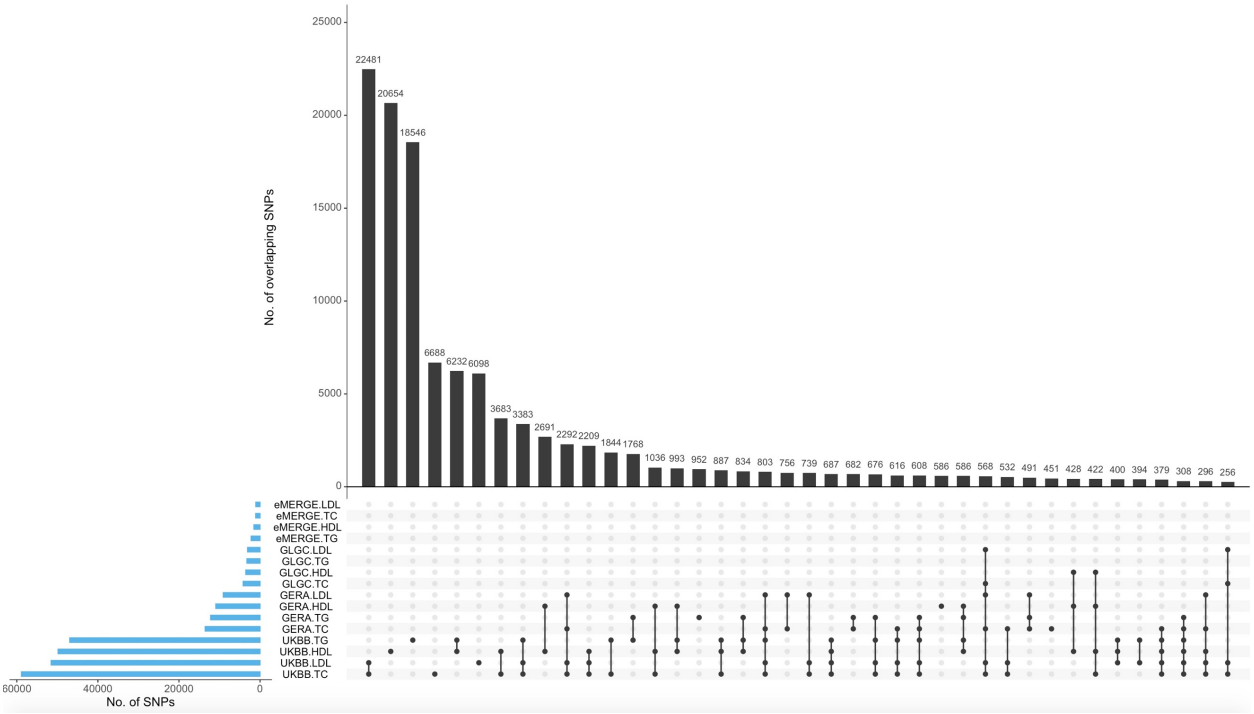
# Supplementary Figures: A unified framework identifies novel links between plasma lipids and diseases from electronic medical records across large-scale cohorts

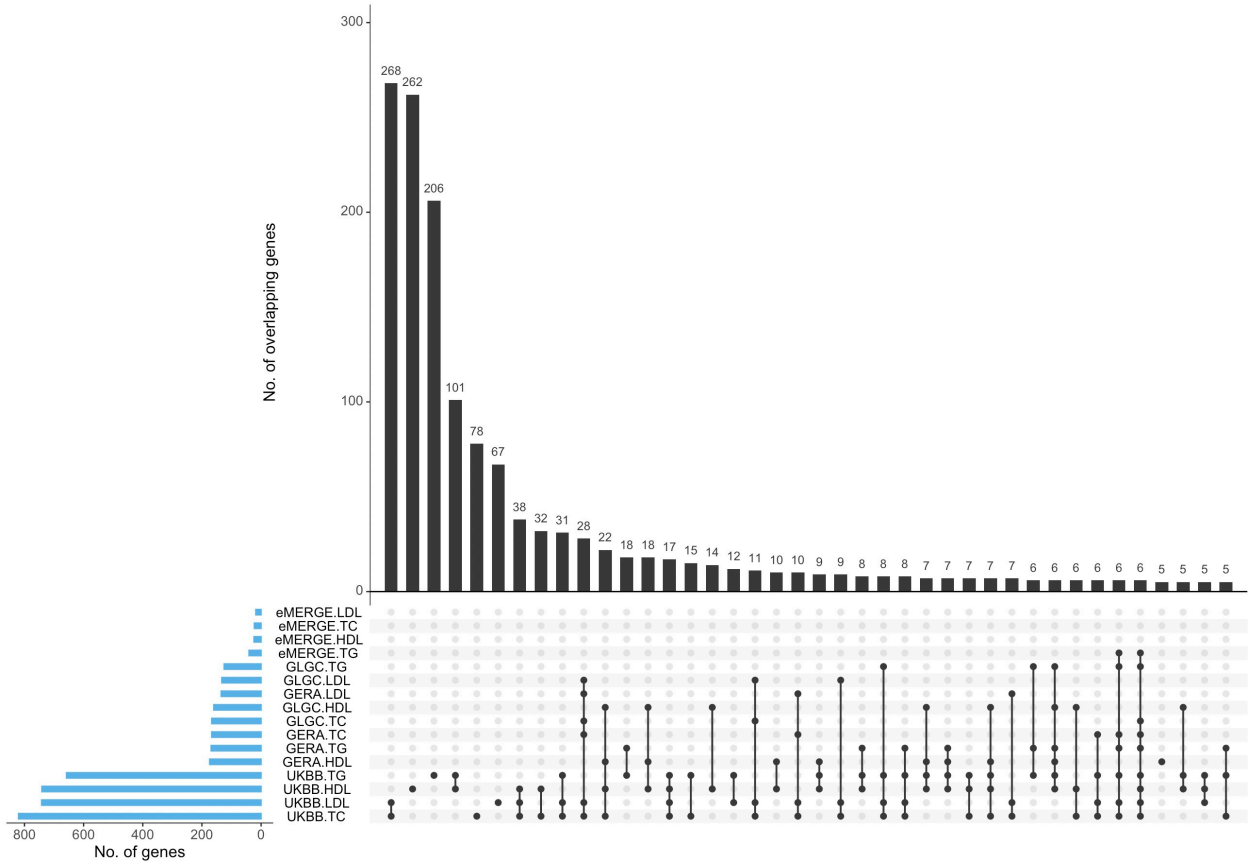
Yogasudha Veturi[1], Anastasia Lucas[1], Yuki Bradford[1], Daniel Hui[1], Scott Dudek[1], Elizabeth Theusch[2], Anurag Verma[1], Jason E. Miller[1], Iftikhar Kullo[3], Hakon Hakonarson[4], Patrick Sleiman[4], Daniel Schaid[5], Charles M. Stein[6], Digna R. Velez Edwards[7,8,9], QiPing Feng[6], Wei-Qi Wei[7], Marisa W. Medina[2], Ronald Krauss[2], Thomas J. Hoffmann[10], Neil Risch[10], Benjamin F. Voight[11,12], Daniel J. Rader[1,11] and Marylyn D. Ritchie[1*]

[1]Department of Genetics and Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
[2]Department of Pediatrics, University of California San Francisco, Oakland, CA, USA.
[3]Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA.
[4]Center for Applied Genomics, Children's Hospital of Philadelphia, PA, USA.
[5]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.
[6]Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.
[7]Department of Biomedical Informatics in School of Medicine, Vanderbilt University, Nashville, TN, USA.
[8]Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA.
[9]Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA.
[10]Institute for Human Genetics, and Department of Epidemiology & Biostatistics, University of California and San Francisco, San Francisco, CA, USA.
[11]Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
[12]Institute of Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

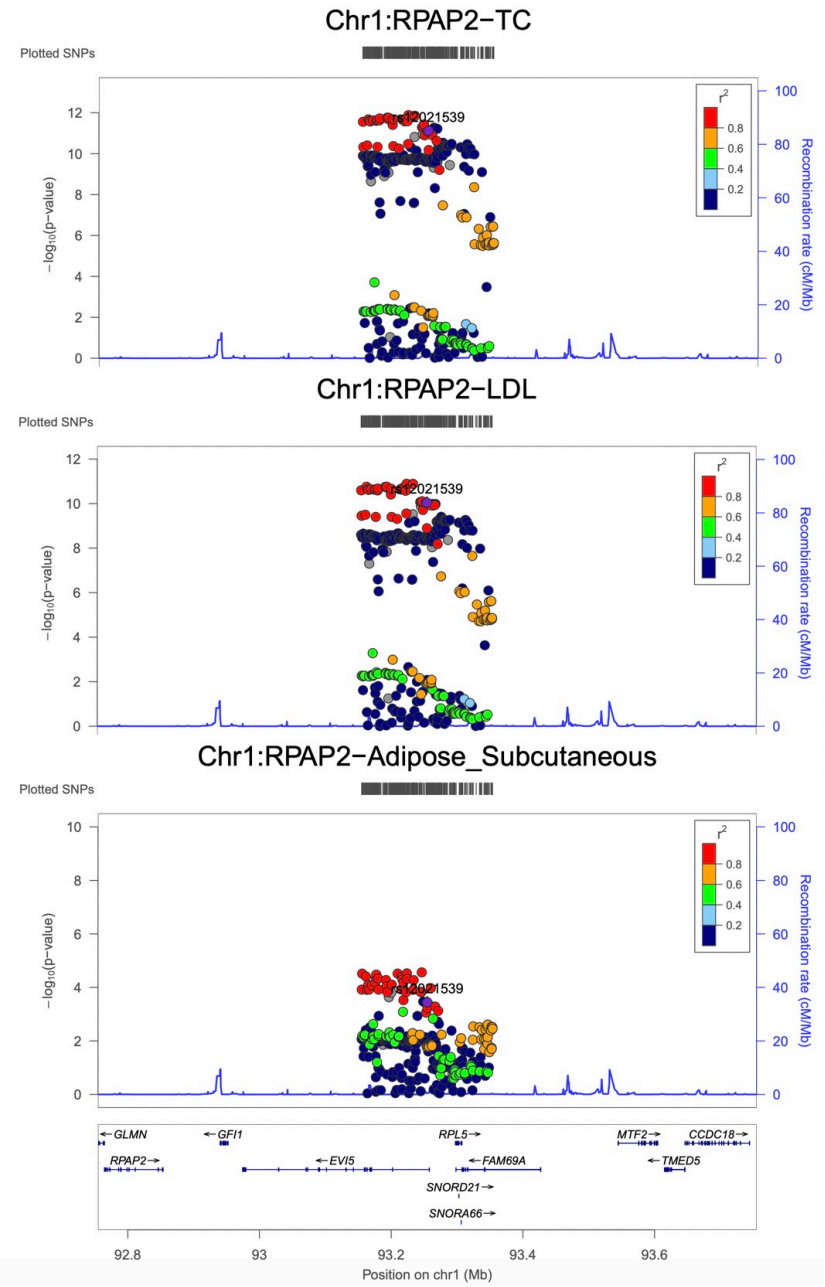
*e-mail: marylyn@pennmedicine.upenn.edu



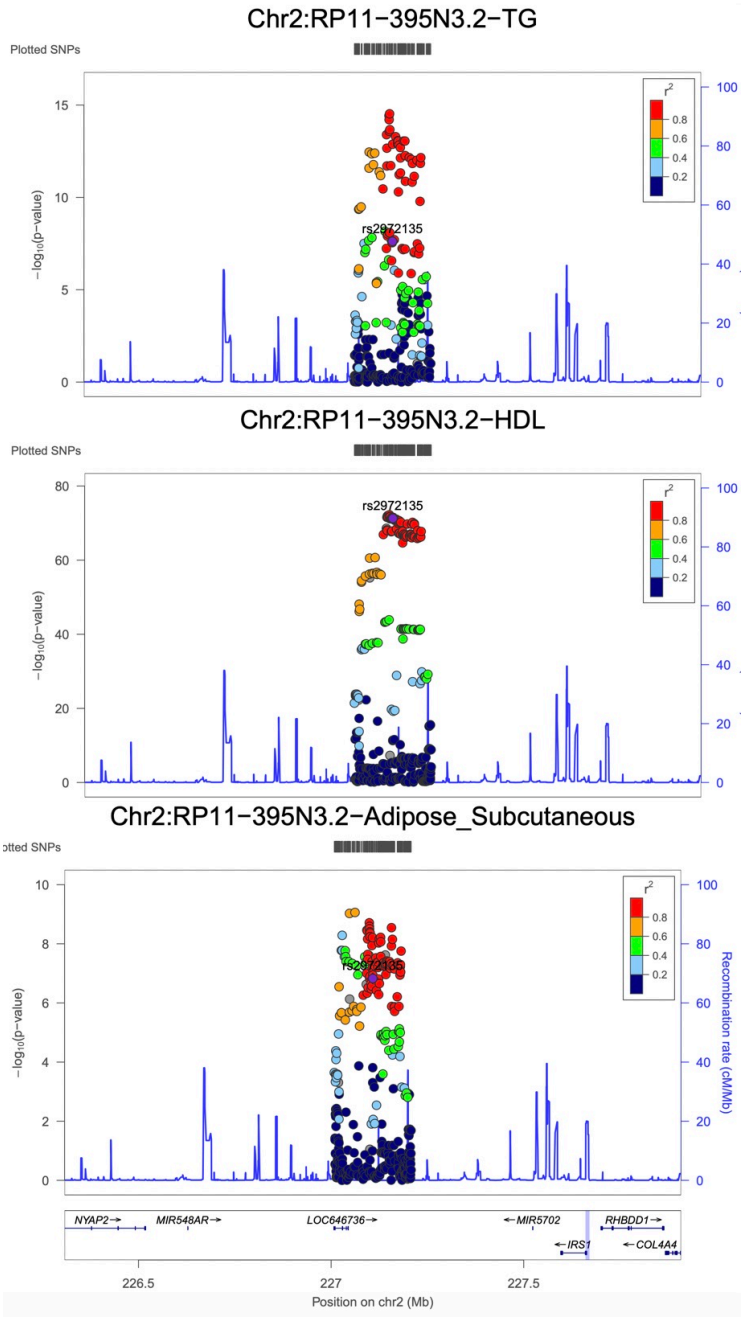

**Supplementary Figure 1. Overlap of lipid GWAS-significant SNPs across multiple cohorts**. The UpSet plot shows the SNPs that overlap at the Bonferroni threshold from GWAS of each of the four plasma lipids (HDL-C, LDL-C, TC, TG) across the four cohorts considered in Phase I of this study (eMERGE, GERA, GLGC and UKB). The numbers on top indicate the GWAS-significant SNPs exclusively detected in an intersecting set.



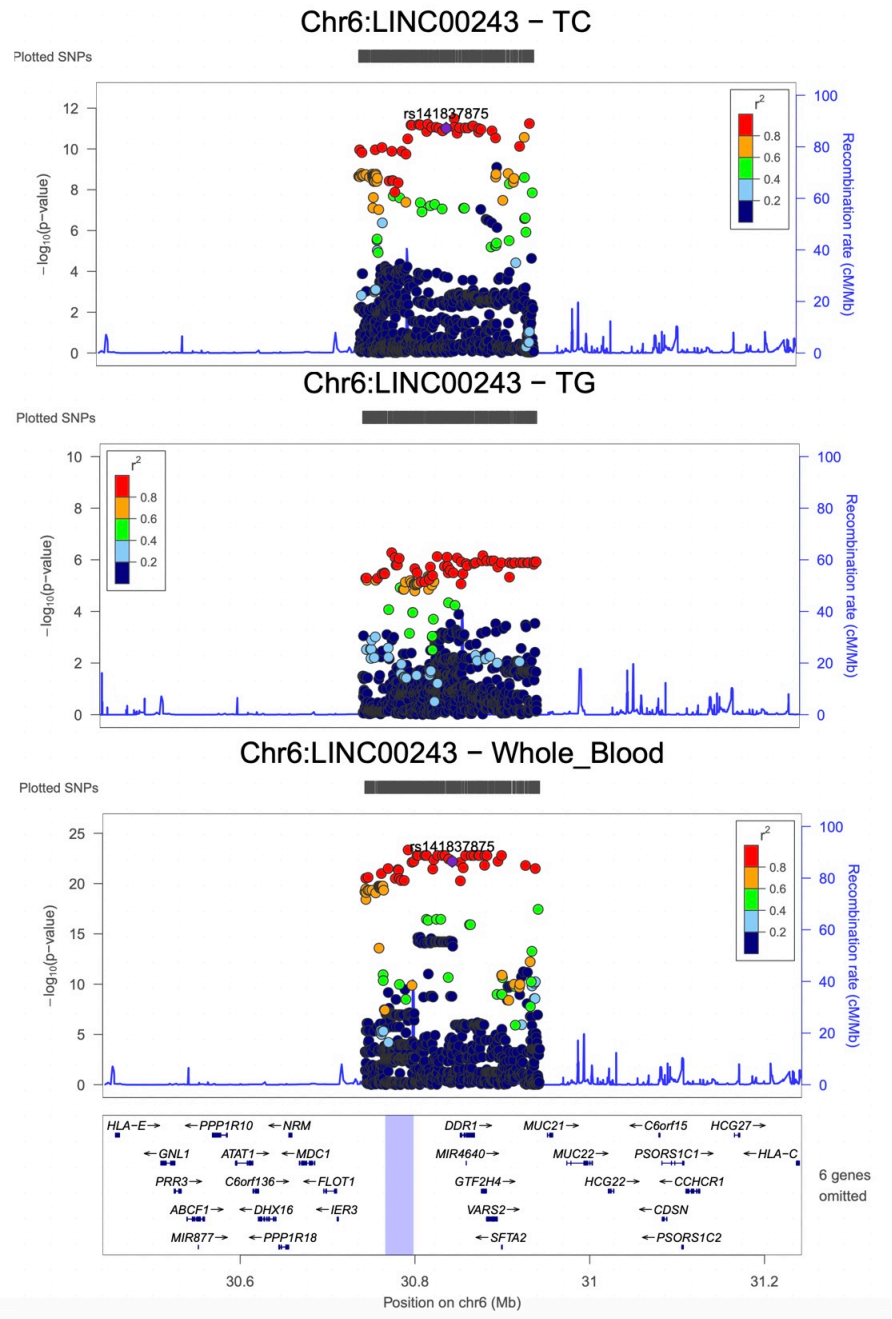

**Supplementary Figure 2. Overlap of lipid TWAS-significant genes across multiple cohorts**. The UpSet plot shows the genes that overlap at the Bonferroni threshold from TWAS of each of the four plasma lipids (HDL-C, LDL-C, TC, TG) across the four cohorts considered in Phase I of this study (eMERGE, GERA, GLGC and UKB). The numbers on top indicate the TWAS-significant genes exclusively detected in an intersecting set prior to applying coloc filters.



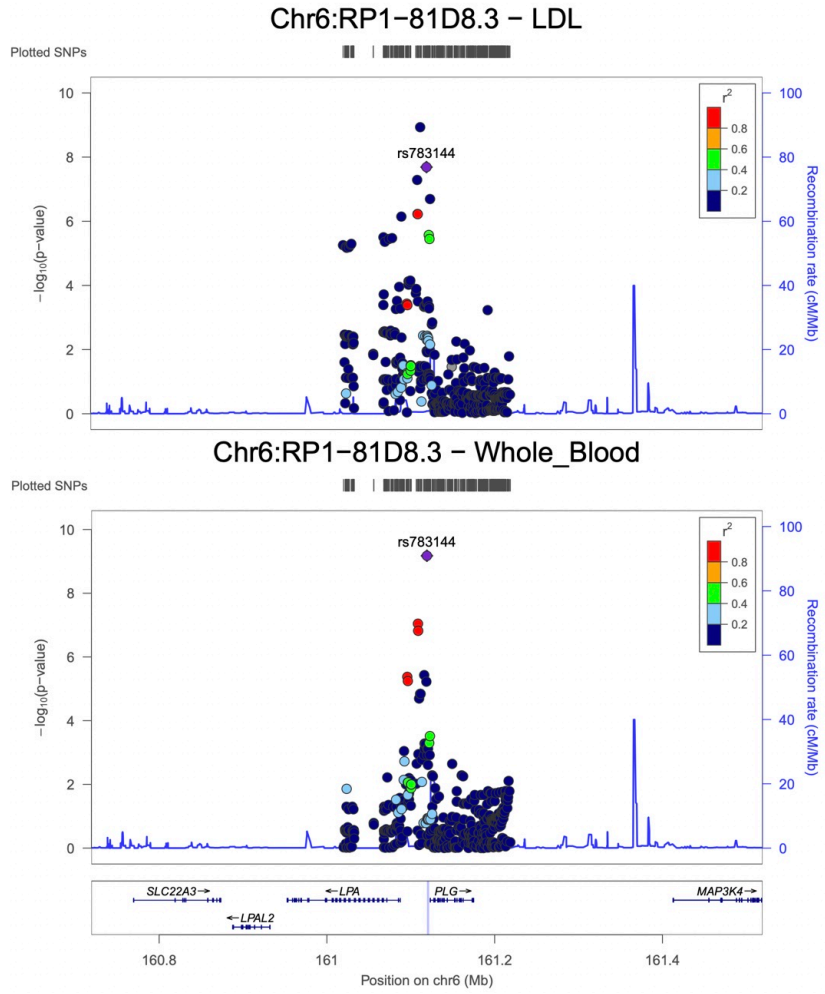

**Supplementary Figure 3. Local association plots for lipids (top and middle) and gene expression (bottom) for the lipid TWAS-significant novel gene *RPAP2* on chromosome 1**. Y-axis on these LocusZoom plots shows -logP values from: (top) GWAS of TC (two-sided linear regression) in UKB, (middle) GWAS of LDL (two-sided linear regression) in UKB, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in adipose subcutaneous tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb). *RPAP2* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



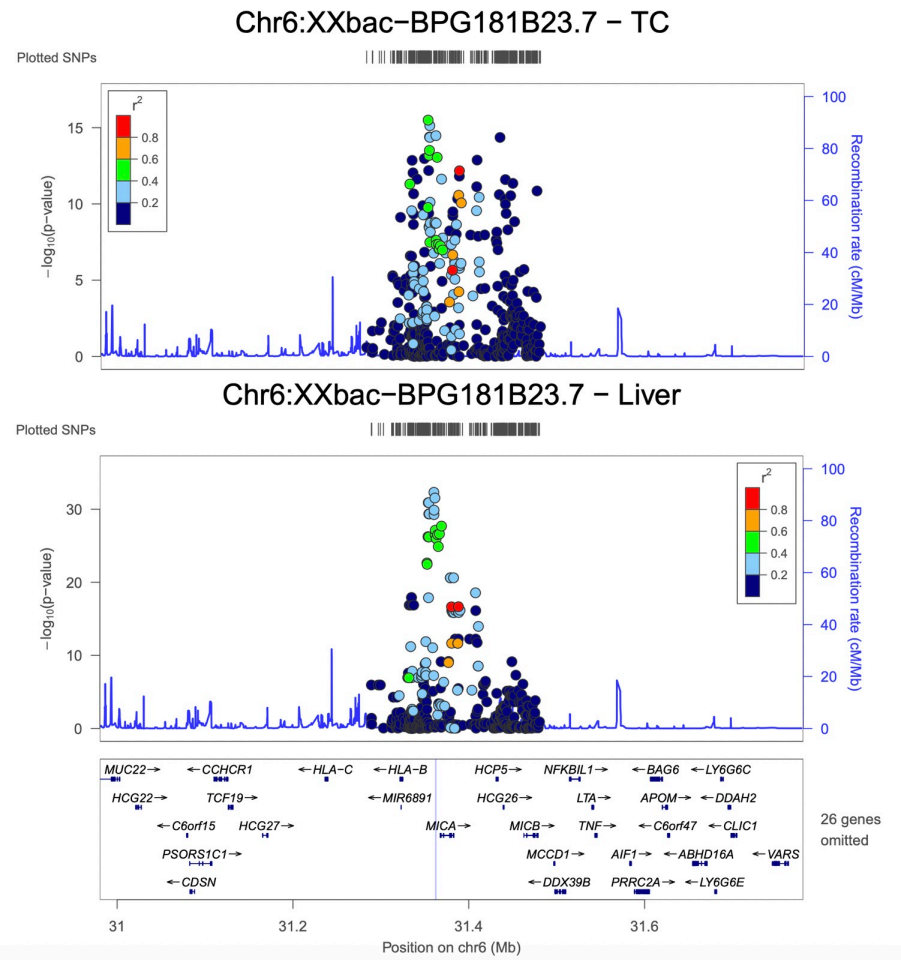

**Supplementary Figure 4. Local association plots for lipids (top and middle) and gene expression (bottom) for the lipid TWAS-significant novel gene *RP11-395N3.2* on chromosome 2**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS of TG (two-sided linear regression) in GLGC, (middle) GWAS of HDL (two-sided linear regression) in UKB, and (middle) all variant-gene cis-eQTL associations (two-sided linear regression) in adipose subcutaneous tissue from GTEx v8 (bottom). X-axis shows physical position on the chromosome (Mb). *RP11-395N3.2* was Bonferroni-significant from lipid TWAS (Phase I).



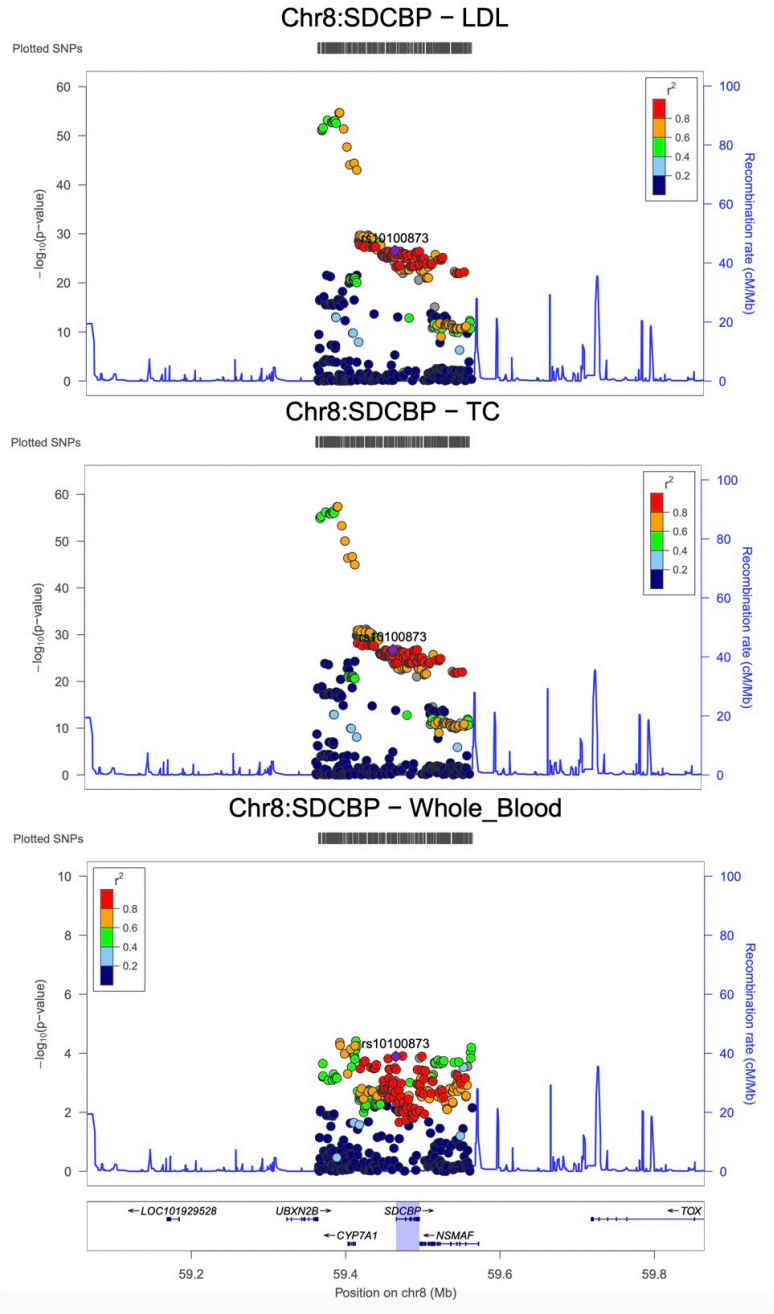

**Supplementary Figure 5. Local association plots for lipids (top and middle) and gene expression (bottom) for the lipid TWAS-significant novel pseudogene *LINC00243* on chromosome 6**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS of TC (two-sided linear regression) in GERA, (middle) GWAS of TG (two-sided linear regression) in GERA, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in whole blood from GTEx v8. X-axis shows physical position on the chromosome (Mb). *LINC00243* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



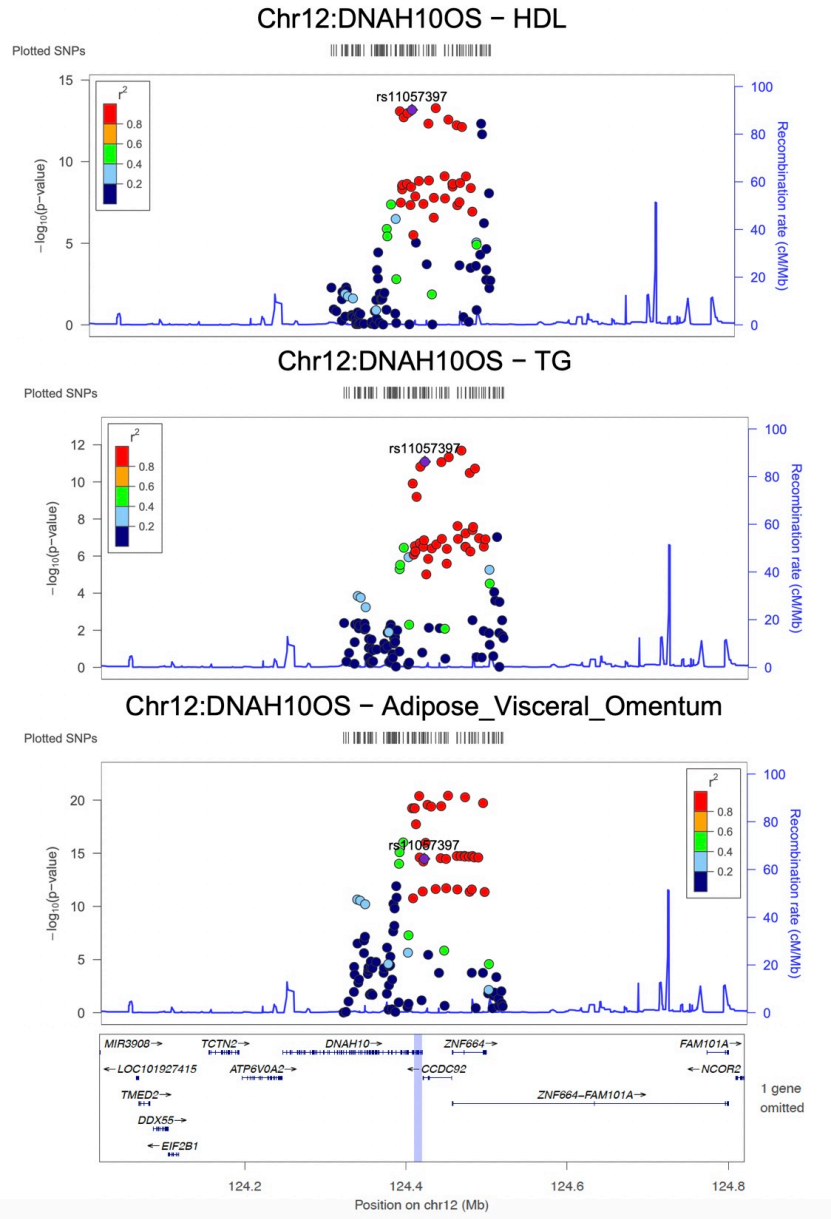

**Supplementary Figure 6. Local association plots for lipids (top) and gene expression (bottom) for the lipid TWAS-significant novel locus *RP1-81D8.3* on chromosome 6**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS of LDL (two-sided linear regression) in eMERGE, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in whole blood tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb). *RP1-81D8.3* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



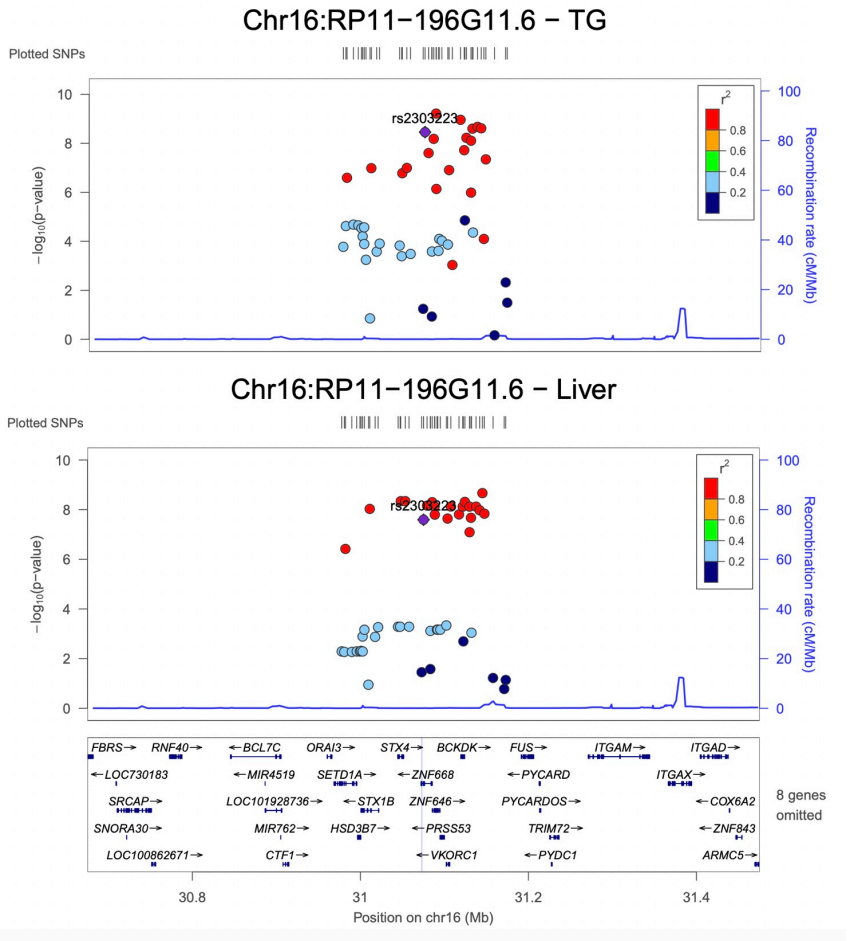

**Supplementary Figure 7. Local association plots for lipids (top) and gene expression (bottom) for the TWAS-significant novel locus *XXbac-BPG181B23.7* on chromosome 6**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS of TC (two-sided linear regression) in GLGC, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in liver tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb**).** *XXbac-BPG181B23.7* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



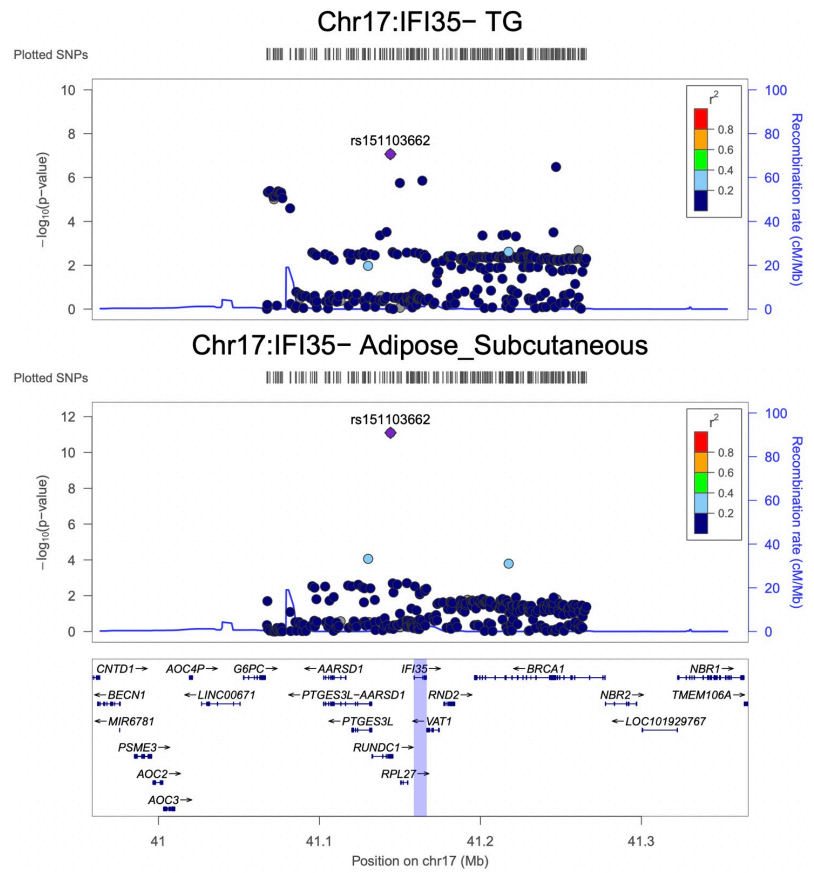

**Supplementary Figure 8. Local association plots for lipids (top and middle) and gene expression (bottom) for the lipid TWAS-significant novel locus *SDCBP* on chromosome 8**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS of LDL (two-sided linear regression) in UKB, (middle) GWAS of TC (two-sided linear regression) in UKB, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in whole blood tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb). *SDCBP* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



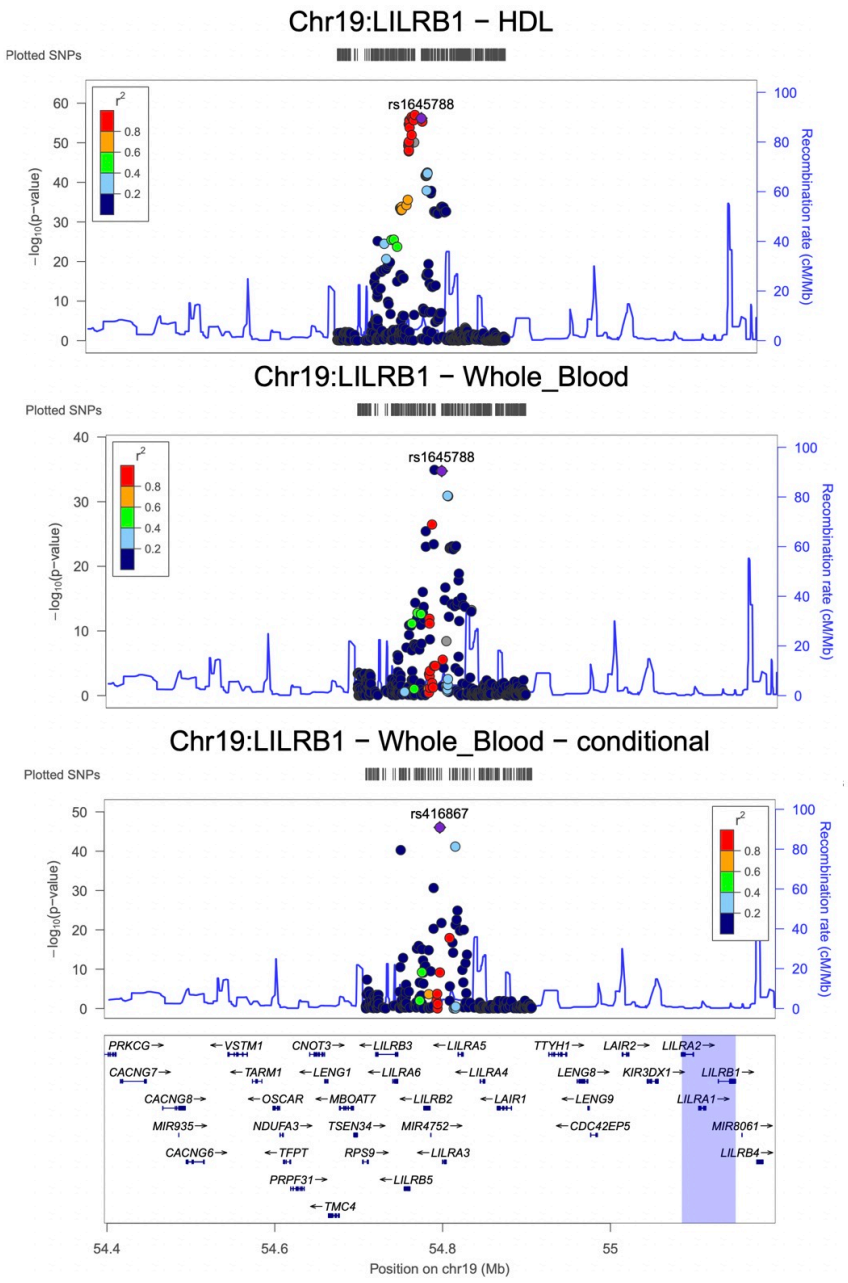

**Supplementary Figure 9. Local association plots for lipids (top and middle) and gene expression (bottom) for the lipid TWAS-significant novel locus *DNAH10OS* on chromosome 12**. Y-axis on these LocusZoom plots indicates -logP values (y-axis) from: (top) GWAS of HDL (two-sided linear regression) in GERA, (middle) GWAS of TG (two-sided linear regression) in GLGC, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in adipose visceral omentum tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb). *DNAH10OS* was Bonferroni-significant from lipid TWAS (Phase I).



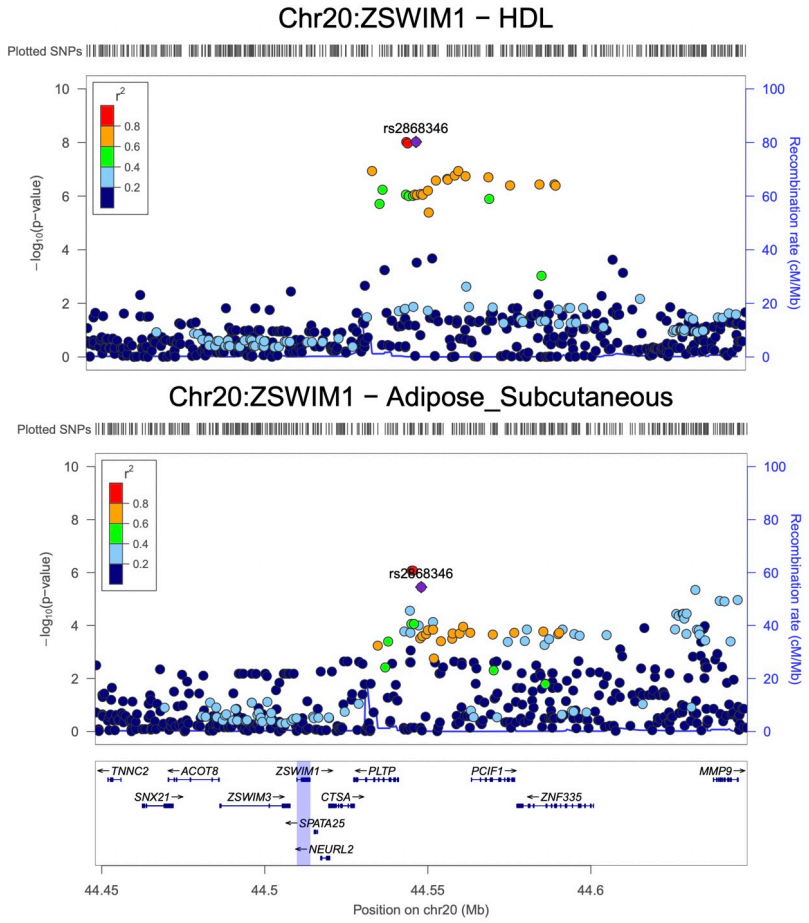

**Supplementary Figure 10. Local association plots for lipids (top) and gene expression (bottom) for the lipid TWAS-significant novel locus *RP11-196G11.6* on chromosome 16**. Y-axis indicates -logP values from: (top) GWAS of TG (two-sided linear regression) in GLGC, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in liver tissue from GTEx v8 (bottom). X-axis shows physical position on the chromosome (Mb). *RP11-196G11.6* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



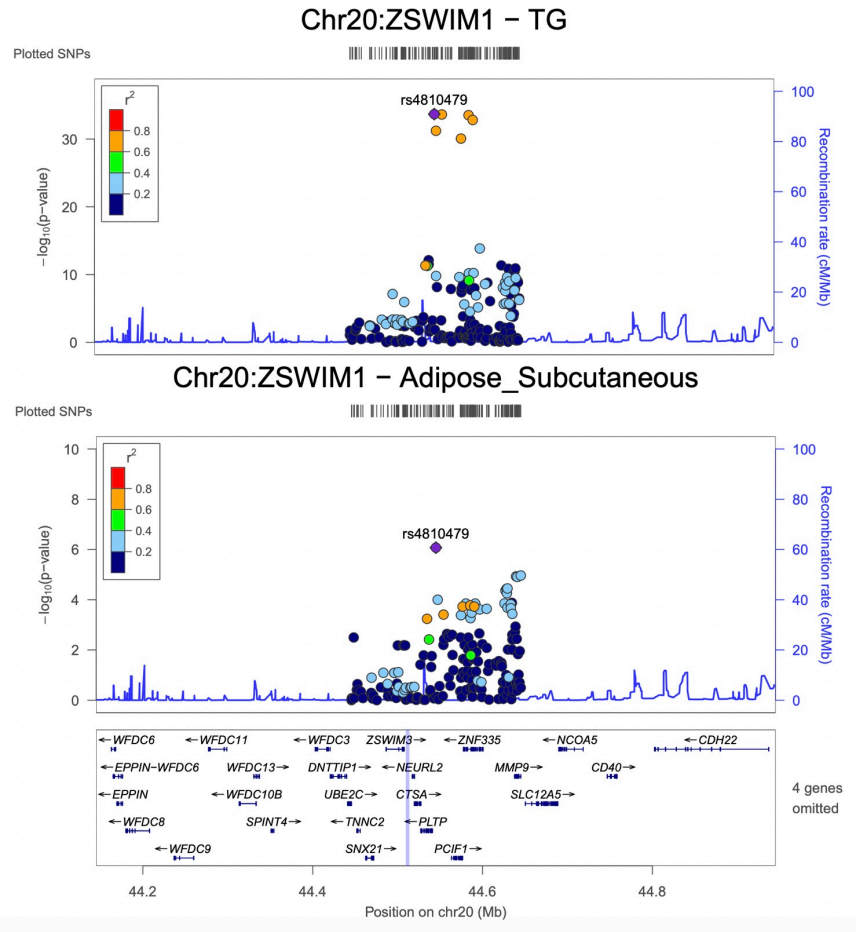

**Supplementary Figure 11. Local association plots for lipids (top) and gene expression (bottom) for the lipid TWAS-significant novel locus *IFI35* on chromosome 17**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS of TG (two-sided linear regression) in UKB, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in adipose subcutaneous tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb). *RP11-196G11.6* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



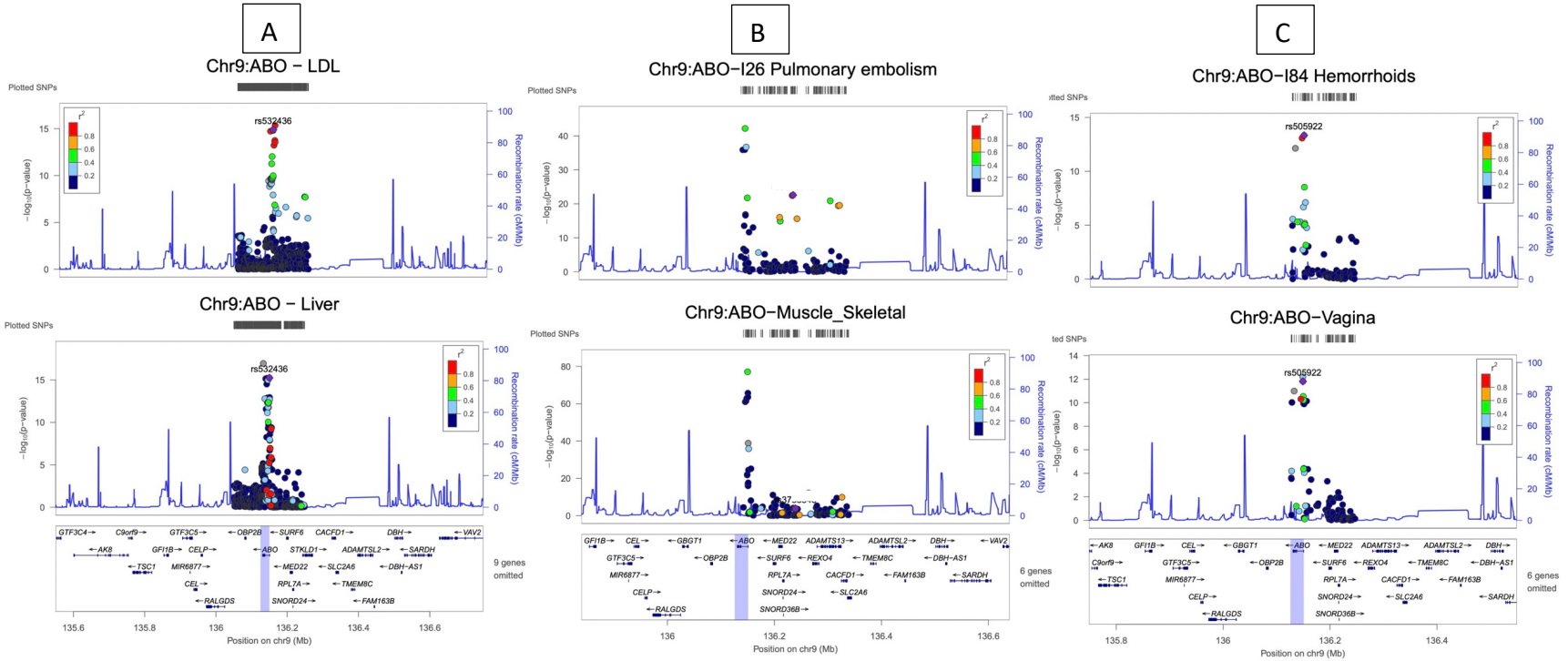

**Supplementary Figure 12. Local association plots for lipids (top) and gene expression (middle and bottom) for the lipid TWAS-significant novel locus *LILRB1* on chromosome 19**. Y-axis indicates -logP values from: (top) GWAS (two-sided linear regression) of HDL in UKB, (middle) all variant-gene cis-eQTL associations (two-sided linear regression) in whole blood tissue from GTEx v8 and (bottom) conditional p-values after conditioning on the top eQTL at the locus. X-axis shows physical position on the chromosome (Mb). *LILRB1* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



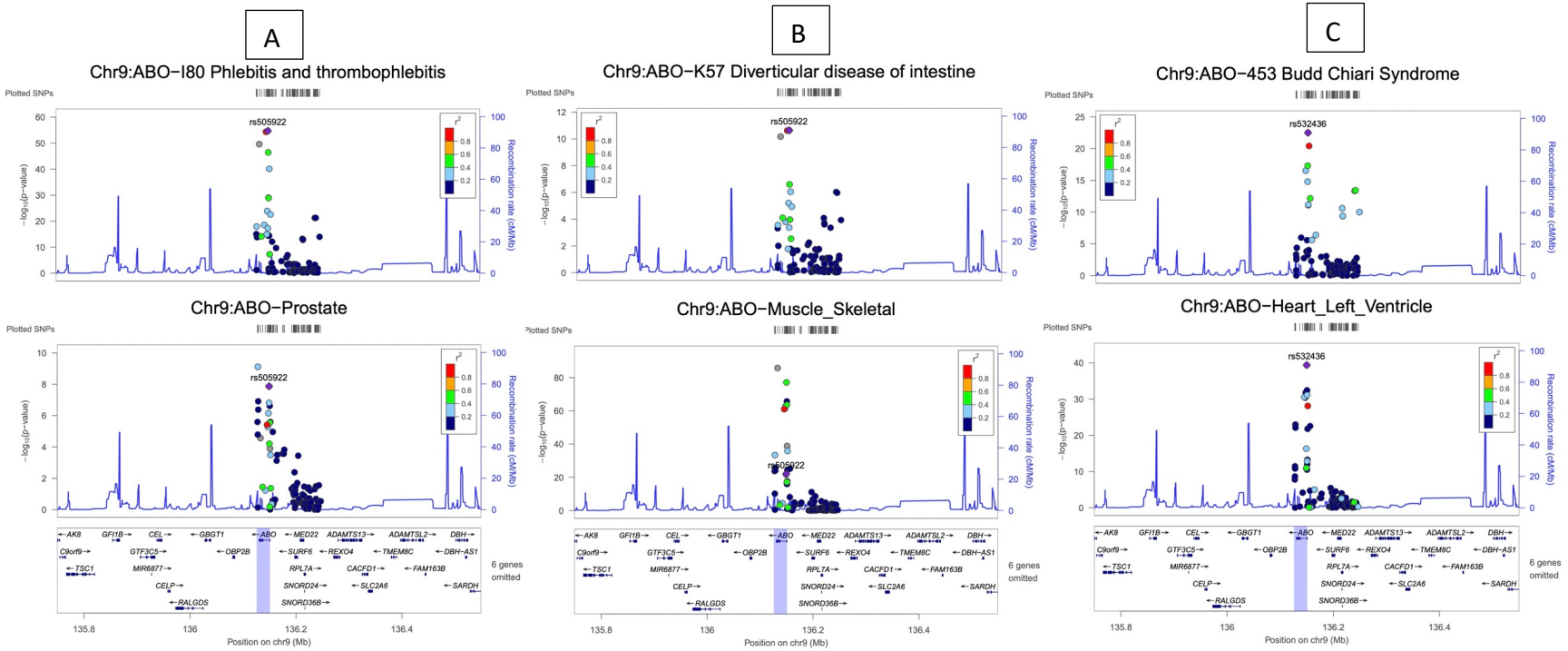

**Supplementary Figure 13. Local association plots for lipids (top) and gene expression (bottom) for the lipid TWAS-significant novel locus *ZSWIM1* on chromosome 20**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS (two-sided linear regression) of HDL in eMERGE, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in adipose subcutaneous tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb). *ZSWIM1* was Bonferroni-significant from lipid TWAS (Phase I).



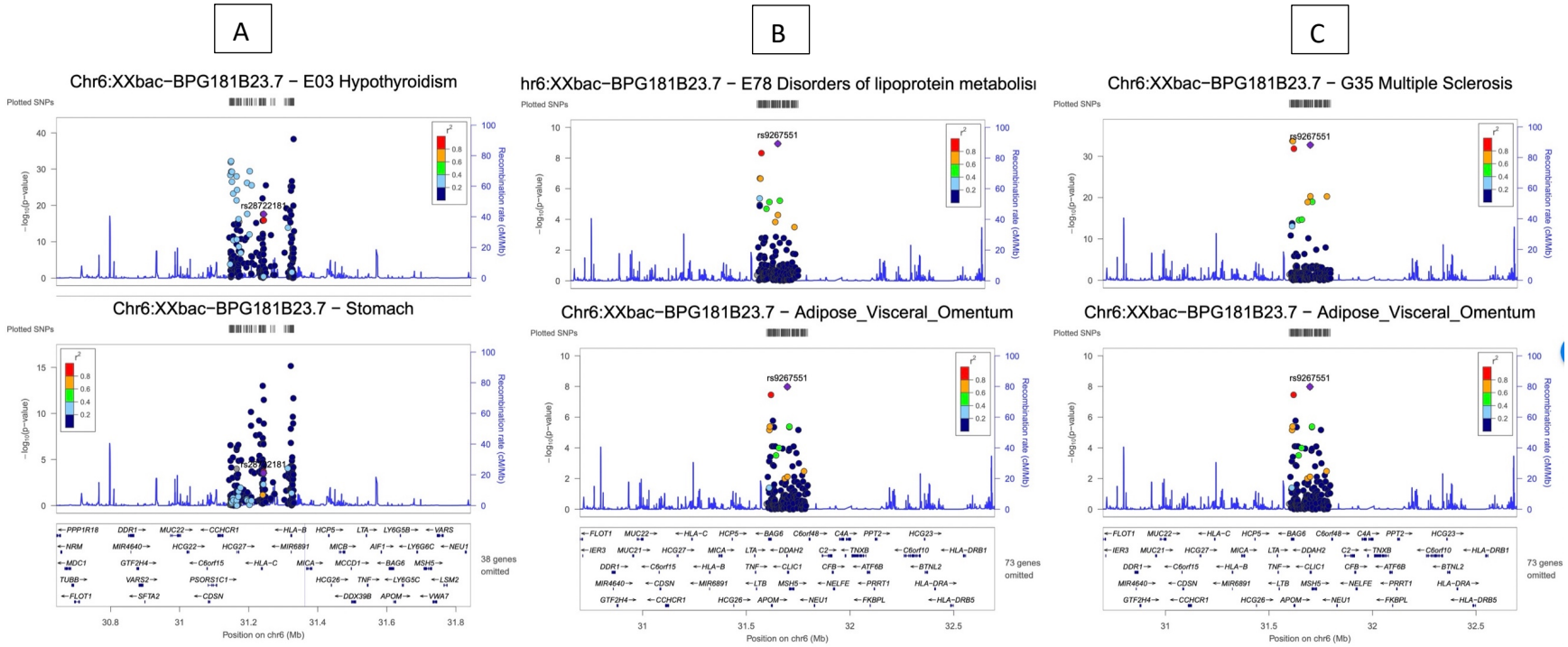

**Supplementary Figure 14. Local association plots for lipids (top) and gene expression (bottom) for the lipid TWAS-significant novel locus *ZSWIM1* on chromosome 20**. Y-axis on these LocusZoom plots indicates -logP values from: (top) GWAS (two-sided linear regression) of TG in GLGC, and (bottom) all variant-gene cis-eQTL associations (two-sided linear regression) in adipose subcutaneous tissue from GTEx v8. X-axis shows physical position on the chromosome (Mb). *ZSWIM1* was significant at the Bonferroni threshold from lipid TWAS (Phase I).



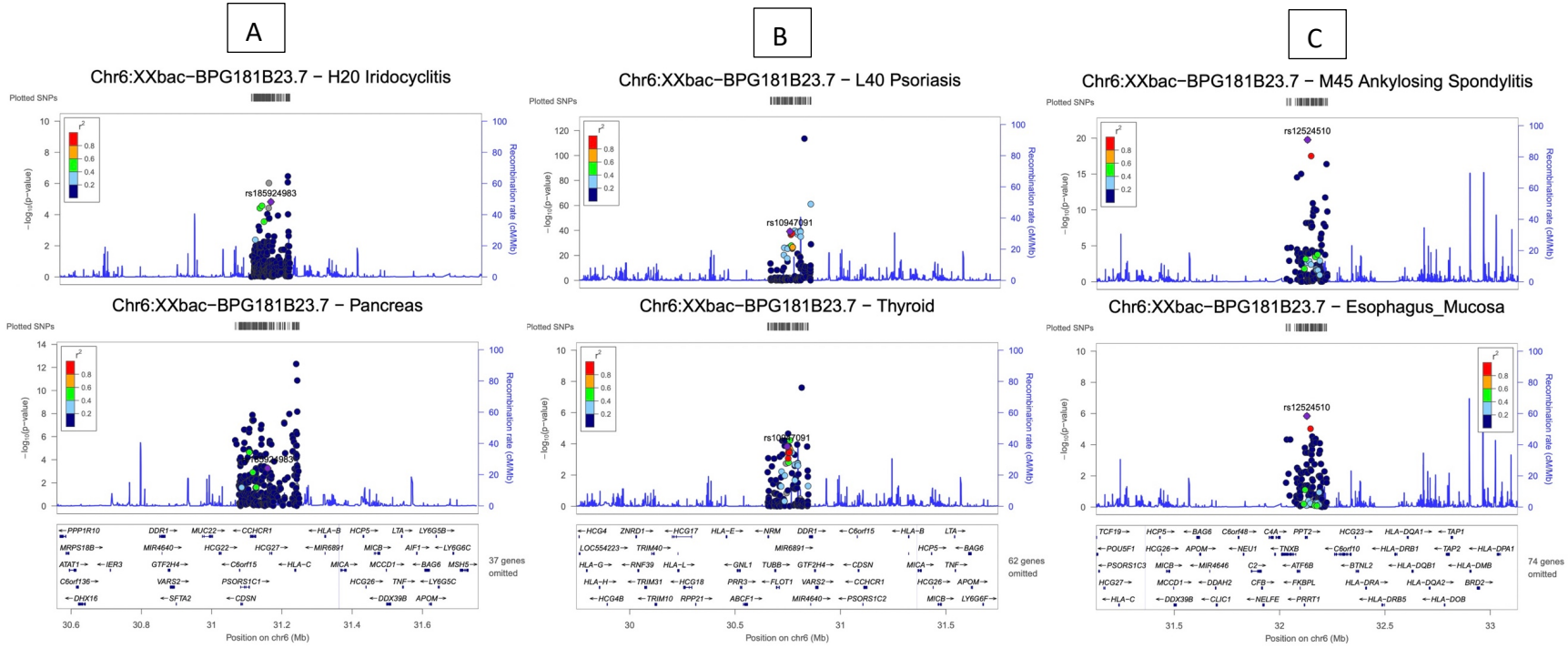

**Supplementary Figure 15. Local association plots for lipids/diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS-significant locus *ABO* on chromosome 9.** In (A) Y-axis shows -logP values from GWAS (two-sided linear regression) on the top whereas in (B) and (C) Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) on the top; in all plots, Y-axis shows all variant-gene cis-eQTL associations (two-sided linear regression) for *VARS2* across tissues of interest on the bottom, specifically, (A) GWAS of LDL in UKB (top), and liver tissue (bottom); (B) Pulmonary embolism in UKB (top), and muscle skeletal tissue (bottom); and (C) Haemorrhoids in UKB (top), and vagina tissue (bottom). X-axis shows physical position on the chromosome (Mb). *ABO* was significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).



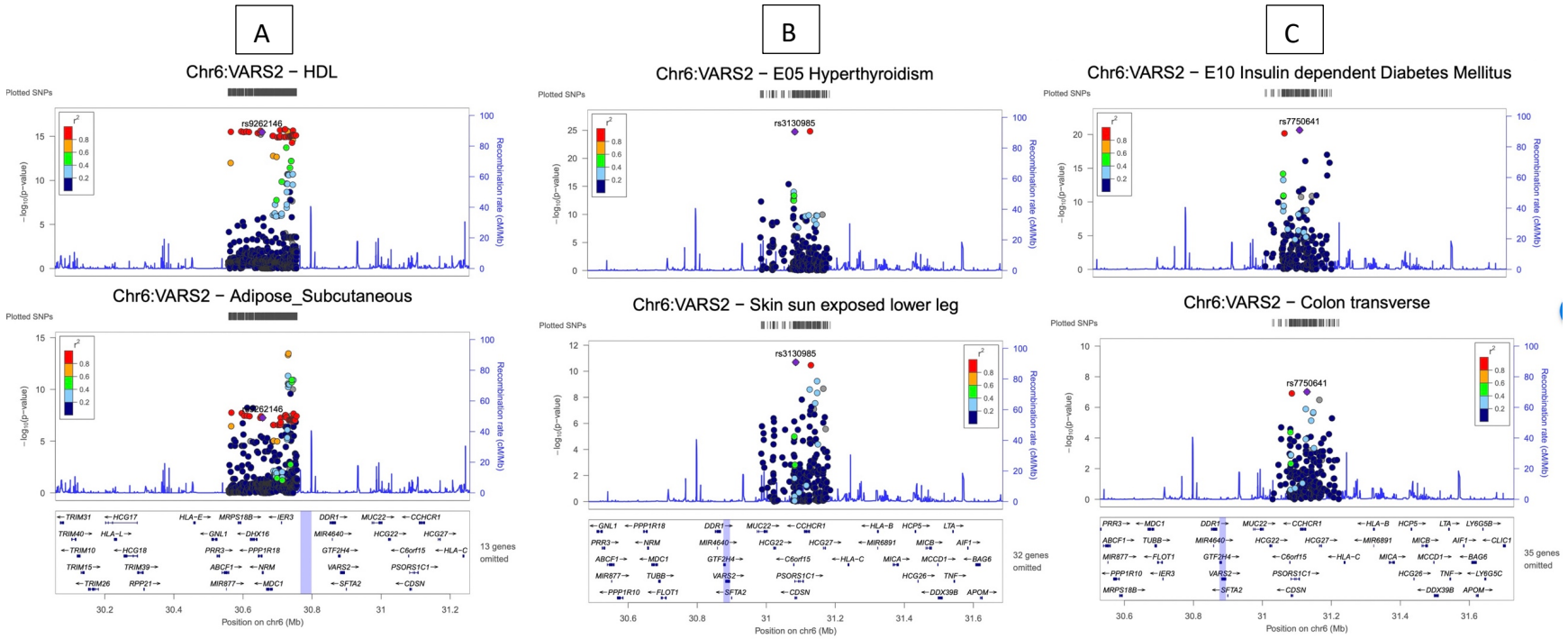

**Supplementary Figure 16. Local association plots for diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS significant locus *ABO* on chromosome 9.** In all plots Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) at the top and all variant-gene cis-eQTL associations (two-sided linear regression) for *ABO* across tissues of interest on the bottom; specifically (A) Phlebitis and thrombophlebitis in UKB (top), and prostate tissue (bottom); (B) Diverticular disease of intestine in UKB (top), and muscle skeletal tissue (bottom); and (C) Budd-Chiari Syndrome in eMERGE (top), and heart left ventricle tissue (bottom). X-axis shows physical position on the chromosome (Mb). *ABO* was significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).



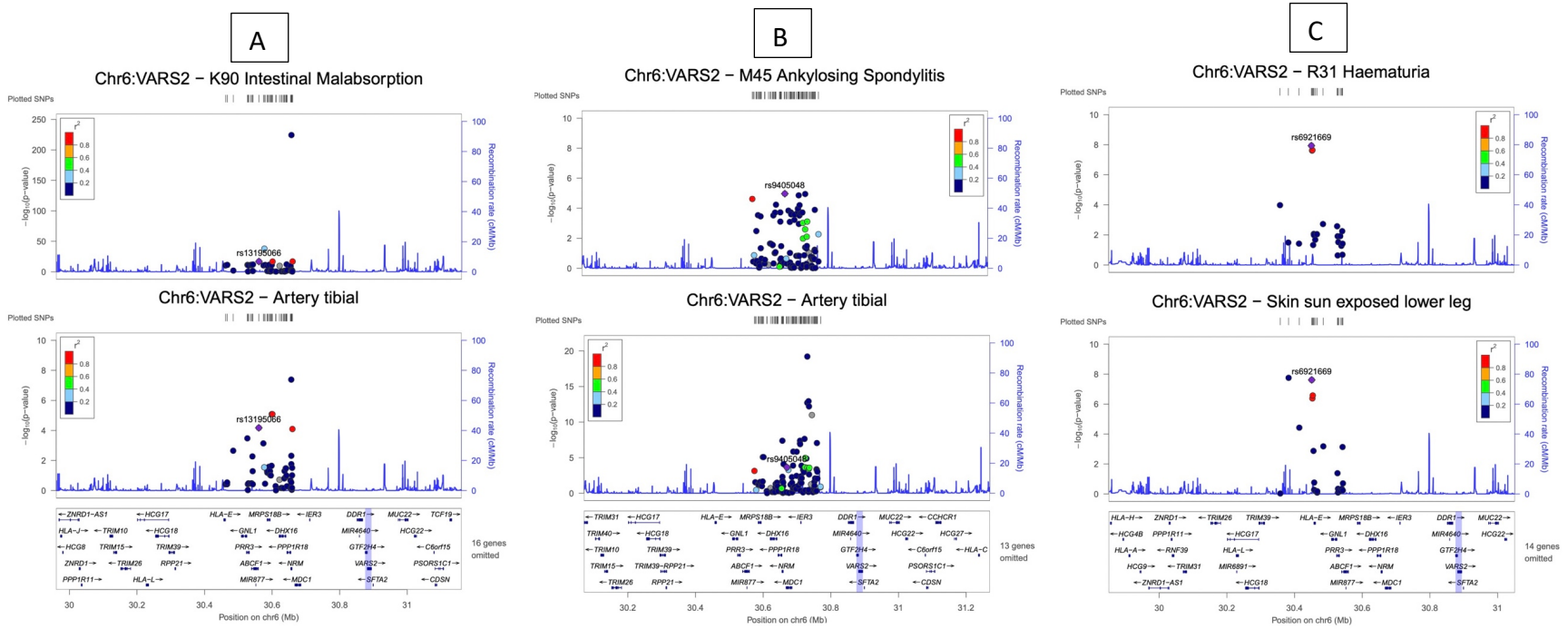

**Supplementary Figure 17. Local association plots for diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS significant locus *XXbac-BPG181B23.7* on chromosome 6.** In all plots Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) at the top and all variant-gene cis-eQTL associations (two-sided linear regression) for *XXbac-BPG181B23.7* across tissues of interest on the bottom; specifically, (A) Other hypothyroidism in UKB (top), and stomach tissue (bottom); (B) Disorders of lipoprotein metabolism in UKB (top), and adipose visceral omentum tissue (bottom); and (C) Multiple sclerosis in UKB (top), and adipose visceral omentum tissue (bottom). X-axis shows physical position on the chromosome (Mb). *XXbac-BPG181B23.7* was Bonferroni significant from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).



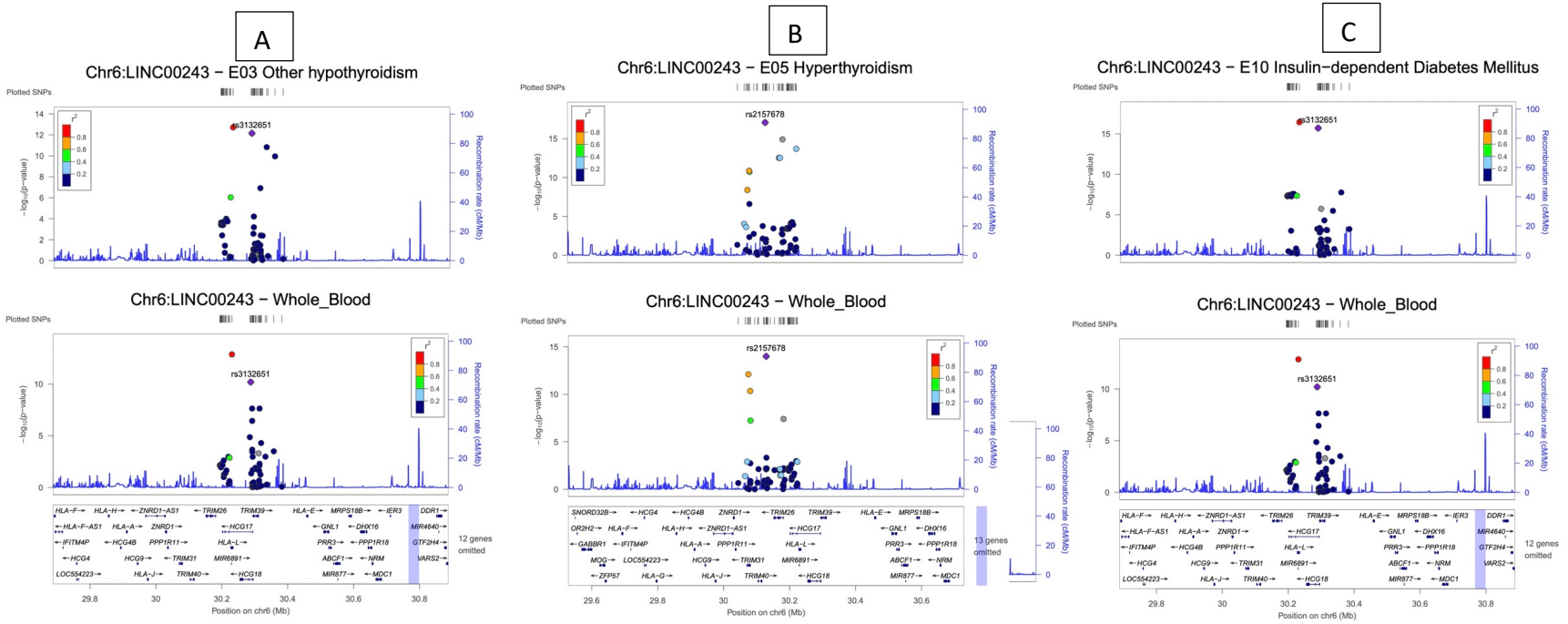

**Supplementary Figure 18. Local association plots for diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS significant locus *XXbac-BPG181B23.7* on chromosome 6.** In all plots Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) at the top and all variant-gene cis-eQTL associations (two-sided linear regression) for *XXbac-BPG181B23.7* across tissues of interest on the bottom; specifically (A) Iridocyclitis in UKB (top), and pancreas tissue (bottom); (B) Psoriasis in UKB (top), and thyroid tissue (bottom); and (C) Ankylosing spondylitis in UKB (top), and esophagus mucosa tissue (bottom). X-axis shows physical position on the chromosome (Mb). *XXbac-BPG181B23.7* was significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).



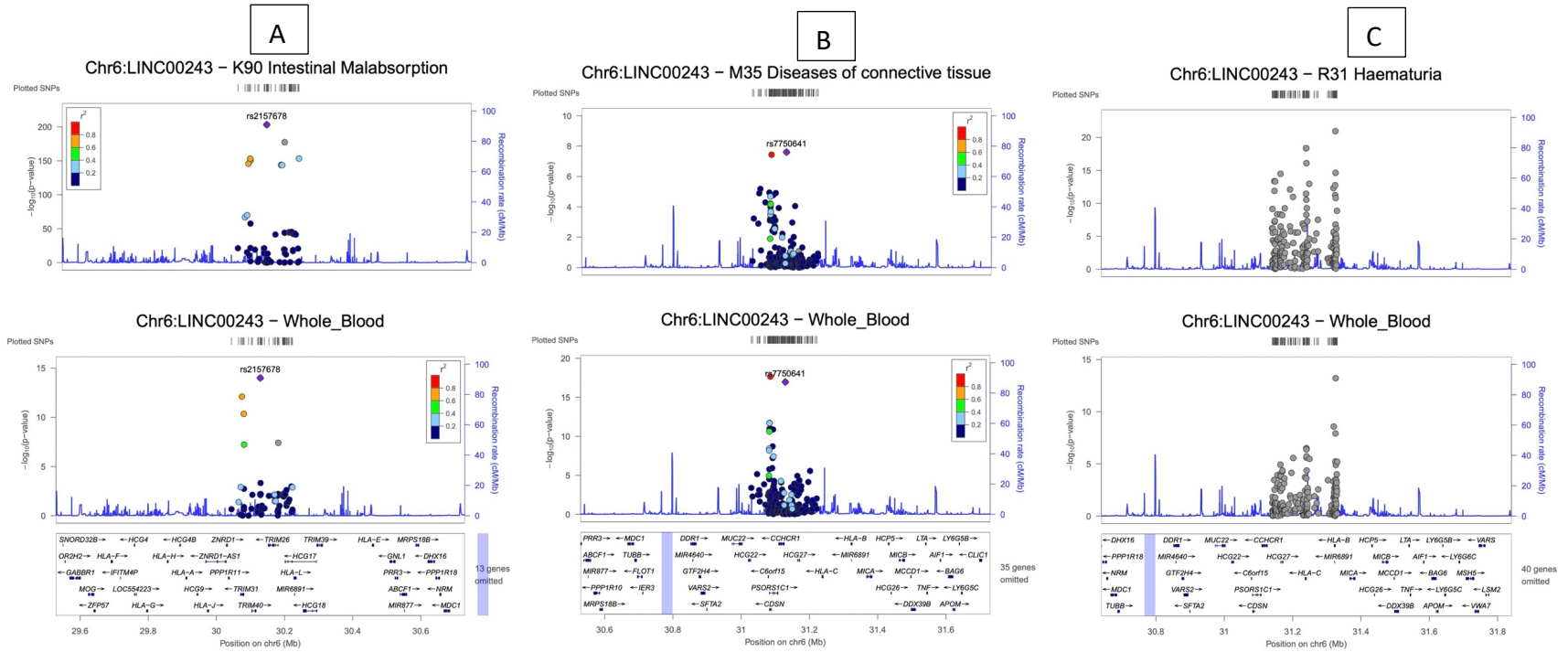

**Supplementary Figure 19. Local association plots for lipids/diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS significant locus *VARS2* on chromosome 6.** In (A) Y-axis shows -logP values from GWAS (two-sided linear regression) on the top whereas in (B) and (C) Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) on the top; in all plots, Y-axis shows all variant-gene cis-eQTL associations (two-sided linear regression) for *VARS2* across tissues of interest on the bottom, specifically (A) HDL (top), and adipose subcutaneous tissue (bottom); (B) Hyperthyroidism in UKB (top), and sun exposed lower leg tissue (bottom); and (C) Insulin dependent diabetes mellitus in UKB (top), and colon transverse tissue (bottom). X-axis shows physical position on the chromosome (Mb). *VARS2* was significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).



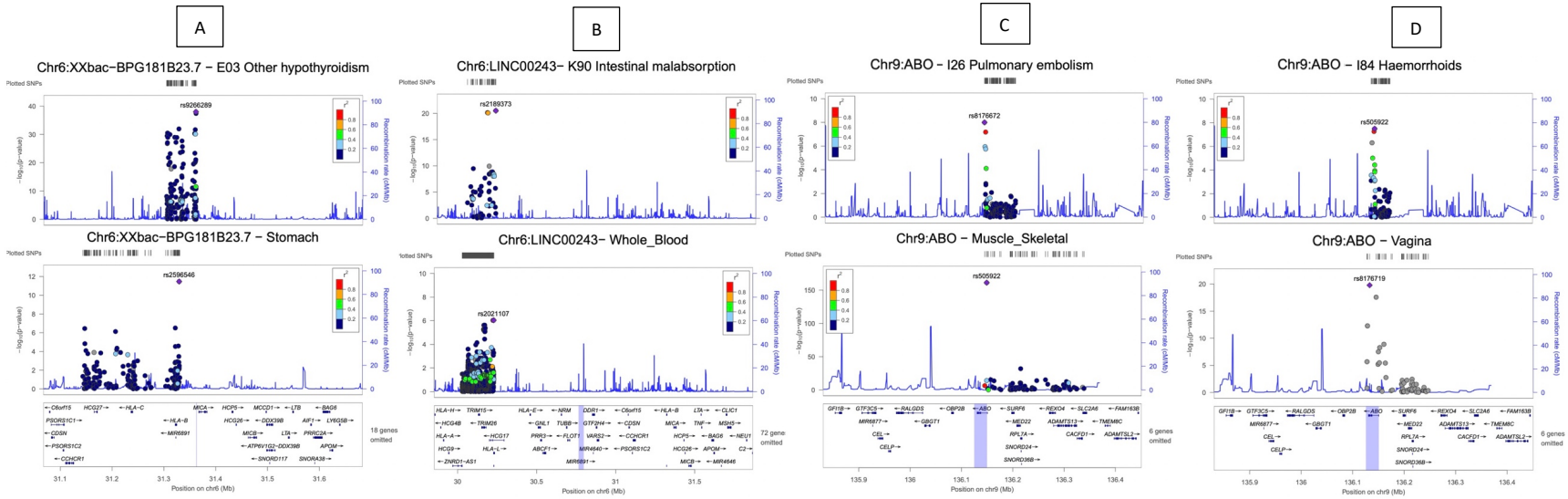

**Supplementary Figure 20. Local association plots for diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS significant locus *VARS2* on chromosome 6.** In all plots Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) at the top and all variant-gene cis-eQTL associations (two-sided linear regression) for *VARS2* across tissues of interest on the bottom; specifically (A) Intestinal malabsorption in UKB (top), and artery tibial tissue (bottom); (B) Ankylosing spondylitis in UKB (top), and artery tibial tissue (bottom); and (C) Haematuria in UKB (top), and skin sun exposed lower leg tissue (bottom). X-axis shows physical position on the chromosome (Mb). *VARS2* was significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).



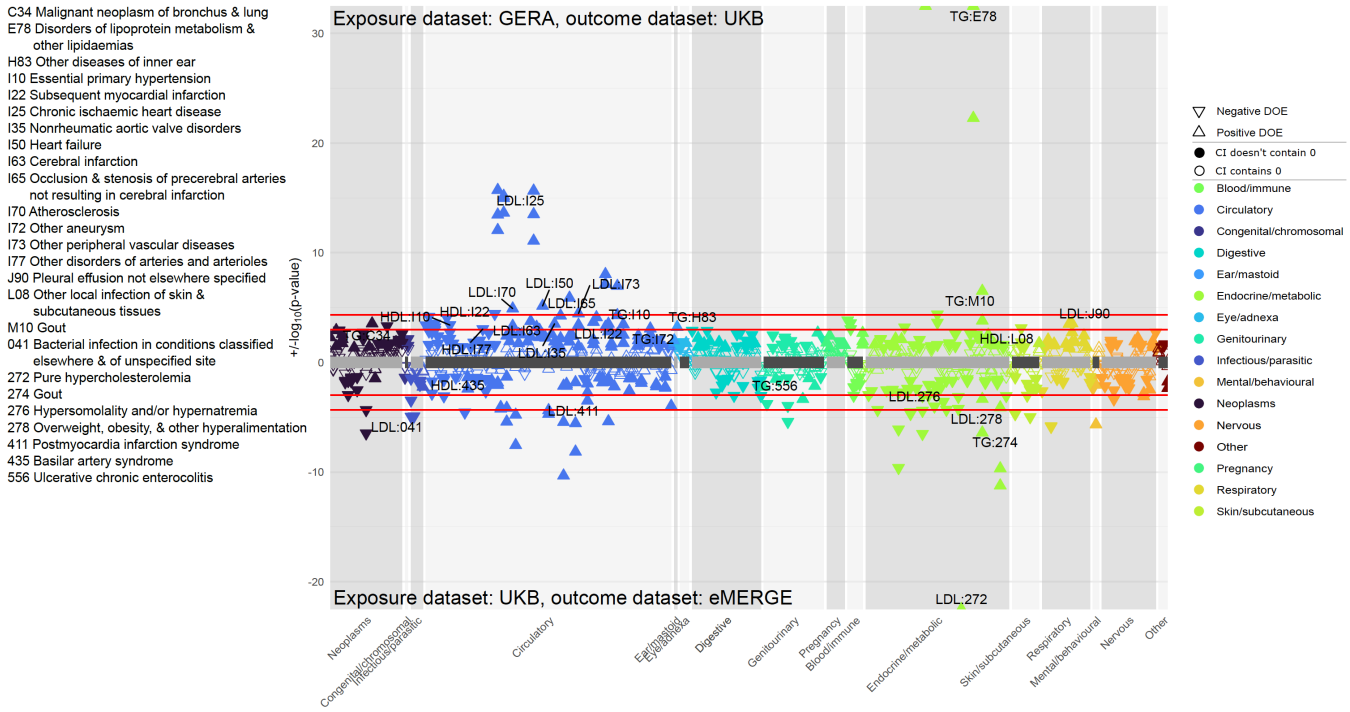

**Supplementary Figure 21. Local association plots for diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS significant locus *LINC00243* on chromosome 6.** In all plots Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) at the top and all variant-gene cis-eQTL associations (two-sided linear regression) for *LINC00243* across tissues of interest at the bottom; specifically (A) Other hypothyroidism in UKB (top), and whole blood tissue (bottom); (B) Hyperthyroidism in UKB (top), and whole blood tissue (bottom); and (C) Insulin dependent diabetes mellitus in UKB (top), and whole blood tissue (bottom). X-axis shows physical position on the chromosome (Mb). *LINC00243* was significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).



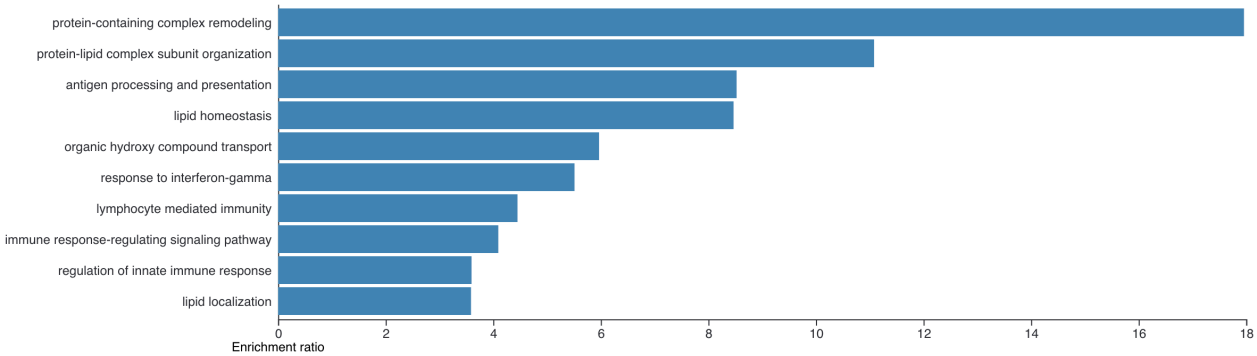

**Supplementary Figure 22. Local association plots for diseases (top) and gene expression from GTEx v8 (bottom) for the lipid TWAS- and Xpress-PheWAS significant locus *LINC00243* on chromosome 6.** In all plots Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) at the top and all variant-gene cis-eQTL associations (two-sided linear regression) for *LINC00243* across tissues of interest on the bottom; specifically (A) Intestinal malabsorption in UKB (top), and whole blood tissue (bottom); (B) Diseases of connective tissue in UKB (top), and whole blood tissue (bottom); and (C) Haematuria in UKB (top), and whole blood tissue (bottom). X-axis shows physical position on the chromosome (Mb). *LINC00243* was significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).




**Supplementary Figure 23. Local association plots for diseases from UKB (top) and gene expression from GTEx v8 (bottom) conditional on the top eQTL at the locus to reveal "secondary" signals**. In all plots Y-axis shows -logP values from lipid-guided PheWAS (two-sided logistic regression) at the top and all variant-gene cis-eQTL associations (two-sided linear regression) after conditioning on the top-eQTL at the locus at the bottom; specifically, (A) Other hypothyroidism in UKB (top), and stomach tissue (bottom) for the locus *XXbac-BPG181B23.7* on chromosome 6; (B) Intestinal malabsorption in UKB (top), and whole blood tissue (bottom) for the locus *LINC00243* on chromosome 6; (C) Pulmonary embolism in UKB (top) and muscle skeletal tissue (bottom) for the locus *ABO* on chromosome 9, and (D) Haemorrhoids in UKB (top), and vagina tissue (bottom) for the locus *ABO* on chromosome 9. X-axis shows physical position on the chromosome (Mb). All three genes were significant at the Bonferroni threshold from both lipid TWAS (Phase I) and Xpress-PheWAS (Phase II).




**Supplementary Figure 24. Two sample univariable Mendelian randomization excluding the *HLA* region on chromosome 6.** Top panel: Exposure dataset is GERA and outcome dataset is UKB. Bottom panel: Exposure dataset is UKB and outcome dataset is eMERGE. The diseases are grouped into different categories; direction of triangle corresponds to direction of MR effect. In each panel, the two red horizontal lines correspond to the Bonferroni and FDR thresholds. We label FDR-significant ICD codes from at least one of three two-sided tests (inverse-variance weighted, Egger, and median-based) with Egger pleiotropy (intercept) p-value > 0.05 as evidence of minimal heterogeneity. Filled points have confidence intervals that don't contain 0 whereas non-filled points have confidence intervals that contain 0.




**Supplementary Figure 25.** Results (FDR<0.05) from over-representation analysis for Bonferroni-significant genes detected from Xpress-PheWAS.