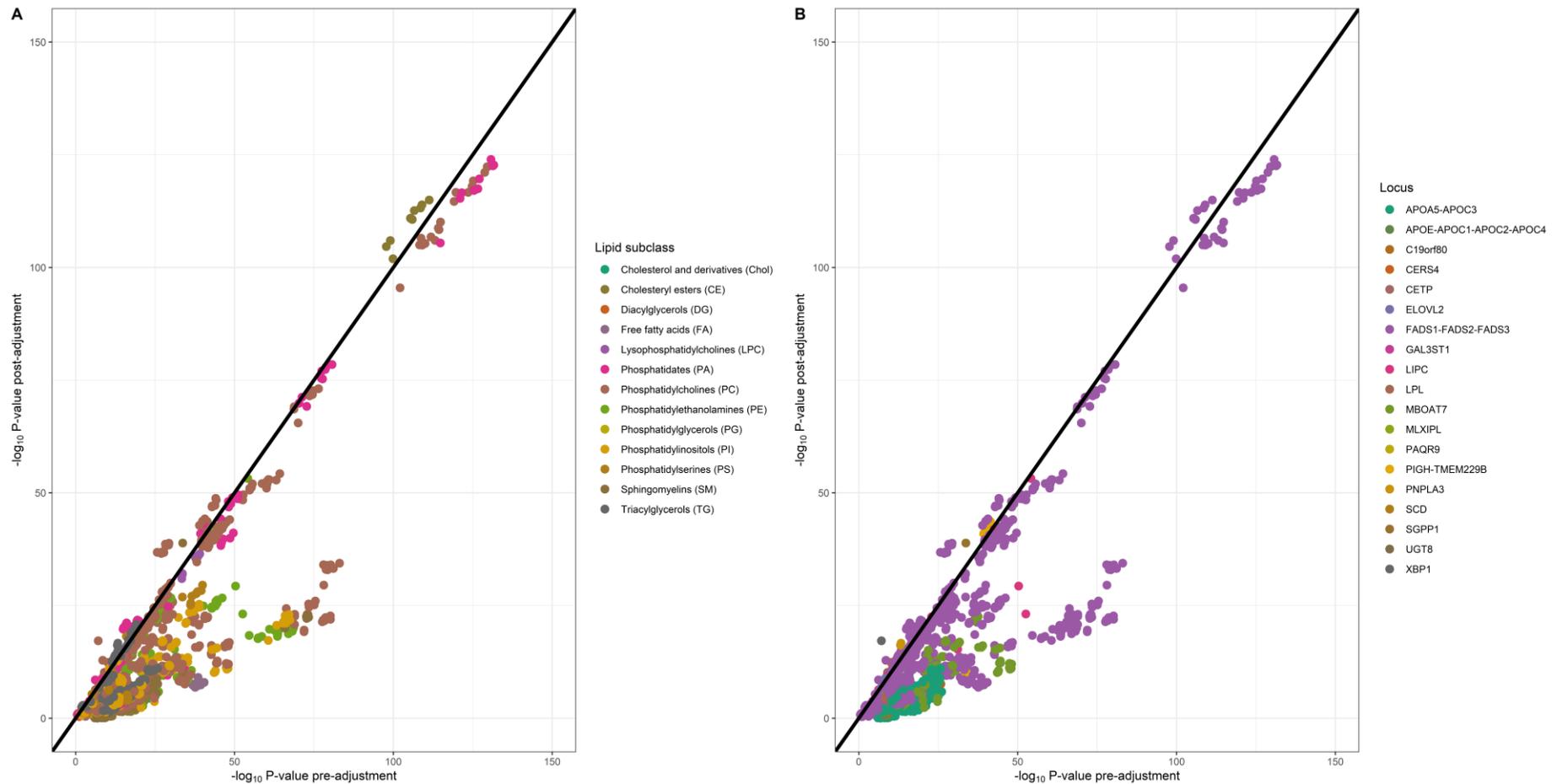


**(x) XBP1**



Forest plots showing the association of the top 20 most significantly associated lipids in PROMIS with the lead variant in each significant locus from the conditional analyses. **Note:** \* =  $P < 0.001$ ; \*\* =  $P < 5 \times 10^{-8}$ ; \*\*\* =  $P < 8.9 \times 10^{-10}$ .

**Supplementary Figure 8.** Comparison of genetic associations with lipids pre- and post-adjustment for clinical lipid measures



Comparison of  $-\log_{10} P$ -values in PROMIS from conditional analyses (pre-adjustment) and after adjustment for several clinical lipid measures (total cholesterol, HDL cholesterol, and triglycerides). Association estimates are shown coloured by (a) lipid subclass and (b) genetic locus.