

A multi-ancestry genome-wide meta-analysis, fine-mapping, and gene prioritization approach to characterize the genetic architecture of adiponectin

Vishal Sarsani¹, Sarah M. Brotman², Yin Xianyong³, Lillian Fernandes Silva⁴, Markku Laakso⁴, Cassandra, N. Spracklen^{5*}.

¹ Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, Massachusetts, USA

² Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

³ Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA

⁴ Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland

⁵ Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, Amherst, Massachusetts, USA

Corresponding Author Contact Information: 715 N. Pleasant Street, 429 Arnold House, Amherst, MA 01003;

* cspracklen@umass.edu

Abstract

Previous genome-wide association studies (GWAS) for adiponectin, a complex trait linked to type 2 diabetes and obesity, identified >20 associated loci. However, most loci were identified in populations of European ancestry, and many of the target genes underlying the associations remain unknown. We conducted a multi-ancestry adiponectin GWAS meta-analysis in ≤46,434 individuals from the METSIM cohort and the ADIPOGen and AGEN consortiums. We combined study-specific association summary statistics using a fixed-effects, inverse variance-weighted approach. We identified 22 loci associated with adiponectin ($P < 5 \times 10^{-8}$), including 15 known and 7 previously unreported loci. Among individuals of European ancestry, GCTA-COJO identified 14 additional distinct signals at the *ADIPOQ*, *CDH13*, *HCAR1*, and *ZNF664* loci. Leveraging the multi-ancestry data, FINEMAP + SuSiE identified 46 causal variants (PP>0.9), which also exhibited potential pleiotropy for cardiometabolic traits. To prioritize target genes at associated loci, we propose a combinatorial likelihood scoring formalism (“GPScore”) based on measures derived from 11 gene prioritization strategies and the physical distance to the transcription start site. With “GPScore”, we prioritize the 30 most probable target genes underlying the adiponectin-associated variants in the multi-ancestry analysis, including well-known causal genes (e.g., *ADIPOQ*, *CDH13*) and novel genes (e.g., *CSF1*, *RGS17*). Functional association networks revealed complex interactions of prioritized genes, their functionally connected genes, and their underlying pathways centered around insulin and adiponectin signaling, indicating an essential role in regulating energy balance in the body, inflammation, coagulation, fibrinolysis, insulin resistance, and diabetes. Overall, our analyses identify and characterize adiponectin association signals and inform experimental interrogation of target genes for adiponectin.

Introduction

Adiponectin, an abundant adipocytokine secreted almost exclusively by adipocytes, is crucial in the interconnection between adiposity, insulin resistance, and inflammation¹⁻³. Previous genetic studies have identified more than 20 loci harboring variants associated with serum levels of adiponectin, including *ADIPOQ* and *CDH13*⁴⁻⁷. Genetic analyses have also indicated shared allelic architecture between adiponectin, type 2 diabetes, and other metabolic traits (e.g., BMI, WHR)^{4,7,8}. Due to its diverse physiological functions in glucose and lipid metabolism, inflammation, and oxidative stress across metabolic and cardiovascular tissues, adiponectin could be a possible therapeutic target for metabolic syndrome, diabetes, and coronary disease⁹. However, to move established loci toward effective clinical and therapeutic targets, the functional variant(s), target/effector gene(s), and the mechanistic direction(s) of effect need to be identified; for most loci identified to date, this information remains largely unknown.

Multiple methods have been proposed to prioritize target genes underlying genome-wide association study (GWAS) signals using expression, functional genomics, and network data¹⁰⁻¹². However, individual approaches often have conflicting findings, making it difficult to interpret or prioritize candidate target genes. Attempts to incorporate integrative and complementary gene prioritization approaches to identify disease-risk genes have been somewhat successful^{13,14}. However, the primary challenge in using prioritization approaches for a complex trait or disease is that the lack of customizability to focus on the integration of the data that is most relevant (e.g., tissue specificity) to the disease or trait of interest. This may result in a researcher choosing results in an ad hoc or post-hoc manner. A researcher may also want to maximize the likelihood of selecting a suitable target gene for experimental follow-up, and considering support from multiple approaches will make the selection(s) more robust.

Leveraging summary statistics from diverse genetic studies, we conducted a multi-ancestry, genome-wide meta-analysis for adiponectin in up to 46,434 individuals from the Metabolic Syndrome in Men (METSIM) cohort and the ADIPOGen and AGEN consortiums^{4,5,15}. Our primary objectives were to: 1) discover previously unreported loci associated with plasma adiponectin levels; 2) narrow putative causal variants underlying the association signals; 3) prioritize target genes systematically by using our proposed “GPScore” approach based on evidence derived from 11 gene prioritization strategies and physical distance to transcription start sites; and 4) perform functional profiling of target genes

and their underlying pathways.

Materials and methods

Ethics statement

The research protocol for all studies was reviewed and approved by the institutional ethics review committees at the involved institutions. Written informed consent was obtained from all study participants.

Study Design and Participants

Our meta-analysis included summary statistics from 16 studies of European-ancestry individuals from the ADIPOGen consortium (n = 29,347 from the discovery phase)⁴, four studies with individuals of East Asian-ancestry from the AGEN consortium (n=7,825)⁵, and the METSIM study (European-ancestry; n=9,262)¹⁵. While most of our analyses were performed in both European-only and all ancestries combined (“multi-ancestry”), we consider the multi-ancestry analyses to be our primary results. A detailed description of participant characteristics, genotype and phenotype information, quality control, and imputation can be found in Table S1.

Genome-wide meta-analyses

We performed the European-ancestry and multi-ancestry adiponectin meta-analyses using a fixed-effects inverse variance-weighted meta-analysis approach with the random-effects model (RE2) implemented in METASOFT¹⁶, which corrects for population structure while allowing for examination of heterogeneity statistics. In the multi-ancestry meta-analysis, we also generated Bayes factors (BF) from MR-MEGA¹⁷, which employs meta-regression to account for heterogeneity in allelic effects associated with ancestry. We defined the “lead” association signal at each locus to be the most significant variant ($P < 5 \times 10^{-8}$) within a 500 kb window. We considered association signals to be previously unreported, or “novel”, if they were located > 500 kb from a previously reported adiponectin signal. To identify distinct association signals at each locus identified in the European-ancestry analysis^{4-6,15,18}, we performed approximate conditional analyses using genome-wide joint conditional analysis (COJO) implemented in the Genome-wide Complex Traits Analysis (GCTA)

software¹⁹. Using data from 10,197 METSIM participants to calculate linkage disequilibrium (LD) and a collinearity threshold of $(R^2) = 0.9$, we considered a variant to represent an additional, distinct signal at a locus if: 1) the variant achieved $(P < 5 \times 10^{-8})$ in the COJO analysis, and 2) the variant was located within +/-1 Mb from the original lead variant at that locus. We use “index variant” to denote the most significant variant within each of the secondary association signals.

Pleiotropic associations with phenotypes from CMDKP We interrogated our lead and index variants with data from the Common Metabolic Diseases Knowledge Portal (www.cmdkp.org) to explore pleiotropic associations between the index variants from the multi-ancestry meta-analysis and other complex traits across five common disease areas (type 1 and type 2 diabetes, cardiovascular and cerebrovascular disease, and sleep disorders). We only considered traits/diseases that achieved $P < 0.05$ in the original analyses. For ease of visualization, we aggregated traits/diseases into 23 broad categories.

Identifying candidate causal variants

We used several approaches to identify candidate causal variants at loci identified in both the multi-ancestry and European ancestry-only meta-analyses. Unless otherwise stated, we used data from 10,197 METSIM participants as the reference for LD calculations.

Statistical fine-mapping We performed statistical fine-mapping using FINEMAP²⁰ and SuSiE²¹. At each adiponectin-associated locus, we computed in-sample dosage LD using LDstore²². We defined a fine-mapping region as the 3 Mb window (+/- 1.5 Mb) around each lead variant. We allowed up to 10 causal variants per window and extracted the posterior inclusion probabilities (PIP) of each variant using each method independently. The variants with a PIP > 0.90 in either of the fine-mapping methods, along with having LD $r^2 > 0.8$ with the lead variant, are considered the final candidate causal variants.

Causal variant annotation The majority of the candidate causal variants reside in non-coding regions of the human genome. We used RegulomeDB²³ to annotate the candidate causal variants from fine-mapping with evidence of regulatory function(s) through functional genomic assays and

computational approaches. RegulomeDB provides a score (range 1-7) indicating its potential to be functional in regulatory elements and the probability of confidence in the score for each variant. We interrogated fine-mapped candidate causal variants from the multi-ancestry fine-mapping using CAUSALdb²⁴, a database containing fine-mapping results from over 3,052 GWAS summary statistics. Whenever multiple variants mapped to the same trait across different studies, we meta-analyzed the different p-values together.

Gene Prioritization

We created the Gene Priority Score (GPScore), which combines evidence from 11 gene prioritization strategies^{11,21,25-33}, along with the physical distance to the transcription start site(s) (TSS), to prioritize target genes underlying our adiponectin association signals. For all approaches, we considered all protein-coding genes within the +/- 1.5 Mb of the original lead variant from the multi- or European-ancestry meta-analyses. For prioritization solely based on gene expression^{21,27-30}, we restricted our analysis to tissues most biologically-relevant to adiponectin (adipose subcutaneous, adipose visceral omentum, adrenal gland, artery aorta, artery coronary, artery tibial, heart atrial appendage, heart left ventricle, kidney cortex, liver, muscle-skeletal, thyroid, and whole blood). Unless otherwise stated, we used data from 10,197 METSIM participants as the reference for LD calculations.

Gene Prioritization Score GPScore is a combinatorial likelihood score constructed by leveraging various gene prioritization strategies described below. The score is unbiased and only involves weighting factors to prioritize signals from a particular tissue. GPScore is defined as

$$GPScore = \frac{(-\text{Log}_{10}(P_g) + S_G) \cdot C_S}{\log_2 TSS_d}$$

Where P_G is a combined P-value for MAGMA²⁵, summary Mendelian randomization (SMR)²⁸, expression quantitative trait loci (eQTL) colocalization (LD-based approach), SNP-heritability enrichment²⁹ and Downstreamer³² (described below), computed by the sum of z via Stouffer's method³⁴. The second term (S_G) represents the combined scores of eQTL colocalization (coloc)³⁵, EMS³⁰, PoPS³¹, EpiMap¹¹, cS2G³³ and Genehancer¹⁸ scores. We normalized scores for EMS, PoPS, and Genehancer scores by scaling between 0 and 1. The third term (C_S) is a score ranging

from 0 to 1, representing the proportion of support for a particular protein-coding gene across all 11 gene prioritization strategies. Each strategy has a maximum support score of 1. For eQTL colocalizations (coloc), SNP-heritability enrichment, EMS score, and SMR, we assigned a score of one if the tissue source was adipose or cardiac-related tissue (given the trait of interest is adiponectin); else, we assigned a score of 0.8. The final term (TSS_d) is the distance from the lead variant to the transcription start site of the gene/transcript, measured in base pairs.

One of the methodological issues in computing GPScore is that some of the gene prioritization strategies (i.e., coloc, EMS, SMR, GeneHancer, cs2G) report values for individual variants/position, whereas other strategies prioritize genes (i.e., MAGMA, Downstreamer, Epimap, PoPS, LDSC). We transformed variant/position level scores into gene-level scores using recommendations from Lehne et al.³⁶. For expression-based enrichments, a variant/position may have a statistic corresponding to a tissue (e.g., colocalization of a variant in adipose tissue). Using the average or highest quartile of a statistic may result in the inability to assign tissue information after the final transformation. Because preserving tissue information in the variant scores is essential for weighing support (C_S), we have opted to use either the lowest p-value (MAGMA, SMR, eQTL colocalization LD-based approach, SNP-heritability enrichment, Downstreamer) or the maximum score (coloc, EMS, PoPS, EpiMap, Genehancer) per gene. Alternatively, in situations where tissue information is less critical, the average or highest quartile of statistics could be used.

MAGMA gene analysis To quantify the degree of association of each gene with adiponectin while incorporating LD structure between variants, we used MAGMA v1.10²⁵ to perform gene-set analyses and obtain gene P-values from the multiple linear principal components regression F-test. The null hypothesis of the F-test is that the gene has no effect on the phenotype, conditional on all covariates. The variant-wise mean model was used, and variants were assigned to one of the 18,383 (GRCh37) protein-coding genes with an annotation window of 50 kb. We used the computed gene P-value to calculate the P_G term in GPScore.

eQTL colocalization (coloc) We used a combination of SuSiE²¹ and coloc²⁷ to assess for evidence of shared association signals/causal variants between our adiponectin GWAS data and *cis*-eQTLs from GTEx³⁷ using the coloc.susie() function³⁵. The coloc + SuSiE approach has been shown to improve the accuracy of colocalization analyses (coloc only) when multiple causal variants exist within a window³⁵. We extracted the posterior probability of the variant being causal for the shared

signal (H4) for each tissue most biologically-relevant to adiponectin and used it in the S_G term for the GPScore.

eQTL colocalization (LD-based) As a supplemental approach to coloc+SuSiE, we further examined for colocalizations between our adiponectin-associated variants and eQTLs in GTEx and in RNA-seq data from 434 METSIM participants using an LD-based approach²⁶. For colocalization with the GTEx data (restricted to adiponectin-related tissues described above), we considered the GWAS and eQTL signals to be colocalized when the lead or index GWAS variant and the variant most strongly associated with the expression level of the corresponding transcript (eSNP) exhibited high pairwise LD ($LD\ r^2 > 0.70$) in European + East Asian ancestry data within 1KGP328. For colocalization with the METSIM RNA-seq data, we considered the GWAS and eQTL signals to be colocalized when the lead or index GWAS variant and the eSNP exhibited high pairwise LD ($LD\ r^2 > 0.80$) calculated using 10,197 METSIM participants. The P-value of association with the eSNP was used to calculate the P_G term in GPScore.

Summary-based Mendelian randomization Summary-based Mendelian randomization (SMR) is another approach that integrates GWAS and eQTL data to identify genes whose expression levels are associated with a complex trait. We applied multi-SNP SMR²⁸ to test for the effect of gene expression (adiponectin-related tissues from GTEx) variation on adiponectin. We included the probes with at least one cis-QTL at $P_{EQTL} < 5 \times 10^{-8}$, and we performed a HEIDI (Heterogeneity in dependent instruments) test to exclude results that may reflect linkage. In GPScore calculations, we considered the P_{SMR} values of each probe corresponding to tissue, excluding probes with strong evidence of heterogeneity ($P_{HEIDI} > 0.01$). We used the P_{SMR} to calculate the P_G term in GPScore.

SNP-heritability enrichment To evaluate whether the variant heritability was enriched in the variants within tissue-specific genes (± 100 kb) compared to other regions, we applied partitioned LD score regression (LDSC)²⁹ with the ‘-overlap-annot’ option. We constructed the annotation list of tissue-specific genes by selecting genes with the top 10% median TPM (transcripts per million) values in each adiponectin-relevant tissue from the expression data. We calculated the P-values from one-sided Z-score coefficients for tissue-specific genes and included them in the P_G term in GPScore.

EMS annotation The EMS annotation is defined as the predicted probability that a variant has a *cis*-regulatory effect on gene expression, which we calculated by training a random forest model on fine-mapped eQTLs and 6,121 features, such as epigenetic marks and sequence-based neural network predictions³⁰. We extracted the computed EMS for all top variant-gene pairs in adiponectin-relevant tissues in GTEx v8³⁷ from the Finucane Lab (<https://www.finucanelab.org/data>). We then extracted the normalized EMS score for all variants from the adiponectin meta-analyses ($P < 0.05$ and minor allele frequency $> 0.50\%$) and then incorporated it into the S_G term in GPScore.

PoPS A gene-level Polygenic Priority Score (PoPS) is a similarity-based gene prioritization approach that leverages both polygenic and locus-specific genetic signals³¹. We used PoPS v0.2 to identify potential target genes from gene-level association statistics (derived from MAGMA) by integrating 57,543 gene features from public bulk and single-cell expression datasets, protein-protein interaction networks, and pathway databases. The PoPs score for each gene is then incorporated into the S_G term in GPScore.

Downstreamer We used Downstreamer³² to calculate co-regulation Z-scores per gene using expression data from 31,499 public RNA-seq samples from many different tissues³⁸. Integrating this information into the calculation of the gene priority score will likely prioritize genes that are co-regulated with many important adiponectin-associated genes in the GWAS. We calculated the P-values from one-sided co-regulation Z-scores per gene and included them in calculating the P_G term in GPScore.

Epimap Human epigenome reference, EpiMap (epigenome integration across multiple annotation projects) uses the correlation of enhancer activity with gene expression across cell types to map regulatory SNPs to their target genes. This information is useful in the complex trait investigation and studies of disease locus mechanisms. We downloaded gene-enhancer links by mark activity-by-gene expression correlation for the adipose tissue from the Epimap repository¹¹ and mapped the link scores for the genes overlapping in the 3Mb region around the lead variant. We then incorporated the link scores into the S_G term in GPScore.

GeneHancer GeneHancer calculates a score derived from: gene-enhancer genomic distances and combined scores of tissue co-expression correlation between genes and enhancer RNAs, enhancer-

targeted transcription factor genes, eQTLs for variants within enhancers, and capture Hi-C¹⁸. The score represents the strength of enhancer association to a particular gene (GeneHancer score), which can be useful in predicting regulatory elements and their target genes. We extracted the GeneHancer scores for enhancer gene pairs for genes physically located within the adiponectin fine-mapping regions and incorporated into the S_G term in GPscore.

Combined SNP-to-gene score annotation The SNP to Gene (cS2G) strategy is a heritability-based framework that combines different SNP-to-gene strategies to link regulatory variants to their target genes. We extracted the cS2G scores, which include 10 main functional SNP-to-gene strategies, as previously described³³, from 9,997,231 variants in the 1000 Genomes Project European reference panel and 19,476,620 variants in the UK Biobank. The cS2G score annotation was performed on selected variants ($P < 0.05$ and minor allele frequency $> 0.50\%$) from the meta-analyses. The cS2G score ranges from 0 to 1, where a score > 0.5 is considered good evidence; we incorporated this score into the S_G term in GPscore.

Disease enrichment and Functional association networks

Disease enrichment analysis of prioritized genes To study mechanisms underlying human diseases related to adiponectin, we investigated the human gene-disease associations of our adiponectin-prioritized genes. We examined for enrichment between the prioritized genes from the multi-ancestry adiponectin meta-analysis and a wide range of disease phenotypes relevant in human genomics by gene-disease enrichment analysis (DisGeNET)³⁹, using enrichr⁴⁰. We considered a DisGeNET term enriched if adjusted $P < 0.01$.

Functional association network construction To investigate potentially functionally-similar genes and their underlying pathways, we constructed a functional association network using GeneMANIA⁴¹. Prioritized genes alone do not contain enough information to build networks that mediate the underlying functional relationship. We used an algorithm⁴² in GeneMANIA that constructs network weights based on the reproducibility of Gene Ontology (GO) biological process co-annotation patterns. We expanded the gene list with functionally similar genes from multiple genomics and proteomics data sources. We primarily used the physical and genetic interaction, predicted protein interaction, and pathway and molecular interaction data available in

GeneMANIA.

Results

Multi-ancestry meta-analysis reveals 15 known and 7 novel risk loci for adiponectin

We conducted a multi-ancestry, genome-wide association meta-analysis for adiponectin using summary statistics from the METSIM cohort¹⁵ and the ADIPOGen⁴ and AGEN consortia⁵ ($N \leq 46,434$; Table S1). We also performed a genome-wide association meta-analysis for adiponectin in data from up to 38,609 individuals of European-ancestry (METSIM cohort and ADIPOGen consortium only)^{4,15}. Each individual cohort performed single-variant association analyses for adiponectin, adjusted for age, sex, BMI, and study-specific covariates as appropriate (e.g., principal components, study site). We meta-analyzed summary statistics using fixed-effects analyses implemented in METASOFT¹⁶. We corrected for population substructure in the multi-ancestry meta-analysis using the computed inflation factors for both the mean effect ($\lambda_{\text{MeanEffect}}$) and heterogeneity portions ($\lambda_{\text{heterogeneity}}$)¹⁶. Because RE2¹⁶ and MR-MEGA¹⁷ are not sufficiently powered in a meta-analysis of only three input files⁴³, we consider the fixed-effects results as our primary results; however, we present results from RE2 and MR-MEGA for comparison.

Multi-ancestry meta-analysis association results We identified 22 loci associated with adiponectin ($P < 5 \times 10^{-8}$), including seven loci that have not previously been associated with adiponectin (Table 1, Figure 1, Figure S1). Previously unreported loci are located at or near *CSF1*, *RGS17*, *ADRB1*, *PDE3B*, *RBMS2*, *HCAR1*, and *PHF23*. Effect direction for all lead variants showed concordance between all three studies, and the effect sizes for each of the lead variants were also consistent between the three datasets (Spearman $r \geq 0.88$; Figure S2). Further, MR-MEGA results for all 22 loci have a $\log_{10}\text{BF} > 30$, further supporting the results from the fixed-effect models. We also confirmed ($P < 0.05$) 83 out of 167 (49.7%) previously reported variants to be associated with adiponectin in our meta-analysis (Table S2).

Lead variants at four loci exhibited evidence of heterogeneity in effect sizes ($I^2 > 80\%$; rs1515108 (*IRS1*), rs17366568 (*ADIPOQ*), rs7978610 (*ZNF664*), rs12051272 (*CDH13*); Table 1). All four loci are well-established associations with adiponectin levels^{4,5,7}. Among the four heterogeneous loci, two

of the loci were still considered statistically significant in RE2 models accounting for the high heterogeneity (rs17366568, rs12051272). Large differences in effect allele frequencies between European- and East-Asian ancestry populations is likely driving the observed heterogeneity (Table 1, Figure S2).

European meta-analysis In a meta-analysis of European-ancestry individuals from METSIM and the ADIPOGen consortium, we identified 19 loci associated with adiponectin (Table 2, Figures 1, S2), 18 of which were also identified in the multi-ancestry results. One additional previously unreported locus, located near *LINC01214* on chromosome 1 (rs7617025) was identified only in the European-ancestry analysis. All the lead variants showed effect direction concordance between METSIM and ADIPOGen.

Distinct association signals At all 19 loci identified in the European-ancestry, we sought to identify additional association signals located within +/- 1 Mb of the lead variant using GCTA-COJO. We detected ($P_{joint} < 5 \times 10^{-8}$) 14 additional signals at four loci (Table S3; Figures 1, S3), including eight signals near *ADIPOQ* (LD $r^2 = 0.001-0.846$), one near *HCAR1* (LD $r^2 = 0.183$), one near *ZNF664* (LD $r^2 = 0.008$), and four near *CMIP/CDH13* (LD $r^2 = 0-0.046$) (Table S4).

Pleiotropic associations with complex traits We evaluated the lead variants at each of the 22 loci reported in the multi-ancestry meta-analysis for association with other complex traits and diseases. For these traits, we examined existing GWAS and genome-wide meta-analysis results in the Common Metabolic Diseases Knowledge Portal (www.cmdkp.org). We found 89 unique lead variant-phenotype combinations that reached Bonferroni significance (Table S5). Lead adiponectin-associated variants are most strongly associated with the lipids, hepatic, hematological, glycemcic, and anthropometric phenotype groups (Figure S4, Table S5).

Putative causal variants implicate shared biological mechanisms

Fine-mapping to identify potential causal variants To identify potential causal variants underlying adiponectin association signals in both the multi- and European-ancestry analyses, we first constructed 3 Mb windows (+/- 1.5 Mb) around the lead variant (Figure S5). The variants with a PIP > 0.90 in either FINEMAP²⁰ or SuSiE²¹ and LD $r^2 > 0.8$ with the lead GWAS variant were identified as candidate causal variants. In fine-mapping the multi-ancestry adiponectin loci, we

nominated 46 putative causal variants at 17 loci (Table S6; Figure S6). Among the 46 putative causal variants, 10 (rs2061155, rs1515108, rs1108842, rs998584, rs596359, rs10787516, rs10886863, rs7978610, rs12051272, rs731839) were also the lead variants from the meta-analysis. In the European-ancestry analyses, fine-mapping analyses nominated putative causal variants at seven loci, including two variants near *CDH13* (Table S7; Figure S7). Among the seven putative causal variants, six (rs16861209, rs998584, rs11045172, rs2925979, rs12051272, rs731839) were the lead variants from the meta-analysis.

Regulatory effects of nominated causal variants To investigate the regulatory effects among our nominated causal variants, we used the RegulomeDB²³ to obtain RegulomeDB scores, which rank the presence of regulatory motifs for variants on a scale from 1a (most evidence) to 7 (no data). In multi-ancestry fine-mapping results, 17 of the 46 putative causal variants were scored as “regulatory” (score < 5, with probability $\text{prob} \geq 0.50$; Table S6). rs683039, located at 12kb 5’ of *RGS17*, had the lowest RegulomeDB score (score=1f; $\text{prob}=0.93$), demonstrating the largest evidence for being in a regulatory region (Figure S1). In European-ancestry fine-mapping results, six of the eight putative causal variants were scored as “regulatory” (score < 5, with $\text{prob} p \geq 0.50$; Table S7). rs2925979, located near *CMIP*, had the lowest RegulomeDB score (score=2b; $\text{prob}=0.75$), demonstrating the largest evidence of being in a regulatory region.

Causal variants enriched for cardiometabolic traits We also examined our nominated causal variants for shared genetic architecture with fine-mapping results of 2,629 unique traits from over 3,052 GWAS summary statistics obtained from the CAUSALdb database²⁴, which includes UK Biobank and other cohorts. When a candidate adiponectin-causal variant was reported in more than one prior study, we meta-analyzed the p-values. In total, we found 306 unique lead variant-trait combinations that are Bonferroni significant (Table S8, Figure S8). We observed that most of our nominated causal variants show prior strong associations for liver-related diseases, type 2 diabetes mellitus, hormones, body size, and body composition measurements (Figure S8).

Gene prioritization highlights genes underlying adiponectin association signals

GPScore to nominate locus-specific target genes Variants underlying genetic association signals may not necessarily regulate the closest gene^{44–48}. To assess and prioritize target genes underlying adiponectin-associated signals, we developed a combinatorial likelihood scoring formalism (“GPScore”) based upon measures derived from 11 gene prioritization strategies (Tables S9-S30), along with the physical distance to the transcription start site(s) (TSS) (Figure 2). We integrated multiple lines of evidence into the score by using four terms: P_G (combined p-value), S_G (combined scaled scores), C_S (proportion of support), and TSS_d (distance from TSS of the gene to the lead GWAS variant). The most likely target gene(s) were selected based on the highest locus-specific GPScore within the 3Mb window. For the four regions with overlapping windows (12q24.31 near *HCAR1* and *ZNF664* and 16q23 near *CMIP* and *CDH13*), the flanking size of 1.5 Mb is reduced to 1 Mb to limit the window overlap.

In multi-ancestry analysis, we prioritized a total of 30 genes with high relative locus-specific GPScore values across 22 locus windows (Table 3, S31; Figure 3). Out of 22 regions, 15 had one prioritized gene, six had two prioritized genes (*PBRM1* and *GNL3* at 3p21.1, *ADIPOQ* and *RFC4* at 3q27.3, *NAPIL5* and *FAM13A* at 4q22.1, *FDGFR2* and *WDR11* at 10q26.12, *CTDNEP1* and *ELP5* at chr17p13.1, and *PEPD*, and *CEBPG* at 17p13.1), and one region had three prioritized genes (*CCDC92*, *DNAH10OS*, and *ZNF664* at 12q24.31) (Table 3, Figure 3). GPScore prioritized well-known adiponectin target genes concerning various cardiovascular/cardiometabolic diseases including *IRS1*, *ADIPOQ*, *PDE3A*, *CCDC92*, *CDH13*^{49–53}. GPScore also prioritized target genes for previously unreported loci that have been shown to play key roles in adipose-related metabolic functions (e.g., *CSFI*, *RGS17*, *ADRBI*, and *CYP2R1*)^{54–57}.

Gene prioritization results using variants from the European-ancestry analysis yielded similar results to the multi-ancestry prioritization (Table S32). At the 18 loci shared between European- and multi-ancestry analyses, 13 loci yielded the same set of prioritized genes. At 4 of the shared loci, the European-ancestry GPScore prioritized additional target genes (Table S32). At the chr17p13.1 locus shared between the European- and multi-ancestry results, GPScore prioritized a different target gene in the European-ancestry analysis (*ACAPI*) than in the multi-ancestry analysis (*CTDNEP1* and *ELP5*). Differences in GPScore prioritization between the two sets of analyses is likely explained by

differences in effect sizes and allele frequencies used as input to the gene prioritization strategies.

To see which of the GPScore terms received the most weight in the score calculation, we quantified each term's contribution to the GPScore (Figures 4a, S9) by fitting a multiple regression model and averaging over orderings proposed by Lindeman et al. and implemented in the *realimpo* package⁵⁸. Across the 22 multi-ancestry regions tested, the average contributions (on a scale of 0-1) were: $S_G=0.133$, $C_S=0.199$, $TSS_d=0.255$, and $P_G=0.411$. The 2q36.3 window where *IRS1* is prioritized has the lowest TSS_d contribution since the transcription start site for *IRS1* is 541,389bp away from the lead adiponectin GWAS variant but has the strongest support for being the effector gene (Table S31). The 3p21.1 window has the highest TSS_d contribution since both the prioritized genes *PBMRI* and *GNL3* are close to the lead variant (Figure 4d). The variants in 5q11.2,6p21.1 and 8q24.13 windows have the lowest contribution of C_S (proportion of support) due to a lack of support from many of the gene prioritization strategies (Figure 4a). The P_G (combined p-value) has an average contribution of 0.411 across all the windows. Regarding the gene prioritization strategies, all prioritized genes have MAGMA P-value $< 5 \times 10^{-8}$, relatively higher POPS scores, and CS2G annotations while supporting the majority of other strategies (Figure S9). Because several prioritization strategies rely on a reference panel for LD calculation, we repeated GPScore using only the METSIM cohort summary statistics; the top prioritized genes for the common windows remain the same (Table S33).

Pinpointing biological functionality through disease enrichment and functional interaction networks

Disease enrichment analysis We evaluated the association of multi-ancestry prioritized genes with disease phenotypes using gene-disease enrichment analysis^{39,40}. We found a strong enrichment (adjusted $P < 0.01$) of DisGeNET terms related to cardiometabolic traits and diseases, including waist-hip ratio, arteriosclerosis obliterans, non-alcoholic fatty liver disease, body mass index, coronary artery disease, hypoadiponectinemia, and high-density lipoprotein measurement (Figure 5a, Tables S34-35). Other highly enriched terms are related to cancer outcomes, including thymoma, endometrial neoplasms, endometrial adenocarcinoma, giant cell tumors, noninfiltrating intraductal carcinoma, and non-hereditary clear cell renal cell carcinoma. These findings support recent epidemiological evidence that links adiponectin to cancer^{59,60}.

Functional association network For the multi-ancestry meta-analysis results, we constructed an interactive functional association network, illustrating the relationships among the connected genes, prioritized genes, and associated pathways (Figures 5, S10; Tables S36-39). In Figure 5, we display the network results where we use the default of adding 20 “related” genes and pathways into the network. With the default network size, 76.6% of our prioritized genes (23/30) were incorporated into functional association networks, with most having multiple connected genes (Figure 5b). A total of 10 pathways were represented in the network, with many genes heavily involved in the adipocytokine signaling pathway, integrins in angiogenesis, leptin pathway, and type 2 diabetes.

Discussion

Adiponectin is an insulin-sensitizing and anti-inflammatory hormone secreted by adipocytes whose signaling is involved in several metabolic processes crucial to type 2 diabetes, neurodegeneration, and cardiovascular disease. In this study, we conducted the largest multi-ancestry adiponectin GWAS meta-analysis to date, in up to 46,434 individuals from the METSIM cohort and the ADIPOGen and AGEN consortiums. We identified 22 loci associated with adiponectin ($P < 5 \times 10^{-8}$), including 15 known and seven novel loci. European-ancestry results revealed 14 additional secondary signals at the *ADIPOQ*, *CDH13*, *HCAR1*, and *ZNF664* loci. Most of the lead variants showed associations with other complex traits/diseases, suggesting pleiotropy. We narrowed the list of putative causal variants to 46 (PP>0.9), 17 of which demonstrated moderate to strong regulatory evidence. Our nominated causal variants also showed potential pleiotropy for other complex cardiometabolic traits, such as liver-related diseases, type 2 diabetes, hormones, body size, and body compositional measurements.

Variants underlying genetic association signals may not necessarily regulate the closest gene⁴⁴⁻⁴⁸ and may impact protein levels thousands or even millions of base pairs away⁶¹. Based on prior estimates⁶², the 3 Mb fine-mapping mapping region (+/- 1.5 Mb around the index variant) should roughly capture more than 99% of such regulatory activity. Multiple methods currently exist to prioritize target, or effector, genes underlying GWAS signals, primarily from gene expression quantitative trait loci (eQTL) data¹⁰, regulatory genomic activity¹¹, and protein-protein interaction networks and pathway databases¹². Two common challenges of existing gene prioritization approaches are 1) inadequate customizability to represent the data which is disease/tissue relevant, and 2) conflicting findings from various approaches due to differences in assumptions, methodology,

and representation models, resulting in the selection of methods/results in an ad hoc or post hoc manner. To address these limitations, we developed combinatorial likelihood scoring formalism, GPScore, based upon measures derived from 11 gene prioritization strategies, along with the physical distance to the transcription start site(s) (TSS). GPScore integrates multiple lines of evidence into the unbiased score that can be customizable to include weighting factors to prioritize signals from a particular tissue and can be broadly applicable to other complex traits that do not have much training data. By utilizing GPScore, we prioritize the 30 most probable target/effector genes for adiponectin in the multi-ancestry analysis. The prioritized gene list includes several well-known causal genes related to serum adiponectin levels. This list includes *ADIPOQ*, which is responsible for encoding the adiponectin protein, and *CDH13*, which encodes an adiponectin receptor protein⁵¹. Insulin receptor substrate-1 (*IRS1*) is a substrate of the insulin receptor tyrosine kinase, and it is central to inducing insulin actions, including binding and activating PI3K and increasing glucose transport⁶³. Acute knockdown in adipocytes was shown to result in increased *Adipoq* mRNA expression, but reduced adiponectin secretion⁶⁴. The treatment of myoblast cell line (C2C12) with adiponectin was also shown to enhance the ability of insulin to stimulate *IRS-1* tyrosine phosphorylation and Akt phosphorylation⁵⁰. *CCDC92* has been shown to be important in insulin resistance and influencing adipocyte differentiation⁶⁵, and knockdown of *CCDC92* in mice was shown to reduce obesity and insulin resistance⁶⁶. We also used GPScore to prioritize target genes among our novel loci from our multi-ancestry meta-analysis, including several for which the biological mechanisms underlying the association with adiponectin may be unclear. The *CSF-1* gene, which was prioritized at 1p13.3, has shown to have therapeutic potential in tissue repair⁶⁷ due to its involvement in macrophage homeostasis and inflammation via the *CSF-1* receptor⁶⁸. There is also some supportive evidence that *CSF-1/CSF-1R* signaling is essential in the pathology of and diabetes⁶⁹. Due to its essential role in G protein signaling, *RGS17* is considered a potential therapeutic target for lung and prostate cancers⁷⁰; however, individuals with type 2 diabetes have been shown to have higher levels of *RGS17* auto antibodies than individuals without type 2 diabetes, and expression levels of *RGS17* are significantly upregulated in adipose tissue during periods of weight loss⁷¹. *ADRB1*, which encodes the adrenergic beta-1 receptor, was prioritized at 10q25.3. While *ADRB1* is most known for the functional role it plays in cardiomyocyte function^{72,73}, visceral adiponectin release was shown to be triggered by cAMP via signaling pathways involving adrenergic beta receptors in mice. The *CYP2R1* protein, a receptor for vitamin D 25-

hydroxylase in the liver ⁷⁴, plays a crucial role in non-alcoholic fatty liver disease ⁷⁵. Epidemiological studies have also demonstrated associations between serum vitamin D and adiponectin levels. While there is less existing evidence to support the mechanism(s) between some of the prioritized genes and adiponectin levels (e.g., *TRIB1*, *RBMS2*), a further experimental examination into the prioritized genes is warranted.

Our functional profiling of prioritized genes with human disease phenotypes revealed the enrichment of terms related to cardiovascular diseases, coronary artery disease, and other cardiometabolic traits. Interestingly, many genes also showed enrichment related to cancer etiopathogenesis, and recent evidence has shown adiponectin's antiangiogenic and tumor growth-limiting properties of adiponectin ⁵⁹. Our functional association network analysis also helped illustrate the complex interactions of prioritized genes, their functionally connected genes, and their underlying pathways. We observed the key underlying pathways centered around insulin signaling (through *IRS1*, *IRS2*, *INS*, *INSR*) and adiponectin signaling (through *ADIPOR1*, *ADIPOR2*) suggesting potential crosstalk and an essential role in regulating energy balance in the body, inflammation, coagulation, fibrinolysis, insulin resistance, and diabetes.

There are several caveats to our current GPScore gene prioritization approach that should be considered. First, while we were able to utilize our approach in both single- and multi-ancestry summary data, our multi-ancestry analysis is still primarily European (83.14%) and includes only one other ancestry population (East-Asian ancestry); thus, more diverse studies are needed to determine if our approach is less robust when including data with additional heterogeneity. Because our data was largely from European-ancestry populations, we used data from 10,197 Finnish METSIM participants as the reference for LD calculations for our fine-mapping and gene prioritization strategies. We acknowledge, however, that this is likely not the most appropriate panel for LD calculation in our multi-ancestry population. We plan to expand our work by utilizing a larger, more diverse ancestry LD reference panel. However, the flexibility of our approach allows researchers to determine the most appropriate LD panel for their study population. The GPScore prioritization procedure depends on the quality and availability of disease/cell-relevant data, the accuracy of the strategy used, and the annotation quality of variants and genes. As additional data becomes available, it could easily be incorporated into our GPScore approach. It is also possible that transforming variant/position-level scores into gene-level scores may result in a loss of information; however, this information is still retained in the initial prioritization approaches and

could still be considered. Lastly, while beneficial to rank genes in a locus-specific manner, the strength of GPScore does not implicate causality for a complex trait or disease, nor is it comparable across locus windows.

In conclusion, we discovered 15 known and seven previously unreported genomic loci associated with adiponectin through multi-ancestry genome-wide meta-analysis. The 46 putative causal variants identified through fine mapping showed potential pleiotropy for cardiovascular/metabolic traits. In addition, we provide a novel and customizable gene prioritization method, GPScore, that integrates multiple lines of evidence, thus increasing the likelihood of identifying the true target gene underlying a GWAS association signal. Our prioritized genes and underlying pathways yield new insights into the genetic architecture of adiponectin. Future research should include phenotyping of adiponectin in large-scale, diverse biobanks and cohorts, thus increasing sample size and ancestral diversity for robust prediction of target genes.

Acknowledgements

We thank the individuals in the METSIM cohort for their continued study participation.

Declaration of Interests

The authors declare no competing interests.

Data and code availability

The summary data of the ADIPOGen consortium, AGEN consortium, and METSIM study may be accessed through their respective websites. The code to reproduce the analyses and figures of this paper is available as Jupyter Notebooks at <https://github.com/vsarsani/GPScore>.

References

- [1] Achari, A. E. and Jain, S. K. (2017). Adiponectin, a Therapeutic Target for Obesity, Diabetes, and Endothelial Dysfunction. *Int J Mol Sci* 18.

- [2] Tilg, H. and Moschen, A. R. (2006). Adipocytokines: mediators linking adipose tissue, inflammation and immunity. *Nat Rev Immunol* 6, 772–783.
- [3] Li, S., Shin, H. J., Ding, E. L., and van Dam, R. M. (2009). Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. *Jama* 302, 179–188.
- [4] Dastani, Z., Hivert, M. F., ..., and O’Leary, D. H. (2012). Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* 8, e1002607.
- [5] Wu, Y., Gao, H., Li, H., Tabara, Y., Nakatochi, M., Chiu, Y. F., Park, E. J., Wen, W., Adair, L. S., Borja, J. B., et al. (2014). A meta-analysis of genome-wide association studies for adiponectin levels in East Asians identifies a novel locus near WDR11-FGFR2. *Hum Mol Genet* 23, 1108–1119.
- [6] Spracklen, C. N., Iyengar, A. K., Vadlamudi, S., Raulerson, C. K., Jackson, A. U., Brotman, S. M., Wu, Y., Cannon, M. E., Davis, J. P., Crain, A. T., et al. (2020). Adiponectin GWAS loci harboring extensive allelic heterogeneity exhibit distinct molecular consequences. *PLoS Genet* 16, e1009019.
- [7] Spracklen, C. N., Karaderi, T., Yaghootkar, H., Schurmann, C., Fine, R. S., Kutalik, Z., Preuss, M. H., Lu, Y., Wittemans, L. B. L., Adair, L. S., et al. (2019). Exome-Derived Adiponectin-Associated Variants Implicate Obesity and Lipid Biology. *Am J Hum Genet* 105, 15–28.
- [8] Mahajan, A., Spracklen, C. N., Zhang, W., Ng, . M. I., M. C. Y., Morris, A. P., eger, S., Della Briotta Parolo, P., Joo, Y. Y., and Hayes, M. G. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* 54, 560–572.
- [9] Fisman, E. Z. and Tenenbaum, A. (2014). Adiponectin: a manifold therapeutic target for metabolic syndrome, diabetes, and coronary disease? *Cardiovasc Diabetol* 13, 103.
- [10] Aguet, F., Anand, S., Ardlie, K. G., Gabriel, S., Getz, G. A., Graubert, A., Hadley, K., Handsaker, R. E., Huang, K. H., Kashin, S., et al. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330.
- [11] Boix, C. A., James, B. T., Park, Y. P., Meuleman, W., and Kellis, M. (2021). Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* 590, 300–307.
- [12] Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, D607–D613.

- [13] Aragam, K. G., Jiang, T., Goel, A., Kanoni, S., Wolford, B. N., ..., de Vries, P. S., and von Scheidt, M. (2022). Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet* 54, 1803–1815.
- [14] Schwartzenuber, J., Cooper, S., Liu, J. Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., Young, A. M. H., Franklin, R. J. M., Johnson, T., Estrada, K., et al. (2021). Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer’s disease risk genes. *Nat Genet* 53, 392–402.
- [15] Laakso, M., Kuusisto, J., Stanek, A., Kuulasmaa, T., Pajukanta, P., Lusa, A. J., Collins, F. S., Mohlke, K. L., and Boehnke, M. (2017). The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res* 58, 481–493.
- [16] Han, B. and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* 88, 586–598.
- [17] MÃ¶ngi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N., McCarthy, M. I., and Morris, A. P. (2017). Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* 26, 3639–3650.
- [18] Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* 2017.
- [19] Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 44, 369–375.
- [20] Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 32, 1493–1501.
- [21] Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society, Series B* pp. forthcoming.

- [22] Benner, C., Havulinna, A. S., Järvelin, M. R., Salomaa, V., Ripatti, S., and Pirinen, M. (2017). Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* *101*, 539–551.
- [23] Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* *22*, 1790–1797.
- [24] Wang, J., Huang, D., Zhou, Y., Yao, H., Liu, H., Zhai, S., Wu, C., Zheng, Z., Zhao, K., Wang, Z., et al. (2020). CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic Acids Res* *48*, D807–D816.
- [25] de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* *11*, e1004219.
- [26] Raulerson, C. K., Ko, A., Kidd, J. C., Currin, K. W., Brotman, S. M., Cannon, M. E., Wu, Y., Spracklen, C. N., Jackson, A. U., Stringham, H. M., et al. (2019). Adipose tissue gene expression associations reveal hundreds of candidate genes for cardiometabolic traits. *The American Journal of Human Genetics* *105*, 773–787.
- [27] Giambartolomei, C. and et. al. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genetics*.
- [28] Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M. E., Wray, N. R., Visscher, P. M., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* *48*, 481–487.
- [29] Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* *47*, 1228–1235.
- [30] Wang, Q. S., Kelley, D. R., Ulirsch, J., Kanai, M., Sadhuka, S., Cui, R., Albors, C., Cheng, N., Okada, Y., Aguet, F., et al. (2021). Leveraging supervised learning for functionally informed fine-mapping of cis-eQTLs identifies an additional 20,913 putative causal eQTLs. *Nat Commun* *12*, 3394.
- [31] Weeks, E. M., Ulirsch, J. C., Cheng, N. Y., Trippe, B. L., Fine, R. S., Miao, J., Patwardhan, T. A., Kanai, M., Nasser, J., Fulco, C. P., et al. (2020). Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *medRxiv*.

- [32] Bakker, O. B., Claringbould, A., Westra, H.-J., Wiersma, H., Boulogne, F., Vösa, U., Symmons, S. M., Jonkers, I. H., Franke, L., and Deelen, P. (2021). Linking common and rare disease genetics through gene regulatory networks. medRxiv.
- [33] Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K. K., Nasser, J., Jagadeesh, K. A., Weiner, D. J., Shi, H., Fulco, C. P., O'Connor, L. J., et al. (2022). Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nat Genet* 54, 827–836.
- [34] Stouffer, S. A., Suchman, E. A., DeVinney, L. C., and Shirley, A. Star, and robin m. williams. 1949. the american soldier: Adjustment during army life. *Studies in social psychology in World War II I*.
- [35] Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet* 17, e1009440.
- [36] Lehne, B., Lewis, C. M., and Schlitt, T. (2011). From SNPs to genes: disease association at the gene level. *PLoS One* 6, e20133.
- [37] Consortium, G. (2020). The gtex consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330.
- [38] Deelen, P., van Dam, S., Herkert, J. C., Karjalainen, J. M., Brugge, H., Abbott, K. M., van Diemen, C. C., van der Zwaag, P. A., Gerkes, E. H., Zonneveld-Huijssoon, E., et al. (2019). Improving the diagnostic yield of exome-sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. *Nat Commun* 10, 2837.
- [39] Pinero, J., Ramirez-Anguita, J. M., Sauch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., and Furlong, L. I. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 48, D845–D855.
- [40] Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128.
- [41] Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G. D., and Morris, Q. (2018). GeneMANIA update 2018. *Nucleic Acids Res* 46, W60–W64.
- [42] Mostafavi, S. and Morris, Q. (2010). Fast integration of heterogeneous data sources for predicting gene function with limited annotation. *Bioinformatics* 26, 1759–1765.

- [43] Willems, E. L., Wan, J. Y., Norden-Krichmar, T. M., Edwards, K. L., and Santorico, S. A. (2020). Transethnic meta-analysis of metabolic syndrome in a multiethnic study. *Genet Epidemiol* *44*, 16–25.
- [44] Claussnitzer, M., Dankel, S. N., Kim, K. H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I. S., Beaudry, J. L., Puvion-Randall, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* *373*, 895–907.
- [45] Farh, K. K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shores, N., Whitton, H., Ryan, R. J., Shishkin, A. A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* *518*, 337–343.
- [46] Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., Jansen, R., de Geus, E. J., Boomsma, D. I., Wright, F. A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* *48*, 245–252.
- [47] Gamazon, E. R., Segrè, A. V., van de Bunt, M., Wen, X., Xi, H. S., Hormozdiari, F., Ong, H., Konkashbaev, A., Derks, E. M., Aguet, F., et al. (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* *50*, 956–967.
- [48] Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* *51*, 592–599.
- [49] Sevillano, J., de Castro, J., Bocos, C., Herrera, E., and Ramos, M. P. (2007). Role of insulin receptor substrate-1 serine 307 phosphorylation and adiponectin in adipose tissue insulin resistance in late pregnancy. *Endocrinology* *148*, 5933–5942.
- [50] Wang, C., Mao, X., Wang, L., Liu, M., Wetzel, M. D., Guan, K. L., Dong, L. Q., and Liu, F. (2007). Adiponectin sensitizes insulin signaling by reducing p70 S6 kinase-mediated serine phosphorylation of IRS-1. *J Biol Chem* *282*, 7991–7996.
- [51] Gao, H., Kim, Y. M., Chen, P., Igase, M., Kawamoto, R., Kim, M. K., Kohara, K., Lee, J., Miki, T., Ong, R. T., et al. (2013). Genetic variation in CDH13 is associated with lower plasma adiponectin levels but greater adiponectin sensitivity in East Asian populations. *Diabetes* *62*, 4277–4283.
- [52] Ding, B., Abe, J. I., Wei, H., Huang, Q., Walsh, R. A., Molina, C. A., Zhao, A., Sadoshima, J., Blaxall, B. C., Berk, B. C., et al. (2005). Functional role of phosphodiesterase 3 in cardiomyocyte apoptosis: implication in heart failure. *Circulation* *111*, 2469–2476.

- [53] Neville, M. J., Wittemans, L. B. L., Pinnick, K. E., Ä, M., Kaksonen, R., inen, K. H., Luan, J., Scott, R. A., Wareham, N. J., Langenberg, C., et al. (2019). Regional fat depot masses are influenced by protein-coding gene variants. *PLoS One* *14*, e0217644.
- [54] Levine, J. A., Jensen, M. D., Eberhardt, N. L., and O'Brien, T. (1998). Adipocyte macrophage colony-stimulating factor is a mediator of adipose tissue growth. *J Clin Invest* *101*, 1557–1564.
- [55] Huang, X., Charbeneau, R. A., Fu, Y., Kaur, K., Gerin, I., MacDougald, O. A., and Neubig, R. R. (2008). Resistance to diet-induced obesity and improved insulin sensitivity in mice with a regulator of G protein signaling-insensitive G184S Gnai2 allele. *Diabetes* *57*, 77–85.
- [56] Riis-Vestergaard, M. J., Richelsen, B., Bruun, J. M., Li, W., Hansen, J. B., and Pedersen, S. B. (2020). Beta-1 and Not Beta-3 Adrenergic Receptors May Be the Primary Regulator of Human Brown Adipocyte Metabolism. *J Clin Endocrinol Metab* *105*.
- [57] Finelli, C. and Tarantino, G. (2013). What is the role of adiponectin in obesity related non-alcoholic fatty liver disease? *World J Gastroenterol* *19*, 802–812.
- [58] GrÄmping, U. (2006). Relative importance for linear regression in r: The package relaimpo. *Journal of Statistical Software* *17*, 1–27.
- [59] Dalamaga, M., Diakopoulos, K. N., and Mantzoros, C. S. (2012). The role of adiponectin in cancer: a review of current evidence. *Endocr Rev* *33*, 547–594.
- [60] Pham, D. V. and Park, P. H. (2022). Adiponectin triggers breast cancer cell death via fatty acid metabolic reprogramming. *J Exp Clin Cancer Res* *41*, 9.
- [61] Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Cortes, A., Koprulu, M., WÄrheide, M. A., Oerton, E., Cook, J., Stewart, I. D., Kerrison, N. D., et al. (2021). Mapping the proteo-genomic convergence of human diseases. *Science* *374*, eabj1541.
- [62] Fauman, E. B. and Hyde, C. (2022). An optimal variant to gene distance window derived from an empirical definition of cis and trans protein QTLs. *BMC Bioinformatics* *23*, 169.
- [63] Rondinone, C. M., Wang, L. M., Lonroth, P., Wesslau, C., Pierce, J. H., and Smith, U. (1997). Insulin receptor substrate (IRS) 1 is reduced and IRS-2 is the main docking protein for phosphatidylinositol 3-kinase in adipocytes from subjects with non-insulin-dependent diabetes mellitus. *Proc Natl Acad Sci U S A* *94*, 4171–4175.

- [64] Groeneveld, M. P., Brierley, G. V., Rocha, N. M., Siddle, K., and Semple, R. K. (2016). Acute knockdown of the insulin receptor or its substrates Irs1 and 2 in 3T3-L1 adipocytes suppresses adiponectin production. *Sci Rep* 6, 21105.
- [65] Zhao, W., Rasheed, A., Tikkanen, E., Lee, J.-J., Butterworth, A. S., Howson, J. M., Assimes, T. L., Chowdhury, R., Orho-Melander, M., Damrauer, S., et al. (2017). Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nature genetics* 49, 1450–1457.
- [66] Yang, M., Qiu, S., He, Y., Li, L., Wu, T., Ding, N., Li, F., Zhao, A. Z., and Yang, G. (2021). Genetic ablation of C-reactive protein gene confers resistance to obesity and insulin resistance in rats. *Diabetologia* 64, 1169–1183.
- [67] Hume, D. A. and MacDonald, K. P. (2012). Therapeutic applications of macrophage colony-stimulating factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling. *Blood* 119, 1810–1820.
- [68] Lin, W., Xu, D., Austin, C. D., Caplazi, P., Senger, K., Sun, Y., Jeet, S., Young, J., Delarosa, D., Suto, E., et al. (2019). Function of CSF1 and IL34 in Macrophage Homeostasis, Inflammation, and Cancer. *Front Immunol* 10, 2019.
- [69] Carrero, J. A., McCarthy, D. P., Ferris, S. T., Wan, X., Hu, H., Zinselmeyer, B. H., Vomund, A. N., and Unanue, E. R. (2017). Resident macrophages of pancreatic islets have a seminal role in the initiation of autoimmune diabetes of NOD mice. *Proc Natl Acad Sci U S A* 114, E10418–E10427.
- [70] Bodle, C. R., Mackie, D. I., and Roman, D. L. (2013). RGS17: an emerging therapeutic target for lung and prostate cancers. *Future Med Chem* 5, 995–1007.
- [71] Chang, D. C., Piaggi, P., Hanson, R. L., Knowler, W. C., Bucci, J., Thio, G., Hohenadel, M. G., Bogardus, C., and Krakoff, J. (2015). Use of a high-density protein microarray to identify autoantibodies in subjects with type 2 diabetes mellitus and an hla background associated with reduced insulin secretion. *PLoS One* 10, e0143551.
- [72] Bäckström, M., Magnusson, Y., Hjalmarson, A., and Andersson, B. (2000). A novel polymorphism in the gene coding for the beta(1)-adrenergic receptor associated with survival in patients with heart failure. *Eur Heart J* 21, 1853–1858.
- [73] Brodde, O. E. (2008). Beta-1 and beta-2 adrenoceptor polymorphisms: functional importance, impact on cardiovascular diseases and drug responses. *Pharmacol Ther* 117, 1–29.

- [74] Cheng, J. B., Levine, M. A., Bell, N. H., Mangelsdorf, D. J., and Russell, D. W. (2004). Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc Natl Acad Sci U S A* *101*, 7711–7715.
- [75] Eliades, M. and Spyrou, E. (2015). Vitamin D: a new player in non-alcoholic fatty liver disease? *World J Gastroenterol* *21*, 1718–1727.

Table 1. Multi-ancestry meta-analysis of adiponectin associations. Lead variant at each locus attaining genome-wide significance ($P < 5 \times 10^{-8}$) in the multi-ancestry meta-analysis (ADIPOGen, AGEN, METSIM).

Lead variant	Chr	hg19 Position	Nearest Gene	N	E/NEA	EAF		Fixed Effects			Random Effects (RE2)		MR-MEGA Bayes Factor		
						(ADIPOGen/METSIM/AGEN)	(ADIPOGen/METSIM/AGEN)	P-value	Beta	SE	I2	Q-value		P-value	Heterogeneity
<i>Novel loci achieving genome-wide significance in multi-ancestry meta-analysis</i>															
rs333947	1	110,470,764	CSF1	46,340	A/G	0.15 / 0.18 / 0.28		1.22E-08	-0.031	0.006	42.5	3.5	9.88E-02	0.0	33.2
rs596359	6	153,457,053	RGS17	46,408	T/C	0.43 / 0.45 / 0.26		2.43E-08	0.023	0.004	76.3	8.4	5.04E-02	1.3	36.9
rs10787516	10	115,813,924	ADRB1	46,406	T/C	0.40 / 0.46 / 0.59		1.24E-09	0.026	0.004	71.3	7.0	6.46E-02	0.3	41.2
rs11023332	11	14,784,110	PDE3B	46,367	C/G	0.43 / 0.44 / 0.35		3.27E-08	-0.022	0.004	7.1	2.2	1.10E-01	0.0	30.0
rs2657888	12	56,938,383	RBMS2	45,417	T/G	0.36 / 0.39 / 0.25		2.60E-09	0.025	0.004	15.6	2.4	8.36E-02	0.0	35.1
rs601339	12	123,174,743	HCAR1	46,388	G/A	0.15 / 0.17 / 0.49		2.33E-14	0.040	0.005	0.0	0.2	2.45E-02	0.0	55.8
rs222852	17	7,140,606	PHF23	46,404	G/A	0.49 / 0.38 / 0.37		3.83E-09	0.026	0.004	67.7	6.1	8.35E-02	0.1	38.2
<i>Previously reported loci achieving genome-wide significance in multi-ancestry meta-analysis</i>															
rs2061155	1	219,665,008	LYPALI-ASI	46,406	T/C	0.40 / 0.37 / 0.37		5.34E-09	-0.024	0.004	70.5	6.8	9.03E-02	0.0	38.2
rs1515108	2	227,123,086	IRS1	46,370	T/C	0.38 / 0.39 / 0.91		2.30E-11	-0.028	0.004	89.6	19.2	1.22E-03	6.8	61.2
rs1108842	3	52,720,080	PBRM1	46,400	C/A	0.46 / 0.47 / 0.57		7.04E-16	0.033	0.004	78.6	9.3	1.08E-02	0.8	71.8
rs17366568	3	186,570,453	ADIPOQ	38,849	A/G	0.09 / 0.14 / 0.03		7.90E-99	-0.168	0.007	89.4	18.9	4.45E-12	7.1	461.5
rs13131633	4	89,739,479	FAM13A	46,349	T/C	0.49 / 0.45 / 0.65		1.85E-08	-0.023	0.004	66.8	6.0	1.03E-01	0.0	35.0
rs6450176	5	53,298,025	ARL15	46,302	A/G	0.23 / 0.23 / 0.44		1.43E-12	-0.033	0.005	69.4	6.5	2.93E-02	0.4	54.0
rs998584	6	43,757,896	VEGFA	45,226	A/C	0.48 / 0.50 / 0.56		8.19E-12	-0.033	0.005	57.5	4.7	4.52E-02	0.0	48.7
rs2980879	8	126,481,475	TRIB1	41,147	T/A	0.38 / 0.28 / 0.28		2.06E-09	0.028	0.005	0.0	1.1	8.15E-02	0.0	34.3
rs10886863	10	122,929,493	WDR11-FGFR2	38,199	T/C	0.008 / 0.07 / 0.36		1.80E-12	0.077	0.011	71.4	7.0	2.65E-02	0.7	54.0
rs11045172	12	20,470,221	PDE3A	44,409	C/A	0.19 / 0.23 / 0.79		4.03E-12	0.040	0.005	31.0	2.9	4.20E-02	0.0	48.3
rs10778506	12	107,143,260	RFX4	46,400	T/C	0.47 / 0.29 / 0.67		1.60E-08	0.024	0.004	55.6	4.5	1.02E-01	0.0	33.7
rs7978610	12	124,468,572	ZNF664	44,745	C/G	0.39 / 0.27 / 0.08		7.94E-16	0.037	0.005	91.6	23.8	1.87E-04	8.4	86.0
rs2925979	16	81,534,790	CMIP	46,408	T/C	0.28 / 0.31 / 0.41		5.59E-28	-0.049	0.004	71.1	6.9	6.88E-04	0.7	124.5
rs12051272	16	82,663,288	CDH13	32,655	T/G	0.009 / 0.11 / 0.29		3.89E-213	-0.358	0.011	94.2	34.3	1.41E-24	15.2	1002.5
rs731839	19	33,899,065	PEPD	46,229	A/G	0.33 / 0.34 / 0.46		3.28E-21	0.042	0.004	75.1	8.0	2.53E-03	1.2	94.7

*Novel locus is ≥ 500 kb from previously reported adiponectin-associated variant achieving genome-wide significance ($P < 5 \times 10^{-8}$; GWAS catalog, March 2022)

Table 2. Meta-analysis of adiponectin associations in European-ancestry population. List of all lead variants achieving genome-wide significance ($P < 5 \times 10^{-8}$) in the meta-analysis of adiponectin associations in the European population (ADIPOGen and METSIM).

Lead variant	Chr	hg19 Position	Nearest Gene	N	EA/NEA (ADIPOGen/METSIM)	Fixed Effects			Random Effects (RE2)				
						EAF	P-value	Beta	SE	I2	Q-value	P-value	Heterogeneity
<i>Novel loci for any adiposity trait achieving genome-wide significance in European ancestry meta-analysis</i>													
rs7617025	3	150,055,958	LINC01214	38,578	A/G	0.09 / 0.07	7.47E-10	-0.045	0.007	0.0	0.6	1.10E-09	0.0
rs10787516	10	115,813,924	ADRB1	38,579	C/T	0.40 / 0.46	3.33E-09	0.026	0.004	85.6	6.9	2.47E-09	1.3
rs11023332	11	14,784,110	PDE3B	38,540	C/G	0.43 / 0.44	3.24E-08	-0.023	0.004	37.8	1.6	4.08E-08	0.00
rs2657888	12	56,938,383	RBMS2	38,562	G/T	0.36 / 0.39	5.92E-09	-0.025	0.004	57.5	2.4	8.44E-09	0.0
rs601339	12	123,174,743	HCAR1	38,561	G/A	0.15 / 0.17	4.88E-13	0.039	0.005	0.0	0.2	8.38E-13	0.0
rs17805277	17	7,201,753	YBX2	35,526	G/T	0.11 / 0.13	6.94E-09	0.042	0.007	0.0	0.7	9.86E-09	0.0
<i>Previously reported loci achieving genome-wide significance in European ancestry meta-analysis</i>													
rs2061155	1	219,665,008	LYPLAL1-AS1	38,579	C/T	0.40 / 0.37	9.10E-09	-0.025	0.004	84.8	6.6	8.00E-09	0.9
rs2943657	2	227,123,439	IRSI	38,543	C/T	0.38 / 0.39	1.71E-10	-0.028	0.004	94.2	17.4	1.93E-12	9.8
rs1108842	3	52,720,080	PBRM1/STAB1	38,574	C/A	0.46 / 0.47	2.00E-15	0.033	0.004	88.7	8.9	1.00E-15	2.6
rs16861209	3	186,563,114	ADIPOQ	38,435	A/C	0.01 / 0.03	1.30E-96	0.204	0.010	96.5	28.3	3.73E-100	19.3
rs13131633	4	89,739,479	FAM13A	38,524	C/T	0.49 / 0.45	3.59E-08	-0.023	0.004	83.2	5.9	3.53E-08	0.5
rs6450176	5	53,298,025	ARL15	38,475	A/G	0.23 / 0.23	2.23E-10	-0.032	0.005	81.1	5.3	3.12E-10	0.2
rs998584	6	43,757,896	VEGFA	37,403	A/C	0.48 / 0.50	2.36E-11	-0.033	0.005	77.9	4.5	3.76E-11	0.0
rs2980879	8	126,481,475	TRIB1	33,320	A/T	0.38 / 0.28	2.36E-09	0.029	0.005	0.0	0.4	3.44E-09	0.0
rs11045172	12	20,470,221	PDE3A	36,583	A/C	0.19 / 0.23	3.97E-11	0.040	0.006	65.3	2.9	6.30E-11	0.0
rs7978610	12	124,468,572	ZNF664	36,945	C/G	0.39 / 0.27	1.66E-13	0.035	0.005	89.8	9.8	5.40E-14	3.4
rs2925979	16	81,534,790	CMIP	38,583	C/T	0.28 / 0.31	5.72E-24	0.047	0.005	78.1	4.6	1.32E-23	0.0
rs12051272	16	82,663,288	CDH13	24,829	G/T	0.01 / 0.11	8.04E-115	-0.331	0.015	96.1	25.4	4.30E-118	18.3
rs731839	19	33,899,065	PEPD	38,402	A/G	0.33 / 0.34	8.69E-18	0.040	0.005	79.6	4.9	1.73E-17	0.0

* Novel locus is ≥ 500 kb from previously reported adiponectin-associated variant achieving genome-wide significance ($P < 5 \times 10^{-8}$; GWAS catalog, March 2022)

Table 3. Adiponectin multi-ancestry meta-analysis prioritized target gene list. Prioritization of 30 likely target/effector genes by using locus-specific gene priority score constructed from eleven gene prioritization strategies along with TSS distance.

Lead Variant	Chr: Position(hg19)	Prioritized Gene	Lead GWAS Variant to gene TSS	GPScore
rs333947	1:110470764	<i>CSF1</i>	17900	1.3351
rs2061155	1:219665008	<i>LYPLAL1</i>	317822	0.2503
rs1515108	2:227123086	<i>IRS1</i>	541389	0.9413
rs1108842	3:52720080	<i>PBRM1</i>	147	3.7674
rs1108842	3:52720080	<i>GNL3</i>	4908	2.5499
rs17366568	3:186570453	<i>ADIPOQ</i>	9990	1.1682
rs17366568	3:186570453	<i>RFC4</i>	45606	0.8853
rs13131633	4:89739479	<i>NAPIL5</i>	120093	0.5508
rs13131633	4:89739479	<i>FAM13A</i>	293070	0.4818
rs6450176	5:53298025	<i>FST</i>	521786	0.3813
rs998584	6:43757896	<i>VEGFA</i>	19975	0.2982
rs596359	6:153457053	<i>RGS17</i>	4669	1.6313
rs2980879	8:126481475	<i>TRIB1</i>	38912	0.3179
rs10787516	10:115813924	<i>ADRB1</i>	10118	0.3602
rs10886863	10:122929493	<i>FGFR2</i>	428479	0.3052
rs10886863	10:122929493	<i>WDR11</i>	318806	0.2006
rs11023332	11:14784110	<i>CYP2R1</i>	129688	0.5779
rs11045172	12:20470221	<i>PDE3A</i>	51958	1.9119
rs2657888	12:56938383	<i>RBMS2</i>	22670	0.5667
rs10778506	12:107143260	<i>RIC8B</i>	25113	0.4618
rs601339	12:123174743	<i>HCAR2</i>	13147	0.9471
rs7978610	12:124468572	<i>CCDC92</i>	11194	3.0608
rs7978610	12:124468572	<i>DNAH10OS</i>	49041	2.0406
rs7978610	12:124468572	<i>ZNF664</i>	12180	2.0127
rs2925979	16:81534790	<i>CMIP</i>	56015	0.8602
rs12051272	16:82663288	<i>CDH13</i>	2880	3.3914
rs222852	17:7140606	<i>CTDNEP1</i>	15204	1.5380
rs222852	17:7140606	<i>ELP5</i>	14129	1.3088
rs731839	19:33899065	<i>PEPD</i>	113635	0.5830
rs731839	19:33899065	<i>CEBPG</i>	34829	0.5403

Figure Legends:

Figure 1. Miami plot of the multi-ancestry (top) and European-ancestry (bottom) genome-wide meta-analysis results for adiponectin. The y-axis represents $-\log_{10}(P)$ value from the fixed-effects analyses. Each blue dot represents a variant tested in the meta-analysis. The transcript closest to the lead variant (triangles) at each locus is listed in rectangle boxes. Index variants from the additional distinct association signals identified in the European-ancestry meta-analyses are denoted by squares and colored based on the linkage disequilibrium with the lead variant at that locus (orange: $0.10 < r^2 < 0.42$; yellow: $r^2 < 0.10$).

Figure 2. Schematic overview of gene prioritization score construction for finding target genes of adiponectin. Flow diagram illustrating various gene prioritization strategies using multiple data sources: expression QTLs, pathways and networks, regulatory elements, and non-functionally informed methods. The gene prioritization score incorporates all this evidence using four terms: PG, SG, CS, and TSSd (distance to the transcription start site). We prioritize locus-specific target genes using the gene prioritization score, with higher values indicating stronger support across approaches.

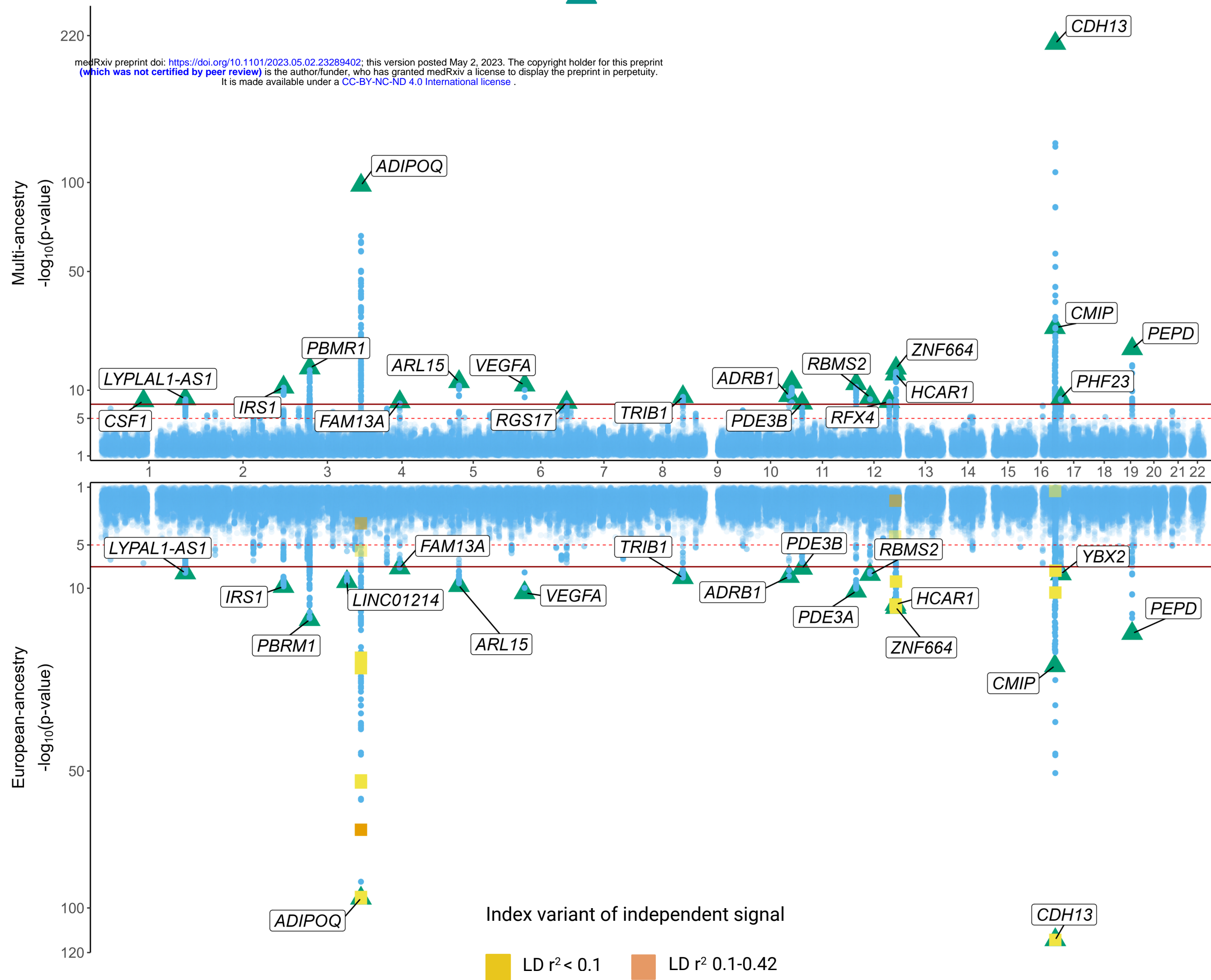
Figure 3. Adiponectin locus-specific prioritized genes and strength. A chromosome ideogram depicts the location and strength of GPScore values, CS (proportion of support) of each prioritized gene identified in an adiponectin multi-ancestry meta-analysis. Red stars highlight the strength of the GPScore measured in percentile, whereas green circles highlight the proportion of support from the 11 prioritization approaches. Some locations have multiple prioritized genes with relatively similar scores.

Figure 4. The relative contribution of terms from the Gene Priority Score and illustrative biological examples. a. Relative importance of four terms: PG, SG, CS, and TSSd for the gene priority score construction across 22 associated regions in multi-ancestry analysis. b-d. Top: Locus zoom plots of multi-ancestry summary statistics, colored by LD with the lead variant. The color of the heatmap represents enrichment for the GPScore term. The top prioritized genes at each exemplary locus are *CSF1* (b), *ADIPOQ* and *RFC4* (c), *GNL3* and *PBRM1* (d).

Figure 5. Disease enrichment and functional interaction network of adiponectin multi-ancestry prioritized genes. a. Select results from the disease enrichment analysis (DisGeNET, enrichr) using the multi-ancestry prioritized genes list. DisGeNET terms related to cardiovascular/cardiometabolic traits show strong enrichment (adjusted $p < 0.01$). b. An interactive functional association network, illustrating the relationships among the connected genes (blue circles), prioritized genes (red triangles), and underlying associated pathways constructed using GeneMANIA. A total of 10 pathways were represented in the network (purple squares). A complete network is shown in Figure S10.

▲ Initial lead variant

medRxiv preprint doi: <https://doi.org/10.1101/2023.05.02.23289402>; this version posted May 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Variants in the 3 Mb window of the lead variant

medRxiv preprint doi: <https://doi.org/10.1101/2023.05.02.23289402>; this version posted May 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Gene Prioritization Strategies

Expression eQTLs

- 13 relevant GTEx tissues
- METSIM Subcutaneous Adipose

Expression
Modifier
Score (EMS)
annotation

Summary
Mendelian
Randomization
(SMR)

eQTL
colocalization
(LD)

eQTL
colocalization
(coloc)

LDSC
enrichment

Polygenic enrichment

- Biological pathways
- PPI networks
- Gene co-regulation
- Expression

Downstreamer

PoPS

Regulatory elements

EpiMap

GeneHancer

Non-functionally informed and target annotation

TSS
distance

MAGMA
gene
analysis

cS2G
score
annotation

Gene Prioritization Score

$$\text{GPScore} = \frac{(-\log_{10}(\mathbf{P}_G) + \mathbf{S}_G) \cdot \mathbf{C}_S}{\log_2(\text{TSS}_d)}$$

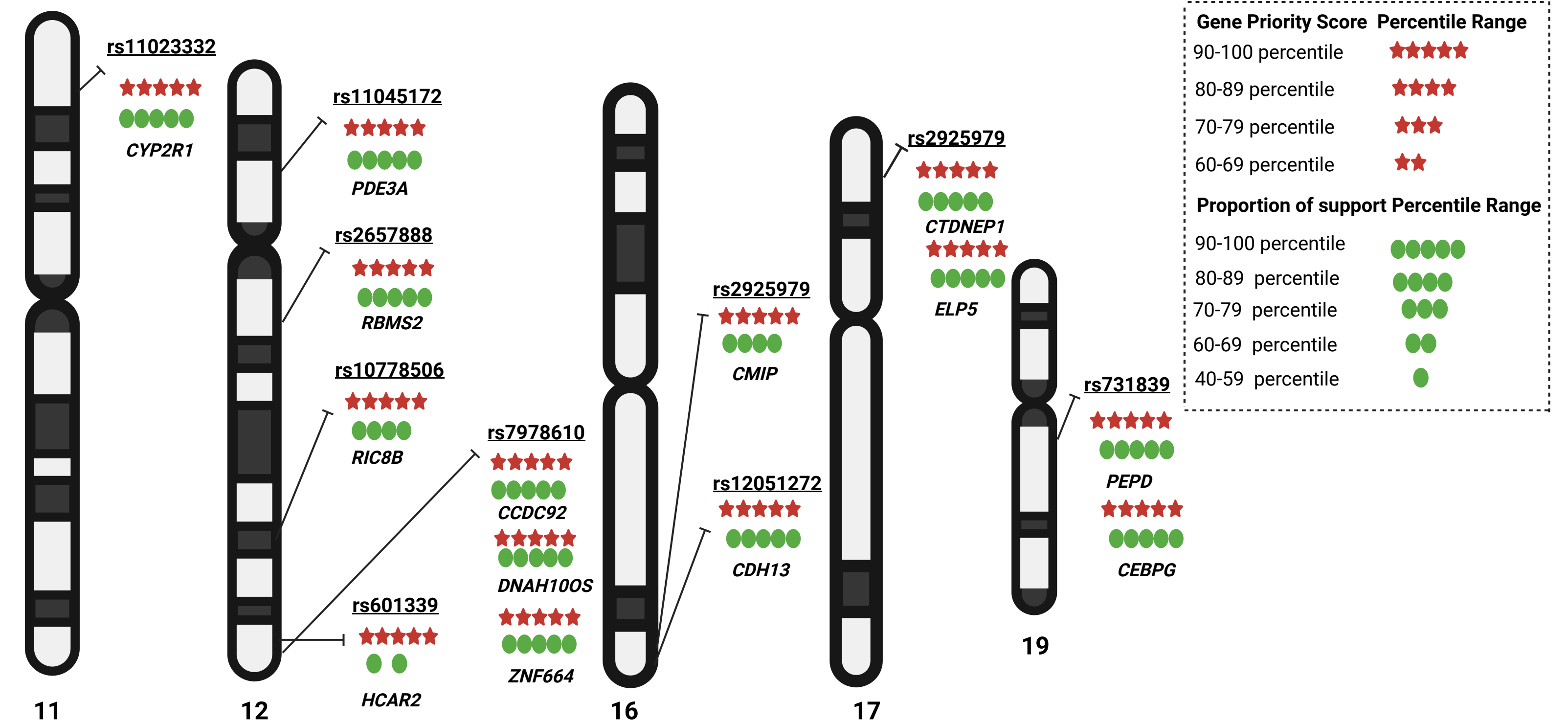
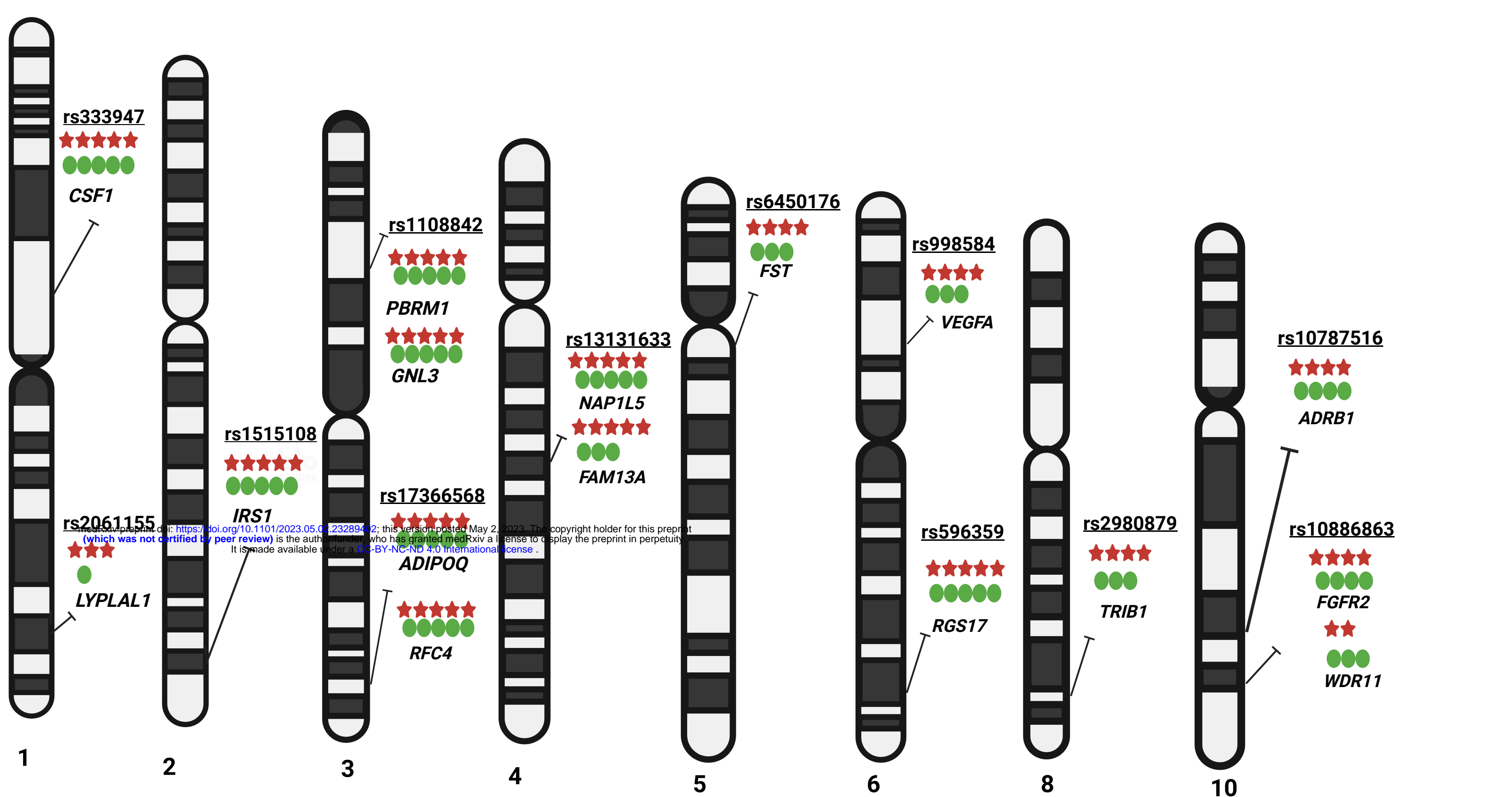
\mathbf{P}_G -----> Combined P-value by the sum of z via Stouffer's method

\mathbf{S}_G -----> Scaled combined scores

TSS_d -----> Distance to the transcription start site of gene-lead variant

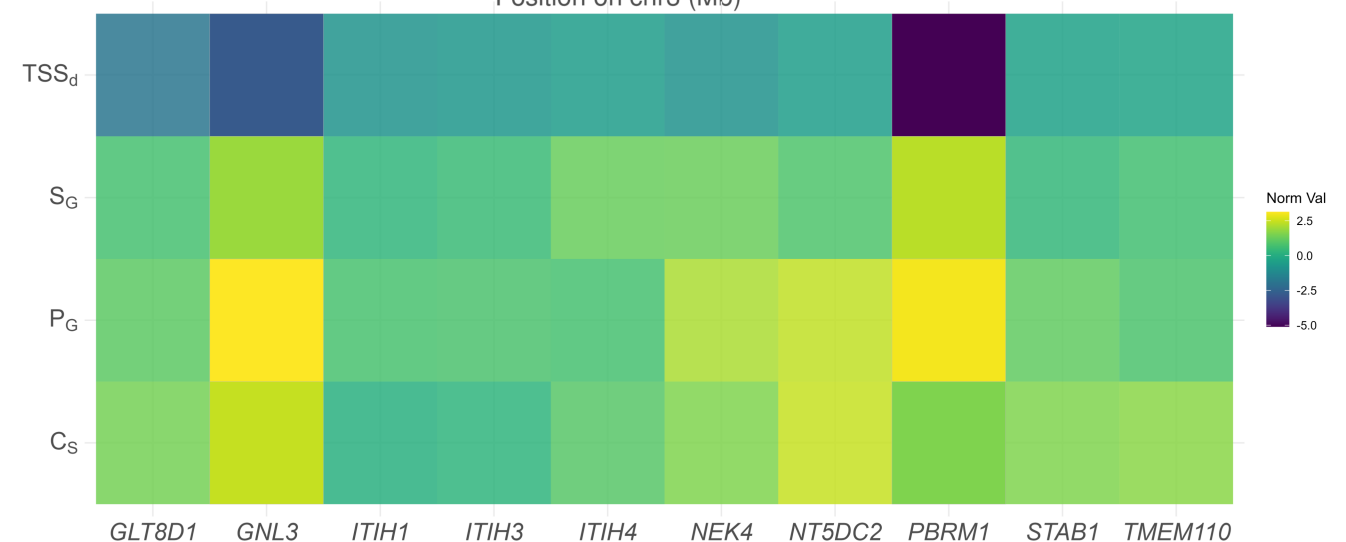
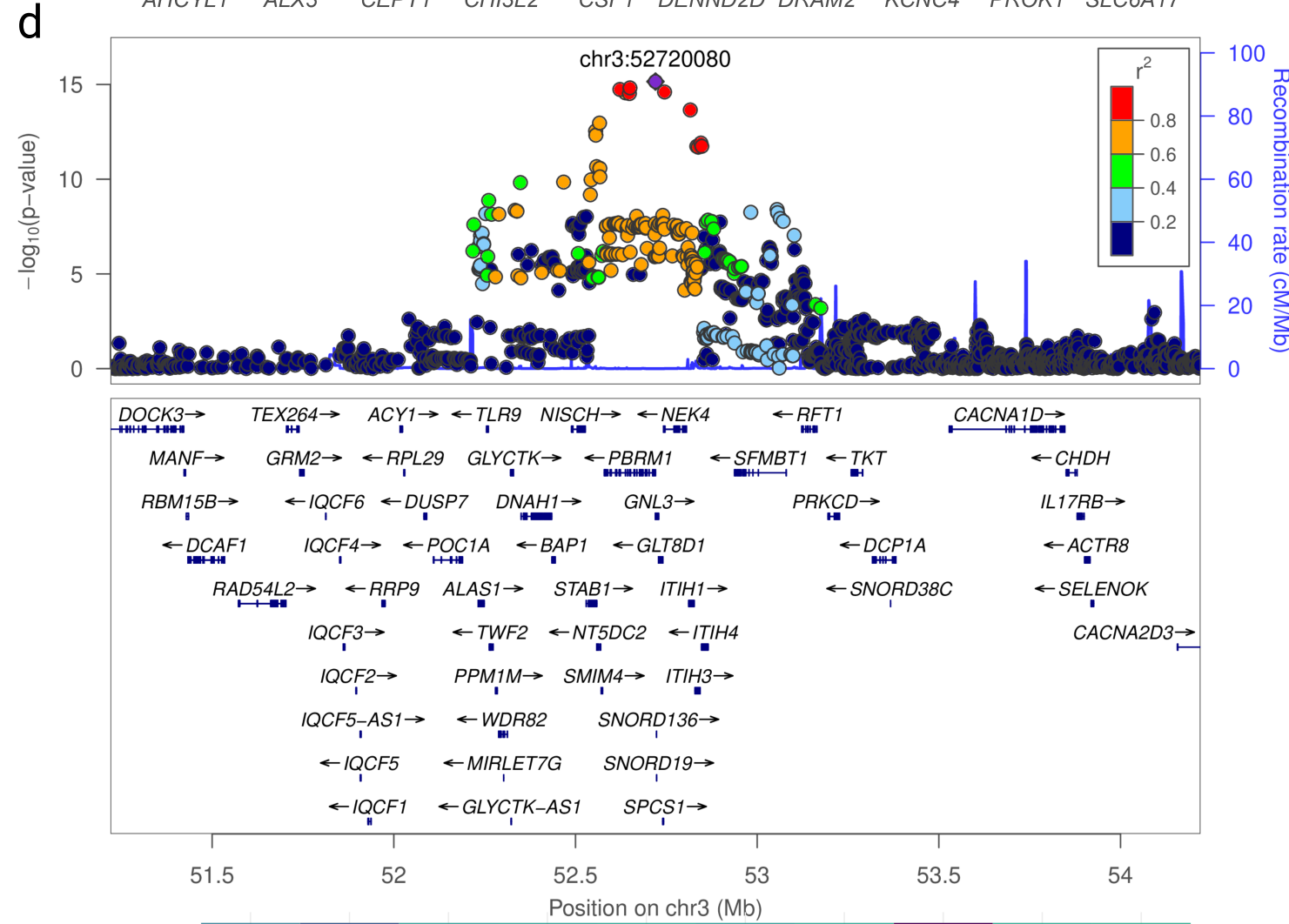
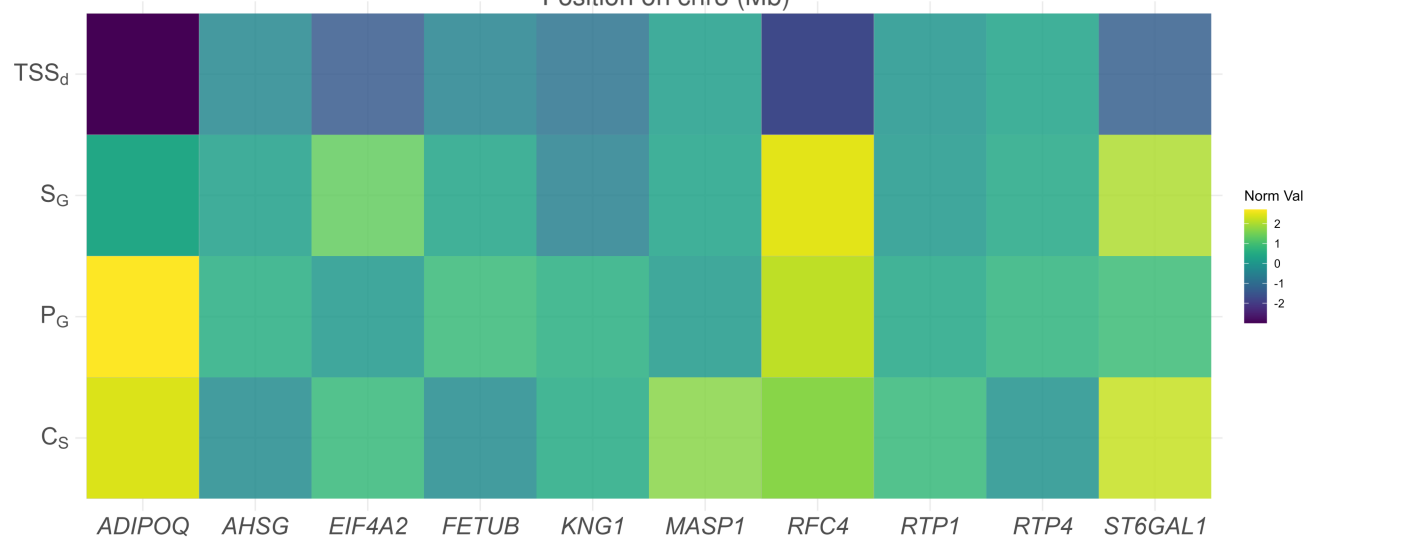
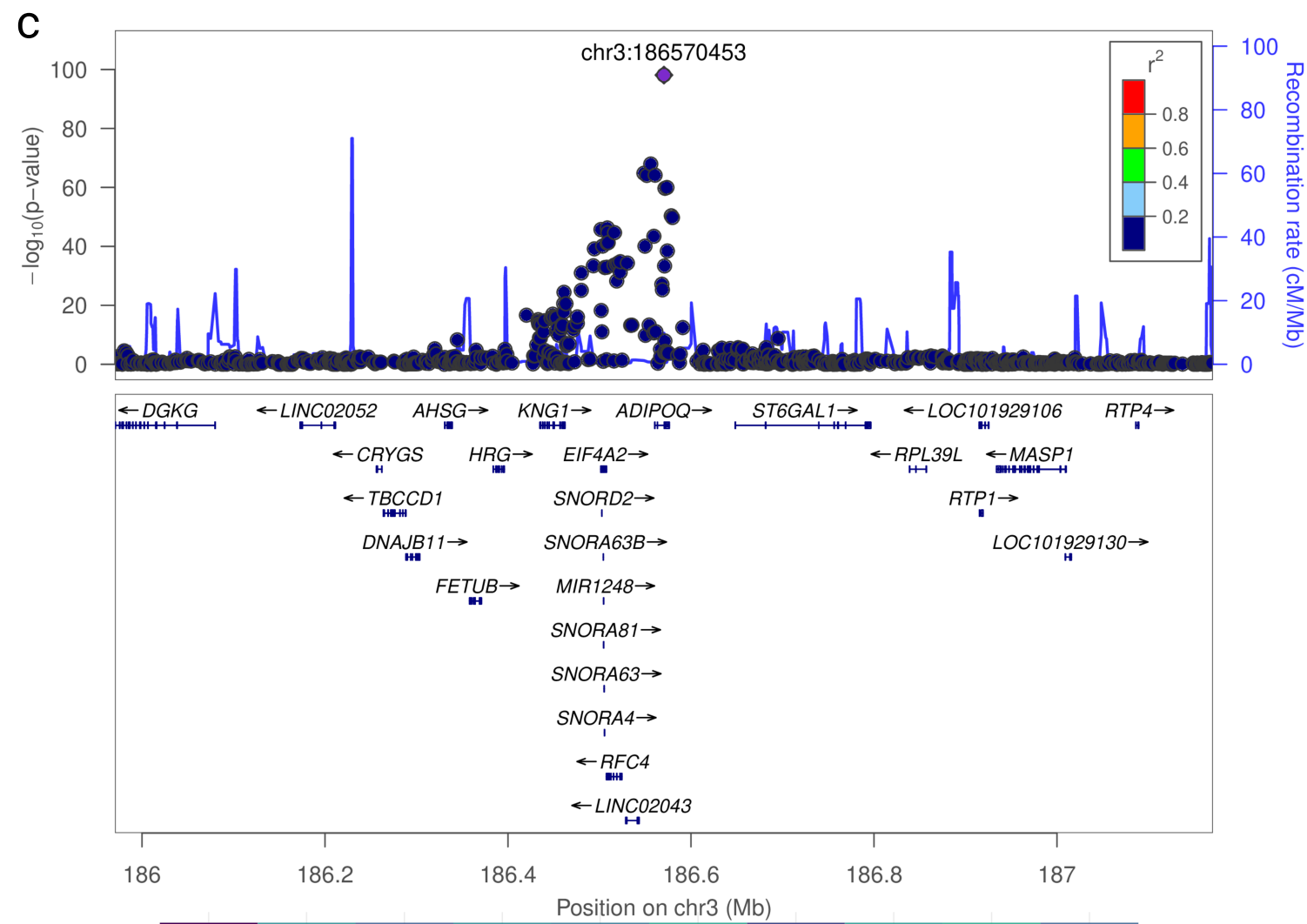
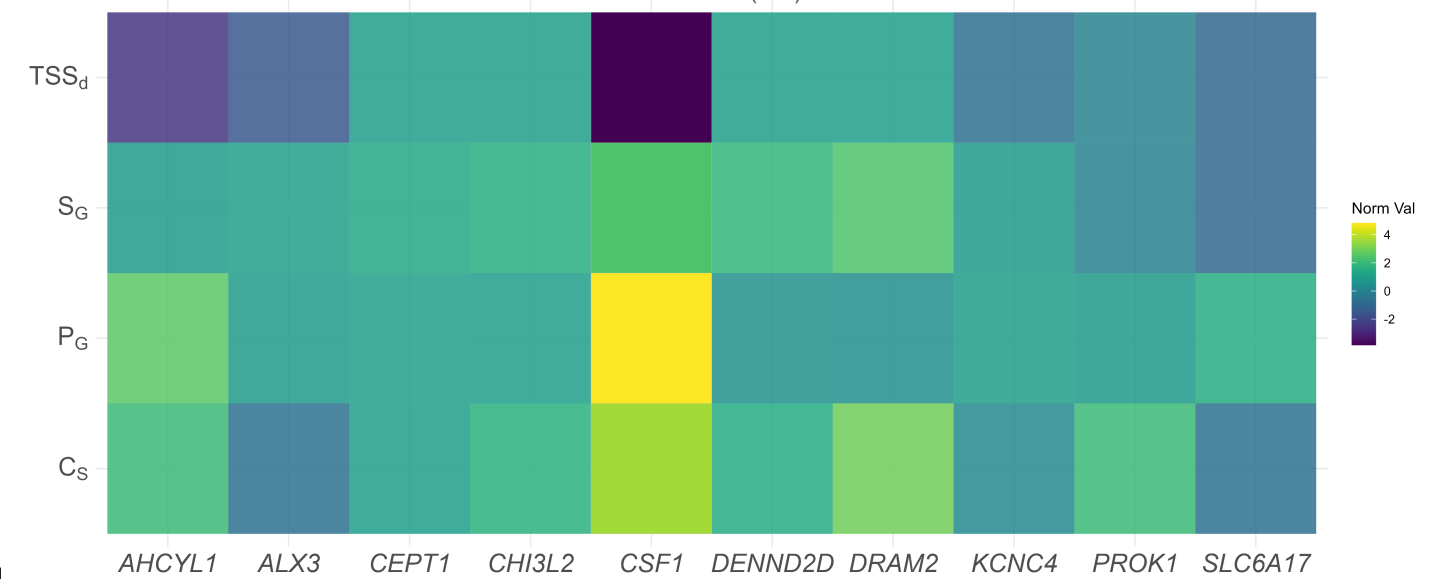
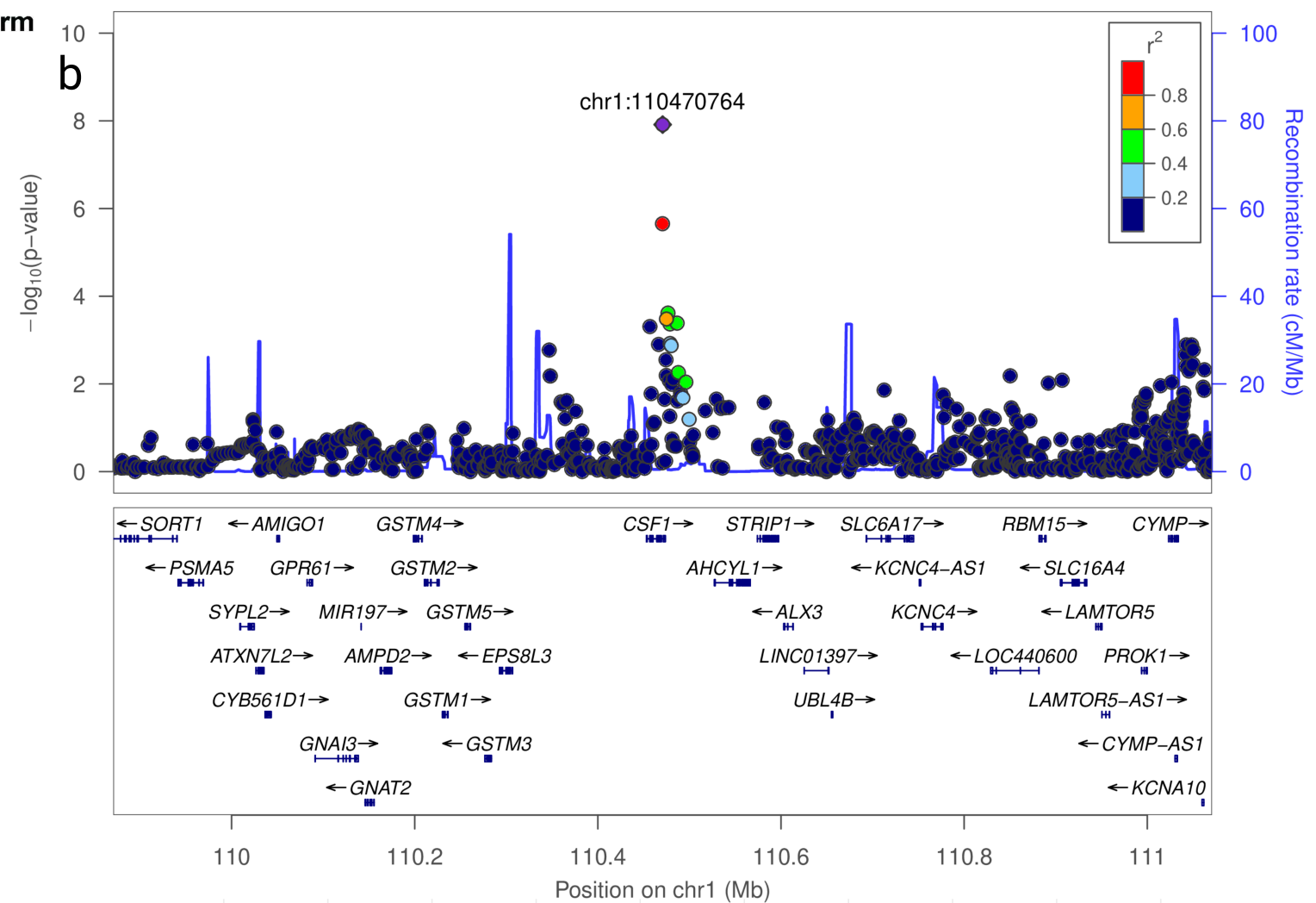
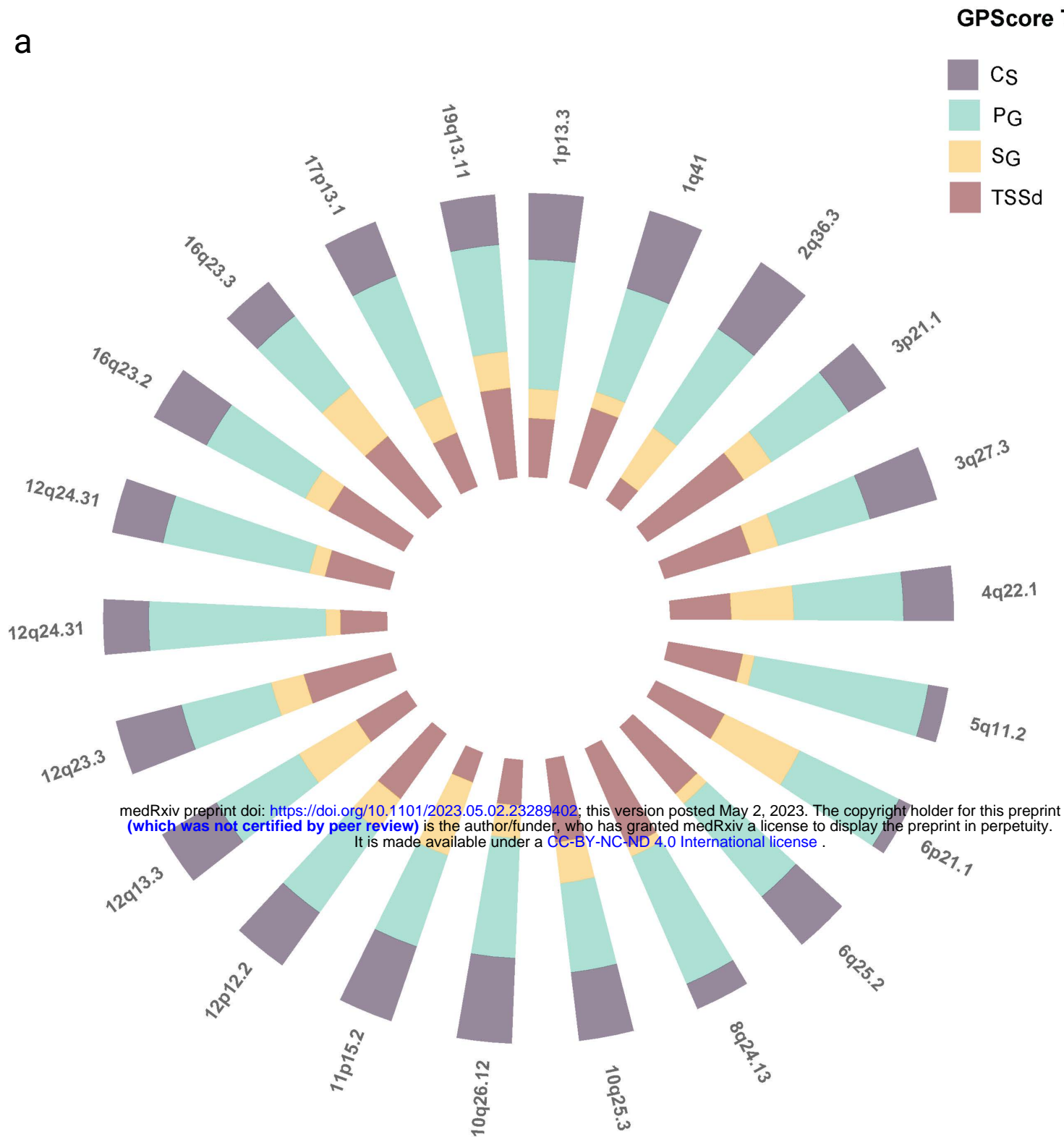
\mathbf{C}_S -----> Proportion of support for all 11 strategies

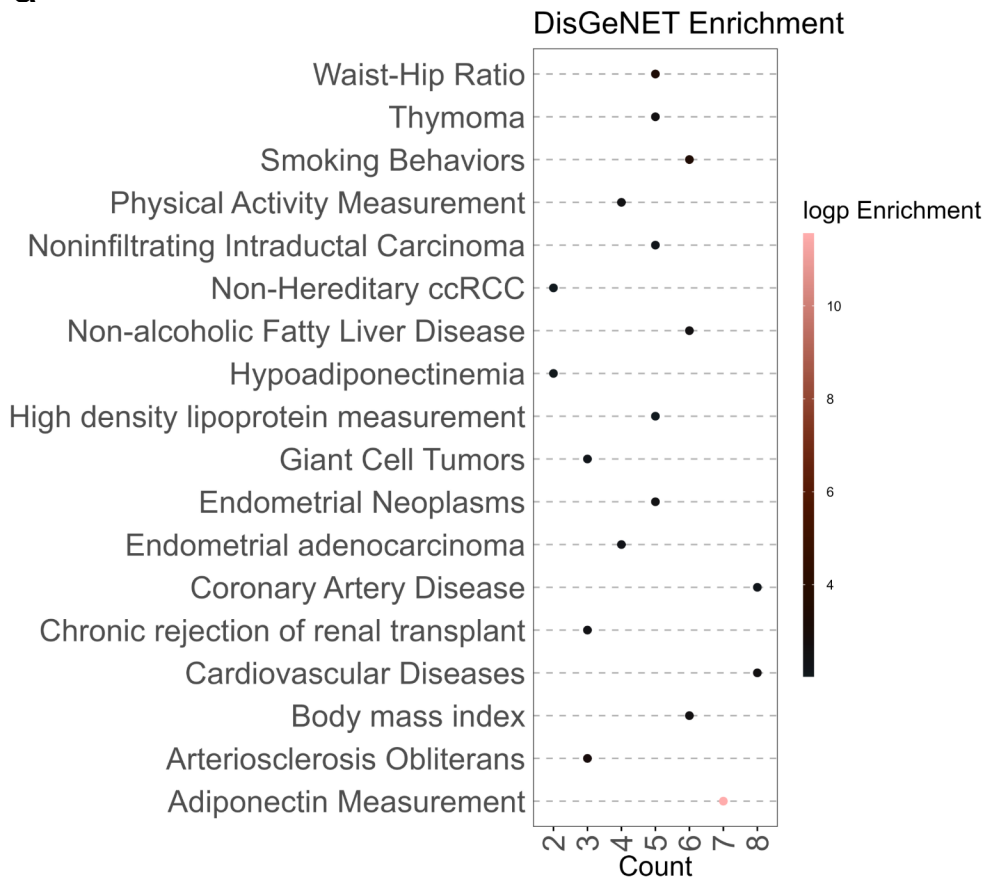
Gene Prioritization List



Gene Priority Score	Percentile Range
★★★★★	90-100 percentile
★★★★	80-89 percentile
★★★	70-79 percentile
★★	60-69 percentile

Proportion of support	Percentile Range
●●●●●	90-100 percentile
●●●●	80-89 percentile
●●●	70-79 percentile
●●	60-69 percentile
●	40-59 percentile



a**b**