

Description of Supplementary Tables

Supplementary Table 1

Description of GWAS summary statistics

Phenotypes explored, i.e., Hypertension (HTN), Heart Failure (HF), Ischaemic Heart Disease (IHD), Mixed Hyperlipidaemia (MHL), Cerebrovascular accident (CVA), Atherosclerosis (AS), Myocardial Infarction (MI), Coronary Heart Disease (CHD), Atrial Fibrillation (AF), and Type 2 Diabetes (T2D).

Supplementary Table 2

SNP heritability and genetic correlation

Estimated using linkage disequilibrium score regression (LDSC) function and LD Scores calculated by *Bulik-Sullivan BK et al.* including only HapMap3 SNPs with minor allele frequency (MAF) > 0.01; and standard errors estimated using a block jackknife over SNPs.

Supplementary Table 3

Exploratory factor analysis (EFA)

Implementing multivariable LDSC, then exploratory factor analysis (EFA) informing subsequent confirmatory factor analysis.

Supplementary Table 4

Confirmatory factor analysis (CFA)

Fitting both common factor and two factor models indicated by EFA, implemented in Genomic SEM.

Supplementary Table 5

Characterization and annotation of lead variants

Characterization and annotation implemented using FUMA GWAS SNP2GENE. Independent significant SNPs identified. Lead SNPs and genomic risk loci defined. Functional consequences on genes (ANNOVAR), Combined Annotation Dependent Depletion score (CADD) score, RegulomeDB (RDB) score, 15 chromatin state (127 tissue/cell types) annotated.

Supplementary Table 6

Lead SNPs annotated in the five input CMD GWASs

Independent significant SNP annotation in five input cardiometabolic disease (CMD) GWASs.

Supplementary Table 7

GWAS Catalog look-up of the mvCMD GWAS lead SNPs

Implemented using FUMA GWAS SNP2GENE.

Supplementary Table 8

GWAS Catalog look-up of the mvCMD GWAS novel lead SNPs

Implemented using FUMA GWAS SNP2GENE.

Supplementary Table 9

ABF, SuSiE and FINEMAP fine-mapping of mvCMD lead SNPs

Fine-mapping implemented to identify most plausible causal variants. SNPs contained within each 95% credible set for mvCMD loci are listed, along with their inclusion probability and functional annotation.

Supplementary Table 10

Transcriptome-wide association study (TWAS)

TWAS implemented to integrate GWAS and gene expression datasets to identify gene-trait associations and prioritize causal genes at GWAS loci.

Supplementary Table 11

Multi-marker analysis of genomic annotation (MAGMA) gene-based results

MAGMA implemented with data from GTEx (version 8) to conduct gene-based and gene-set analyses. SNPs were mapped to 19,060 protein-coding genes located within 10 kb of the lead SNPs, while accounting for LD between SNPs using the 1000 Genomes Project reference panel.

Supplementary Table 12

Gene-set enrichment analysis (GSEA) using MAGMA-derived genes

FUMA GENE2FUNC ("gene to function") gene-set analyses implemented in FUMA GWAS SNP2GENE (<https://fuma.ctglab.nl/>) using genes identified with MAGMA to evaluate potential relationships between mvCMD and lists of mapped genes from MSigDB gene sets, i.e., Reactome, and Gene Ontology (GO).

Supplementary Table 13

Mendelian disease enrichment

To investigate potential relationships of mvCMD with Mendelian disease genes and associated pathways with MendelVar (<https://mendelvar.mrcieu.ac.uk/>), using lead SNPs as input to MendelVar, performing analyses using intervals based on a ± 0.5 Mbp window around lead SNPs (1000 Genomes Phase 3 European reference panel). Within MendelVar, INRICH run using "Gene" enrichment mode and default setting for the target gene set filter and minimum observed threshold, and included gene sets from the Disease Ontology (do) (<https://disease-ontology.org/>) database.

Supplementary Table 14

Mendelian pathway enrichment

To investigate potential relationships of mvCMD with Mendelian disease genes and associated pathways with MendelVar (<https://mendelvar.mrcieu.ac.uk/>), using lead SNPs as input to MendelVar, performing analyses using intervals based on a ± 0.5 Mbp window around lead SNPs (1000 Genomes Phase 3 European reference panel). Within MendelVar, INRICH run using "Gene" enrichment mode and default setting for the target gene set filter and minimum observed threshold, and included gene sets from the Human Phenotype Ontology (hpo) (<https://hpo.jax.org/app/>) database.

Supplementary Table 15

Mendelian randomization analysis investigating the causal role of different variables on mvCMD

Single variable Mendelian randomization (SVMR) implemented using TwoSampleMR package. Effect estimates are standardised i.e. effects correspond to standard deviation (SD) change in outcome per SD increase in exposure.

Supplementary Table 16

Polygenic Mendelian randomization risk factors and biomarkers

Description and sources of risk factors and biomarkers used in Mendelian randomization analysis exploring causal role on mvCMD, including but not limited to lipids, lung function, blood cells and inflammatory markers, liver function, glycemic markers, other markers, kidney function, blood pressure, anthropometric, substance use, sleep, well-being, education and physical activities.

Supplementary Table 17

Metformin target genes

Seven primary metformin targets (AMPK, MCI, GPD1, GPD2, PEN2, FBP1, and GLP1) were identified from literature. ChEMBL database used to identify genes related

to the mechanism of action for the seven metformin target genes.

Supplementary Table 18

Drug-target MR IVs for metformin targets extracted from GWAS data of HbA1c

Variants within 100 kb of gene boundaries (cis-IVs) extracted from the GWAS of circulating HbA1c levels used for the MR (OpenGWAS ID: ukb-d-30750_irnt, N=344,182).

Supplementary Table 19

Drug-target Mendelian randomization of metformin target genes on mvCMD

MR IVW (random-effects analysis performed when there were more than three variants) also MR Egger and Wald Ratio analyses, both accounting for correlation between IVs (correlation matrices generated using 1000 Genomes Phase 3 EUR reference panel).

Supplementary Table 20

Drug-target MR IVs for antidiabetics targets extracted from GWAS data of HbA1c

Supplementary Table 21

Drug-target Mendelian randomization of antidiabetics target genes on mvCMD

Supplementary Table 22

Drug-target MR IVs for lipid-modulating targets from GWAS data of circulating lipid levels

Variants within 100 kb of gene boundaries (cis-IVs) extracted from the GWAS of circulating lipid levels used for the MR (low-density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C), triglycerides (TG), and lipoprotein A Lp(A)) from UK Biobank participants of European ancestry (N range: 361,194 to 441,016).

Supplementary Table 23

Drug-target Mendelian randomization of lipid-modulating target genes on mvCMD

MR IVW (random-effects analysis performed when there were more than three variants) also MR Egger analyses, accounting for correlation between IVs (correlation matrices generated using 1000 Genomes Phase 3 EUR reference panel).

Supplementary Table 24**Protein coding genes within 50kb of HbA1c lead SNPs**

For cis-IVs MR scan of protein coding genes associated with biomarkers identified previously in polygenic and drug-target MR; having extracted lead variants with P-values $< 5.00E-08$ ($LD R^2 < 0.1$, threshold 10000 kb) associated with HbA1c, Bioconductor BiomaRt package (interface to BioMart databases) used to identify and curate protein-coding genes.

Supplementary Table 25**Protein coding genes within 50kb of HDL-C lead SNPs**

For cis-IVs MR scan of protein coding genes associated with biomarkers identified previously in polygenic and drug-target MR; having extracted lead variants with P-values $< 5.00E-08$ ($LD R^2 < 0.1$, threshold 10000 kb) associated with HDL-C, Bioconductor BiomaRt package (interface to BioMart databases) used to identify and curate protein-coding genes.

Supplementary Table 26**Protein coding genes within 50kb of LDL-C lead SNPs**

For cis-IVs MR scan of protein coding genes associated with biomarkers identified previously in polygenic and drug-target MR; having extracted lead variants with P-values $< 5.00E-08$ ($LD R^2 < 0.1$, threshold 10000 kb) associated with LDL-C, Bioconductor BiomaRt package (interface to BioMart databases) used to identify and curate protein-coding genes.

Supplementary Table 27**Protein coding genes within 50kb of TG lead SNPs**

For cis-IVs MR scan of protein coding genes associated with biomarkers identified previously in polygenic and drug-target MR; having extracted lead variants with P-values $< 5.00E-08$ ($LD R^2 < 0.1$, threshold 10000 kb) associated with TG, Bioconductor BiomaRt package (interface to BioMart databases) used to identify and curate protein-coding genes.

Supplementary Table 28**Colocalization results for genes near HbA1c lead SNPs on mvCMD**

Genes with $PP.H4.abf > 0.8$ considered suggestive evidence of a shared causal variant between mvCMD and HbA1c within the gene locus.

Supplementary Table 29

Colocalization results for genes near HDL-C lead SNPs on mvCMD

Genes with $PP.H4.abf > 0.8$ considered suggestive evidence of a shared causal variant between mvCMD and HDL-C within the gene locus.

Supplementary Table 30

Colocalization results for genes near LDL-C lead SNPs on mvCMD

Genes with $PP.H4.abf > 0.8$ considered suggestive evidence of a shared causal variant between mvCMD and LDL-C within the gene locus.

Supplementary Table 31

Colocalization results for genes near TG lead SNPs on mvCMD

Genes with $PP.H4.abf > 0.8$ considered suggestive evidence of a shared causal variant between mvCMD and TG within the gene locus.

Supplementary Table 32

Cis-IVs scan of protein coding genes near HbA1c lead SNPs

Taking forward HbA1c protein coding genes with posterior probabilities ($PP.H4 > 0.8$) from (Supplementary Table 28) to cis-IVs for MR analysis.

Supplementary Table 33

Cis-IVs scan of protein coding genes near HDL-C lead SNPs

Taking forward HDL-C protein coding genes with posterior probabilities ($PP.H4 > 0.8$) from (Supplementary Table 29) to cis-IVs for MR analysis.

Supplementary Table 34

Cis-IVs scan of protein coding genes near LDL-C lead SNPs

Taking forward LDL-C protein coding genes with posterior probabilities ($PP.H4 > 0.8$) from (Supplementary Table 30) to cis-IVs for MR analysis.

Supplementary Table 35

Cis-IVs scan of protein coding genes near TG lead SNPs

Taking forward TG protein coding genes with posterior probabilities ($PP.H4 > 0.8$) from (Supplementary Table 31) to cis-IVs for MR analysis.

Supplementary Table 36

Druggable genes list

To evaluate potential therapeutic actionability of top protein-coding genes, downloaded druggable genes defined by *Finan et al.*

Supplementary Table 37

Cis-IVs genes drug-gene interaction information from the Drug-Gene Interaction database (DGIdb)

Searching list of protein coding genes against a compendium of drug-gene interactions and potentially 'druggable' genes to assess potential therapeutic actionability of these protein coding genes.

Supplementary Table 38

A detailed summary of genetic variants in significant genes

Fully statistically significant cis-eQTLs (false discovery rate < 0.05 , ± 1 Mb from each probe) were obtained from the eQTLGen consortium and a meta-analysis of 31,684 individual peripheral blood eQTLs.

Supplementary Table 39

Cis-eQTLs scan of significant genes in biomarker near lead SNPs

Taking forward significant protein coding genes associated with biomarkers from (Supplementary Table 38) to cis-eQTLs for MR analysis.

Supplementary Table 40

Univariate cis-eQTLs scan of significant genes in biomarker near lead SNPs

Given the effect of six significant genes on the cardiometabolic diseases, HF, CVA, IHD, MHL, and HTN GWAS (input for mvCMD GWAS) were analyzed. Additionally, other related diseases (AS, MI, CHD, AF, and T2D) are also involved in MR analysis.

Supplementary Table 41

Cis-IVs for circulating proteins from the SCALLOP

Exploring causal role of circulating proteins from SCALLOP (Systematic and Combined AnaLysis of Olink Proteins) consortia (N = 30,931, European ancestry), which used Olink Proteomics platform to perform protein quantitative trait loci (pQTLs) mapping of plasma proteins. Selected pQTLs associated with plasma protein at genome wide significance within cis-acting loci of target gene boundaries (± 100 kilobases).

Supplementary Table 42**Results of cis-pQTLs scan of circulating proteins from the SCALLOP**

Exploring causal role of circulating proteins from (Supplementary Table 41) to cis-pQTLs for MR analysis.

Supplementary Methods

Overview of Genomic SEM

Genomic SEM is a two-stage structural equation modeling methodology (1-3). In the first stage, the empirical genetic covariance matrix along with its corresponding sampling covariance matrix are estimated. In the second stage, a SEM is specified and parameters are estimated by minimizing the discrepancy between the model-implied genetic covariance matrix and the empirical covariance matrix obtained in the previous stage. Model parameters (θ) are estimated to minimize the discrepancy between the model-implied covariance matrix ($\Sigma(\theta)$) and the empirical covariance matrix (S). A model is considered well-fitting when $\Sigma(\theta)$ closely approximates S . We report results using weighted least squares (WLS) estimation, which weights the discrepancy function by the inverse of the diagonal elements of the sampling covariance matrix and computes standard errors using the full sampling covariance matrix. Genomic-SEM has demonstrated robustness to variations in sample sizes and sample overlap across input GWAS studies, thereby enhancing its applicability and enabling improved statistical power through increased effective sample sizes (4). Furthermore, in contrast to other GWAS meta-analysis methods, we employed Genomic SEM to initially model the joint genetic architecture of cardiometabolic diseases by jointly analyzing GWAS data from five genetically correlated CMD-related phenotypes. Subsequently, we generated a unified GWAS to identify individual SNP associations for this general latent CMD factor. Below, we outline the pertinent principles and detailed information regarding SEM, drawing from the genomic SEM methods paper by *Grotzinger et al* (4).

We applied Genomic SEM to estimate the genetic covariance structure among traits. Genomic SEM partitions the model into a measurement model and a structural model. In the measurement model, genetic components of k phenotypes are expressed as linear functions of m latent variables:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

where \mathbf{y} is a $k \times 1$ vector of observed traits, $\boldsymbol{\eta}$ is an $m \times 1$ vector of latent variables, \mathbf{A} is a $k \times m$ matrix of factor loadings, and $\boldsymbol{\varepsilon}$ is a $k \times 1$ vector of residuals. The model-implied covariance matrix for a confirmatory factor analysis (CFA) is:

$$\Sigma(\theta) = \mathbf{A}\boldsymbol{\Psi}\mathbf{A}' + \boldsymbol{\Theta}$$

where $\boldsymbol{\Psi}$ is an $m \times m$ latent variable covariance matrix and $\boldsymbol{\Theta}$ is a $k \times k$ residual covariance matrix, typically assumed to be diagonal. To model relationships among latent variables, we incorporated a structural model, defined as:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}$$

where \mathbf{B} is an $m \times m$ matrix of regression coefficients and $\boldsymbol{\zeta}$ represents latent variable residuals. The full SEM-implied covariance structure is:

$$\Sigma(\theta) = \mathbf{A}(\mathbf{I} - \mathbf{B})^{-1} - \boldsymbol{\Psi}[(\mathbf{I} - \mathbf{B}')^{-1}]\mathbf{A}' + \boldsymbol{\Theta}$$

where \mathbf{I} is an $k \times k$ identity matrix. This framework allows for estimation of latent genetic factors and their interrelations while accounting for measurement error and

residual covariances.

In the first stage, Genomic SEM uses a multivariable version of cross-trait LDSC (5) to estimate the genetic covariance matrix:

$$S_{LDSC} = \begin{bmatrix} h_1^2 & & & \\ \sigma_{g1,g2} & h_2^2 & & \\ \vdots & & \ddots & \\ \sigma_{g1,gk} & \sigma_{g2,gk} & \cdots & h_k^2 \end{bmatrix}$$

(k is number of observed phenotypes - $k = 5$ in our study). To ensure unbiased standard error (SE) estimates and test statistics, we compute the sampling covariance matrix, $V_{S_{LDSC}}$, which contains all nonredundant elements of the S_{LDSC} matrix. This symmetric matrix has $k^*(k^* + 1)/2$ nonredundant elements. The diagonal elements of $V_{S_{LDSC}}$ represent sampling variances (squared SEs), while the off-diagonal elements capture sampling covariances. These covariances reflect the overlap in sample distributions contributing to the variance and covariance estimates in S_{LDSC} . This $V_{S_{LDSC}}$ matrix can be written as:

$$V_{S_{LDSC}} = \begin{bmatrix} SE(h_1^2)^2 & & & & \\ cov(h_1^2, \sigma_{g1,g2}) & SE(\sigma_{g1,g2})^2 & & & \\ \vdots & \vdots & \ddots & & \\ cov(h_1^2, \sigma_{g1,gk}) & cov(\sigma_{g1,g2}, \sigma_{g1,gk}) & & SE(\sigma_{g1,gk})^2 & \\ \vdots & \vdots & & \vdots & \\ cov(h_1^2, h_j^2) & cov(\sigma_{g1,g2}, h_j^2) & cov(\sigma_{g1,gk}, h_j^2) & SE(h_j^2)^2 & \\ \vdots & \vdots & \vdots & \vdots & \\ cov(h_1^2, \sigma_{gj,gk}) & cov(\sigma_{g1,g2}, \sigma_{gj,gk}) & cov(\sigma_{g1,gk}, \sigma_{gj,gk}) & cov(h_j^2, \sigma_{gj,gk}) & SE(\sigma_{gj,gk})^2 \\ cov(h_1^2, h_k^2) & cov(\sigma_{g1,g2}, h_k^2) & cov(\sigma_{g1,gk}, h_k^2) & cov(h_j^2, h_k^2) & cov(\sigma_{gj,gk}, h_k^2) & SE(h_k^2)^2 \end{bmatrix}$$

Diagonal elements in $V_{S_{LDSC}}$ are estimated using a jackknife resampling procedure from the LDSC package extended in the GenomicSEM package. Next the effects of the individual SNPs are incorporated: first, the initial input genetic covariance matrix is extended to model covariances between individuals SNPs and each observed phenotype by incorporating a vector of SNP-phenotype covariances, denoted S_{SNP} to S_{LDSC} :

$$S_{Full} = \begin{bmatrix} \sigma_{SNP}^2 & & & & \\ \sigma_{SNP,g1} & h_1^2 & & & \\ \sigma_{SNP,g2} & \sigma_{g1,g2} & h_2^2 & & \\ \sigma_{SNP,g3} & \sigma_{g1,g3} & \sigma_{g2,g3} & h_3^2 & \\ \vdots & \vdots & & \ddots & \\ \sigma_{SNP,gk} & \sigma_{g1,gk} & \sigma_{g2,gk} & \sigma_{g3,gk} & \cdots & h_k^2 \end{bmatrix}$$

The sampling covariance matrix, $V_{S_{Full}}$, associated with the expanded S_{Full} covariance matrix, consists of multiple components. One block, $V_{S_{LDSC}}$, includes the sampling variances and covariances of SNP heritabilities and genetic covariances, derived using the multivariable LDSC method. Another block, $V_{S_{SNP}}$, represents the sampling covariance matrix of SNP effects on phenotypes. SNP variance, treated as fixed, has its sampling variance and covariance with other terms set to zero (or a negligible value for computational efficiency). Sampling covariances among SNP-

genotype covariances are estimated using cross-trait LDSC intercepts, rescaled to their respective variances. A final block captures the covariances of SNP-genotype covariances with genetic variances and covariances, fixed to zero due to independence from test statistics in other LD blocks. Overall, the $\mathbf{V}_{\mathbf{S}_{Full}}$ matrix is structured as:

$$\mathbf{V}_{\mathbf{S}_{Full}} = \begin{bmatrix} \mathbf{V}_{\mathbf{S}_{SNP}} & \\ 0 & \mathbf{V}_{\mathbf{S}_{LDSC}} \end{bmatrix}$$

In the second stage, parameters of the user-specified SEM model are estimated with either WLS or maximum likelihood (ML) estimators using the \mathbf{S}_{LDSC} matrix (from the first stage). WLS and ML estimators weigh the matrix information differently, but both estimators minimize the fit error between the model-implied and empirical genetic covariances. WLS optimizes the fit function by leveraging the diagonal elements of the $\mathbf{V}_{\mathbf{S}_{LDSC}}$ matrix and adjusting the standard errors of the estimates using the off-diagonal elements, which account for the correlations among the sampling errors of the summary statistics.

We evaluated model fit using conventional SEM indices, including the model χ^2 statistics, the Akaike information criterion (AIC), comparative fit index (CFI), and the standardized root mean square residual (SRMR). Model χ^2 is an index of exact fit of a SEM. It indexes whether the model-implied genetic covariance matrix, $\Sigma(\theta)$, differs from the empirical genetic covariance matrix, \mathbf{S} . Model χ^2 can also be used as a relative fit index for comparing nested models. CFI indexes the extent to which the proposed model fits better than a model that allows all phenotypes to be heritable, but assumes that they are genetically uncorrelated. Akaike information criterion (AIC) is a relative fit index that balances fit with parsimony, and can be used to compare models regardless of whether they are nested. AIC is calculated as:

$$AIC = \chi^2 + 2 \times fp$$

where fp is the number of free parameters in the model. SRMR is an index of approximate model fit that is calculated as the standardized root mean squared difference between the model-implied and observed correlations in $\Sigma(\theta)$ and \mathbf{S} , respectively. The goodness of fit for both the confirmatory and exploratory models was assessed using standard fit statistics and recommended criteria: a lower Akaike information criterion (AIC), a comparative fit index (CFI) within the range of 0.97 to 1.00 (indicating good fit; 0.95-0.97, acceptable fit), and a standardized root mean square residual (SRMR) less than 0.05 (indicating good fit; 0.05-0.10, acceptable fit) (2,6). CFA confirmed that our common-factor model latent factor was a good fit of the CMD-related GWAS data used as indicator/observed variables: the mvCMD AIC was 35.64, CFI was 0.994 and the SRMR was 0.039 (**Supplementary Table 4**).

The Genomic SEM method is robust to sample overlap, making it applicable to our analyses given the overlap across cohorts included in the five CMD-related GWASs included (**Supplementary Table 1**). We calculated the effective sample sizes for each SNP included in the mvCMD GWAS following the previously established procedures (7). Given the multivariate GWAS effect estimate for SNP J ,

$$\beta_J = \frac{Z_J}{\sqrt{n_J \times 2 \times \text{MAF}_J(1 - \text{MAF}_J)}}$$

where Z_J is the multivariate GWAS association statistic for SNP J , n_J is the effective sample size for SNP J that we aim to calculate, MAF_J is the MAF of SNP J ; the SNP J variance (σ_J^2) is $2 \times \text{MAF}_J(1 - \text{MAF}_J)$, and rearranging,

$$n_J = \frac{(Z_J/\beta_J)^2}{\sigma_J^2}$$

Effective sample size (Neff) is taken to be approximately equal to the mean n_J for m SNPs meeting the MAF thresholds. In our analyses, following recommendations, we restrict the MAF to between 10% and 40%, because the effective calculations are inflated:

$$\text{Neff} \approx \frac{1}{m} \sum_{\text{MAF}=a}^b n_J$$

Q_{SNP} Heterogeneity

A key strength of the Genomic-SEM pipeline lies in its capability to evaluate whether multivariate GWAS SNP associations are best explained by shared causal pathways, which operate on observed GWAS results through a common latent factor, or by trait-specific pathways that are independent of the shared multivariate latent structure. SNP-level heterogeneity test statistics are estimated for each lead mvCMD SNP (an independent SNP associated with mvCMD, with a P-value $< 5 \times 10^{-8}$). The Q_{SNP} statistic follows a χ^2 -squared distribution under the null hypothesis that the SNP effect is entirely mediated by a common pathway. A statistically significant Q_{SNP} indicates that the SNP effect is most likely mediated through specific pathways independent of the shared mvCMD factor. Based on previous multivariate GWASs conducted using Genomic SEM, we assessed Q_{SNP} heterogeneity employing a Bonferroni-adjusted P-value threshold of 2.56×10^{-4} , correcting for 195 lead SNPs (7,8). Among the 195 lead SNPs, 29 exhibited heterogeneity, indicating that the majority of the mvCMD lead SNP associations can be best explained by a common causal pathway. Among the mvCMD lead variants with statistically significant Q_{SNP} statistics (P-value $< 2.56 \times 10^{-4}$), the ncRNA_intronic variant rs1556516, located near the gene CDKN2B-AS1, exhibited the strongest evidence for its association being explained by specific and independent pathways (P-value = 4.12×10^{-96}) (**Supplementary Table 5, 6**). Notably, the four variants - rs118039278, rs2119690, rs6657811 and rs964184 - which are located near the LPA, LPL, CELSR2, and ZNF259 genes, showed statistically significant Q_{SNP} values (**Supplementary Table 5, 6**). SNPs showing strong statistical association with one or more lipid traits at the CELSR2, LPL and ZNF259 loci were also associated with coronary artery disease (CAD) risk (9). Furthermore, research has demonstrated that lipoprotein(a) is a genetically determined, causative, and pervasive risk factor for atherosclerotic cardiovascular disease (10). The human plasma lipidome has been recognized as a critical indicator for assessing the risk of cardiometabolic diseases. Multivariate GWAS studies have identified novel genetic loci with roles in lipid metabolism, indicated functional effects on detailed circulating lipid measures,

and shown associations with cardiometabolic and related diseases (11). Given its close relationship with lipid metabolism, the aforementioned variant is undoubtedly associated with cardiometabolic disease. However, according to the Q_{SNP} analysis, these variants are not involved in our more general CMD-related multivariate mvCMD genetic signature, possibly because it is primarily associated with lipid metabolism or affects the phenotype through other pathways.

Fine-mapping

We performed fine-mapping to identify the most plausible causal variants in the genomic loci associated with mvCMD. We utilized SuSIE (12), ABF (13), and FINEMAP (14) through the R package echolocator (15) version 2.0.3 to identify the most plausible causal variants associated with mvCMD. Fine-mapping methods represent a robust approach for identifying causal variants underlying specific phenotypes, yet their application remains limited due to the technical complexities involved in implementation. echolocator is an R package designed to automate the entire workflow of genomics fine-mapping, annotation, and visualization, thereby facilitating the identification of the most probable causal variants associated with a given phenotype.

SuSIE (Sum of Single Effects) extends Bayesian Variable Selection in Regression (BVSR) for fine-mapping by implementing an Iterative Bayesian Stepwise Selection (IBSS) process, as opposed to traditional stepwise selection methods (12). This approach generates credible sets of variants that quantify the uncertainty associated with selecting a specific variant from a group of highly correlated variants (12). The Bayesian false discovery probability (BFDP) serves as a robust metric for evaluating the significance of observed associations. Similar to the false positive report probability (FPRP), BFDP is straightforward to compute, yet it offers a naturally defined significance threshold grounded in the relative costs of false discoveries and non-discoveries (13). This approach also benefits from a well-established methodological foundation. Moreover, in the context of multiple testing, the expected number of false discoveries and false non-discoveries can be directly estimated. FINEMAP employs a Shotgun Stochastic Search (SSS) algorithm, which is grounded in the Markov Chain Monte Carlo (MCMC) methods commonly utilized in Bayesian inference applications (14). We used the same 1000 Genomes Phase 3 reference panel that was used to generate the mvCMD summary data in Genomic SEM. SuSIE, ABF, and FINEMAP methods each estimate a posterior probability that the variant is causal for mvCMD. We used a 250 kb window and a stringent probability threshold of 0.95 to define credible sets of potentially causal variants. echolocator defines a "consensus SNP" as a variant that is included in SuSIE, ABF, and FINEMAP (15). It calculates the average posterior probability ("mean.PP") and determines an average credibility set ("mean.CS"). Specifically, the mean.CS is 1 if the mean.PP from SuSIE, ABF, and FINEMAP exceeds 0.95; otherwise, it is 0 (**Supplementary Table 9**).

Transcriptome-Wide Association Study (TWAS)

We subsequently conducted a transcriptome-wide association study (TWAS) to further explore gene-level associations of the mvCMD genetic signature, employing the TWAS FUSION method (16). For our TWAS analysis, we obtained 37,920 pre-computed expression quantitative trait loci (eQTL) features from GTEx v8 via the FUSION website (http://gusevlab.org/projects/fusion/weights/sCCA_weights_v8_2.zip). These features were derived using cross-tissue sparse canonical correlation (sCCA) models, which integrate genetic data across multiple tissues to enhance the statistical power of TWAS. Furthermore, we utilized the 1000 Genomes Project Phase 3 European subpopulation for LD estimation. We excluded genes located within the major histocompatibility complex (MHC) region on chromosome 6, a region renowned for its intricate LD structure. After munging the mvCMD summary statistics, 37,190 out of 37,920 sCCA features were available in the mvCMD data for analysis (**Supplementary Table 10**). The main process of TWAS FUSION is introduced as follows (16). First, identify cis-instrument gene expression signatures, that is, genetic variants associated with gene expression within or in close proximity to a genomic locus. Second, a linear predictor is constructed for each gene using its associated SNP genetic features. Third, TWAS test statistics and summary-level GWAS Z-scores were calculated by incorporating the SNP-based prediction weights. TWAS FUSION employs a variety of linear regression models and Bayesian sparse linear mixed models with regularization techniques (such as top1, LASSO regression, and Elastic-net regression) and calculates out-of-sample R^2 statistics to determine the best model through cross-validation of each gene-GWAS model. Furthermore, we employed a Bonferroni-corrected statistical threshold, which was calculated based on the number of genes tested: $P\text{-value} < 1.34 \times 10^{-6}$ ($0.05/37,190$ sCCA signatures evaluated in mvCMD), as the evaluation criterion for interpreting the TWAS results.

Colocalization analysis employs a Bayesian framework to quantify the strength of SNP associations between mvCMD and TWAS gene expression across various loci, driven by shared causal SNPs (17). This approach facilitates the determination of whether observed associations are attributable to horizontal pleiotropy, wherein a single SNP influences both gene expression and mvCMD (PP.H4), or LD, where independent SNPs in LD impact both gene expression and the GWAS signal (PP.H3). We assessed colocalization evidence between mvCMD-associated genes ($P\text{-value} < 1.34 \times 10^{-6}$) by incorporating all variants within a 500 kb region surrounding the lead variant in each gene (17). We classified the results based on a PP.H4 threshold of > 0.8 , indicating a high probability of a shared causal variant between the eQTL and GWAS signals at this specific genomic locus (17) (**Supplementary Table 10**).

Polygenic Mendelian Randomization (MR)

Mendelian Randomization Assumptions

Mendelian randomization (MR) leverages SNPs as instrumental variables to investigate the causal relationship between genetic predisposition to an exposure trait

and its corresponding outcomes (18). The primary assumptions underpinning MR are as follows: firstly, SNP instruments are associated with the exposure trait of interest ("relevance assumption"). Secondly, SNP instruments influence the outcome trait exclusively through the exposure trait ("exclusion restriction assumption"). Thirdly, the distribution of SNP instruments remain uninfluenced by confounding factors that also affect the outcome trait ("independence assumption") (19). In the subsequent exposition, we delineate the MR methodologies employed in our study and their implementation to investigate the associations of risk factors and biomarkers with mvCMD.

Polygenic MR Phenotype Inclusion

To examine the potential causal influence of diseases, lifestyle factors, and biomarkers on mvCMD, we conducted a MR analysis incorporating 67 different variables derived from GWAS involving participants of European ancestry (**Supplementary Table 15, 16**). These risk factors and biomarkers were pre-selected based on their established associations with CMD (20). It is crucial to identify modifiable CMD risk factors that may inform the development of targeted interventions and prevention strategies. We present below the rationale for selecting these specific risk factors and biomarkers.

1) Lifestyle

Given that the incidence of CMD is predominantly observed in the elderly population, their lifestyle choices and exposure to risk factors significantly influence cardiovascular health outcomes in this demographic (21). Furthermore, given that our subsequent analyses aimed to identify potential common genetic loci and pharmacological intervention strategies for CMD, we concentrated our MR analyses of lifestyle risk factors on common exposures. Therefore, our study incorporated a comprehensive set of exposures including smoking habits, alcohol intake, sleep patterns, educational background, and physical activities. Identifying causal biomarkers associated with cardiovascular health status may provide valuable insights for evaluating the risk of cardiovascular and cerebrovascular diseases during human aging and pinpoint potential targets for pharmacological interventions (22).

2) Lipids and Glycemic Markers

Disturbances in glucose and lipid metabolism have been extensively investigated as a critical mechanism underlying the pathogenesis of CMDs. However, there remains a paucity of research regarding the therapeutic efficacy of drugs targeting genetic loci involved in regulating glucose and lipid metabolism across various CMDs. Longitudinal studies to identify pathogenic biomarkers for CMD are also affected by cohort selection bias and disease heterogeneity (22). Given that MR serves as a crucial analytical methodology to enhance causal inference through the utilization of exposure and outcome data, we conducted polygenic MR analysis of multiple biomarkers employing quality-controlled genetic instrumental variables to evaluate their potential causal role on mvCMD.

Lipids have emerged as pivotal regulators of cardiovascular health. Beyond their influence on energy metabolism, lipid storage, and peroxidation in CMD, recent studies have highlighted that the regulation of lipid metabolism is intricately linked to various metabolic disorders, including non-alcoholic fatty liver disease (23) and diabetes (24). Therefore, we incorporated analyses assessing the effects of several major lipid subfractions, including Apolipoprotein A, Apolipoprotein A-I, Apolipoprotein B, Lipoprotein A, Cystatin C, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), and total cholesterol on mvCMD.

Metabolic dysfunction, often characterized by poorly controlled glucose regulation, is frequently observed in CMDs. Additionally, many therapies aimed at treating CMDs have been shown to improve glucose homeostasis (25). However, the causal relationship between metabolic dysfunction and cardiovascular function has yet to be fully elucidated. Furthermore, considering the pivotal role of metabolic regulation in CMD and type 2 diabetes, modulating glucose homeostasis may represent a viable strategy for mitigating disease risk (26). Consequently, we incorporated glycemic markers, including glucose, fasting insulin, and glycosylated hemoglobin (HbA1c) levels, into our biomarker MR analysis. Lipid and glucose markers, as central hubs of metabolic regulation, serve as common downstream biomarkers reflecting physiological responses to pharmacological modulation by frequently prescribed medications. We investigated the causal effects of drugs that modulate glucose and lipid metabolism on mvCMD and evaluated the potential relationship between gene expression and mvCMD to provide a evidence for selecting novel drug targets.

3) Blood Cells and Inflammatory Markers

Emerging evidence indicates that the crosstalk between immune cells and cardiac parenchymal cells, including cardiomyocytes and fibroblasts, is governed by intricate cellular metabolic pathways (27). Consequently, this study aims to investigate the causal roles of specific immune cell populations, such as basophil, eosinophil, lymphocyte, monocyte and neutrophil, as well as inflammatory markers like C-reactive protein (CRP) in relation to mvCMD.

4) Liver Function

Impaired hepatic function, characterized by alterations in liver function enzyme (LFE) levels as observed in both population-based and longitudinal studies, has been previously associated with an increased risk of diabetes and cardiovascular disease (28). Consequently, we incorporated several LFEs, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), Gamma glutamyltransferase (GGT), and alkaline phosphatase (ALP), into our biomarker MR analysis.

5) Other Markers

We also extended our biomarker analyses to include several other important biomarkers, such as lung function, anthropometric, kidney function, calcium, IGF-1, vitamin D, etc., to identify potential causal markers for mvCMD.

Drug-target MR

We acquired summary-level data for glycated hemoglobin (HbA1c) and circulating lipid levels, including low-density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C), triglycerides (TG), and lipoprotein A (Lp(a)), from UK Biobank participants of European ancestry (N range: 273,896 to 441,016) (29) (<http://www.nealelab.is/uk-biobank/>). These data were utilized to construct genetic instruments serving as proxies for pharmacological modulation of the respective biomarkers.

Antidiabetics Gene Instrument Selection

We conducted an extensive review of antidiabetic medication classes and identified six distinct classes of antidiabetic agents, beyond metformin, that have established drug targets and may be suitable for genetic instrumentation (30). These classes include dipeptidyl peptidase-4 (DPP-4) inhibitors, insulin analogs, glucagon-like peptide-1 receptor (GLP-1R) analogs, thiazolidinediones, sodium-glucose cotransporter-2 (SGLT2) inhibitors, and sulfonylureas. Similarly, we utilized the ChEMBL databases to identify the pharmacologically active targets of DPP-IV inhibitors (DPP4), insulin analogs (INSR), GLP-1R analogs (GLP1R), thiazolidinediones (PPARG), SGLT2 inhibitors (SGLT2), and sulfonylureas (ABCC8 and KCNJ11) (31,32). For each target gene, we evaluated the presence of suitable genetic variants within a 100 kb region flanking the gene boundaries using the HbA1c GWAS dataset. For genetic instruments proxying GLP-1R analogs, sulfonylureas, SGLT2 inhibitors, and thiazolidinediones, we selected variants at conventional genome-wide statistical significance ($P\text{-value} < 5 \times 10^{-8}$). For DPP-IV inhibitors and INSR analogs, given the absence of variants meeting the conventional threshold in either the INSR or DPP4 loci, we adopted a relaxed threshold ($P\text{-value} < 5 \times 10^{-4}$) (Supplementary Table 20).

Lipid-modulating Gene Instrument Selection

We further searched for genetic targets for lipid-modulating therapies (33). Specific genes were selected as follows: PCSK9, HMGCR, NPC1LC, ACLY, and ABCG8 SNP effect estimates were extracted from LDL-C (ieu-b-110, N=440,546); CETP and APOA1 SNP effect estimates were extracted from HDL-C (ieu-b-109, N=403,943); ANGPTL3, ANGPTL4, APOC3, PPARG, and LPL SNP effect estimates were extracted from TG (ieu-b-111, N=441,016); and LPA variants were extracted from the Lp(A) GWAS (ukb-d-30790_irnt, N=273,896) (29). We proxy the pharmacological modulation of these drug targets by identifying lipid-associated SNP cis-acting loci within ± 100 kb of gene boundaries, which represent the primary physiological responses to pharmacological interventions such as LDL-C-lowering therapy, TG-lowering therapy, and HDL-C-raising therapy. For 13 of the remaining 12 lipid-modulating drug targets, we identified variants achieving conventional genome wide statistical significance ($P\text{-value} < 5 \times 10^{-8}$). For the instrumental variable analysis of ACLY in LDL-C data, we used a relaxed P-value threshold ($P\text{-value} < 5 \times 10^{-4}$) due to

the absence of variants meeting conventional genome-wide significance within the *ACLY* locus (**Supplementary Table 22**).

Drug-target MR Methods

We clumped drug targets as described for the metformin analyses ($LD R^2 < 0.2$ threshold within a 250 kb window, based on the 1000 Genomes Phase 3 EUR reference panel), and calculated F-statistics to evaluate instrument strength to assess the MR assumption. We used an F-statistic > 10 to minimal bias from weak IVs, which is particularly critical for drug-target IVs comprising variants with relaxed P-value thresholds because no genome-wide significant variants exist at those loci in the corresponding biomarker GWAS data, such as *DPP4* and *ACLY* (34). After harmonization with mvCMD, we performed correlated MR IVW (random-effects analysis performed when there were more than three variants) and MR Egger analyses accounting for the correlation between our instrument variants, with correlation matrices generated using the 1000 Genomes Phase 3 EUR reference panel. We employed MR IVW as the primary method to enhance precision by integrating additional, partially independent IVs into the drug-target MR analysis (35). Furthermore, we conducted heterogeneity testing, and when heterogeneity tests indicated significant heterogeneity in the MR estimates (Cochran's Q P-value < 0.05), we repeated the analyses using a more stringent $LD R^2$ threshold of 0.1 to remove potential heterogenous IVs.

To elucidate the effects of the aforementioned glucose- and lipid-modulating drugs on mvCMD, we adjusted the MR effect estimates to align with the anticipated physiological responses resulting from pharmacological modulation of the drug target. We applied a Bonferroni-corrected threshold with a P-value of 0.00263, accounting for 19 total drug targets examined in both antidiabetic and lipid-modulating therapies, to systematically guide follow-up analyses on a manageable number of significant findings.

Supplementary Checklist: STROBE-MR Reporting Guidelines

1. TITLE and ABSTRACT

Q: Indicate Mendelian randomization as the study's design in the title and/or the abstract if that is a main purpose of the study.

A: Not applicable for title. MR discussed in abstract.

INTRODUCTION

2. Background

Q: Explain the scientific background and rationale for the reported study. Is causality between exposure and outcome plausible? Justify why MR is a helpful method to address the study question.

A: Addressed in the methods and supplementary methods.

3. Objectives

Q: State specific objectives clearly, including pre-specified causal hypotheses (if any).

A: Addressed in the methods.

METHODS

4. Study Design and Data Sources

Q: Present key elements of study design early in the paper. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:

a) Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.

b) Participants: Give the eligibility criteria, and the sources and methods of selection of participants. Report the sample size, and whether any power or sample size calculations were carried out prior to the main analysis.

c) Describe measurement, quality control, and selection of genetic variants.

d) For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases.

e) For each exposure, outcome and other relevant variables, describe methods of assessment.

and, in the case of diseases, the diagnostic criteria used.

f) Provide details of ethics committee approval and participant informed consent, if relevant.

A: Addressed in the methods and supplementary methods.

5. Assumptions

Q: Explicitly state the three core instrumental variable assumptions for the main analysis (relevance, independence, and exclusion restriction), as well assumptions for

any additional or sensitivity analysis.

A: Addressed in the supplementary methods.

6. Statistical Methods: Main Analysis

Q: Describe statistical methods and statistics used.

a) Describe how quantitative variables were handled in the analyses (that is, scale, units, model)

b) Describe the process for identifying genetic variants and weights to be included in the analyses (i.e, independence and model). Consider a flow diagram.

c) Describe the MR estimator, e.g. two-stage least squares, Wald ratio, and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples.

d) Explain how missing data were addressed.

e) If applicable, say how multiple testing was dealt with.

A: Addressed in the methods and supplementary methods.

7. Assessment of Assumptions

Q: Describe any methods used to assess the assumptions or justify their validity.

A: Addressed in the methods and supplementary methods.

8. Sensitivity Analyses

Q: Describe any sensitivity analyses or additional analyses conducted, including comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, and simulations.

A: Addressed in the methods and supplementary methods.

9. Software and Pre-registration

Q:

a) Name statistical software and package(s), including version and settings used.

b) State whether the study protocol and details were pre-registered (as well as when and where).

A: Addressed in the methods and supplementary methods.

RESULTS

10. Descriptive Data

Q:

a) Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow-diagram.

b) Report summary statistics for phenotypic exposure(s), outcome(s) and other relevant variables (e.g. means, standard deviations, proportions).

c) If the data sources include meta-analyses of previous studies, provide the number of studies, their reported ancestry, if available, and assessments of heterogeneity across these studies. Consider using a supplementary table for each data source.

d) For two-sample Mendelian randomization:

- i. Provide information on the similarity of the genetic variant-exposure associations between the exposure and outcome samples.*
- ii. Provide information on extent of sample overlap between the exposure and outcome data sources.*

A: Addressed in the results, supplementary methods, and supplementary tables.

11. Main Results

Q:

- a) Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale.*
- b) Report causal effect estimate between exposure and outcome, and the measures of uncertainty from the MR analysis. Use an intuitive scale, such as odds ratio, or relative risk, per standard deviation difference.*
- c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time-period.*
- d) Consider any plots to visualize results, such as forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure.*

A: Addressed in the results and supplementary tables.

12. Assessment of Assumptions

Q:

- a) Assess the validity of the assumptions.*
- b) Report any additional statistics (e.g., assessments of heterogeneity, such as I^2 , Q statistic, or E value).*

A: Addressed in the results, discussion, and supplementary tables.

13. Sensitivity and Additional Analyses

Q:

- a) Use sensitivity analyses to assess the robustness of the main results to violations of the assumptions.*
- b) Report results from other sensitivity analyses, such as replication study with different dataset, analyses of subgroups, validation of instrument(s), and simulations.*
- c) Report any assessment of direction of causality (e.g., bidirectional MR).*
- d) When relevant, report and compare with estimates from non-MR analyses.*
- e) Consider any additional plots to visualize results (e.g., leave-one-out analyses).*

A: Addressed in the results and supplementary tables.

DISCUSSION

14. Key Results

Q: Summarize key results with reference to study objectives.

A: Addressed in the discussion.

15. Limitations

Q: Discuss limitations of the study, taking into account the validity of the MR assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias, and any efforts to address them.

A: Addressed in the discussion.

16. Interpretation

Q:

a) Meaning: Give a cautious overall interpretation of results considering objectives and limitations.

Compare with results from other relevant studies.

b) Mechanism: Discuss underlying biological mechanisms that could be modelled by using the genetic variants to assess the relationship between the exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable.

c) Clinical relevance: Evaluate the clinical and public policy relevance of the results, and determine the extent to which they inform the effect sizes of potential interventions.

A: Addressed in the discussion.

17. Generalizability

Q: Discuss the generalizability of the study results (a) to other populations (i.e. external validity), (b) across other exposure periods/timings, and (c) across other levels of exposure.

A: Addressed in the discussion.

OTHER INFORMATION

18. Funding

Q: Describe the sources of funding and the roles of funders in the present study. If applicable, also detail the sources of funding for the databases and the original studies that form the basis of the current research.

A: Addressed in the funding.

19. Data and Data Sharing

Q: Provide the datasets utilized for all analyses, clearly reporting their sources and access methods. Reference these data sources appropriately within the article. Additionally, supply the statistical code necessary to replicate the results presented in the article, or specify whether the code is publicly accessible and provide details on its location.

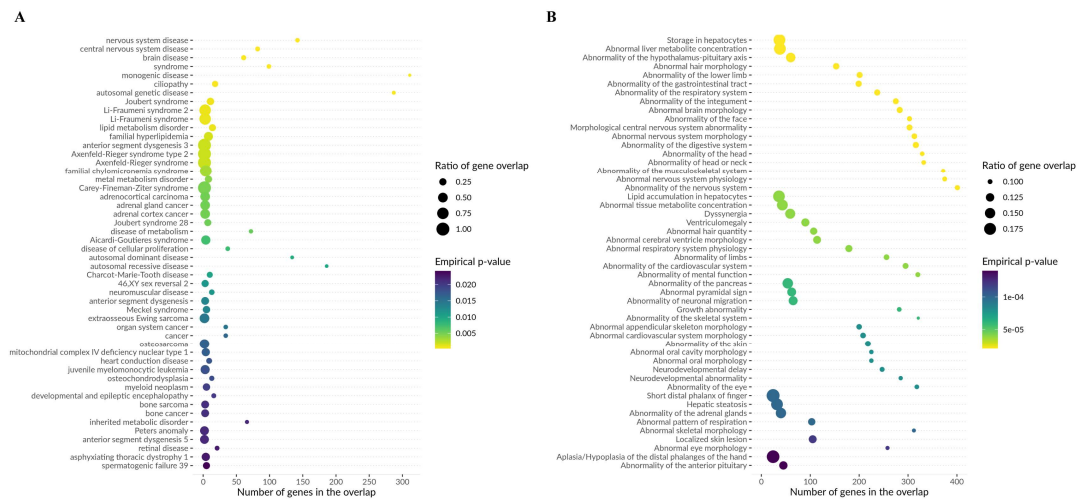
A: Addressed in the methods.

20. Conflicts of Interest

Q: All authors should declare all potential conflicts of interest.

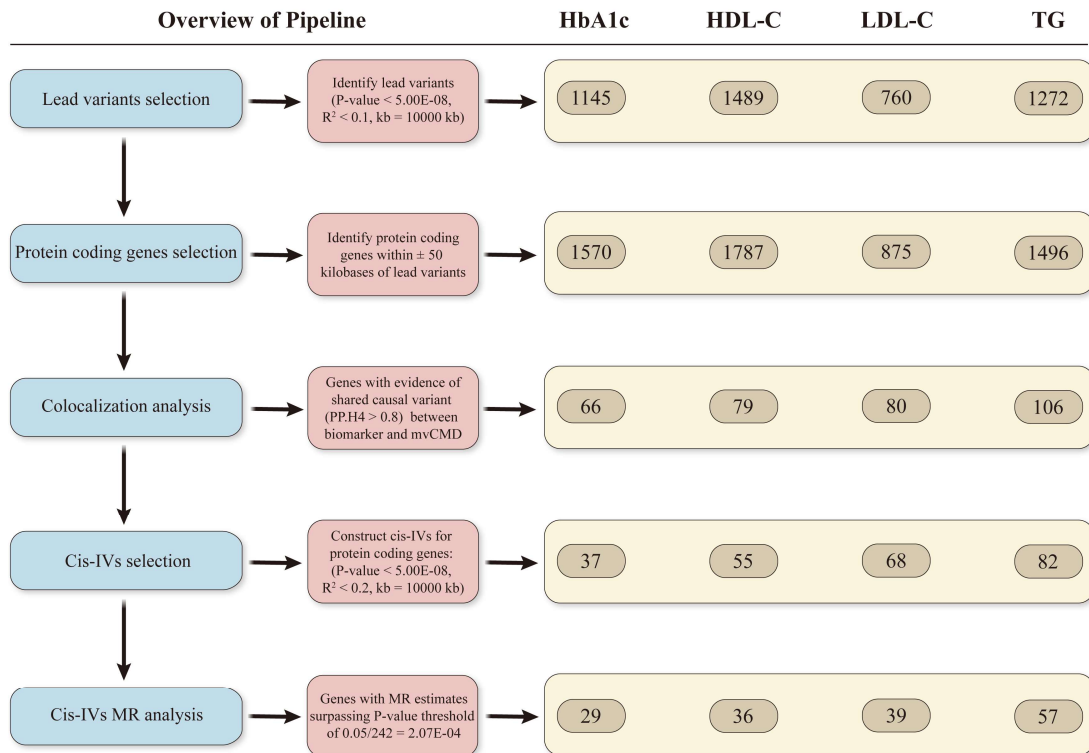
A: Addressed in the declaration of competing interest.

Supplementary Figures



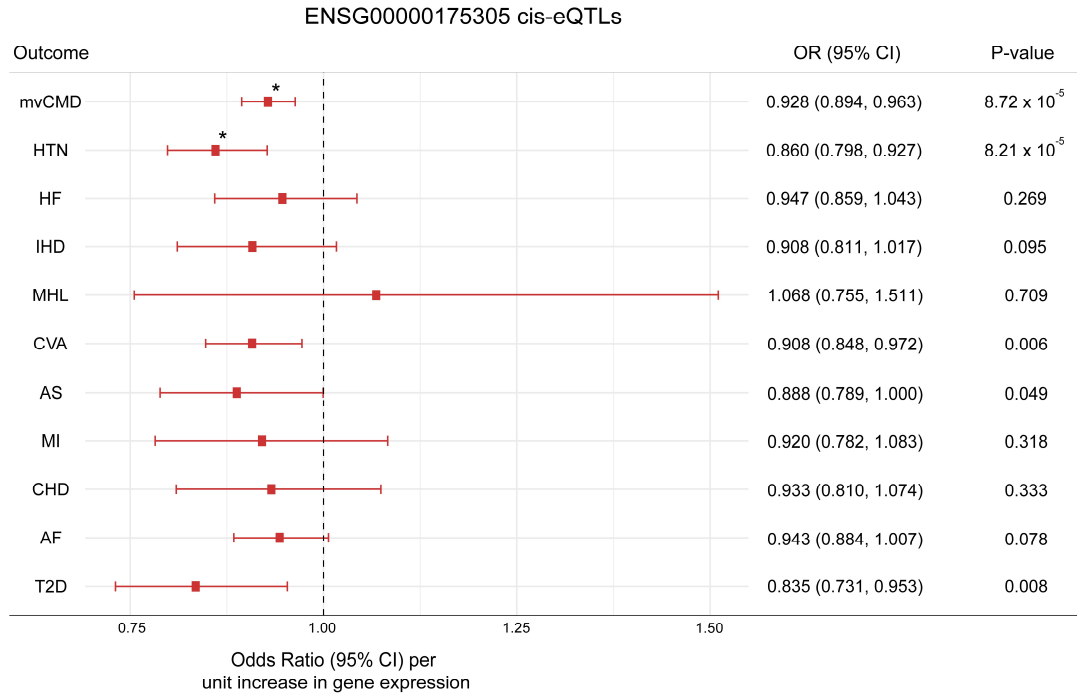
Supplementary Fig. 1 Mendelian Diseases and Phenotypes Enrichment Results

Implemented to investigate potential relationships of mvCMD with Mendelian disease genes and associated pathways with MendelVar (<https://mendelvar.mrcieu.ac.uk/>). Within MendelVar, INRICH was executed in "Gene" enrichment mode with default settings for the target gene set filter and minimum observed threshold. Additionally, gene sets from the Human Disease Ontology (do) (<https://disease-ontology.org/>) and the Human Phenotype Ontology (hpo) (<https://hpo.jax.org/app/>) databases were incorporated. The figure shows the top 50 items in the enrichment analysis results.



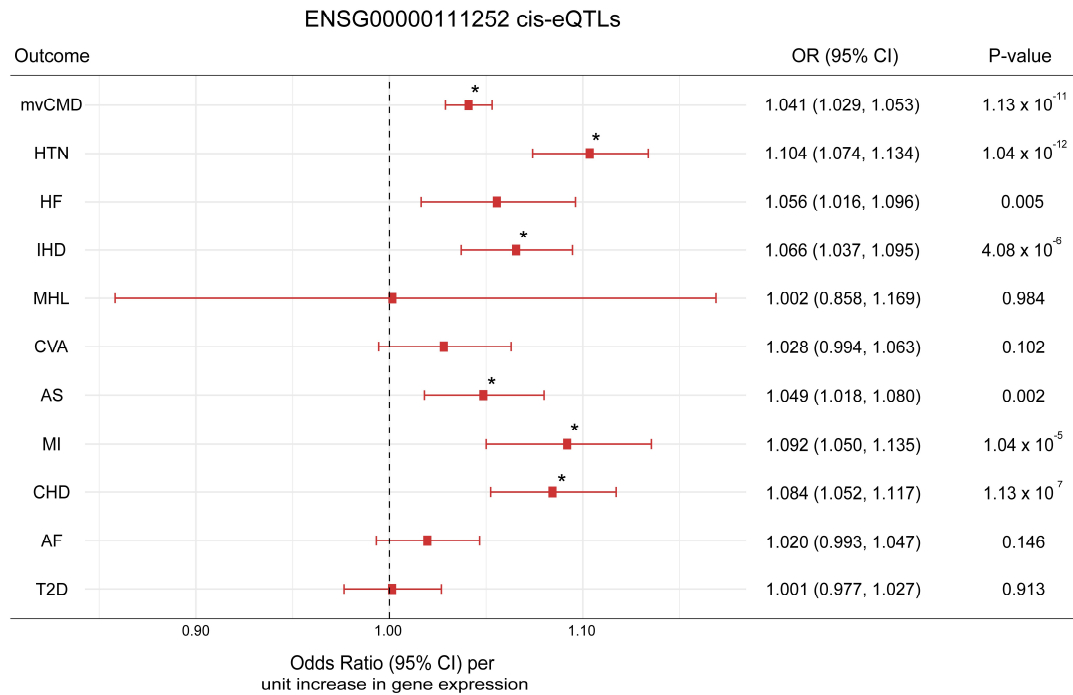
Supplementary Fig. 2 Analysis Overview of Cis-IVs MR of Genes in Biomarker

Results show number of protein-coding genes including in the stages of the colocalization and cis-IVs MR analysis. Exposure GWAS data for each of the biomarkers came from HbA1c (ukb-d-30750_irnt, N=344,182); several lipid subfractions including HDL-C (ieu-b-109, N=403,943), LDL-C (ieu-b-110, N=440,546), and TG (ieu-b-111, N=441,016). Outcome data was mvCMD (N = 932,442). HbA1c, glycated hemoglobin; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglycerides.



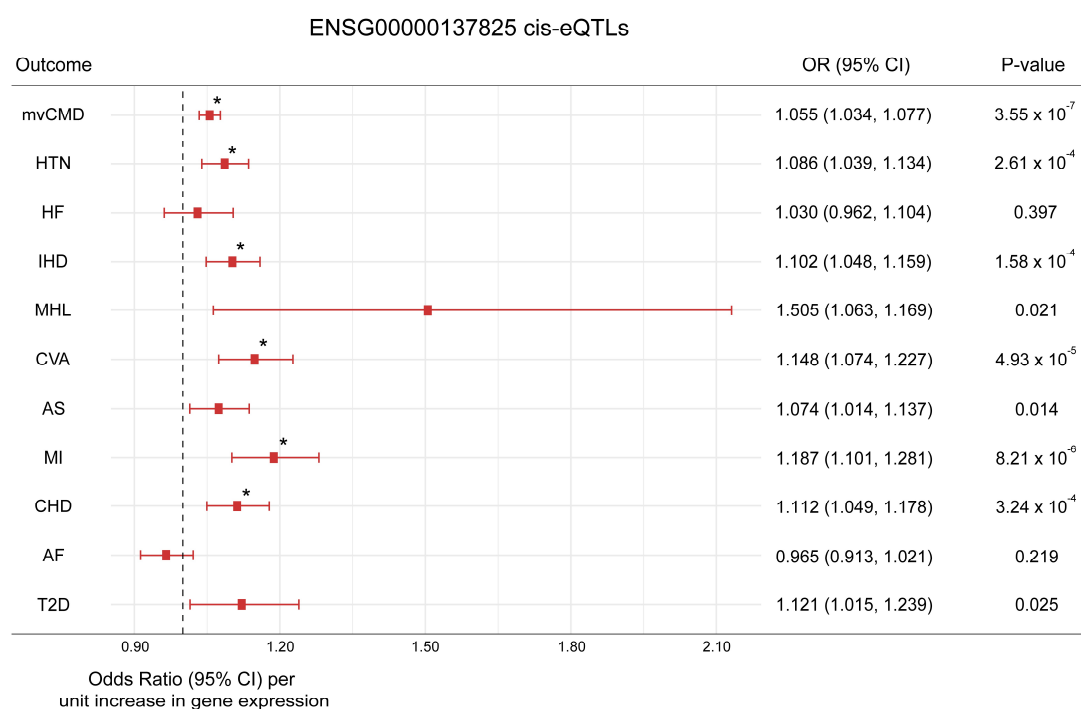
Supplementary Fig. 3 Associations between Genetically Predicted ENSG00000175305 and CMD

Data presented are odds ratio (calculated from the MR estimates) for the IVW MR method and the corresponding 95% CIs. The vertical line at the center of the forest plots represents a value of 0, which corresponds to no change in the IVW estimate of the effect of ENSG00000175305 on mvCMD. Full results are presented in **Supplementary Table 39-40**. * indicates that the MR estimate surpasses the Bonferroni-adjusted P-value threshold = 0.005, corrected for the 10 variables compared. P-values are derived from two-sided Wald tests.



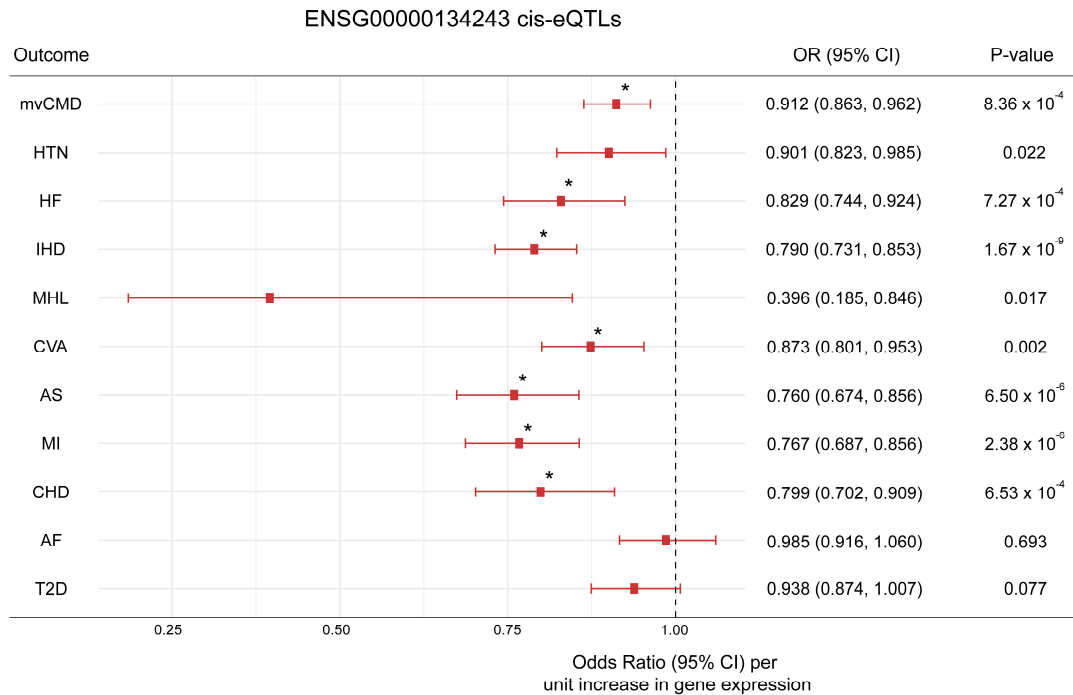
Supplementary Fig. 4 Associations between Genetically Predicted ENSG00000111252 and CMD

Data presented are odds ratio (calculated from the MR estimates) for the IVW MR method and the corresponding 95% CIs. The vertical line at the center of the forest plots represents a value of 0, which corresponds to no change in the IVW estimate of the effect of ENSG00000111252 on mvCMD. Full results are presented in **Supplementary Table 39-40**. * indicates that the MR estimate surpasses the Bonferroni-adjusted P-value threshold = 0.005, corrected for the 10 variables compared. P-values are derived from two-sided Wald tests.



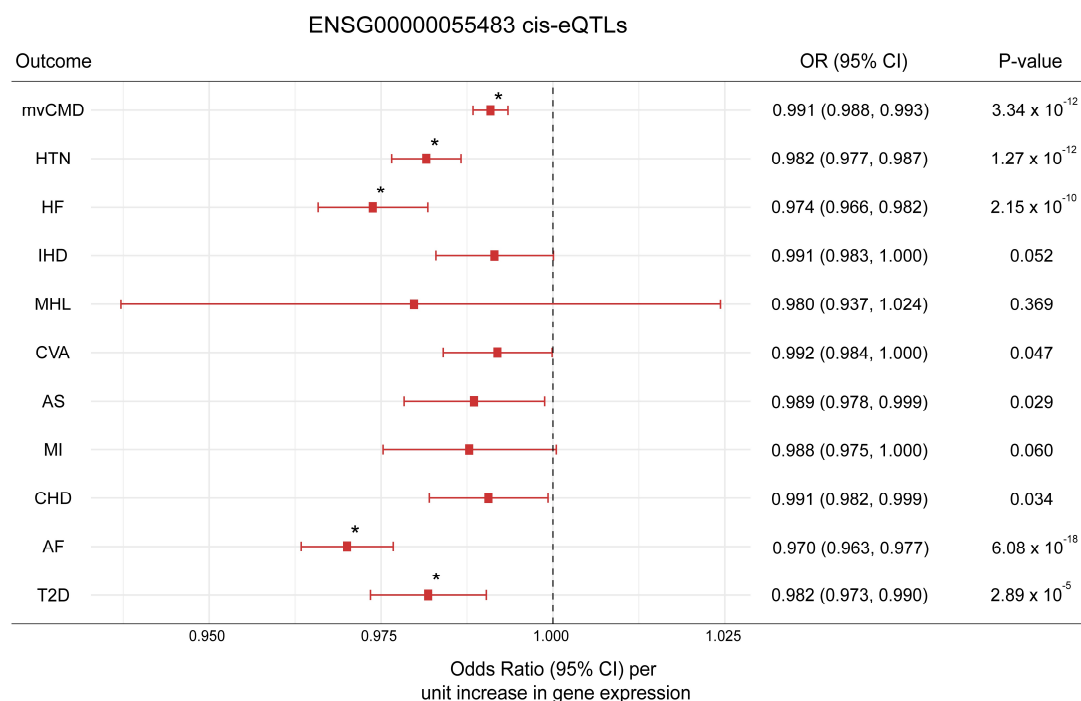
Supplementary Fig. 5 Associations between Genetically Predicted ENSG00000137825 and CMD

Data presented are odds ratio (calculated from the MR estimates) for the IVW MR method and the corresponding 95% CIs. The vertical line at the center of the forest plots represents a value of 0, which corresponds to no change in the IVW estimate of the effect of ENSG00000137825 on mvCMD. Full results are presented in **Supplementary Table 39-40**. * indicates that the MR estimate surpasses the Bonferroni-adjusted P-value threshold = 0.005, corrected for the 10 variables compared. P-values are derived from two-sided Wald tests.



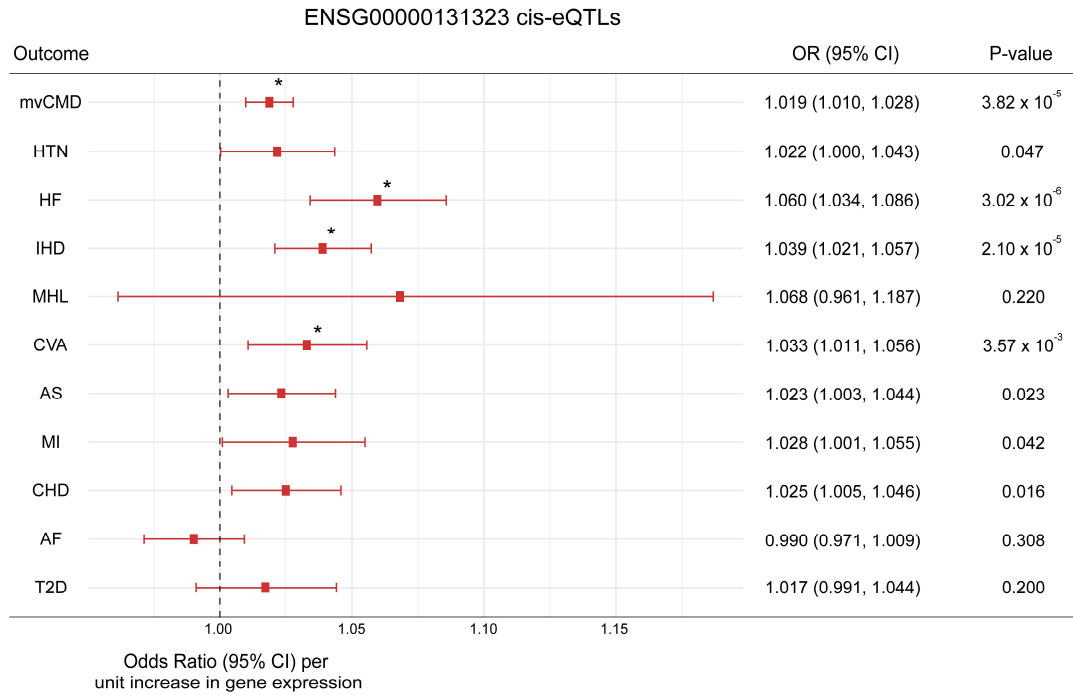
Supplementary Fig. 6 Associations between Genetically Predicted ENSG00000134243 and CMD

Data presented are odds ratio (calculated from the MR estimates) for the IVW MR method and the corresponding 95% CIs. The vertical line at the center of the forest plots represents a value of 1, which corresponds to no change in the IVW estimate of the effect of ENSG00000134243 on mvCMD. Full results are presented in **Supplementary Table 39-40**. * indicates that the MR estimate surpasses the Bonferroni-adjusted P-value threshold = 0.005, corrected for the 10 variables compared. P-values are derived from two-sided Wald tests.



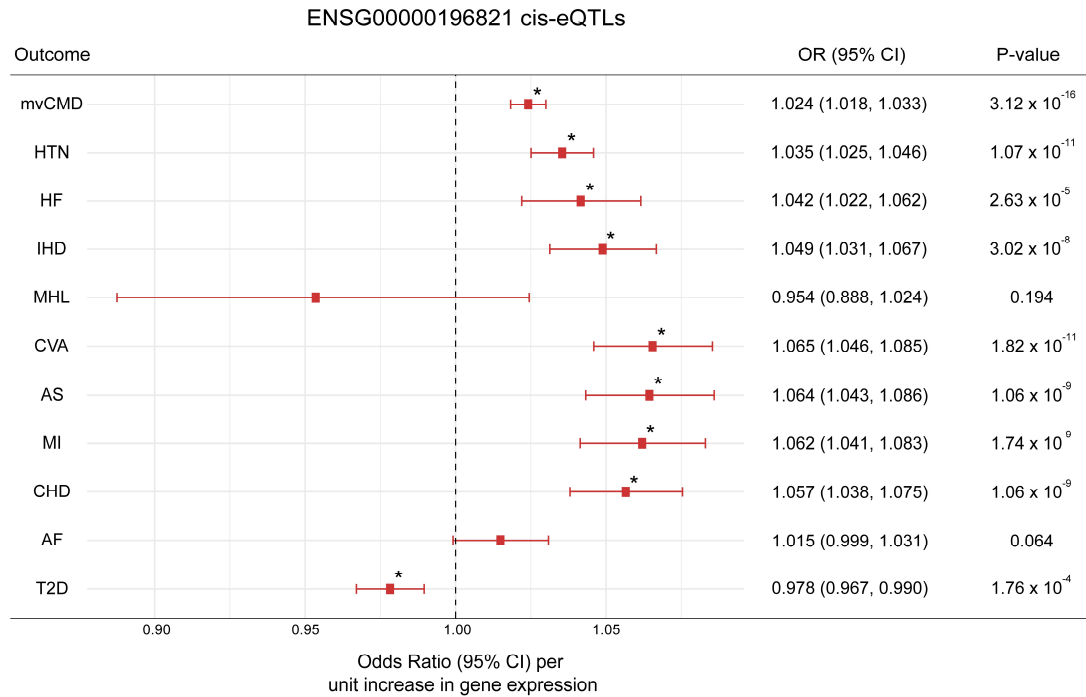
Supplementary Fig. 7 Associations between Genetically Predicted ENSG00000055483 and CMD

Data presented are odds ratio (calculated from the MR estimates) for the IVW MR method and the corresponding 95% CIs. The vertical line at the center of the forest plots represents a value of 0, which corresponds to no change in the IVW estimate of the effect of ENSG00000055483 on mvCMD. Full results are presented in **Supplementary Table 39-40**. * indicates that the MR estimate surpasses the Bonferroni-adjusted P-value threshold = 0.005, corrected for the 10 variables compared. P-values are derived from two-sided Wald tests.



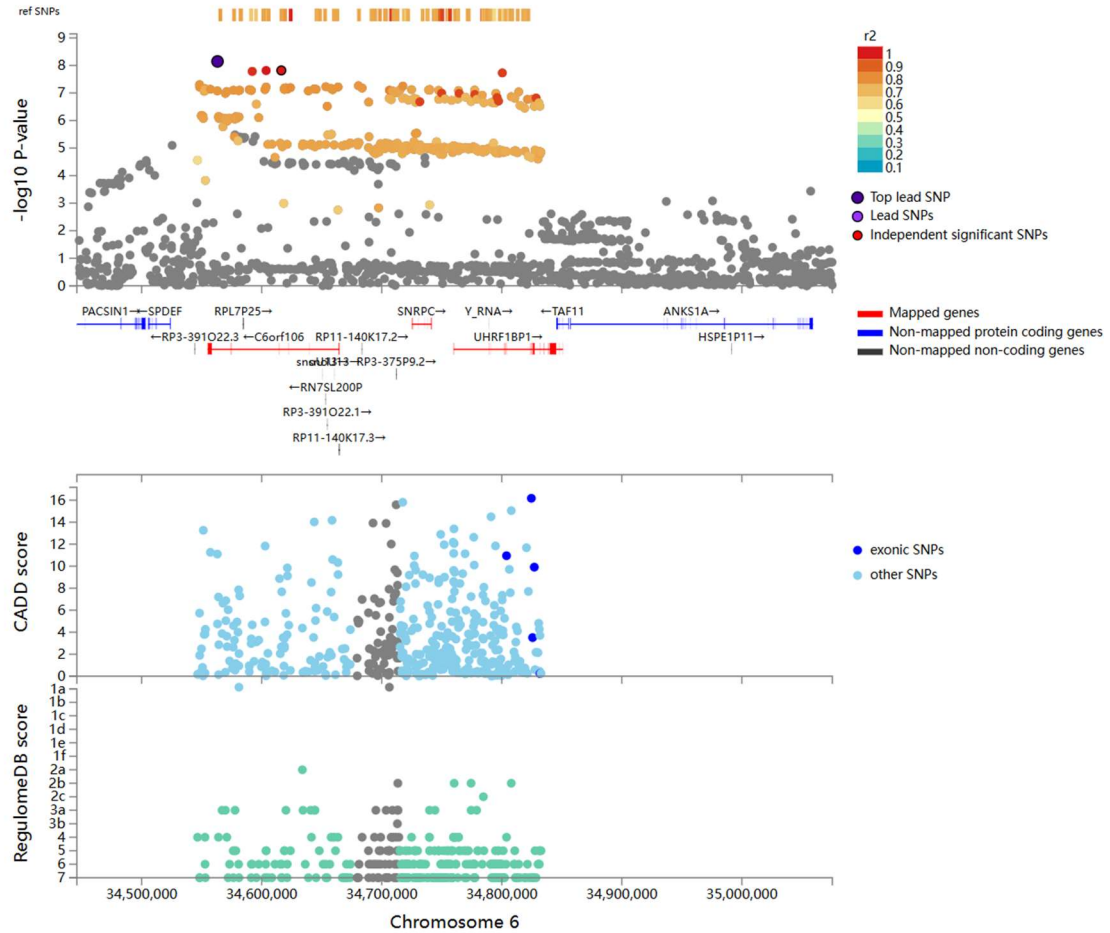
Supplementary Fig. 8 Associations between Genetically Predicted ENSG00000131323 and CMD

Data presented are odds ratio (calculated from the MR estimates) for the IVW MR method and the corresponding 95% CIs. The vertical line at the center of the forest plots represents a value of 0, which corresponds to no change in the IVW estimate of the effect of ENSG00000131323 on mvCMD. Full results are presented in **Supplementary Table 39-40**. * indicates that the MR estimate surpasses the Bonferroni-adjusted P-value threshold = 0.005, corrected for the 10 variables compared. P-values are derived from two-sided Wald tests.



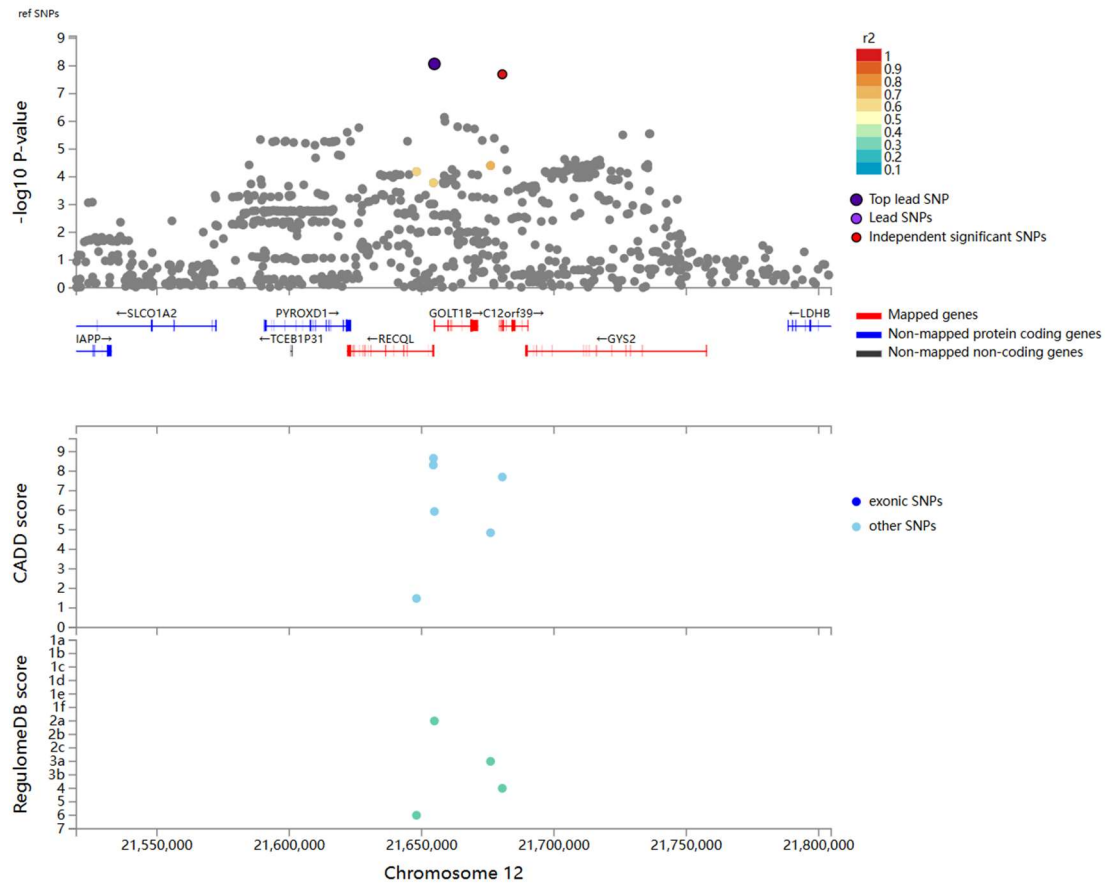
Supplementary Fig. 9 Associations between Genetically Predicted ENSG00000196821 and CMD

Data presented are odds ratio (calculated from the MR estimates) for the IVW MR method and the corresponding 95% CIs. The vertical line at the center of the forest plots represents a value of 1, which corresponds to no change in the IVW estimate of the effect of ENSG00000196821 on mvCMD. Full results are presented in **Supplementary Table 39-40**. * indicates that the MR estimate surpasses the Bonferroni-adjusted P-value threshold = 0.005, corrected for the 10 variables compared. P-values are derived from two-sided Wald tests.



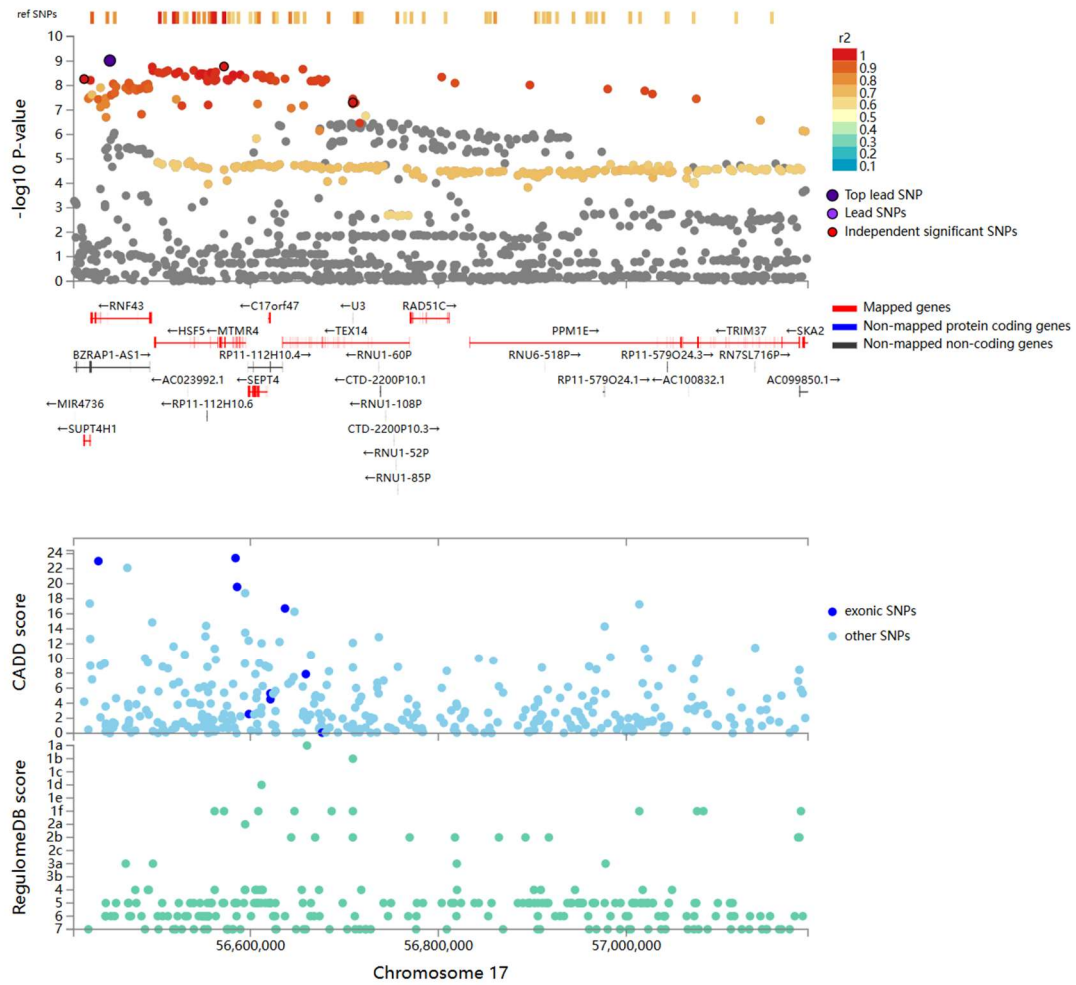
Supplementary Fig. 10 Regional Association Plot for SNP rs205262

Regional plot of the locus around the novel lead variant rs205262 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



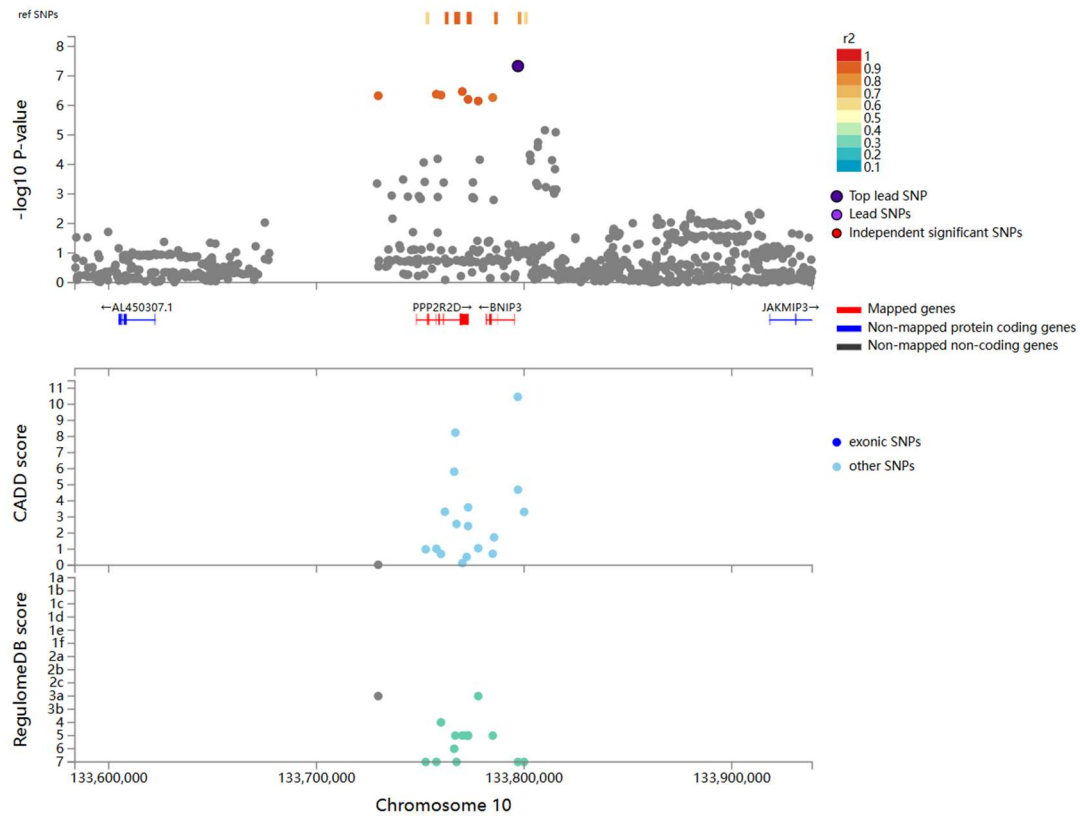
Supplementary Fig. 11 Regional Association Plot for SNP rs1060126

Regional plot of the locus around the novel lead variant rs1060126 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



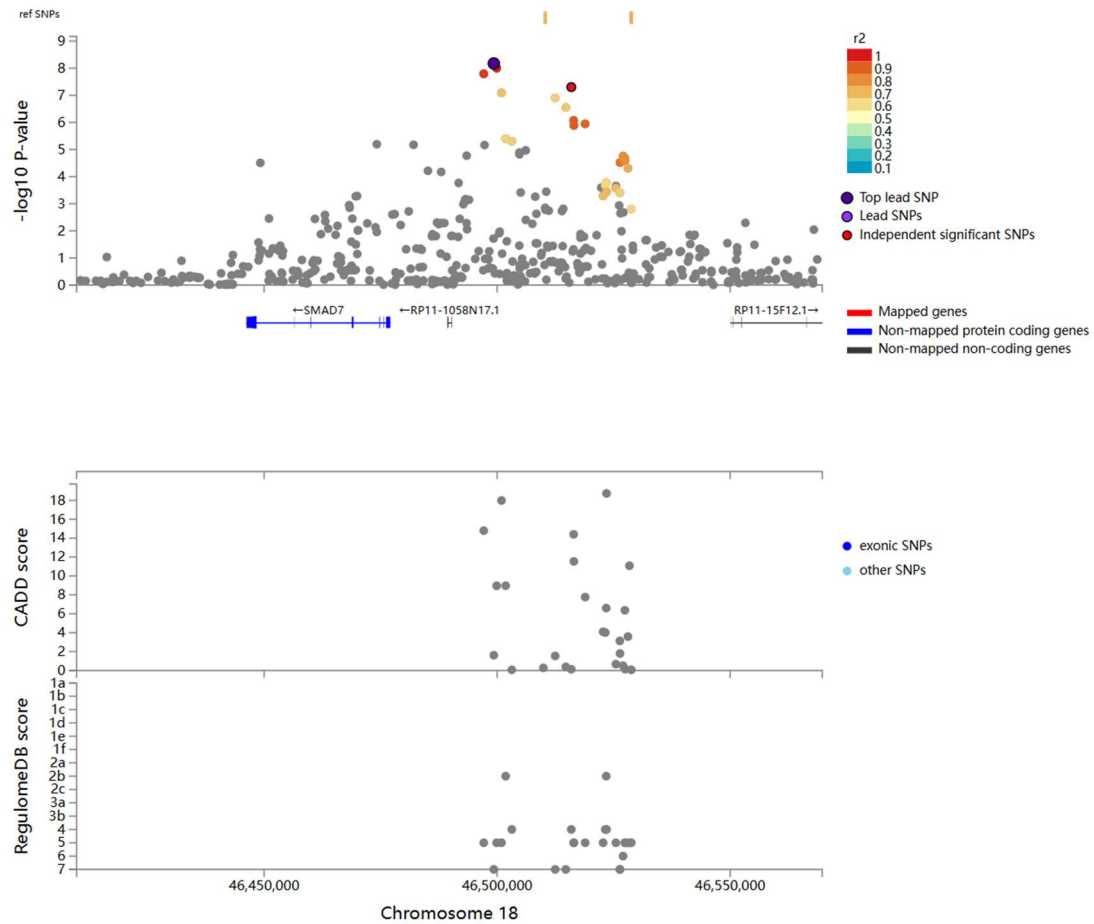
Supplementary Fig. 12 Regional Association Plot for SNP rs2680688

Regional plot of the locus around the novel lead variant rs2680688 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



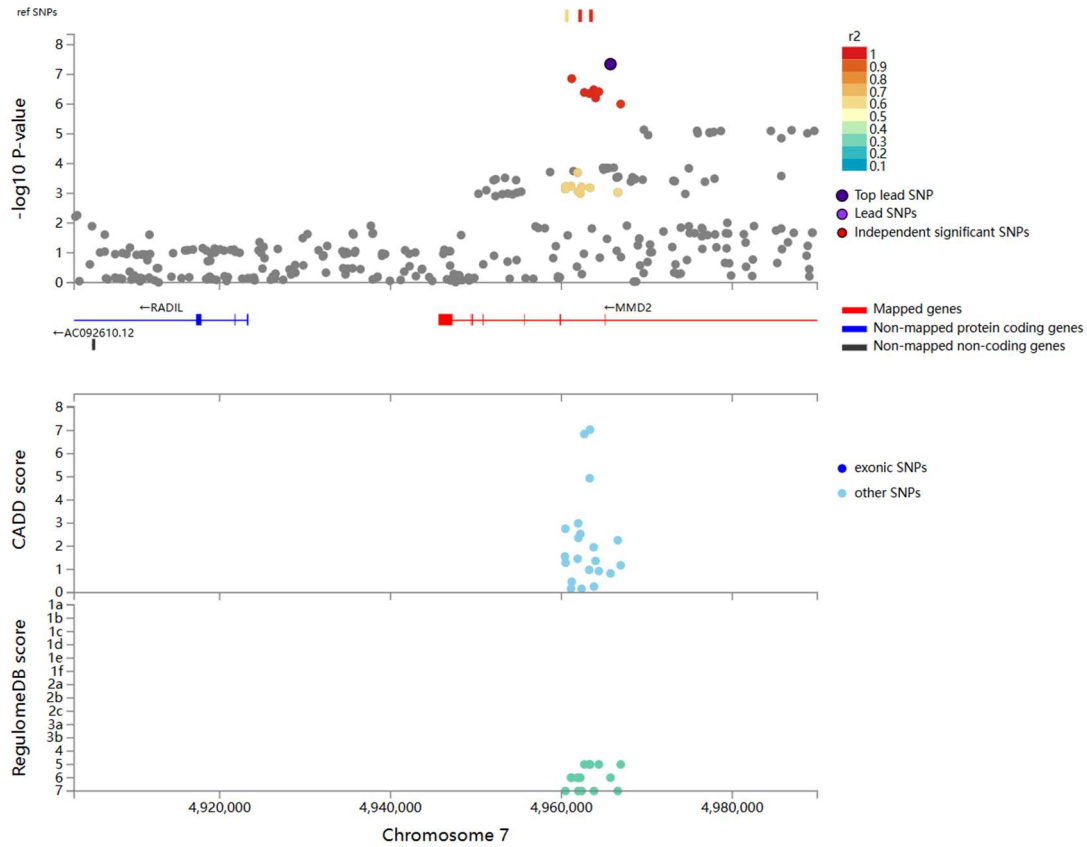
Supplementary Fig. 14 Regional Association Plot for SNP rs3793742

Regional plot of the locus around the novel lead variant rs3793742 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



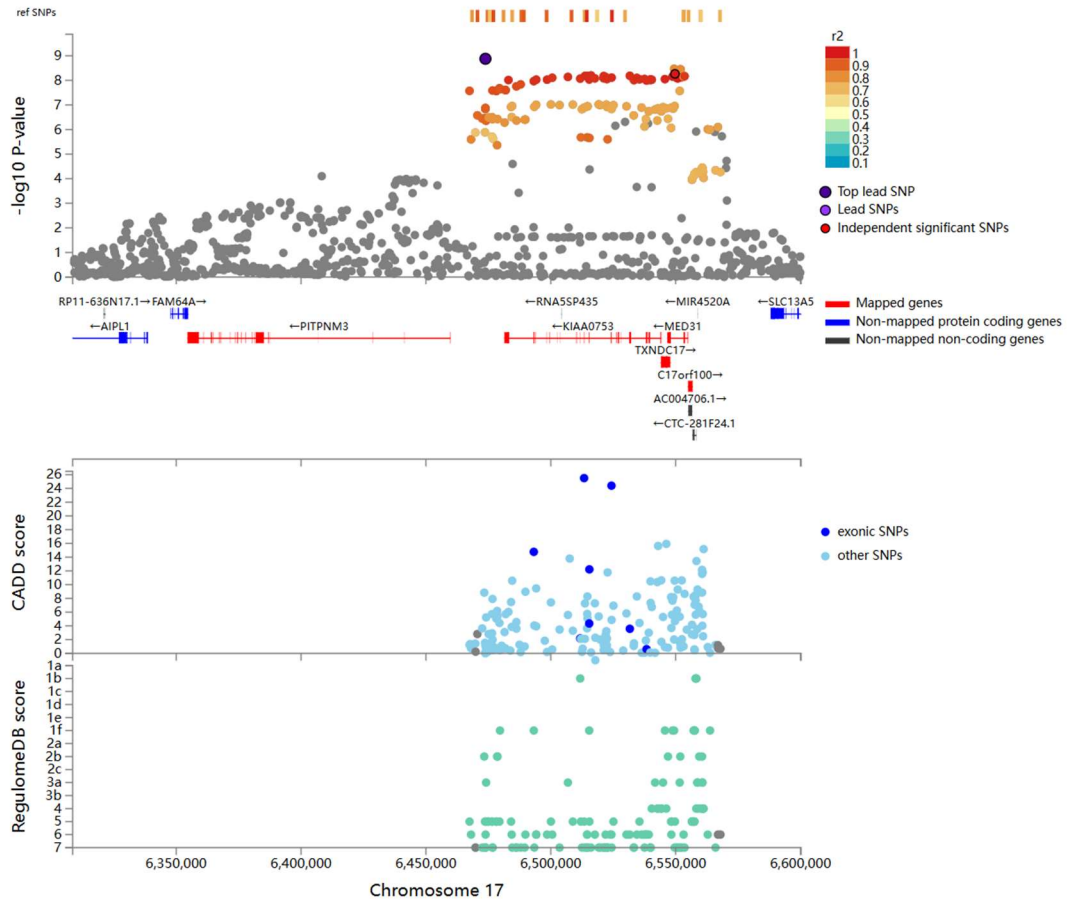
Supplementary Fig. 15 Regional Association Plot for SNP rs6507881

Regional plot of the locus around the novel lead variant rs6507881 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



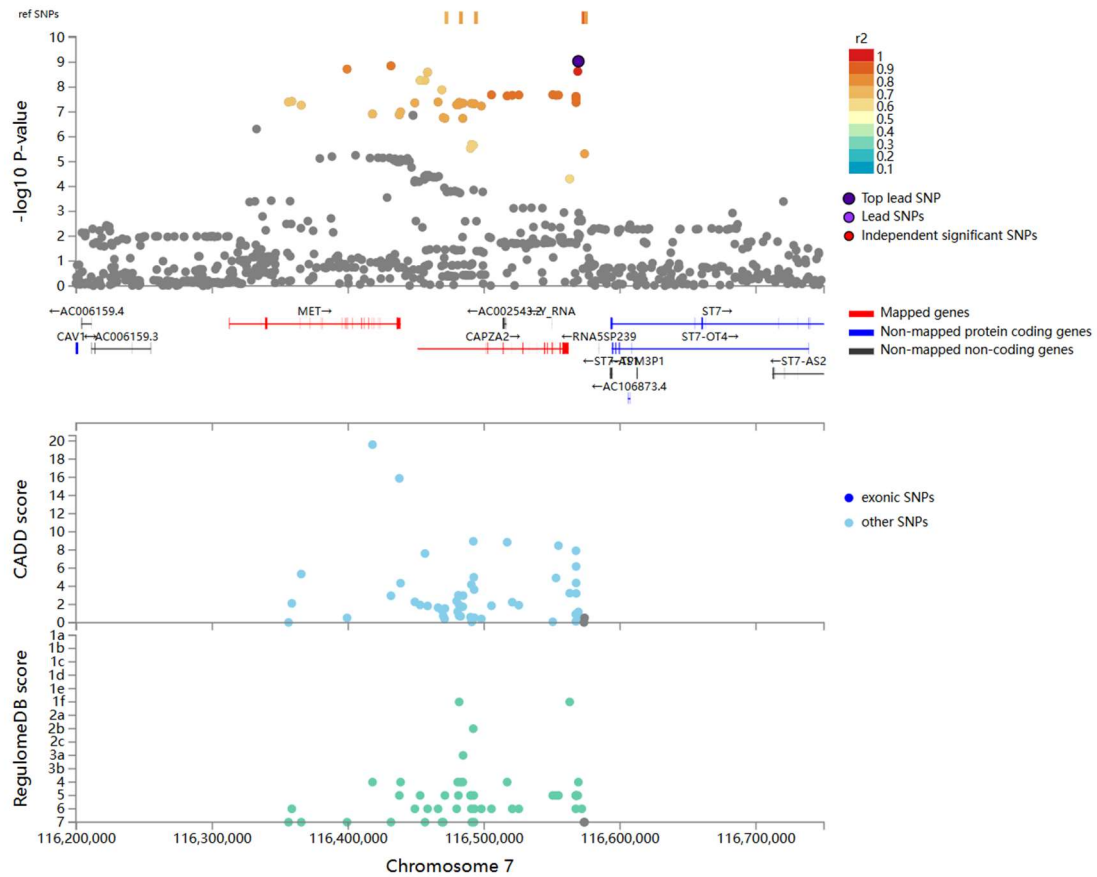
Supplementary Fig. 16 Regional Association Plot for SNP rs6955307

Regional plot of the locus around the novel lead variant rs6955307 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



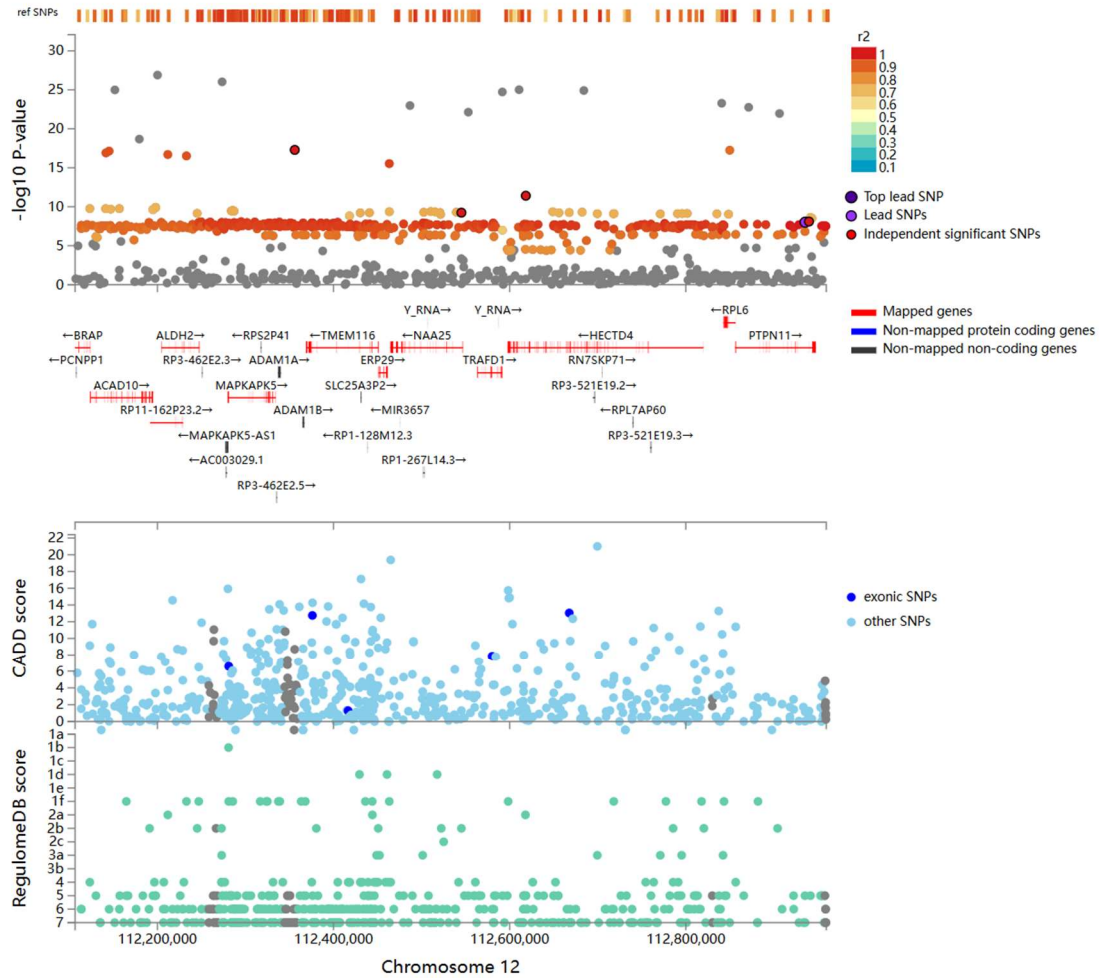
Supplementary Fig. 17 Regional Association Plot for SNP rs7226020

Regional plot of the locus around the novel lead variant rs7226020 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



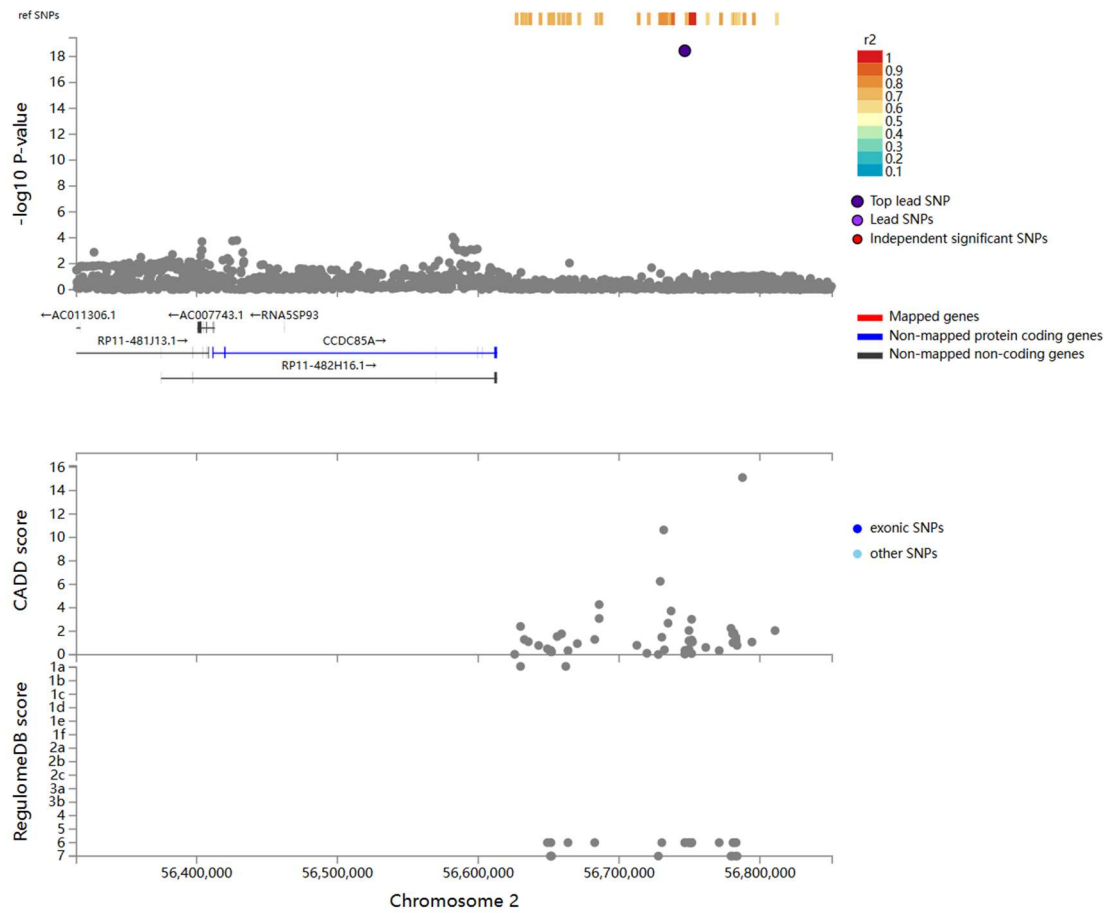
Supplementary Fig. 18 Regional Association Plot for SNP rs7807976

Regional plot of the locus around the novel lead variant rs7807976 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



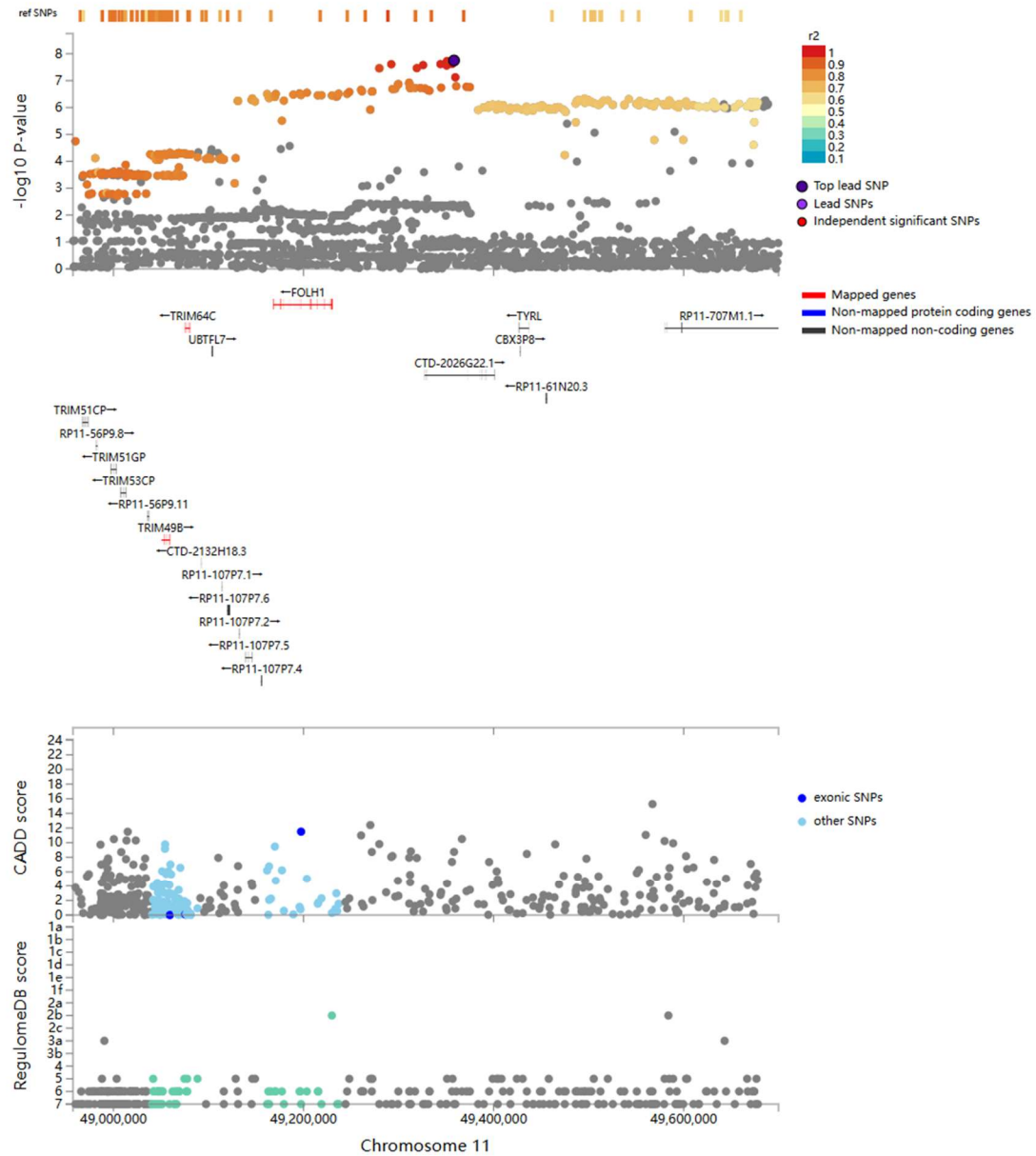
Supplementary Fig. 19 Regional Association Plot for SNP rs7958372

Regional plot of the locus around the novel lead variant rs7958372 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



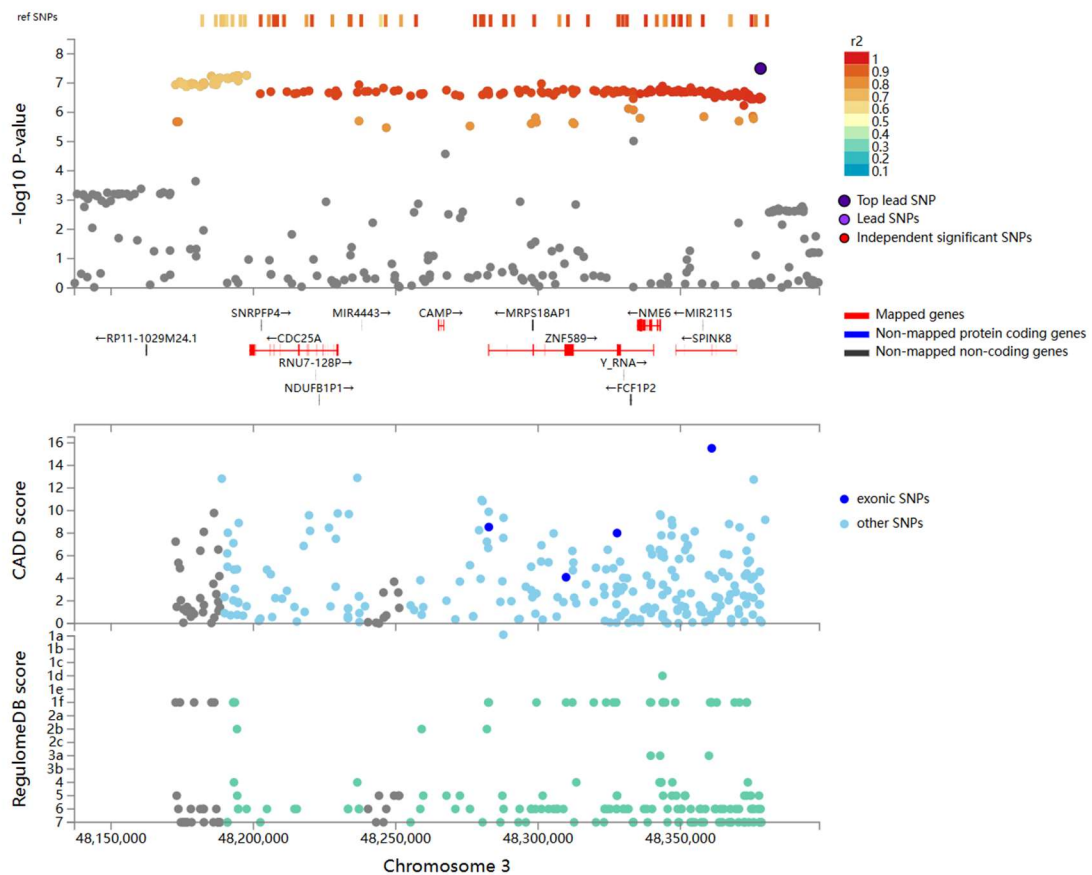
Supplementary Fig. 20 Regional Association Plot for SNP rs10204808

Regional plot of the locus around the novel lead variant rs10204808 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}$ (P-value) results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



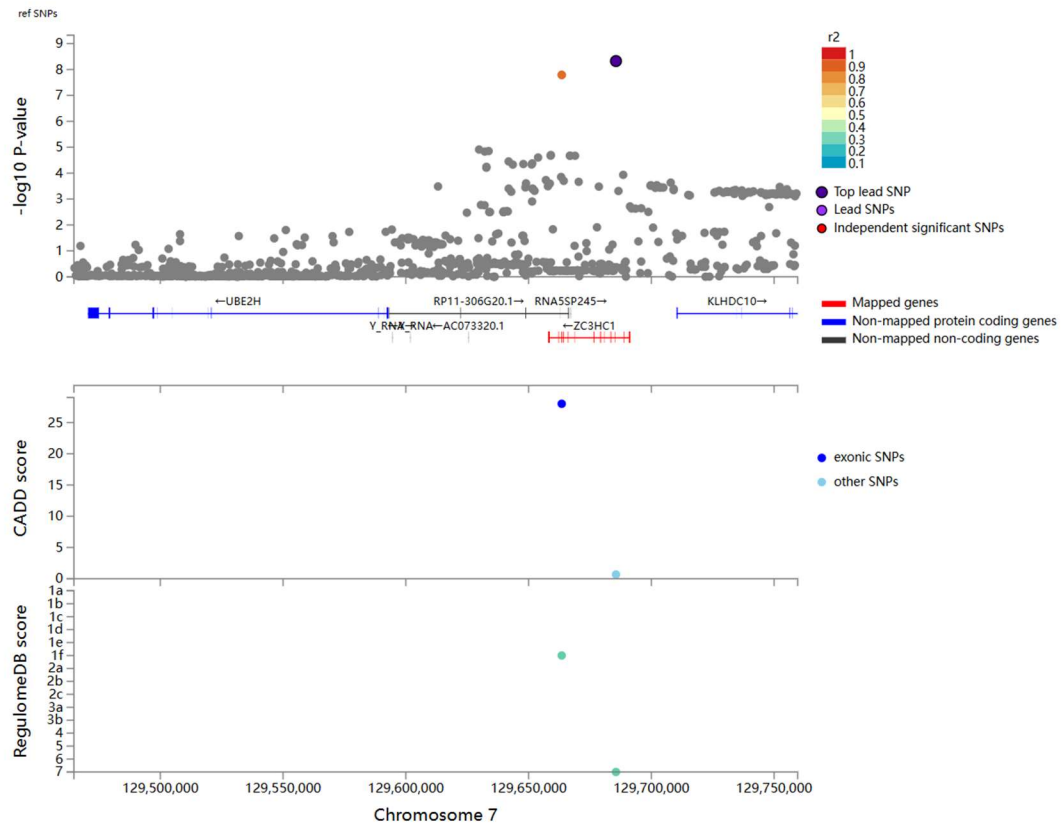
Supplementary Fig. 21 Regional Association Plot for SNP rs11607791

Regional plot of the locus around the novel lead variant rs11607791 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



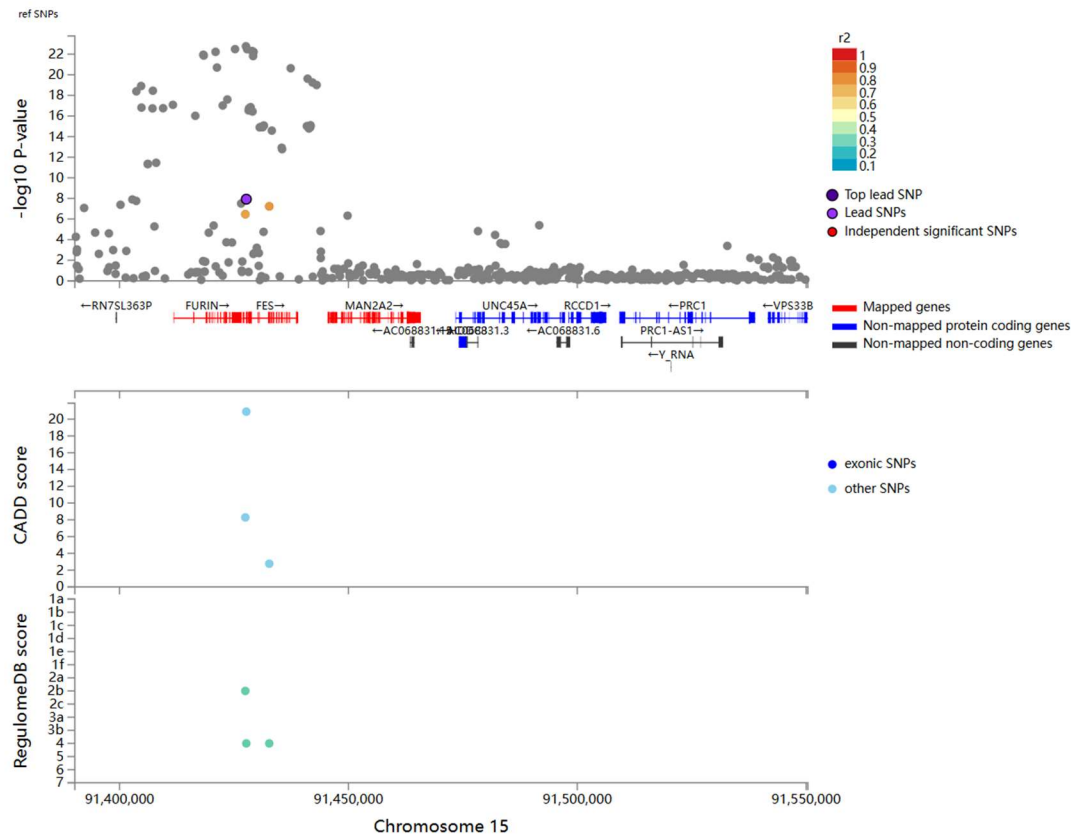
Supplementary Fig. 22 Regional Association Plot for SNP rs35297486

Regional plot of the locus around the novel lead variant rs35297486 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



Supplementary Fig. 23 Regional Association Plot for SNP rs56179563

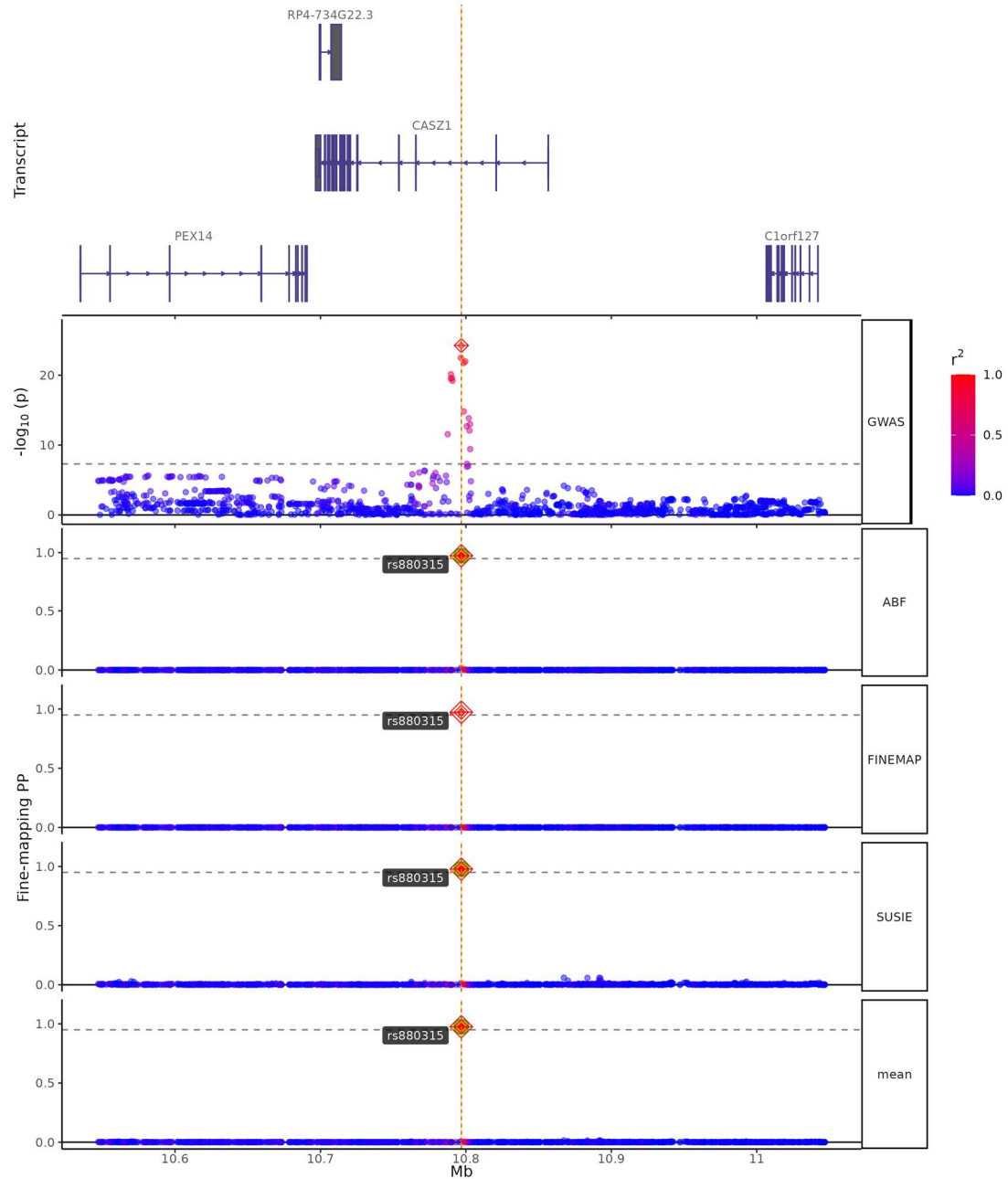
Regional plot of the locus around the novel lead variant rs56179563 identified in the mvCMD (Neff = 932,442). Top panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.



Supplementary Fig. 24 Regional Association Plot for SNP rs138682554

Regional plot of the locus around the novel lead variant rs138682554 identified in the mvCMD (Neff = 932,442). Top panel displays the -log₁₀(P-value) results of two-sided Wald tests for each variant on mvCMD and LD R² information in the locus (variants are colored by LD R²) and the genes prioritized by FUMA are highlighted in red on the track below. The bottom panel illustrates the CADD (combined annotation dependent depletion) scores and RegulomeDB scores, presented in the top and bottom tracks, respectively.

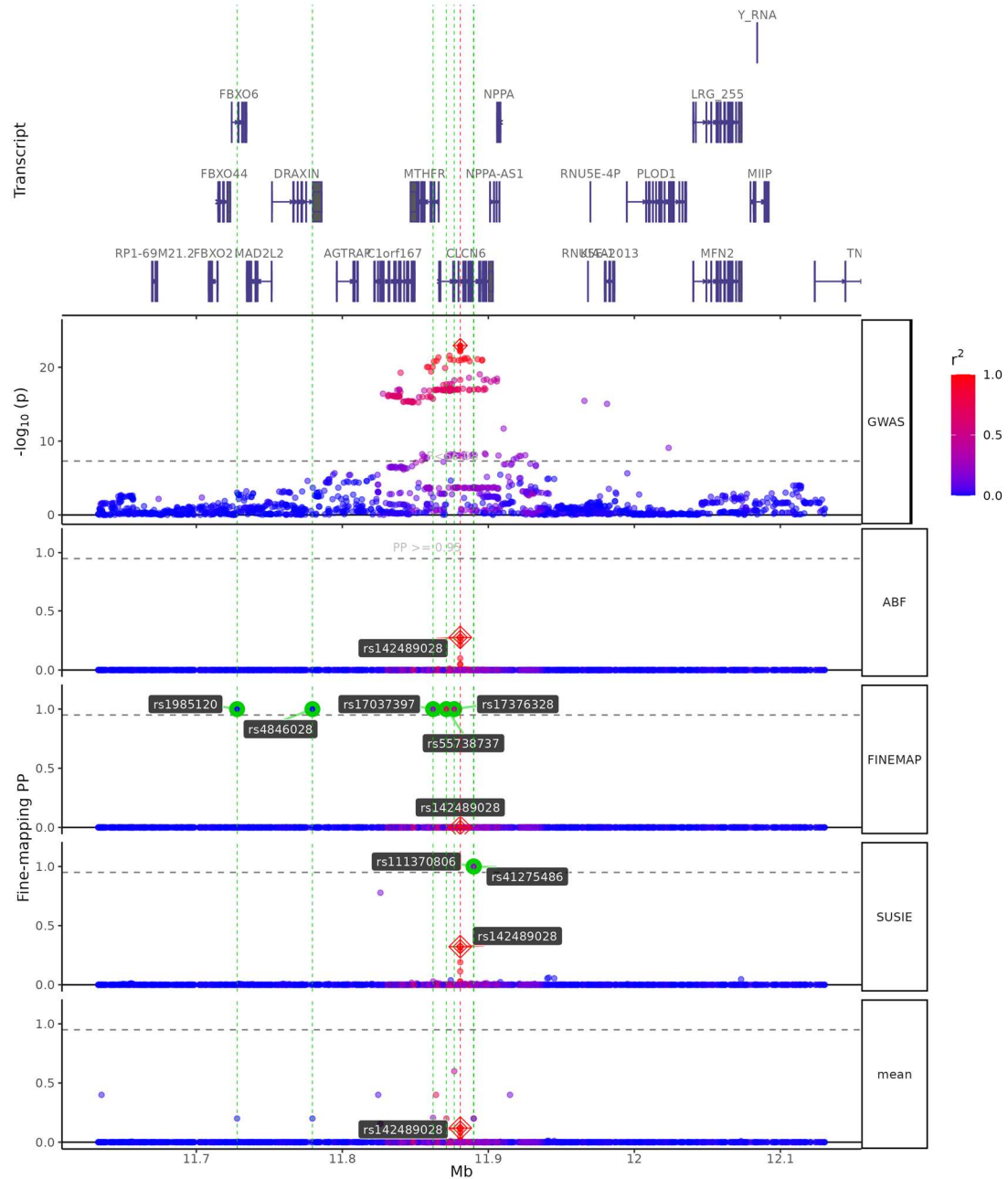
Locus: CASZ1 (SNPs=1,313; zoom=1x)



Supplementary Fig. 25 Fine-mapping Results of CASZ1

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

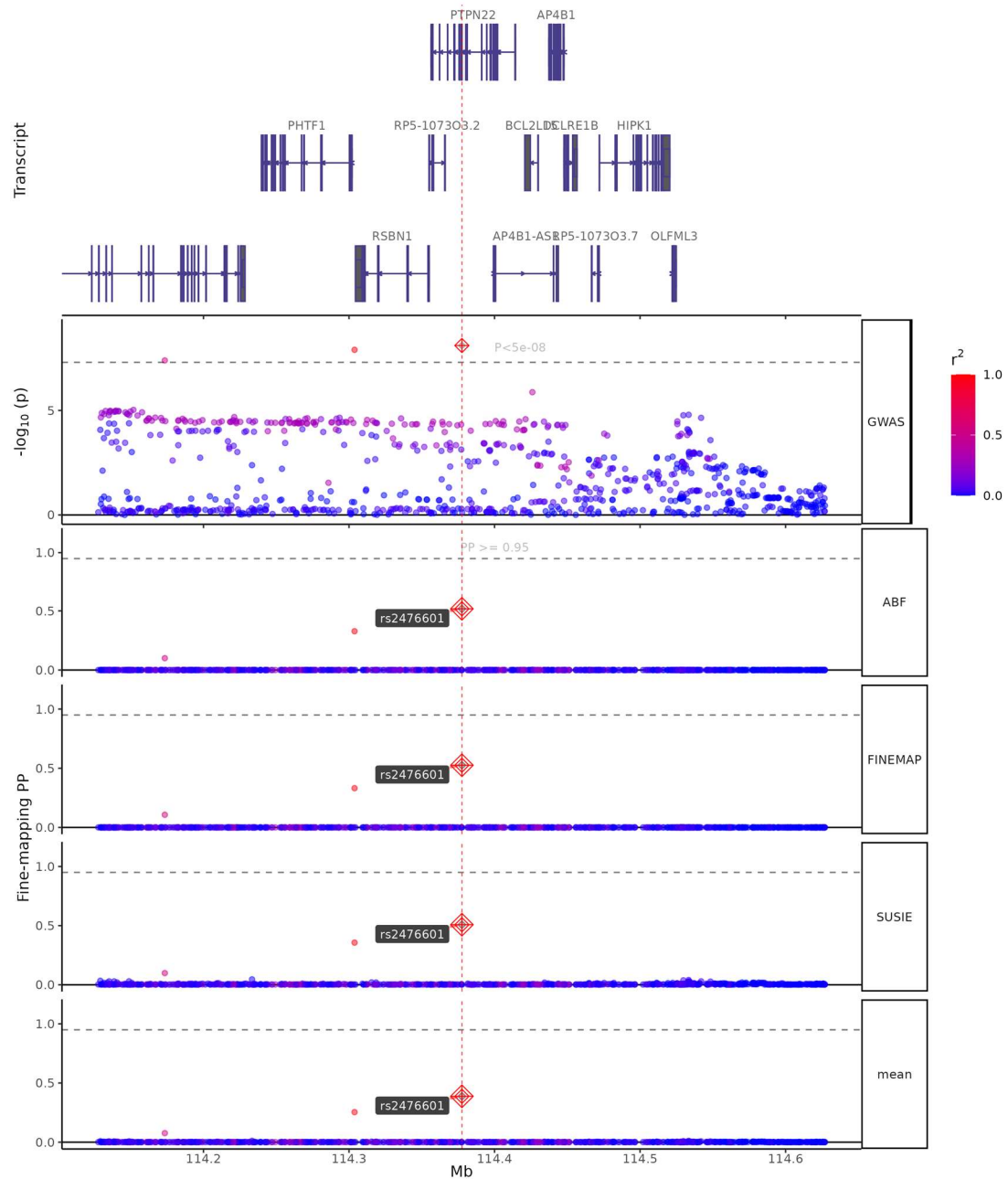
Locus: CLCN6 (SNPs=1,602; zoom=1x)



Supplementary Fig. 26 Fine-mapping Results of CLCN6

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P)$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

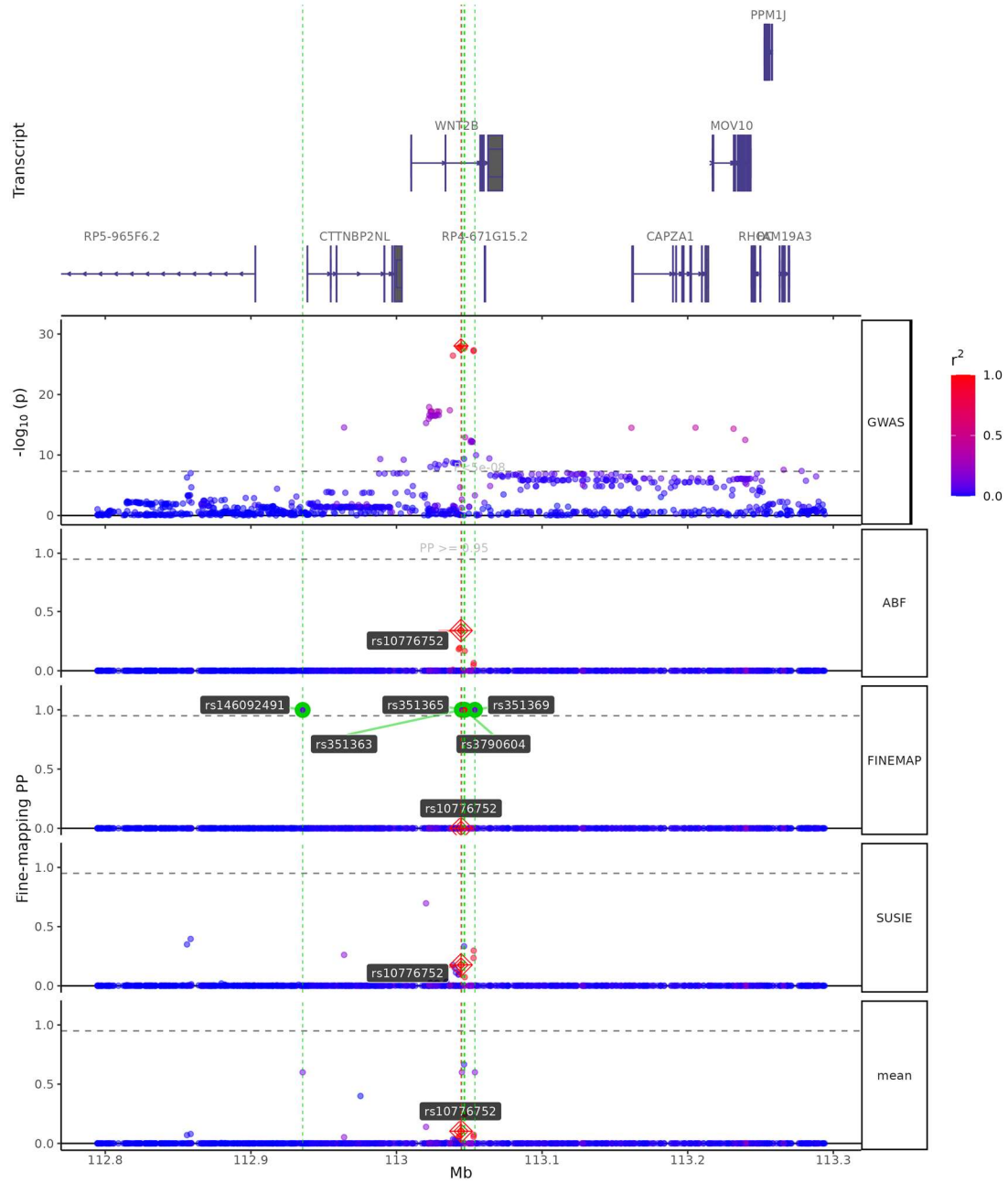
Locus: PTPN22 (SNPs=859; zoom=1x)



Supplementary Fig. 27 Fine-mapping Results of PTPN22

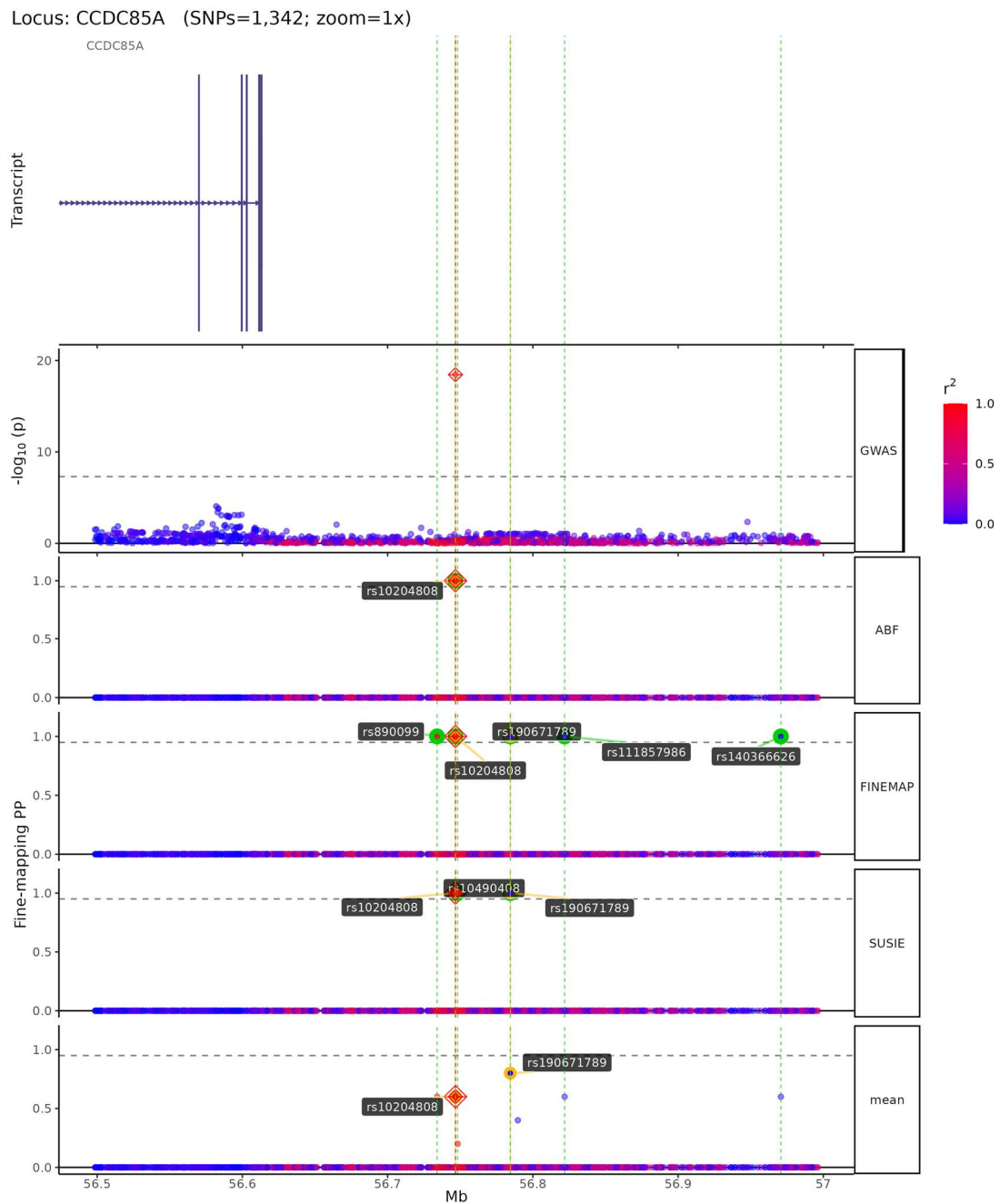
The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

Locus: WNT2B (SNPs=967; zoom=1x)



Supplementary Fig. 28 Fine-mapping Results of WNT2B

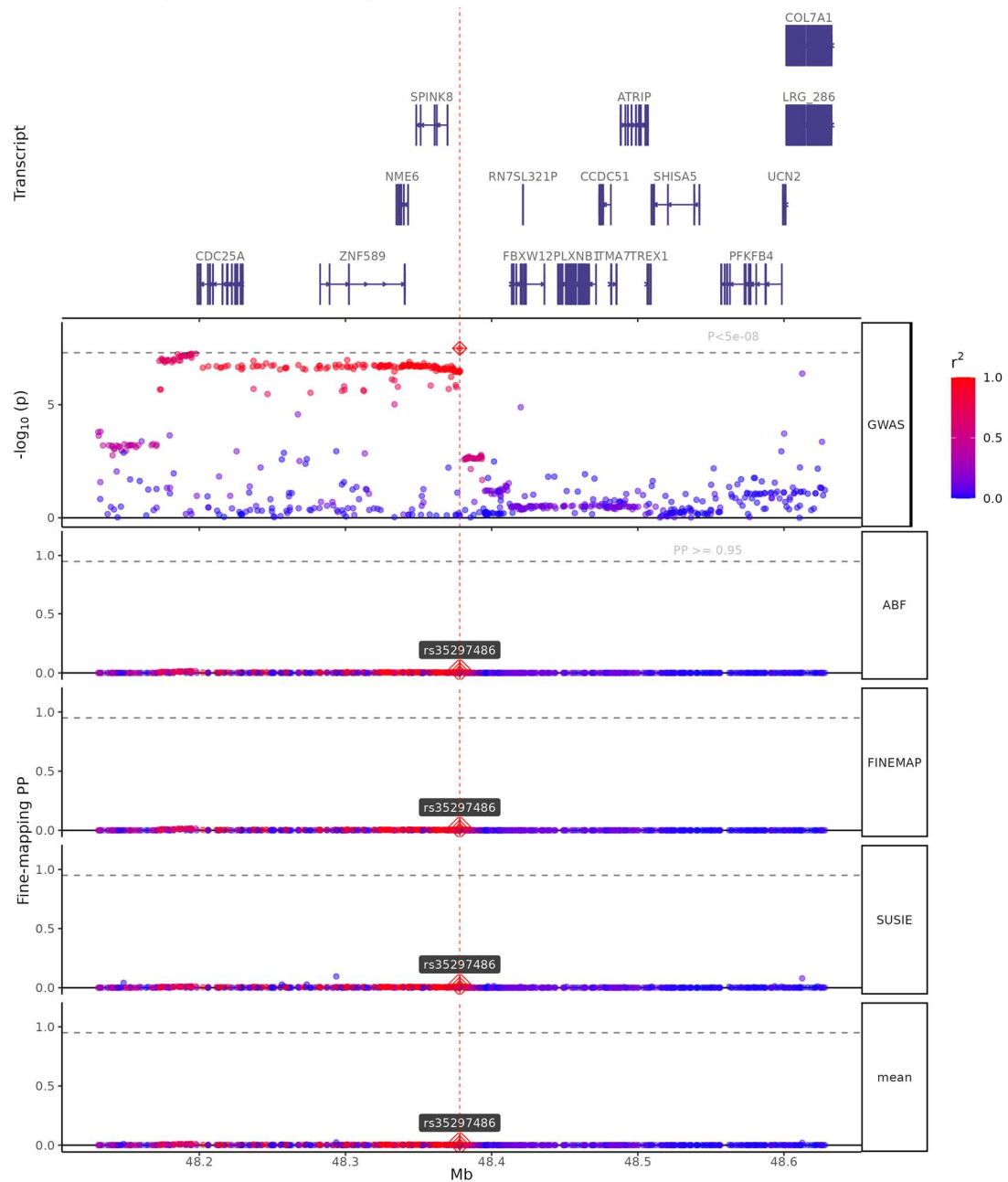
The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.



Supplementary Fig. 29 Fine-mapping Results of CCDC85A

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

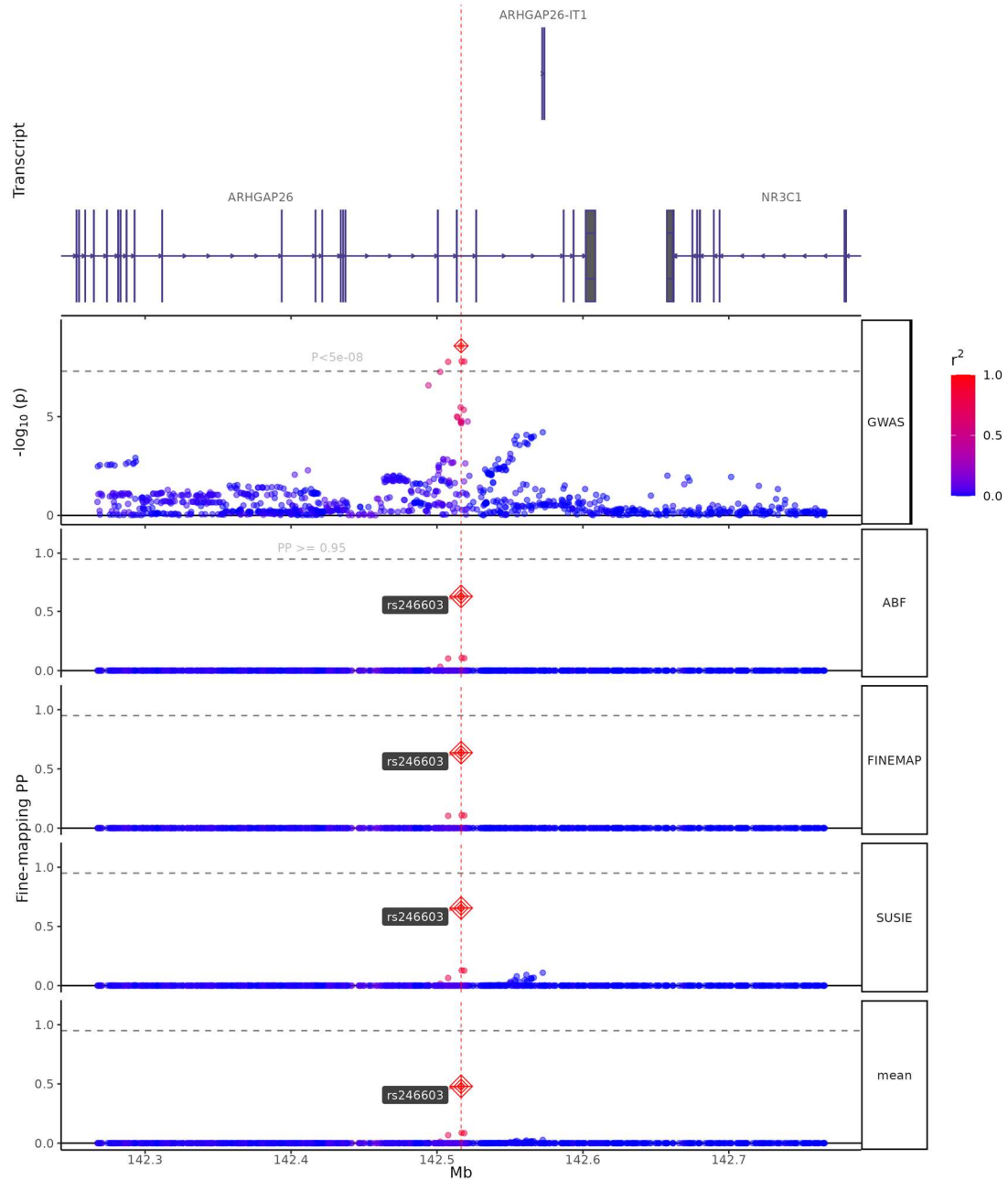
Locus: SPINK8 (SNPs=668; zoom=1x)



Supplementary Fig. 30 Fine-mapping Results of SPINK8

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

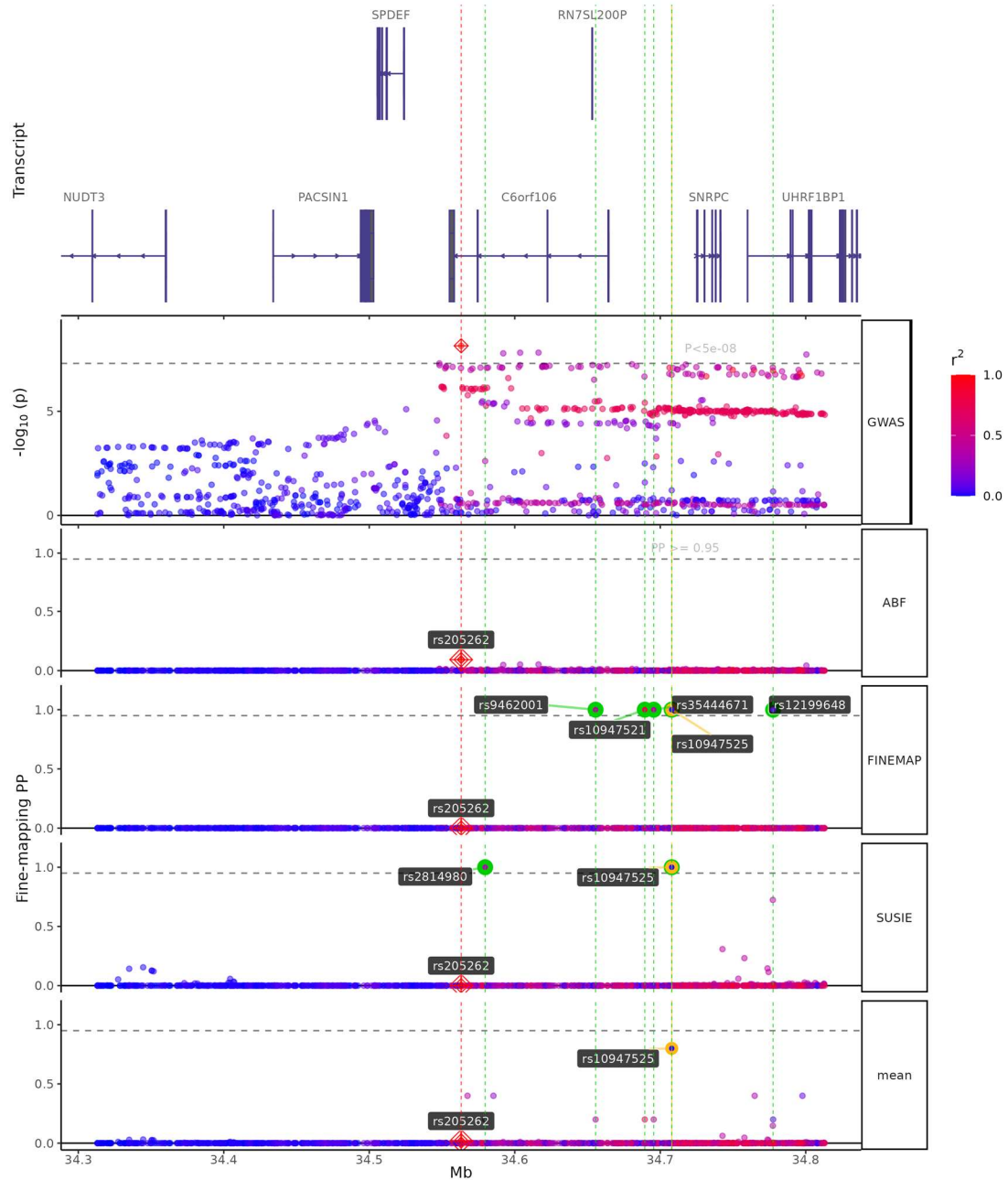
Locus: ARHGAP26 (SNPs=1,019; zoom=1x)



Supplementary Fig. 31 Fine-mapping Results of ARHGAP26

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

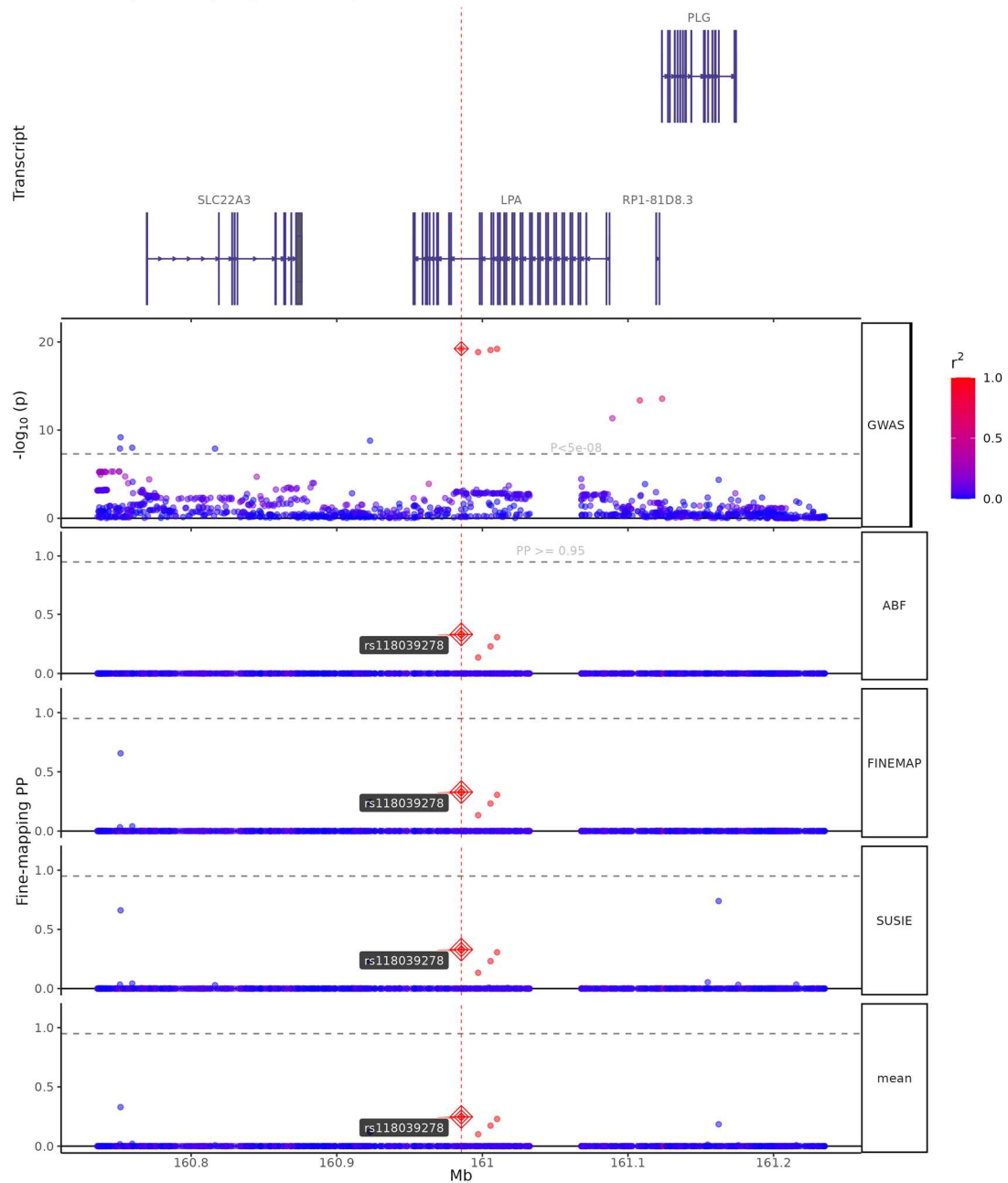
Locus: C6orf106 (SNPs=1,076; zoom=1x)



Supplementary Fig. 32 Fine-mapping Results of C6orf106

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

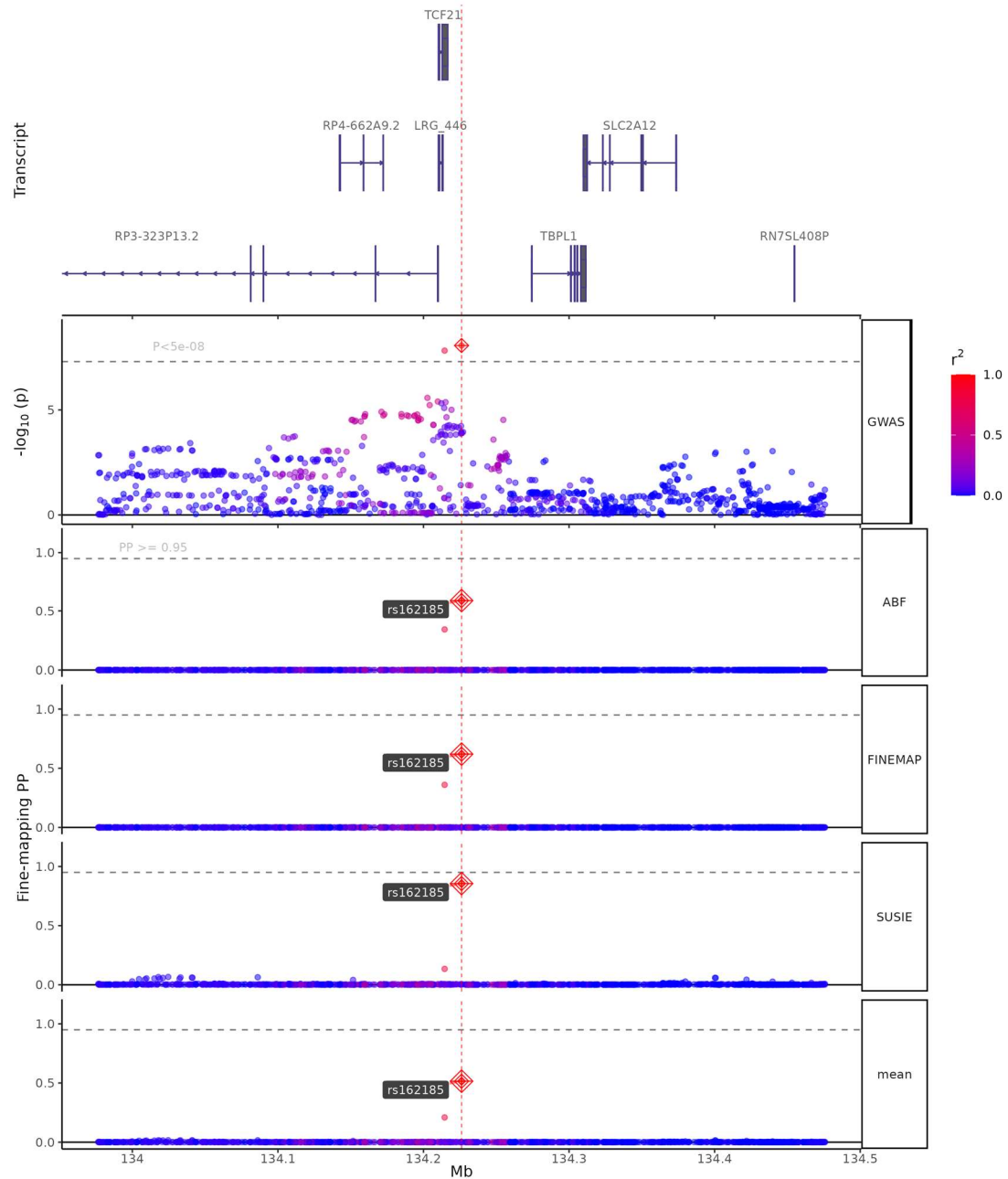
Locus: LPA (SNPs=1,127; zoom=1x)



Supplementary Fig. 33 Fine-mapping Results of LPA

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

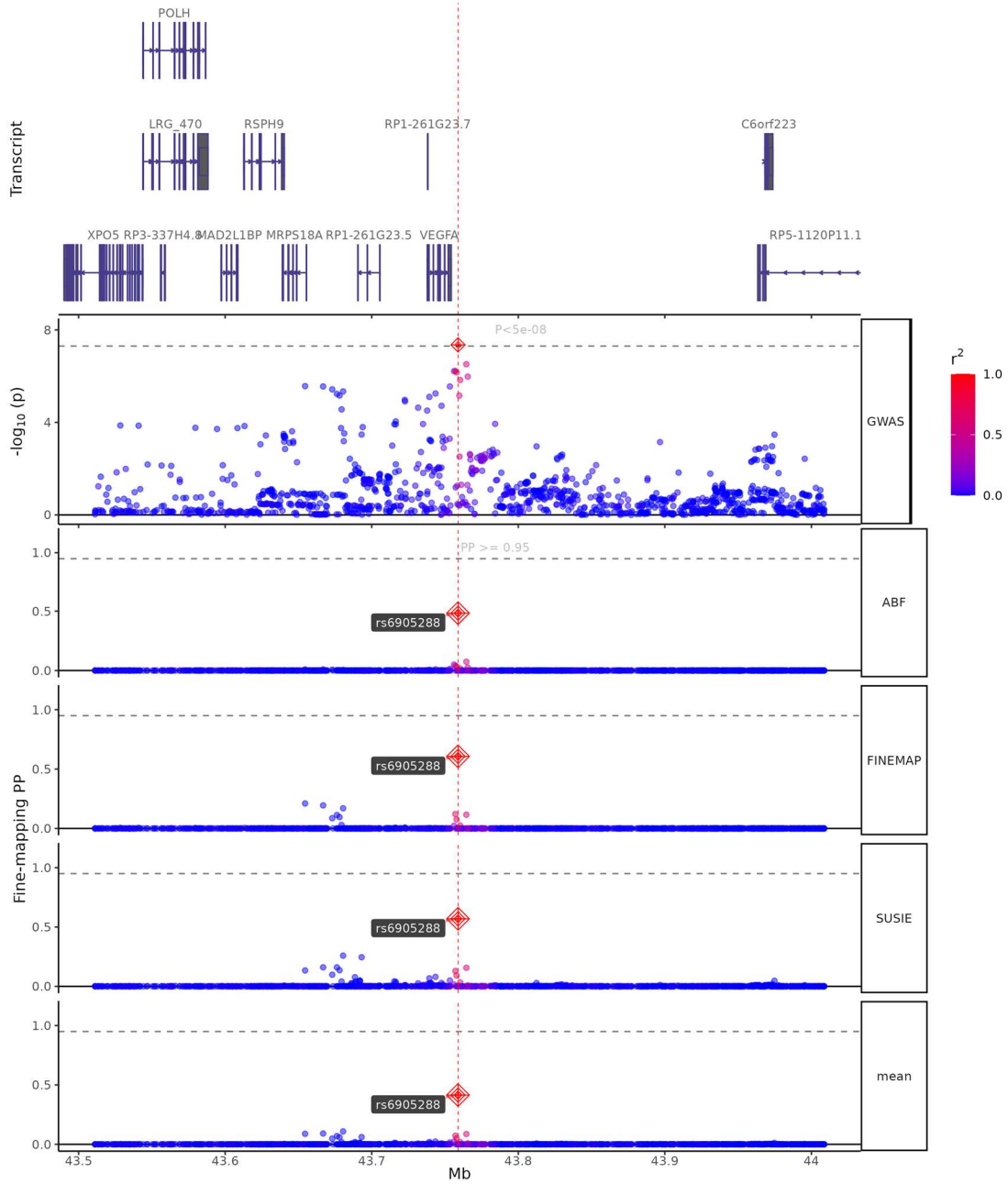
Locus: TCF21 (SNPs=1,117; zoom=1x)



Supplementary Fig. 34 Fine-mapping Results of TCF21

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

Locus: VEGFA (SNPs=1,263; zoom=1x)



Supplementary Fig. 35 Fine-mapping Results of VEGFA

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

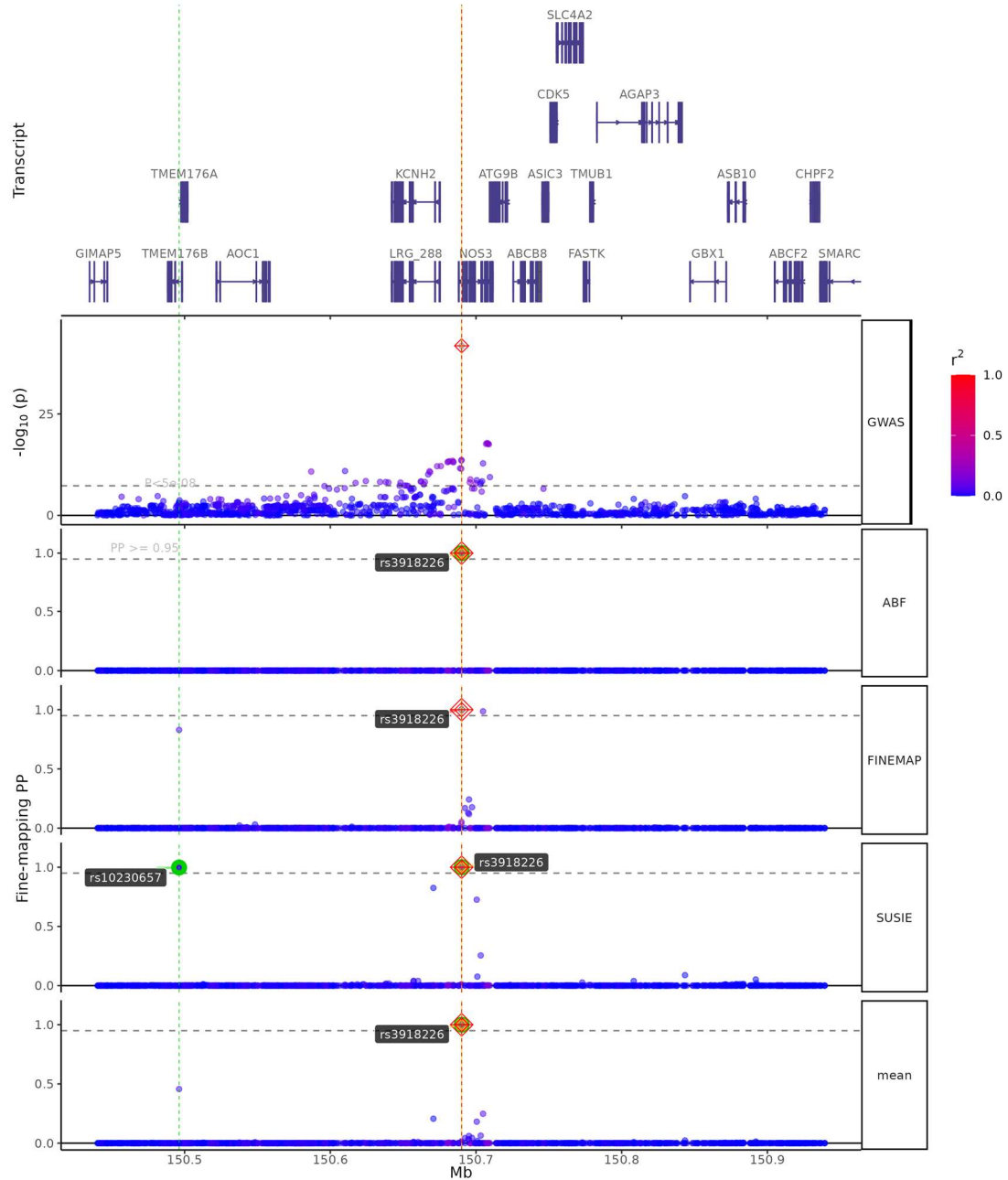
The figure displays five genomic tracks for the RBAK locus on chromosome 10, spanning from approximately 4.7 Mb to 5.2 Mb. The top track shows gene annotations: FOXK1, AP5Z1Y_RNA, PAPOLB, MMD2, RNFB216P1, RBAKDN, and ZNF890P. A red dashed vertical line marks the position of the lead SNP, rs6955307.

- GWAS:** The first track shows the association p-values ($-\log_{10}(p)$) for SNPs across the region. The color of the dots indicates the r^2 value with the lead SNP, ranging from 0.0 (blue) to 1.0 (red). A significant peak is observed at the RBAK locus, reaching a $-\log_{10}(p)$ value of approximately 7.5.
- ABF:** The second track shows the Aberrant Base Frequency (ABF) scores. The lead SNP, rs6955307, is highlighted with a diamond marker.
- FINE-MAP:** The third track shows the Fine-MAP posterior probabilities (PP) for SNPs. The lead SNP, rs6955307, is highlighted with a diamond marker.
- SUSIE:** The fourth track shows the SUSIE posterior probabilities (PP) for SNPs. The lead SNP, rs6955307, is highlighted with a diamond marker.
- mean:** The fifth track shows the mean values for the SNPs. The lead SNP, rs6955307, is highlighted with a diamond marker.

A color scale for r^2 is provided on the right side of the figure, ranging from 0.0 (blue) to 1.0 (red).

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

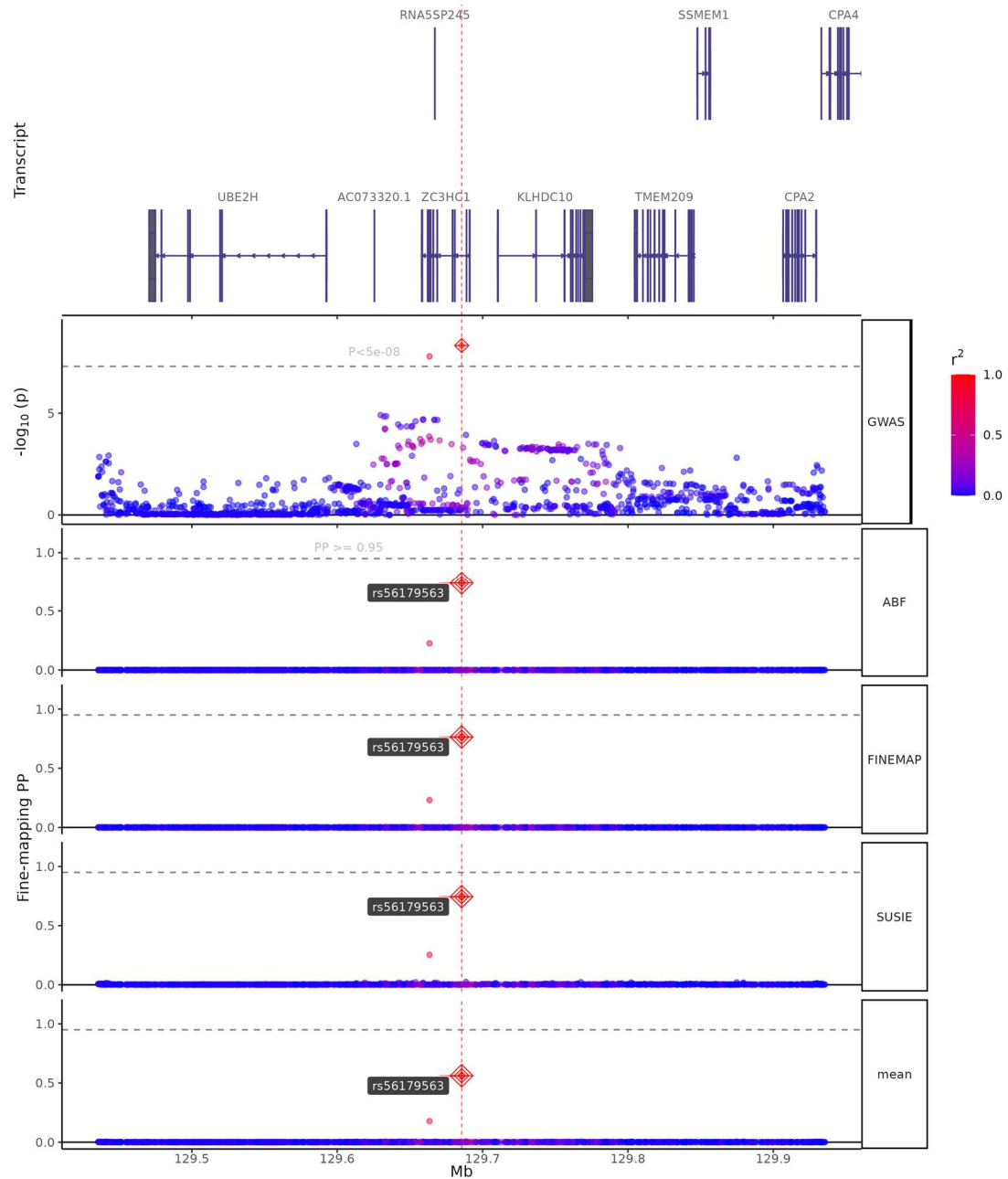
Locus: NOS3 (SNPs=1,399; zoom=1x)



Supplementary Fig. 37 Fine-mapping Results of NOS3

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

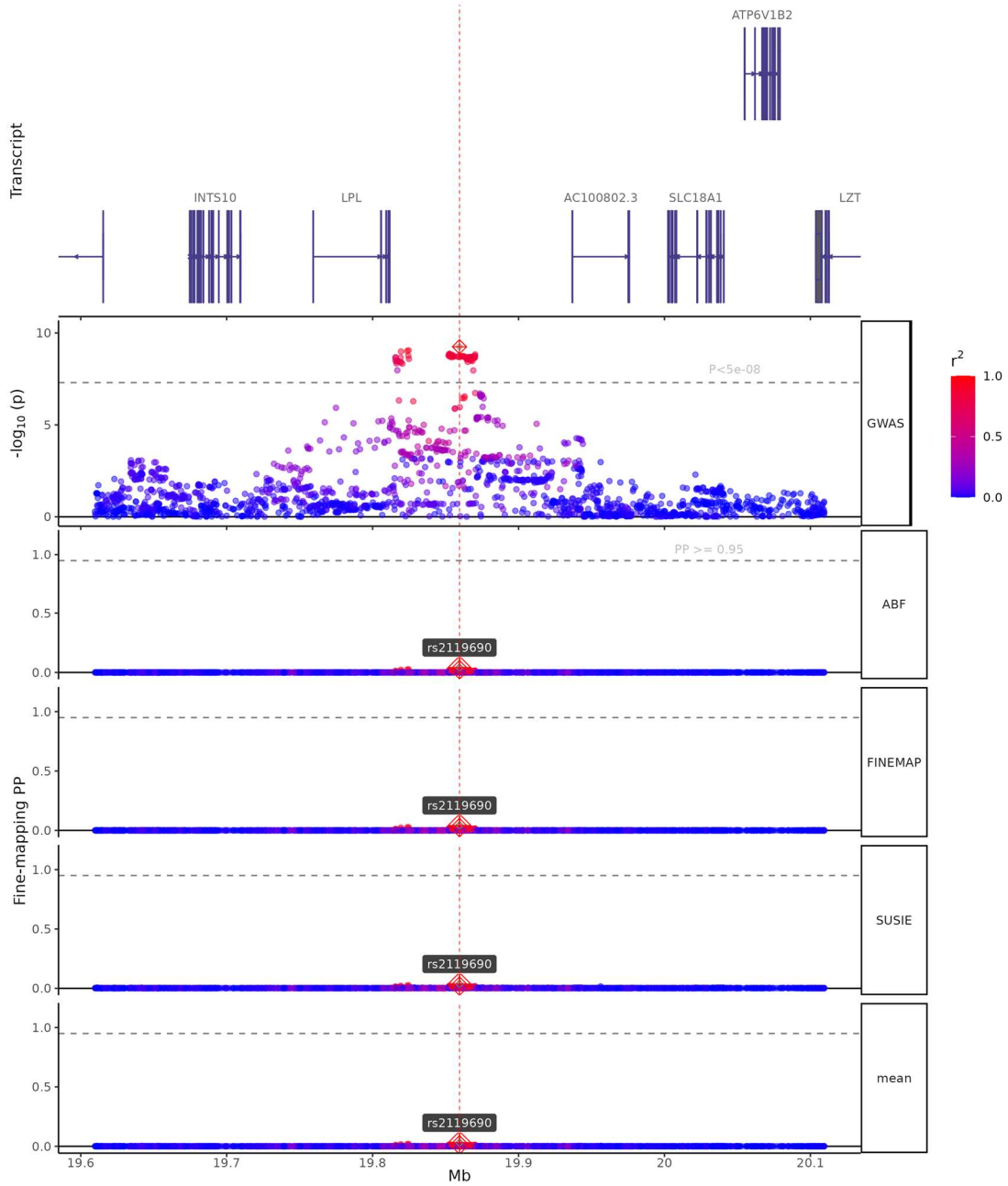
Locus: ZC3HC1 (SNPs=1,357; zoom=1x)



Supplementary Fig. 38 Fine-mapping Results of ZC3HC1

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

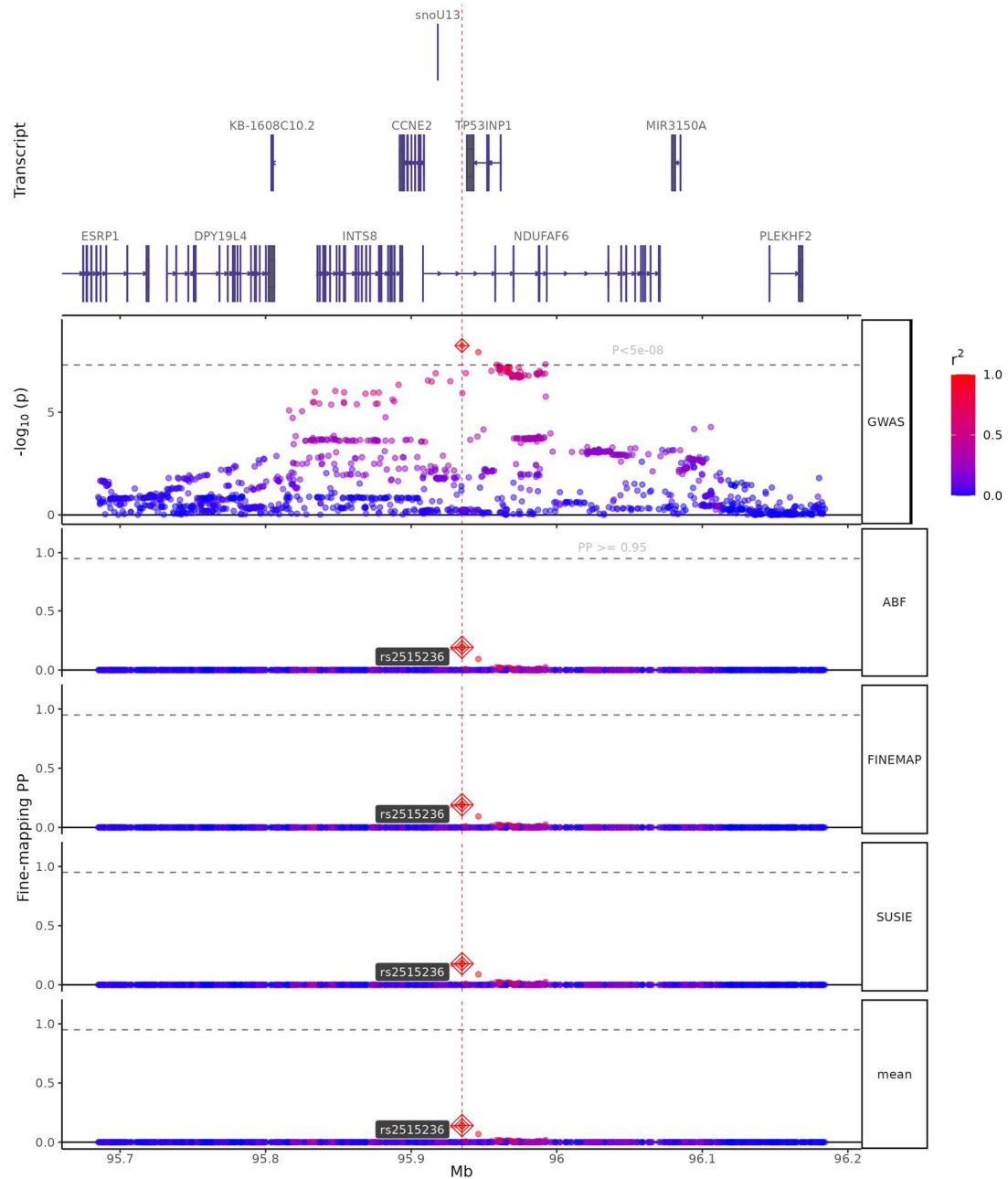
Locus: LPL (SNPs=1,869; zoom=1x)



Supplementary Fig. 39 Fine-mapping Results of LPL

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

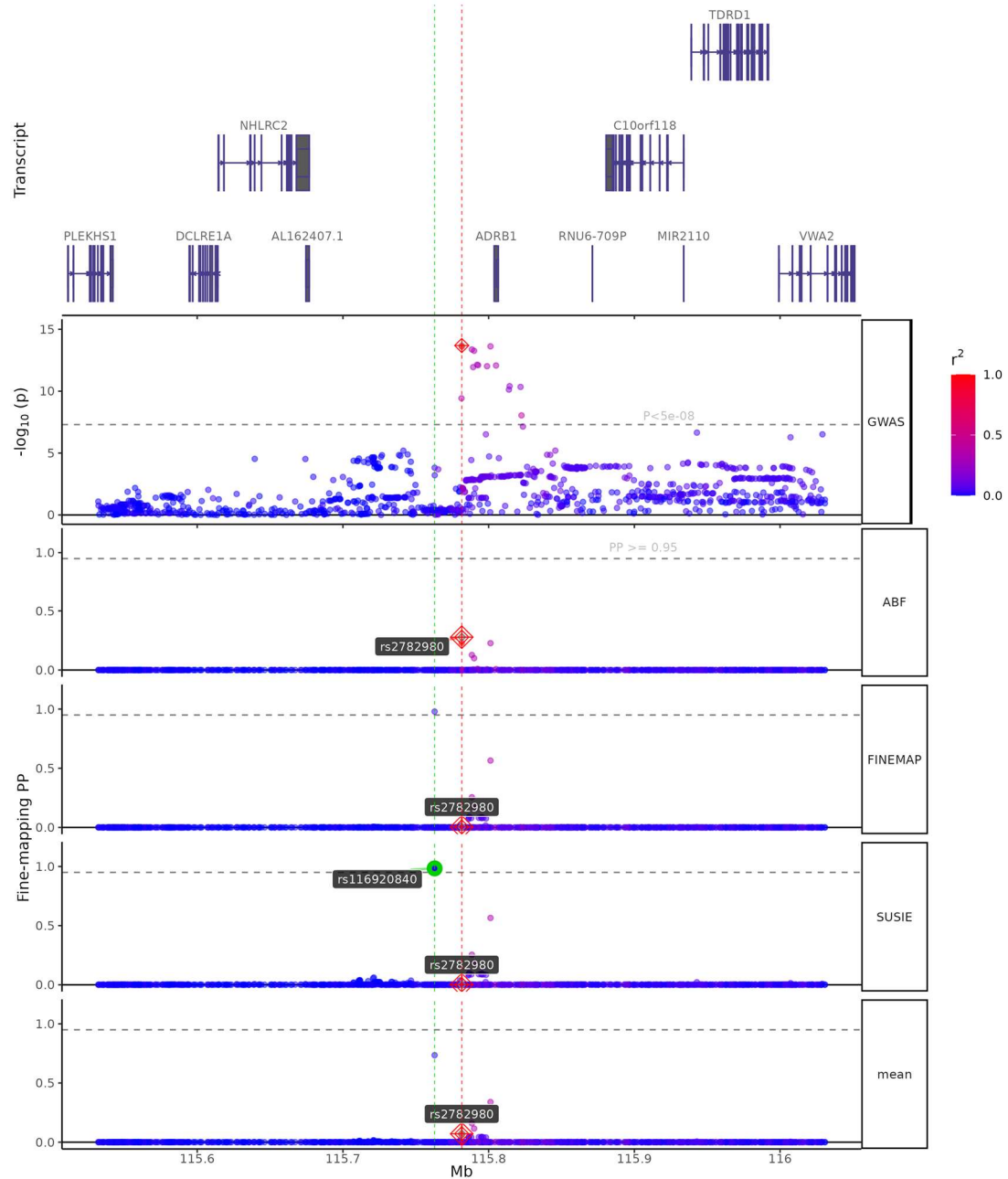
Locus: NDUFAF6 (SNPs=1,387; zoom=1x)



Supplementary Fig. 40 Fine-mapping Results of NDUFAF6

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

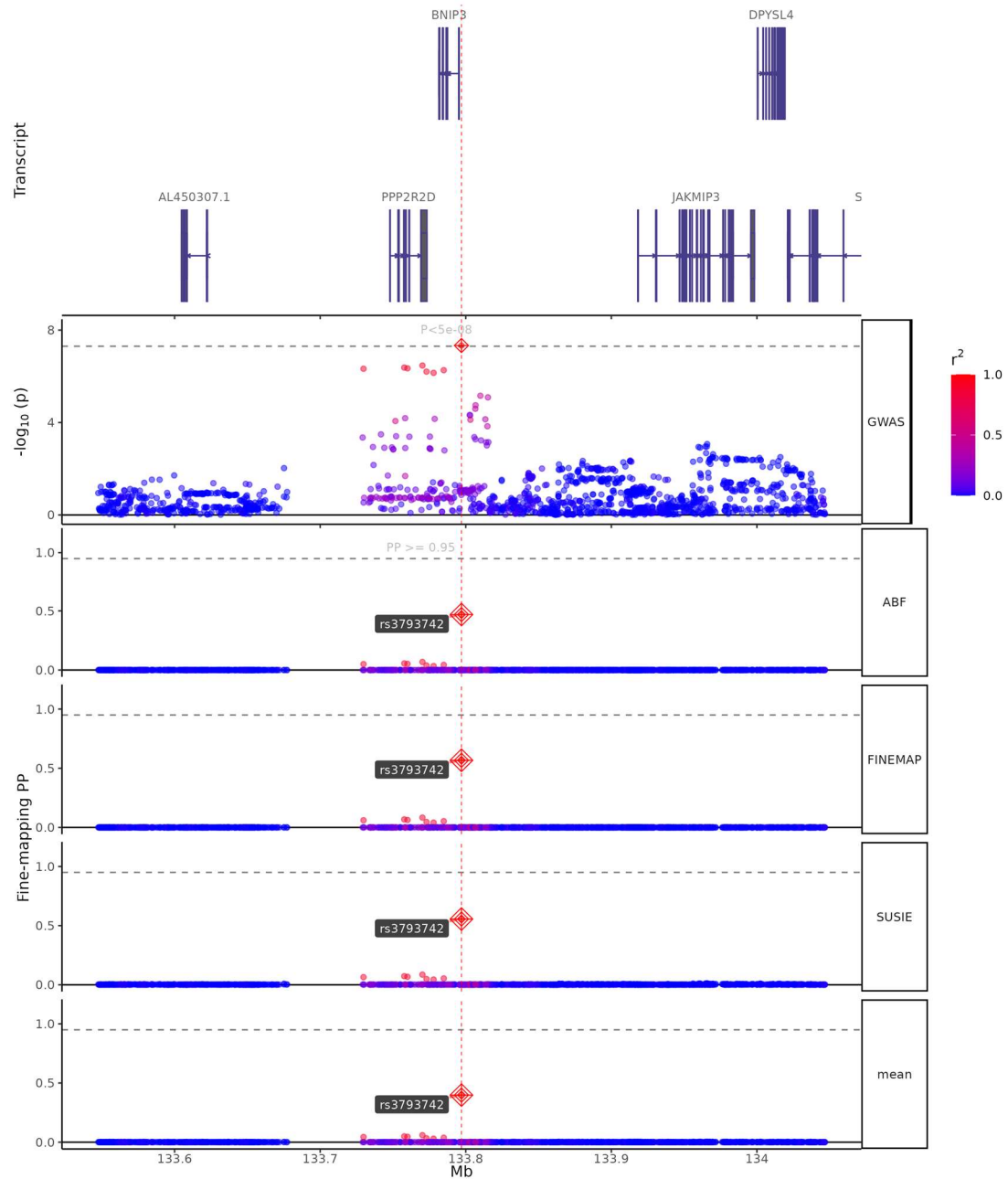
Locus: ADRB1 (SNPs=1,159; zoom=1x)



Supplementary Fig. 41 Fine-mapping Results of ADRB1

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

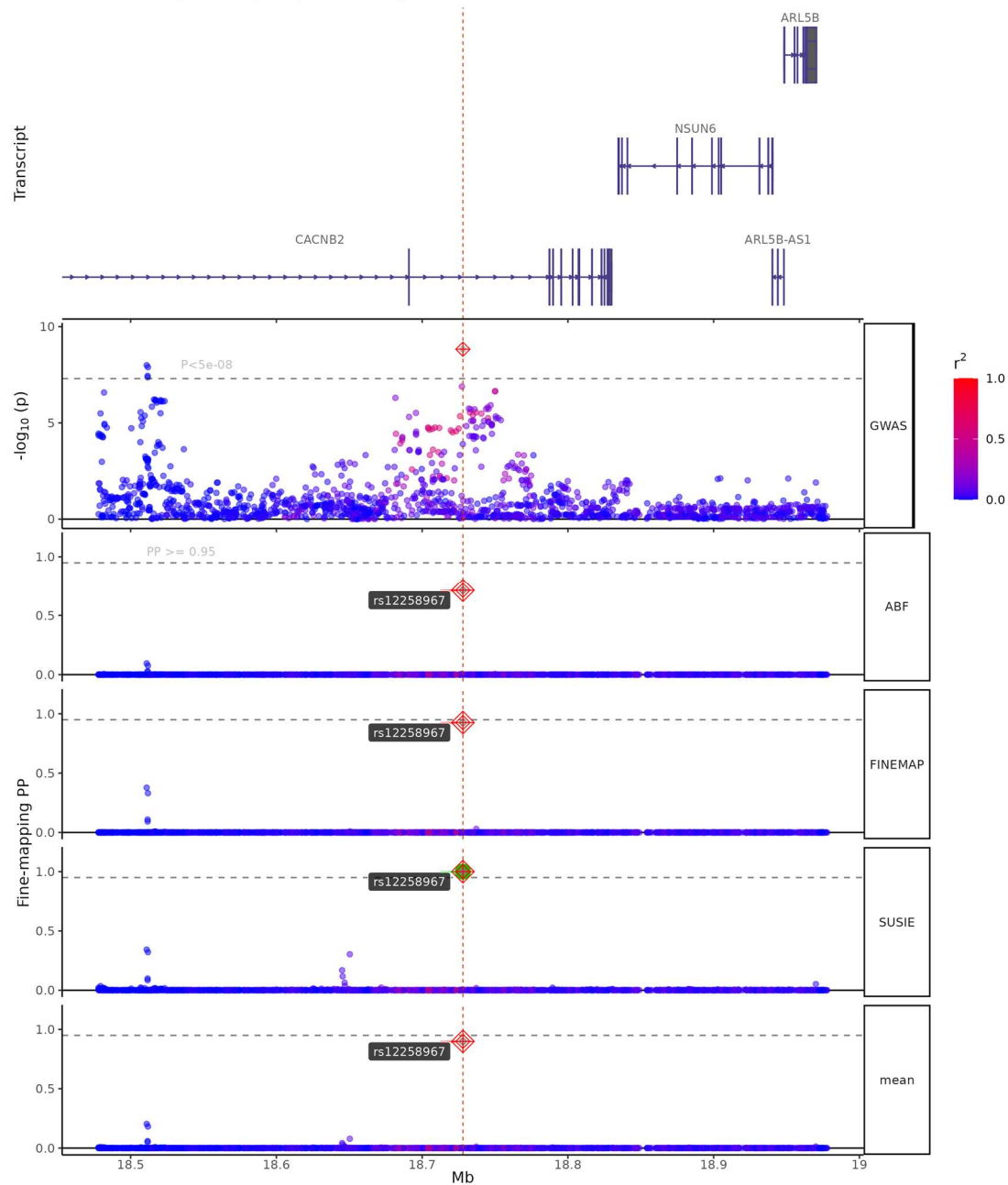
Locus: BNIP3 (SNPs=1,622; zoom=1x)



Supplementary Fig. 42 Fine-mapping Results of BNIP3

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

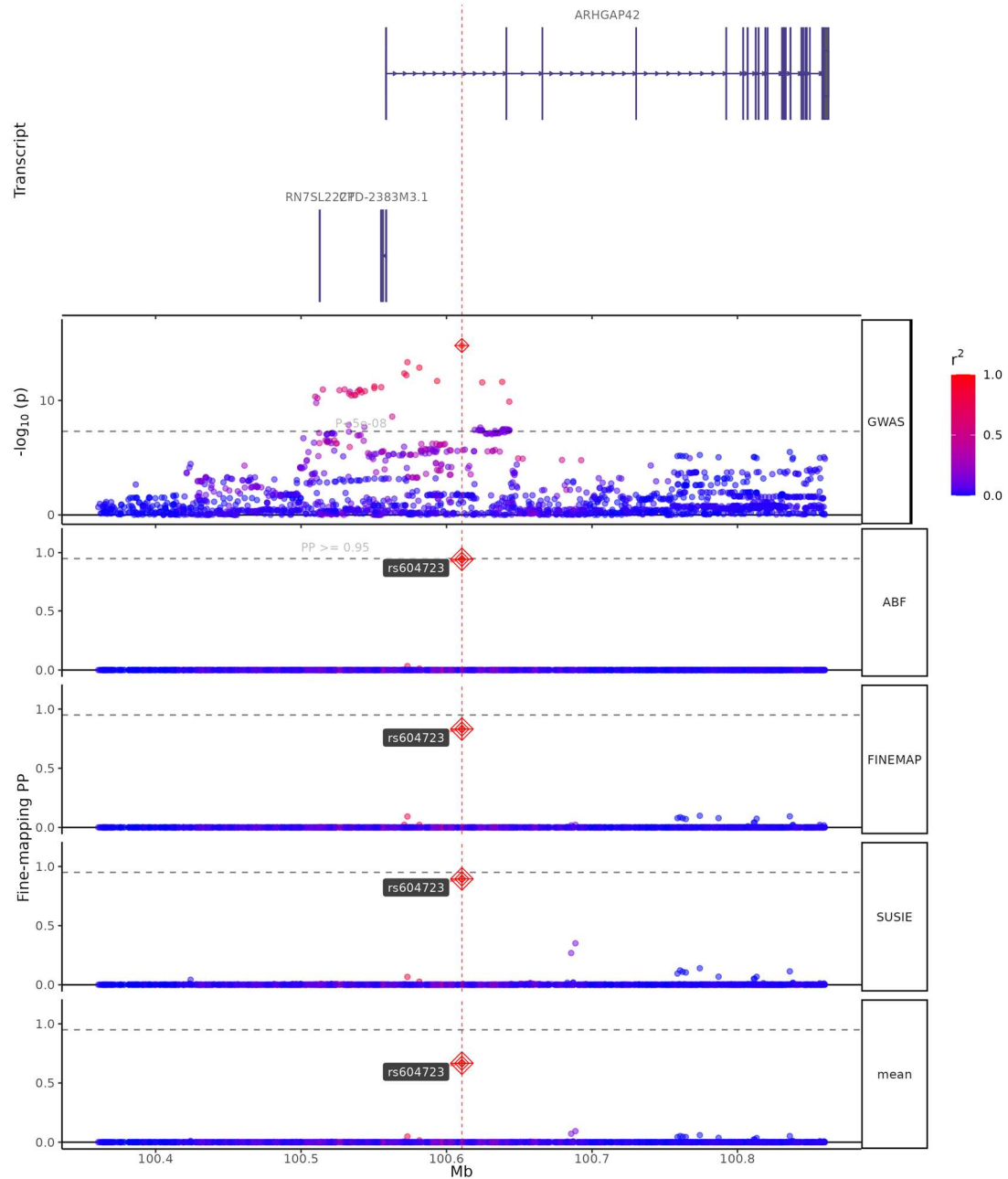
Locus: CACNB2 (SNPs=1,688; zoom=1x)



Supplementary Fig. 43 Fine-mapping Results of CACNB2

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

Locus: ARHGAP42 (SNPs=1,893; zoom=1x)

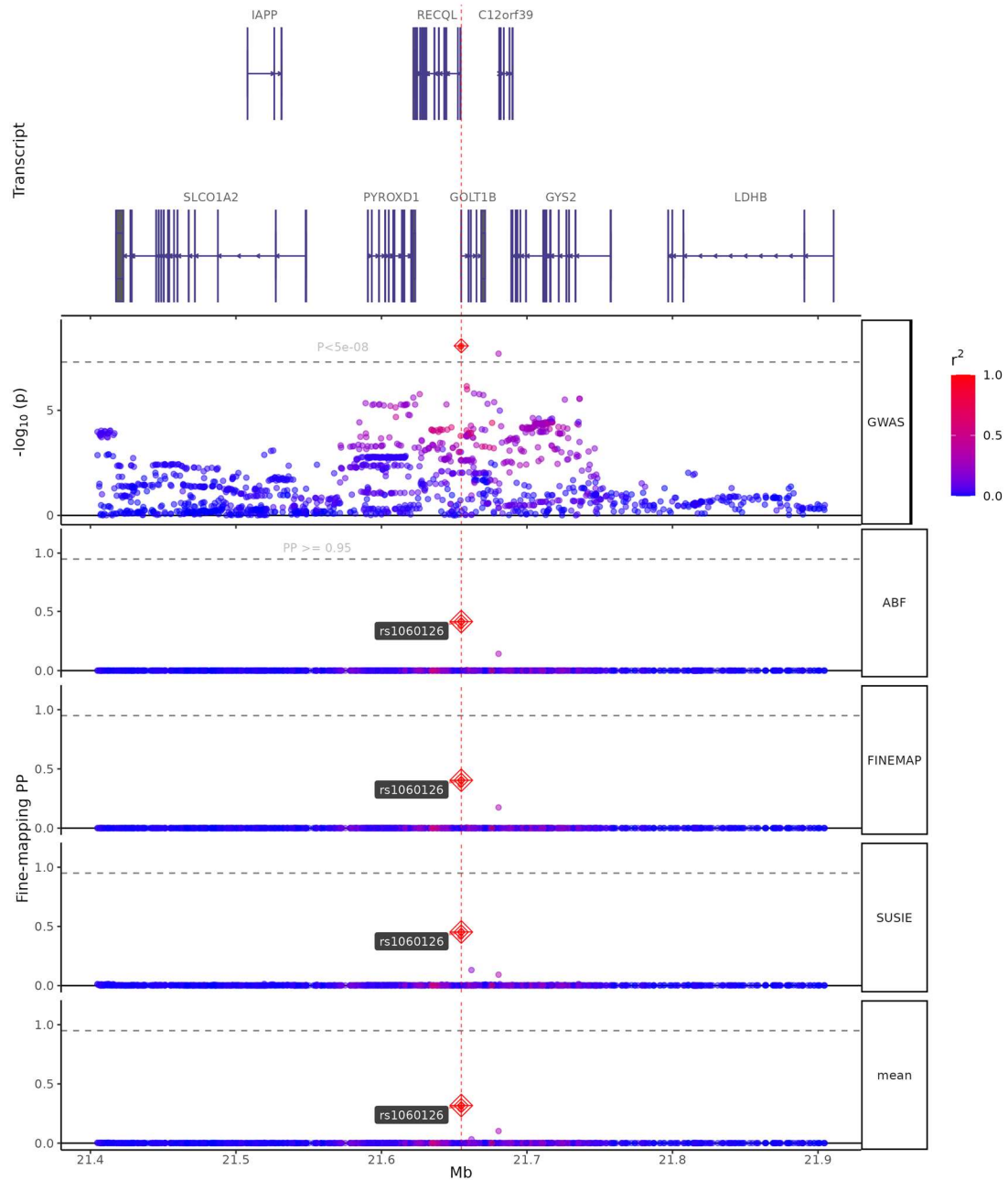


Supplementary Fig. 44 Fine-mapping Results of ARHGAP42

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

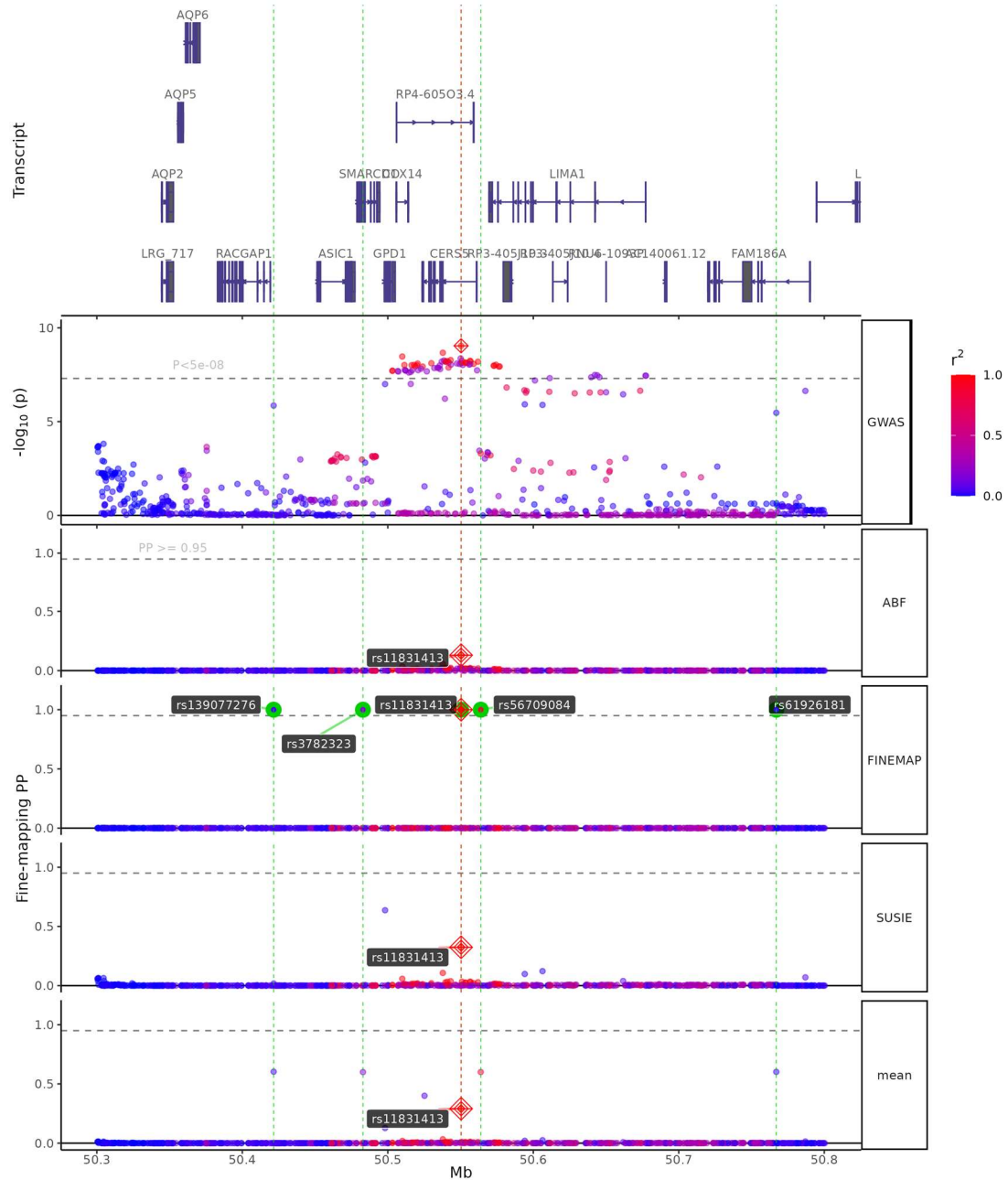
Locus: GOLT1B (SNPs=1,526; zoom=1x)



Supplementary Fig. 46 Fine-mapping Results of GOLT1B

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

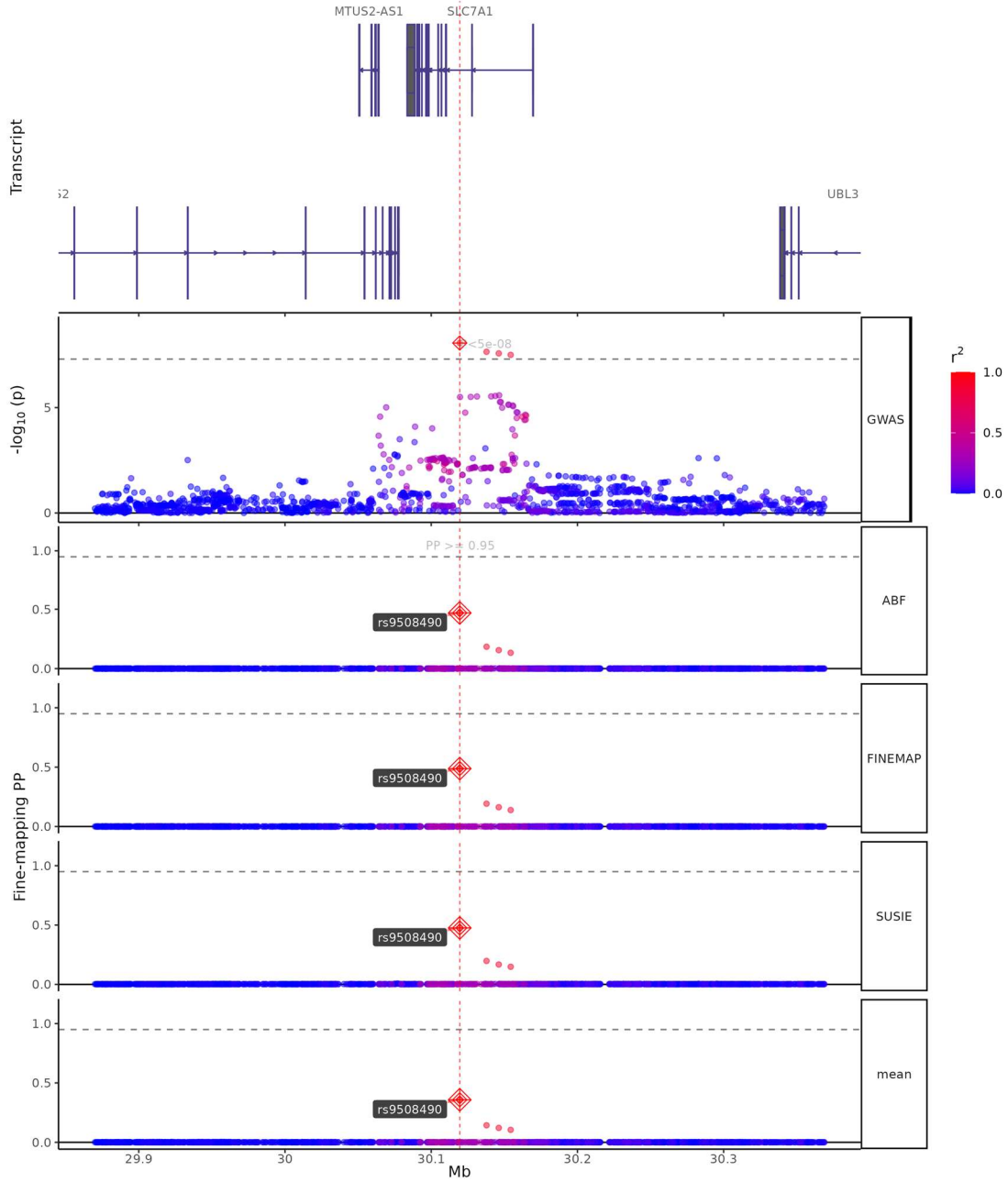
Locus: RP4-605O3.4_CERS5 (SNPs=760; zoom=1x)



Supplementary Fig. 47 Fine-mapping Results of RP4-605O3.4_CERS5

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

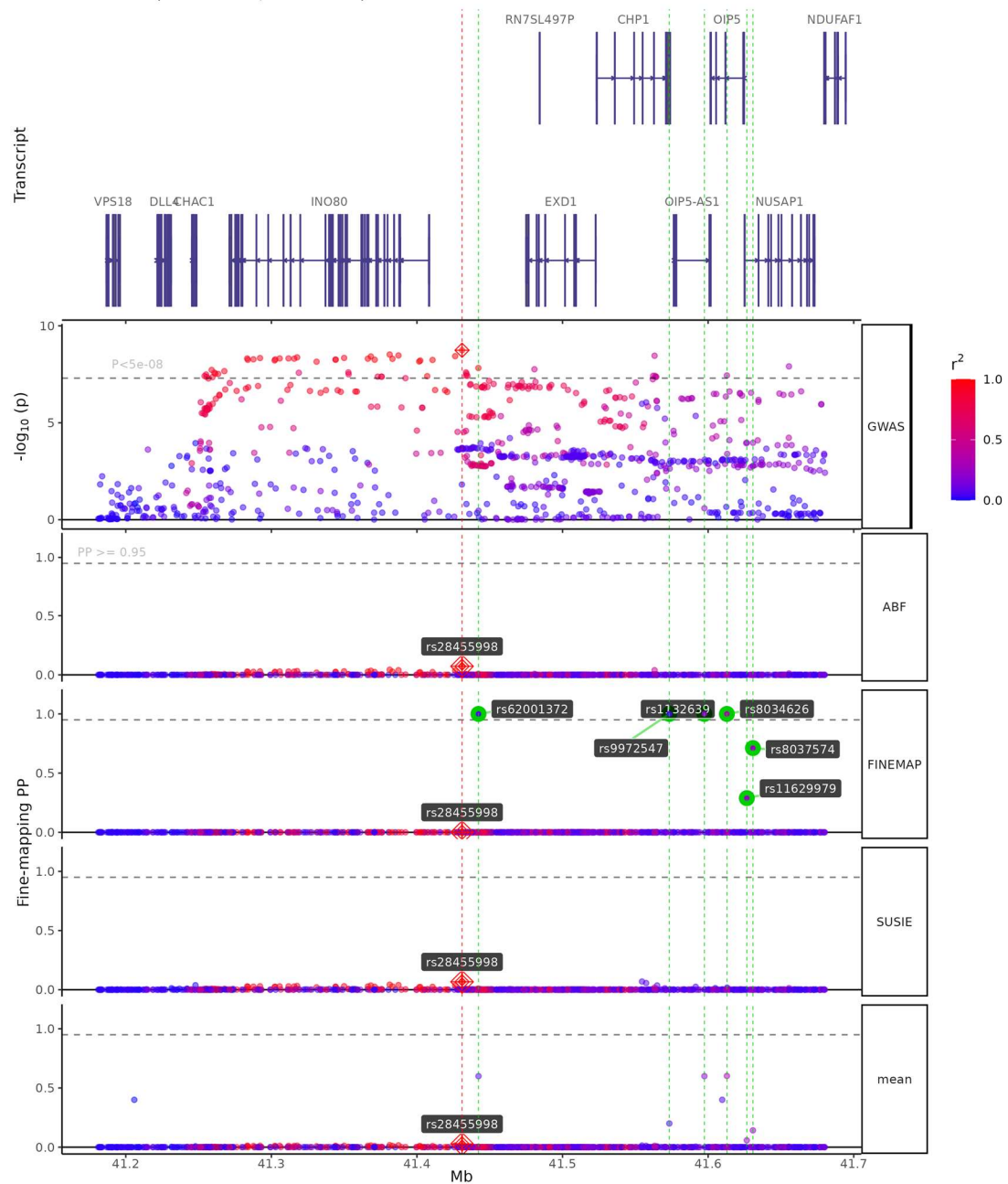
Locus: SLC7A1 (SNPs=1,673; zoom=1x)



Supplementary Fig. 48 Fine-mapping Results of SLC7A1

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

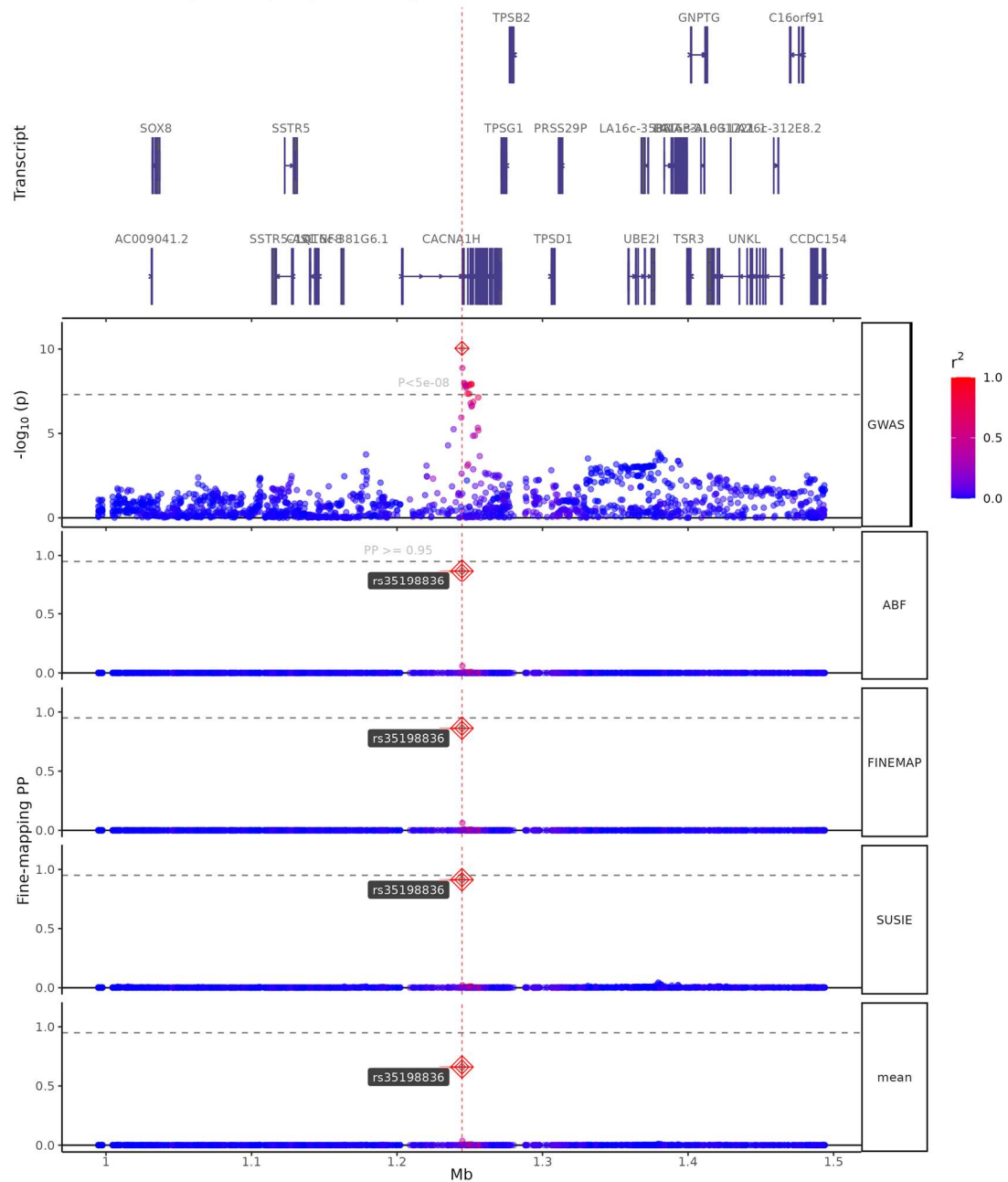
Locus: INO80 (SNPs=927; zoom=1x)



Supplementary Fig. 49 Fine-mapping Results of INO80

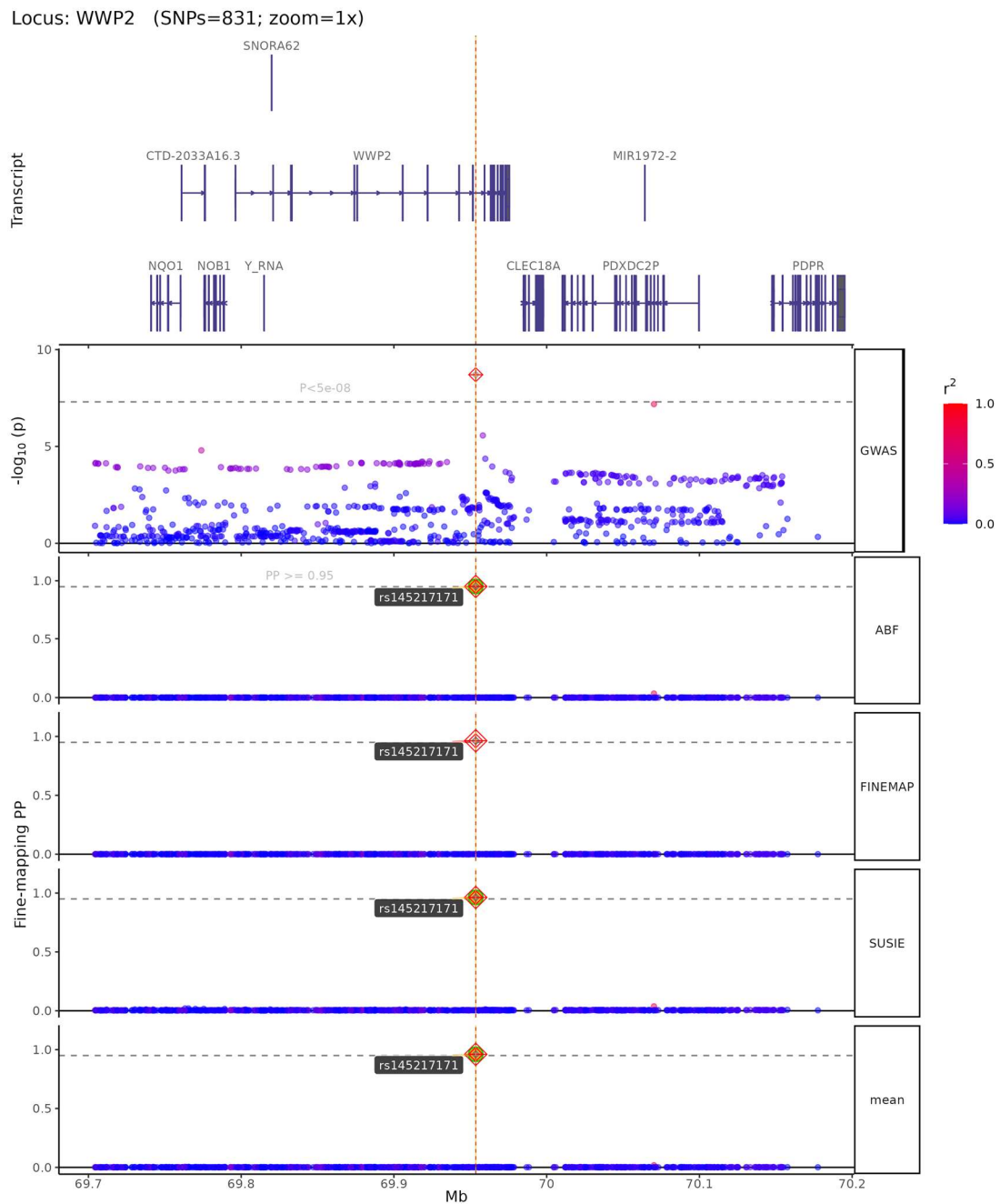
The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

Locus: CACNA1H (SNPs=1,746; zoom=1x)



Supplementary Fig. 50 Fine-mapping Results of CACNA1H

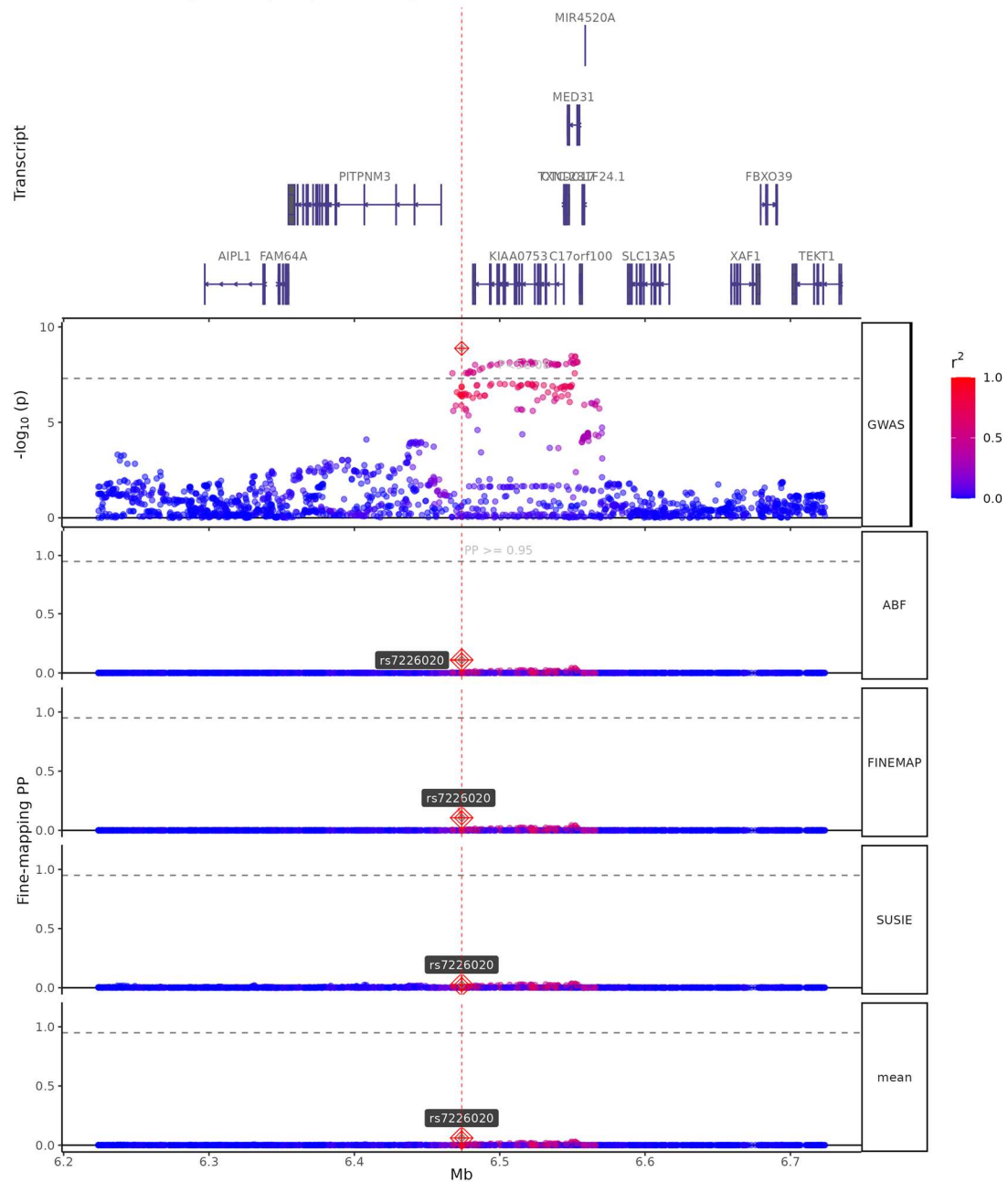
The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.



Supplementary Fig. 51 Fine-mapping Results of WWP2

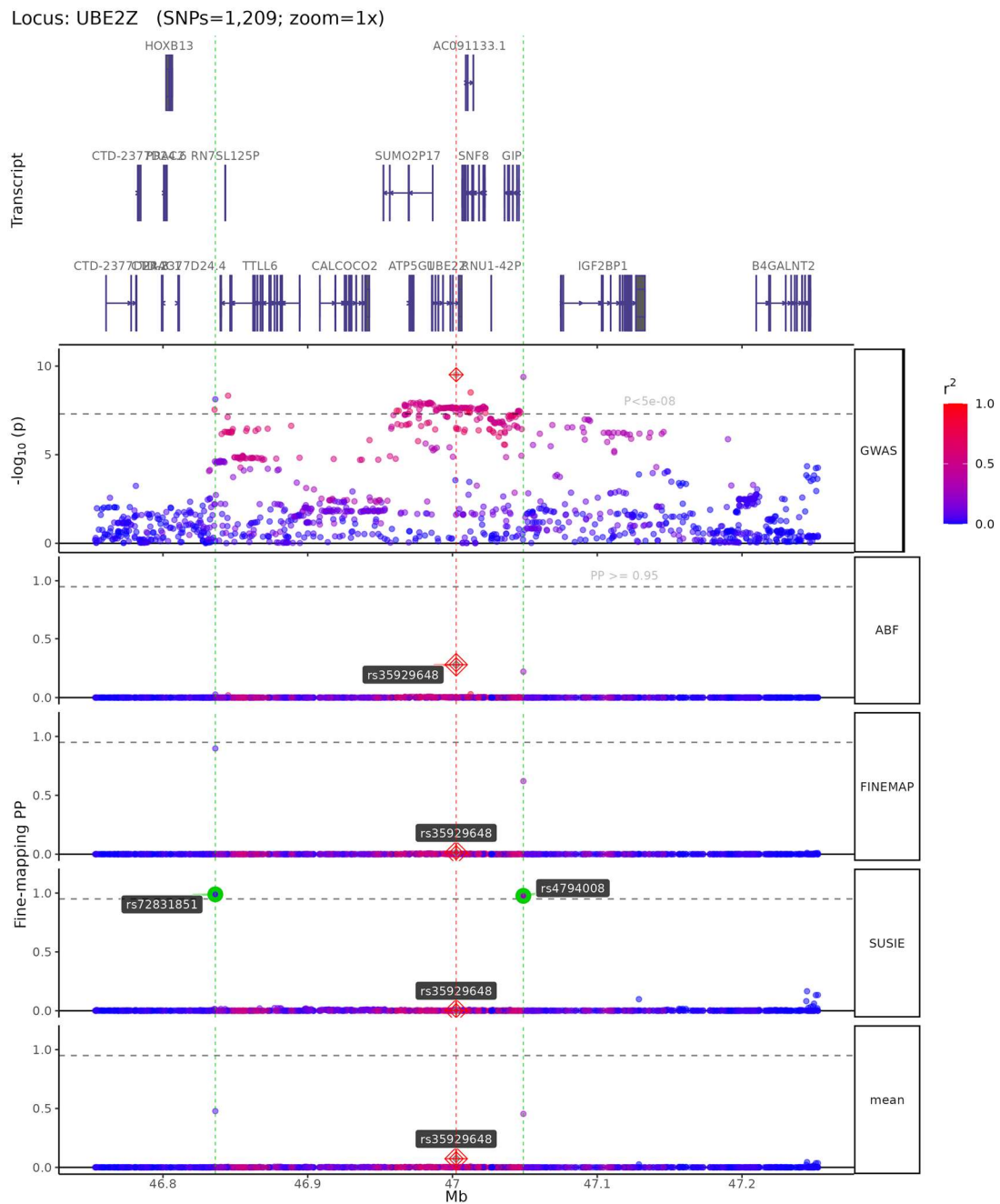
The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

Locus: KIAA0753 (SNPs=1,724; zoom=1x)



Supplementary Fig. 52 Fine-mapping Results of KIAA0753

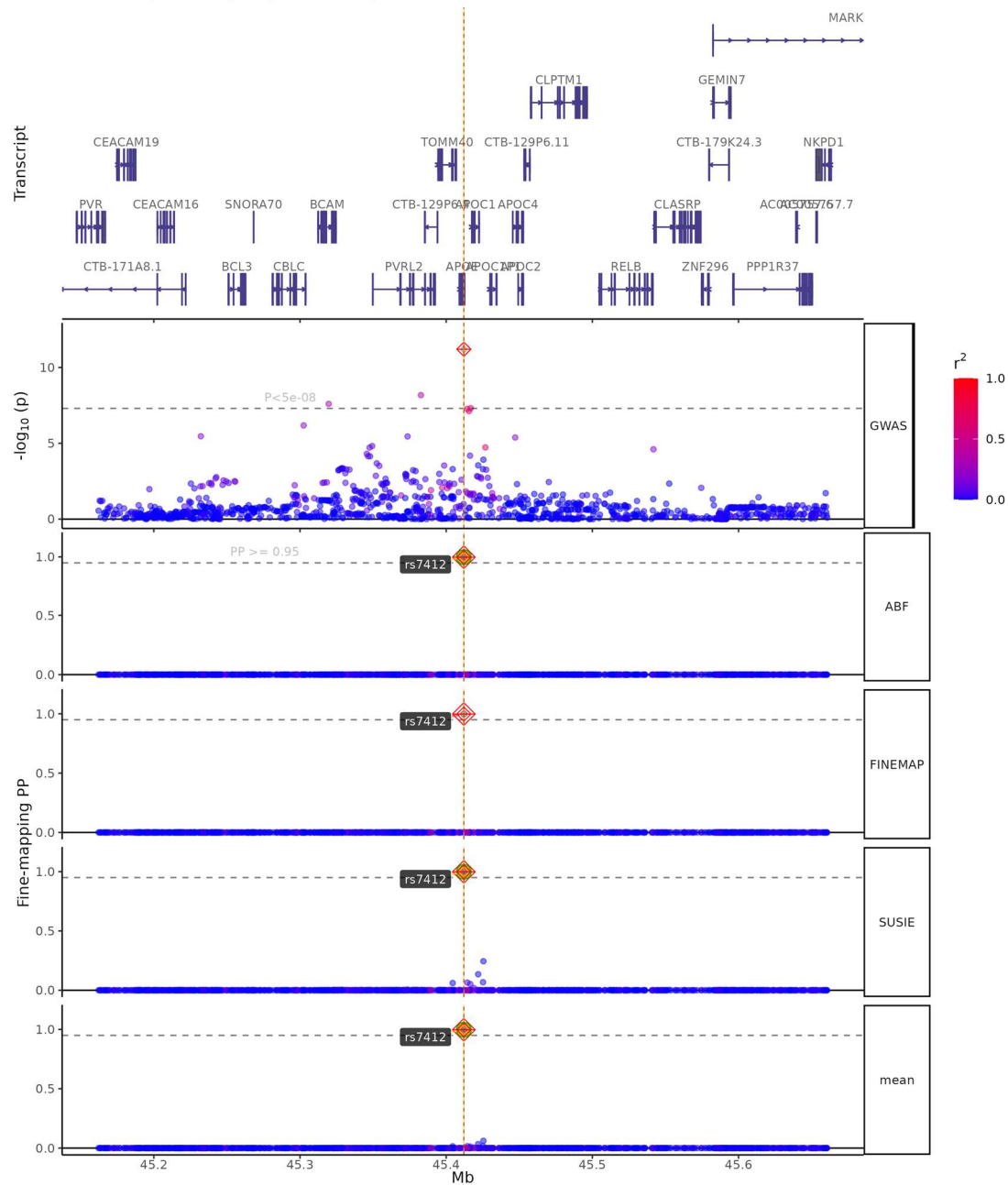
The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P)$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.



Supplementary Fig. 53 Fine-mapping Results of UBE2Z

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

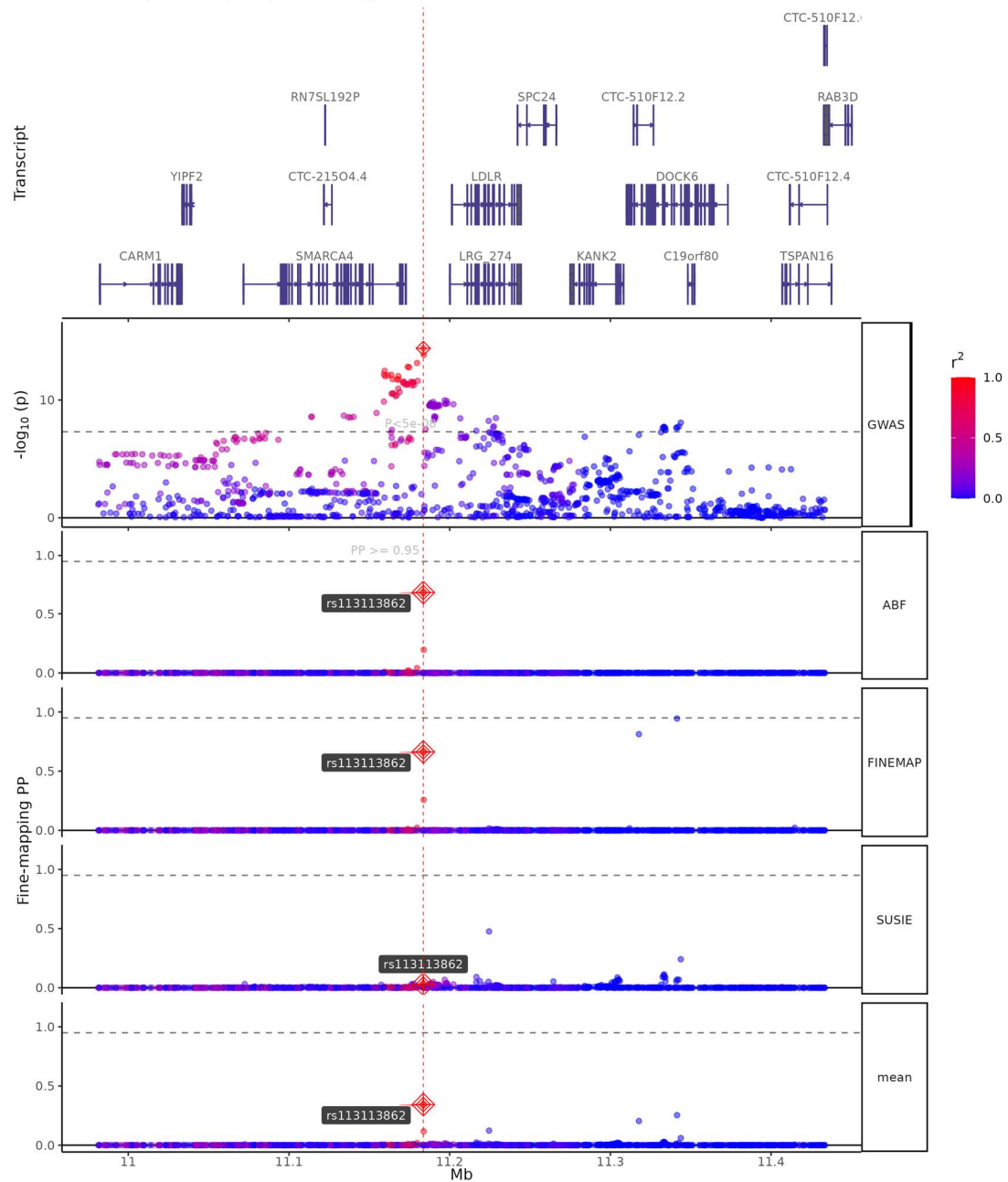
Locus: APOE (SNPs=1,323; zoom=1x)



Supplementary Fig. 54 Fine-mapping Results of APOE

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

Locus: LDLR (SNPs=1,200; zoom=1x)



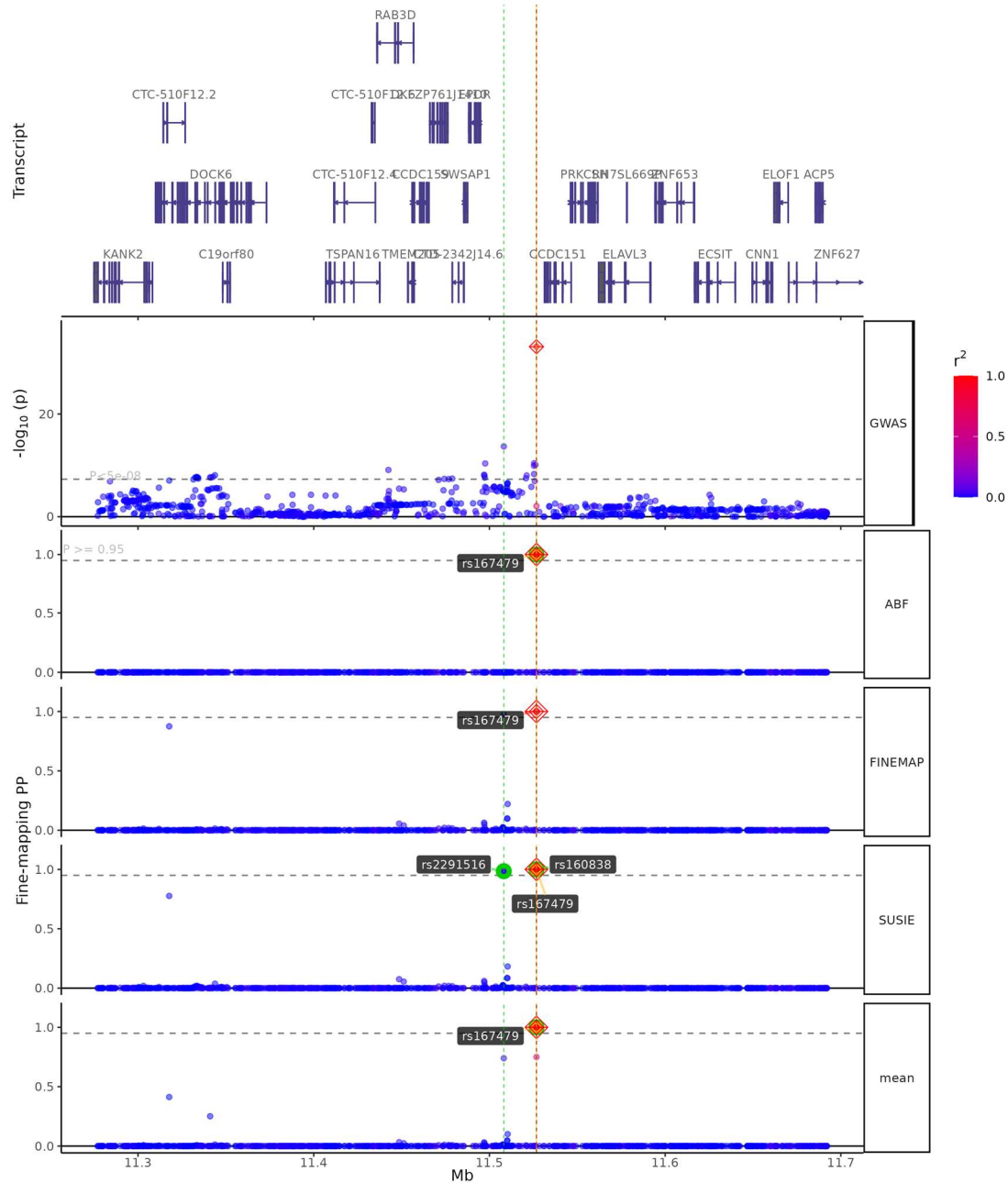
Supplementary Fig. 55 Fine-mapping Results of LDLR

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

[illegible]

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

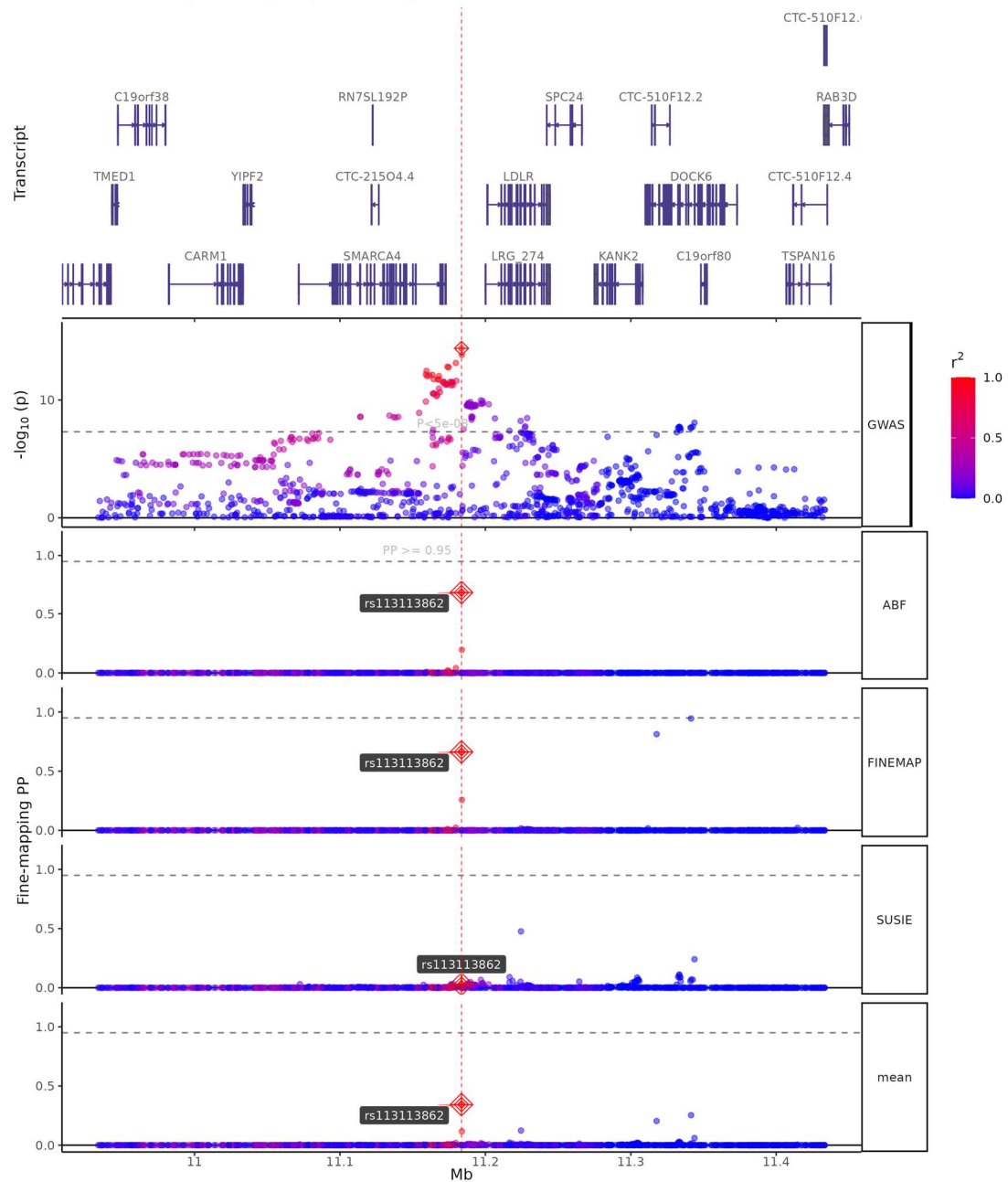
Locus: RAB3D (SNPs=978; zoom=1x)



Supplementary Fig. 57 Fine-mapping Results of RAB3D

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(\text{P-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

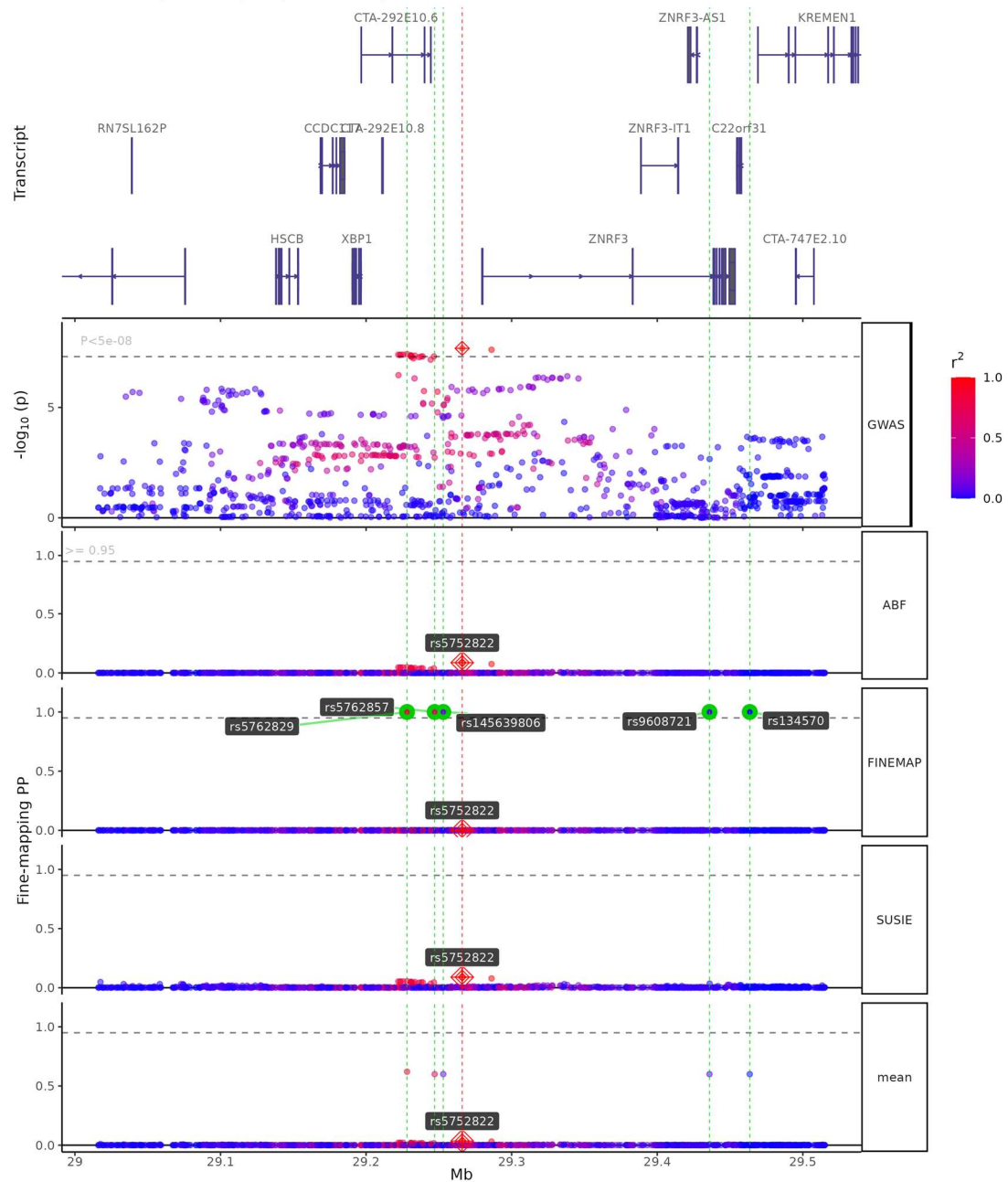
Locus: SMARCA4 (SNPs=1,260; zoom=1x)



Supplementary Fig. 58 Fine-mapping Results of SMARCA4

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

Locus: ZNRF3 (SNPs=1,074; zoom=1x)



Supplementary Fig. 59 Fine-mapping Results of ZNRF3

The top panel displays genes located within 250 kilobases of the lead variant. The second panel displays the $-\log_{10}(P\text{-value})$ results of two-sided Wald tests for each variant on mvCMD and LD R^2 information in the locus (variants are colored by LD R^2). The third panel shows the fine-mapping results of the ABF method. The fourth panel displays the fine-mapping results from the FINEMAP method. The fifth panel exhibits the fine-mapping results of the SuSIE methods. A vertical red line indicates the location of the GWAS lead SNP. See supplementary methods for further details.

References

1. Cheung MW. metaSEM: an R package for meta-analysis using structural equation modeling. *Front Psychol*. 2014;5:1521.
2. Savalei V, Bentler PM. A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal*. 2009;16(3):477-497.
3. Yuan K-H, Bentler PM. Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika*. 2000;65:43-58.
4. Grotzinger AD, Rhemtulla M, de Vlaming R, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav*. 2019;3(5):513-525.
5. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
6. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*. 2003;8(2):23-74.
7. Karlsson Linnér R, Mallard TT, Barr PB, et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat Neurosci*. 2021;24(10):1367-1376.
8. Rosoff DB, Mavromatis LA, Bell AS, et al. Multivariate genome-wide analysis of aging-related traits identifies novel loci and new drug targets for healthy aging. *Nat Aging*. 2023;3(8):1020-1035.
9. Waterworth DM, Ricketts SL, Song K, et al. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol*. 2010;30(11):2264-2276.
10. Reyes-Soffer G, Ginsberg HN, Berglund L, et al. Lipoprotein(a): A Genetically Determined, Causal, and Prevalent Risk Factor for Atherosclerotic Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Arterioscler Thromb Vasc Biol*. 2022;42(1):48-60.
11. Ottensmann L, Tabassum R, Ruotsalainen SE, et al. Genome-wide association analysis of plasma lipidome identifies 495 genetic associations. *Nat Commun*. 2023;14(1):6934.
12. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol*. 2020;82(5):1273-1300.
13. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet*. 2007;81(2):208-227.
14. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*. 2016;32(10):1493-1501.
15. Schilder BM, Humphrey J, Raj T. echolocator: an automated end-to-end statistical and functional genomic fine-mapping pipeline. *Bioinformatics*. 2022;38(2):536-539.
16. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48(3):245-252.

17. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383.
18. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):89-98.
19. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj.* 2018;362:k601.
20. Zeljkovic A, Vekic J, Stefanovic A. Obesity and dyslipidemia in early life: Impact on cardiometabolic risk. *Metabolism.* 2024;156:155919.
21. Jiang M, Tian S, Liu S, et al. Accelerated biological aging elevates the risk of cardiometabolic multimorbidity and mortality. *Nat Cardiovasc Res.* 2024;3(3):332-342.
22. Stitzel NO. Human genetic insights into lipoproteins and risk of cardiometabolic disease. *Curr Opin Lipidol.* 2017;28(2):113-119.
23. Deprince A, Haas JT, Staels B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. *Mol Metab.* 2020;42:101092.
24. Islam MA, Amin MN, Siddiqui SA, Hossain MP, Sultana F, Kabir MR. Trans fatty acids and lipid profile: A serious risk factor to cardiovascular disease, cancer and diabetes. *Diabetes Metab Syndr.* 2019;13(2):1643-1647.
25. Honigberg MC, Zekavat SM, Pirruccello JP, Natarajan P, Vaduganathan M. Cardiovascular and Kidney Outcomes Across the Glycemic Spectrum: Insights From the UK Biobank. *J Am Coll Cardiol.* 2021;78(5):453-464.
26. Chen J, Yin D, Dou K. Intensified glycemic control by HbA1c for patients with coronary heart disease and Type 2 diabetes: a review of findings and conclusions. *Cardiovasc Diabetol.* 2023;22(1):146.
27. DeBerge M, Chaudhary R, Schroth S, Thorp EB. Immunometabolism at the Heart of Cardiovascular Disease. *JACC Basic Transl Sci.* 2023;8(7):884-904.
28. Stefan N, Häring HU, Cusi K. Non-alcoholic fatty liver disease: causes, diagnosis, cardiometabolic consequences, and treatment strategies. *Lancet Diabetes Endocrinol.* 2019;7(4):313-324.
29. Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med.* 2020;17(3):e1003062.
30. Chaudhury A, Duvoor C, Reddy Dendi VS, et al. Clinical Review of Antidiabetic Drugs: Implications for Type 2 Diabetes Mellitus Management. *Front Endocrinol (Lausanne).* 2017;8:6.
31. Davies M, Nowotka M, Papadatos G, et al. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* 2015;43(W1):612-620.
32. Zdrazil B, Felix E, Hunter F, et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* 2024;52(D1):1180-1192.
33. Ference BA, Kastelein JJP, Ray KK, et al. Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. *Jama.* 2019;321(4):364-373.
34. Burgess S, Thompson SG. Avoiding bias from weak instruments in Mendelian

- randomization studies. *Int J Epidemiol.* 2011;40(3):755-764.
35. Schmidt AF, Finan C, Gordillo-Marañón M, et al. Genetic drug target validation using Mendelian randomisation. *Nat Commun.* 2020;11(1):3255.