

Disentangling genetic risks for metabolic syndrome

Authors

Eva S. van Walree^{1,2}, Iris E. Jansen², Nathaniel Y. Bell², Jeanne E. Savage², Christiaan de Leeuw², Max Nieuwdorp³, Sophie van der Sluis⁴, Danielle Posthuma^{2,4}

¹ Department of Clinical Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

² Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU Amsterdam, Amsterdam, the Netherlands.

³ Department of Internal and Vascular Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

⁴ Department of Child and Adolescent Psychology and Psychiatry, section Complex Trait Genetics, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, The Netherlands.

Table of contents:

1. Supplementary information
 - 1A. Structural equation modelling
 - 1B. Is metabolic syndrome truly a syndrome?
2. Supplementary methods
 - 2A. GWAS summary statistics of indicator phenotypes included in this study, quality control and factor analyses
 - 2B. Cross-trait genetic correlations
 - 2C. Identifying independent and unique loci (FUMA)
 - 2D. Functional annotation (FUMA)
 - 2E. Gene-based analysis (MAGMA)
 - 2F. Cell type analyses
 - 2G. Correcting for sample overlap for PRS analysis
 - 2H. Framingham Heart Study quality control
 - 2I. Drug gene set analysis
3. Supplementary results
 - 3A. Common factor GWAS without WC
 - 3B. Additional drug gene-set (repurposing) results
4. Supplementary figures
 - 4A. Exploratory factor analysis scree plot (R nFactors package)
 - 4B. MAGMA gene-based Manhattan plot
 - 4C. SNP-based QQ plot
 - 4D. Gene-based QQ plot

1. Supplementary information

1A. Structural equation modelling

Factor analysis assumes that the correlation between observed indicators (in this case, the MetS components) to be due to one or more unmeasured ‘latent’ constructs or factors. Exploratory factor analysis investigates how many latent constructs these indicators can be condensed into. In a path model a hypothesized structural model reflecting relationships between latent factors and their indicators can be visualized. This model can then be tested mathematically with a confirmatory factor analysis, which tests how well this model fits the data and can compute factor loadings (standardized regression coefficients representing the variance of an indicator explained by the latent factor).

1B. Is metabolic syndrome truly a syndrome?

Whether MetS is truly a syndrome, or rather an observed clustering of independent risk factors is under debate. A syndrome is ‘a group of signs and symptoms that occur together and characterize a particular abnormality or condition’⁷¹. The former, symptoms that occur together, is true for MetS. Yet, they do not characterize one particular abnormality.⁷² Also, we have shown that the genetic overlap between these drivers of MetS is modest, and their functional effects show different hierarchy. However, the PRS results presented in Figure 5 support the idea that the whole is greater than the sum of its parts. Furthermore, using the term MetS raises awareness for the comorbidity between the individual symptoms. The syndrome also has clinical implications, as the presence of one risk factor warrants the investigation to other risk factors, and emphasizes the need to properly control metabolic disturbances through life style changes and drug treatment.

2. Supplementary methods

2A. GWAS summary statistics of indicator phenotypes included in this study

Publicly available GWAS summary statistics of fasting glucose (FG), high-density lipoprotein cholesterol (HDL-C), systolic blood pressure (SBP), triglycerides (TG) and waist circumference (WC) were used. Since SBP and DBP are highly correlated genetically (r_G 0.81), we choose to include only SBP and not both SBP and DBP.³²

Fasting glucose

Fasting glucose (FG) was measured in 46,186 nondiabetic participants of European descent in a study by Dupuis et al.¹ Individuals with a fasting glucose ≥ 7 mmol/l, with known diabetes or on anti-diabetic treatment were excluded. Fasting glucose was measured from whole blood, serum or plasma, or a combination of those. Genetic association analyses were corrected for sex, age, study site (if applicable), geographical covariates (if applicable) and age squared (if applicable).

HDL-C and triglycerides

Both high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG) were measured in a study by Teslovich et al., in 99,000 and 96,598 European individuals, respectively.² Individuals that were on lipid-lowering therapy were excluded. Genetic association analyses were corrected for sex, age and age squared. Additional covariates were added on a per-cohort basis.

Systolic blood pressure and waist circumference

The GWAS on systolic blood pressure (SBP) and waist circumference (WC) were done in UK Biobank by Watanabe et al. on 361,402 and 385,932 European individuals, respectively.³ Genetic association analyses were corrected for age, sex, the first 20 principal components, array, assessment centre, and Townsend deprivation index (a proxy for socioeconomic status).

In the first stage, the genetic covariance matrix, derived from LDSC, and sampling matrix for the five MetS components were estimated in genomic SEM. Quality control for this step consisted of removing SNPs with a minor allele frequency (MAF) <1% (when available), INFO-score <0.9 (when available), SNPs from the MHC region, and SNPs not present in HapMap3. MAF and INFO-scores were not always available in the individual GWAS summary statistics, but filtering SNPs to HapMap3 should ensure a set of relatively common SNPs of good quality.

First, an exploratory factor analysis (EFA) was performed to investigate how many factors were needed to describe the observed genetic covariance matrix between the five MetS components. For this EFA, promax rotation was used in the R *factanal* package⁴. A scree plot was generated with the R *nFactors* package.⁵ As a 1-factor model was suggested, a confirmatory factor analysis (CFA) was run in genomic SEM to establish how well this 1-factor model fitted the data, using the default diagonally weighted least square estimation. Model fit, which evaluates to which extent the model implied covariance matrix approximates the empirical, observed covariance matrix, is considered good with CFI values >0.95.⁶ Furthermore, SRMR values <0.10 are considered acceptable fit and <0.05 good fit.⁷

2B. Cross-trait genetic correlations

Genetic correlations (r_G) between 1) the MetS components, 2) the MetS factor and the largest GWAS on MetS by Lind¹⁵, 3) the MetS factor and diseases assumed to be phenotypically associated with MetS in literature, and 4) the MetS factor and a broad spectrum of diseases that were selected to reflect human health in a broad sense were estimated with LDSC.²⁰ Studies with a heritability z-score <4 (calculated as LDSC-derived h^2 divided by SE) were removed, as those are considered too noisy to generate robust genetic correlations (Supplementary Table 2).²¹ LD scores from the European cohort of the 1000 Genomes study Phase 3 (URLS) were used for analyses with LDSC and genomic SEM (see below).

2C. Identifying independent and unique loci with FUMA

The *SNP2GENE* function was used to identify near-independent ($r^2 < 0.6$) significant ($P \leq 5 \times 10^{-8}$) SNPs, as well as independent ($r^2 < 0.1$) significant lead SNPs. All SNPs available in either the individual GWAS or the 1000 Genomes European reference panel that were in linkage

disequilibrium of $r^2 \geq 0.6$ with a lead SNP (candidate SNPs) were then selected for further annotation. Genomic risk loci were defined by assigning SNPs which are dependent on each other at $r^2 \geq 0.1$ to the same genomic risk locus, and subsequently merging any physically close SNPs (linkage disequilibrium (LD) blocks <250 kb apart) into one genomic risk locus. To identify overlapping and unique genomic risk loci, we overlaid the loci identified in the common factor MetS GWAS with those identified in the individual GWAS of the five MetS components.

2D. Functional annotation (FUMA)

SNP location and alleles were matched to databases that contain known functional annotations to assess consequences for these SNPs, including Combined Annotation (CADD) scores (a measure of deleteriousness)¹⁰, RegulomeDB scores (a score that indicates whether a SNP has a regulatory function, based on information from expression quantitative trait loci (eQTLs) and chromatin databases)¹¹, ANNOVAR categories (which estimates the functional consequence of a SNP based on its position)¹², and 15-core chromatin states (which represent the accessibility of genomic regions)^{13,14}.

FUMA maps candidate SNPs to genes in three ways: 1) positional mapping, which maps a SNP to its closest protein coding gene based on the physical distance according to the human reference assembly (GRCh37/hg19), 2) eQTL mapping, which maps SNPs to genes if it shows an association with the expression level of that gene, and 3) chromatin interaction mapping, which maps a SNP region to a gene if it has a 3D DNA-DNA interaction to that gene. eQTL mapping employed cis- and trans-eQTL (up to 1 Mb) data from tissues from default datasets, which includes all tissues from GTEx v8¹⁵, the eQTL catalogue¹⁶, PsychENCODE¹⁷, van der Wijst et al. scRNA eQTLs¹⁸, DICE¹⁹, eQTLGen²⁰, Blood eQTLs²¹, BIOS QTLs²², MuTHER²³, xQTLServer²⁴, CommonMind Consortium²⁵, and BRAINEAC²⁶. A false discovery rate of 0.05 was used to define significant eQTL associations. FUMA uses chromatin interaction data from four datasets: Hi-C data from Schmitt et al.²⁷, Giusti-Rodriguez et al. (see FUMA²⁸ tutorial), PsychENCODE¹⁷, and FANTOM5²⁹. To further prioritize candidate genes, we filtered chromatin interaction annotations where the genomic risk locus overlapped with an enhancer and the gene-mapped region with a promoter (up to 250 bp upstream and 500 bp upstream from the transcription start site), predicted by the Roadmap Epigenomics Project¹⁴.

2E. Gene based analysis (MAGMA)

MAGMA maps SNPs to protein-coding genes based on their genomic location (where a SNP is assigned to a gene if its location falls within the gene), and then jointly tests the association between SNPs within each gene and the MetS factor. Next, results from the gene-based analyses were used by MAGMA to conduct competitive gene-set analyses (testing for enrichment of genetic signal within gene sets relative to the other protein-coding genes in the data) in gene sets from MsigDB v6.2.^{30,31} To identify independently associated gene sets, a conditional gene-set analysis was performed. First, a gene-set analysis was run while conditioning on the most significant gene set. The remaining significant gene sets were subsequently conditioned on the most significant gene set within these resulting gene sets, until only one significant gene set was left. Finally, gene-property analyses (similar to gene-set analysis, but instead of a binary variable indicating whether a gene is part of a gene set, it uses a continuous variable representing gene expression values in a specific tissue or cell type) were used to test whether 54 tissue specific differential gene expression levels from GTEx v8¹⁵ were predictive of associations with the MetS factor. The significance threshold for the number of genes, gene sets, or tissues tested was $P < 0.05/17,706 \text{ genes} = 2.82\text{e-}6$, $P < 0.05/15,481 \text{ gene sets} = 3.23\text{-}06$, and $P < 0.05/54 \text{ detailed tissues} = 9.26\text{e-}4$, respectively.

2F. Cell type analyses

FUMA uses the MAGMA gene-property analysis to test cell-type specificity. Two cell type analyses were run, one with all available mouse tissue cells ($n = 805$) in FUMA and one with all available human brain tissue cells ($n = 255$) (all tissues from Mouse Cell Atlas³² were used; if a tissue was missing from Mouse Cell Atlas but was available in another dataset, the largest available sample was selected). Per dataset and within dataset conditional analyses were run, but because different groups of datasets were analysed separately, no cross datasets conditional analysis was performed. Cell types showed significant enrichment at $P < 0.05/805 = 6.21\text{e-}05$ for mouse cell type analyses, and $P < 0.05/255 = 1.96\text{e-}4$ for human brain cell type analyses.

Cell types included in mouse cell type specific analyses

GSE98816_Mouse_Brain_Vascular, GSE99235_Mouse_Lung_Vascular,
MouseCellAtlas_all, TabulaMuris_FACS_Aorta, TabulaMuris_FACS_Diaphragm,
TabulaMuris_FACS_Fat, TabulaMuris_FACS_Heart, TabulaMuris_FACS_Large_Intestine,

TabulaMuris_FACS_Skin, TabulaMuris_FACS_Thymus, TabulaMuris_droplet_Trachea (see <https://fuma.ctglab.nl/tutorial#datasets> for references).

Cell types included in Human brain cell type analysis:

PsychENCODE_Developmental, PsychENCODE_Adult, Allen_Human_LGN_level1, Allen_Human_LGN_level2, Allen_Human_MTG_level1, Allen_Human_MTG_level2, DroNc_Human_Hippocampus, GSE104276_Human_Prefrontal_cortex_per_ages, GSE67835_Human_Cortex, Linnarsson_GSE101601_Human_Temporal_cortex, Linnarsson_GSE76381_Human_Midbrain (see <https://fuma.ctglab.nl/tutorial#datasets> for references).

2G. Correcting for sample overlap for PRS analysis

EraSOR uses the degree of sample overlap estimated from the LDSC intercept to correct summary statistics of base data for sample overlap with target data. EraSOR requires summary statistics of the target samples which were generated by running a GWAS on FG, HDL-C, and TG in the FHS samples with PLINK2 using cleaned, called, genotyped data (Supplementary Note).^{33,34} Covariates used were sex, age, and age² for FG, and sex, age, age², and the first 20 principal components for HDL-C and TG, mirroring covariates used in the original GWAS.^{1,2} The MetS factor GWAS was rerun for PRS prediction purposes using the adjusted summary statistics for FG, HDL-C, and TG, and non-adjusted summary statistics for SBP and WC.

2H. Framingham Heart Study quality control

The FHS is a longitudinal community-based cohort study, within which data have been collected on a range of cardiovascular diseases and risk factor as well as genetic data. Our target sample for polygenic prediction was FHS second ('offspring', recruited between 1971 and 1975) and third generation (recruited between 2002 and 2005) subjects.^{24,25} Genotyping in FHS was done with Affymetrix 500k array. Quality control with PLINK^{33,34} consisted of removing variants with a minor allele frequency <0.05, missingness >0.1 and deviation of Hardy-Weinberg equilibrium of $P < 1e-6$, and samples with missingness >0.1, heterozygosity coefficients of >3SD from the mean, and discordant sex. One of each pair of first- or second-degree relatives ($\hat{\pi} > 0.125$) was removed. We extracted only individuals that clustered together with Europeans from 1000 Genomes in an MDS plot.⁸ Finally, we selected samples that had

measurements of FG, HDL-C, TG, SBP, or WC (measured during abdominal CT). After quality control, 371,325 variants and 2,095 samples were available for analysis. Individuals on glucose lowering medication ($n = 102$) were removed for FG analyses, individuals on lipid lowering medication ($n = 268$) were removed for HDL-C and TG analyses, and individuals on blood pressure lowering medication ($n = 200$) were removed for SBP analyses. MetS was defined as 3/5 of hypertension (RR ≥ 130 mmHg (systolic) or antihypertensive treatment), elevated waist circumference (women ≥ 80 cm and men ≥ 94 cm), elevated triglycerides (≥ 150 mg/dl or lipid lowering treatment), reduced HDL-C (≤ 50 mg/dl in women and ≤ 40 mg/dl in men or lipid lowering treatment), and elevated fasting glucose (≥ 100 mg/dl or glucose lowering treatment).³⁵ For MetS prediction, only samples with non-missing information on indicator phenotypes were used. Per phenotype sample size was 1,816 for FG, 1,825 for HDL-C, 1,891 for SBP, 1,812 for TG, 748 for WC, and 666 for MetS (254 cases and 412 controls).

2I. Drug gene set analysis

To identify drugs associated with MetS that may be candidates for treatment, genetically informed drug repurposing was performed drug-gene set analysis via the DRUGSETS software pipeline³⁶. DRUGSETS tests drug-phenotype associations using competitive gene-set analysis in MAGMA³⁷. Gene sets are created for every drug ($n = 1150$) using drug gene targets and interactions gathered from the Clue Repurposing Hub³⁸ and the Drug Gene Interaction database³⁹ (DGIdb). Significant drugs are tested while conditioning on all drug target genes to ensure that associations are driven by effects unique to each drug pathway. Additionally, several types of drug groups, namely Anatomical Therapeutic Classification (ATC) III codes ($n = 85$), clinical indications (i.e., the disease/disorder that a drug is approved to treat; $n = 118$) and mechanism of action categories ($n = 79$) were tested for association with the phenotype using a modified multiple linear regression model. The linear model assesses if the MAGMA Z statistic is higher for drugs in a group versus all drugs not in that group, while accounting for the covariance due to overlapping genes and drugs. This test was performed for each drug group, and Bonferroni correction was used to correct for the number of drugs groups in each category. Drugs showed enrichment at $P < 4.35\text{e-}05$ (0.05/1150), drugs from ATC III codes at $P < 5.10\text{e-}4$ (0.05/89), from clinical indication at $P < 4.24\text{e-}4$ (0.05/118) and from mechanism of action at $P < 6.33\text{e-}4$ (0.05/79). For a full overview of DRUGSETS and methods see³⁶.

3. Supplementary results

3A. Common factor GWAS without WC

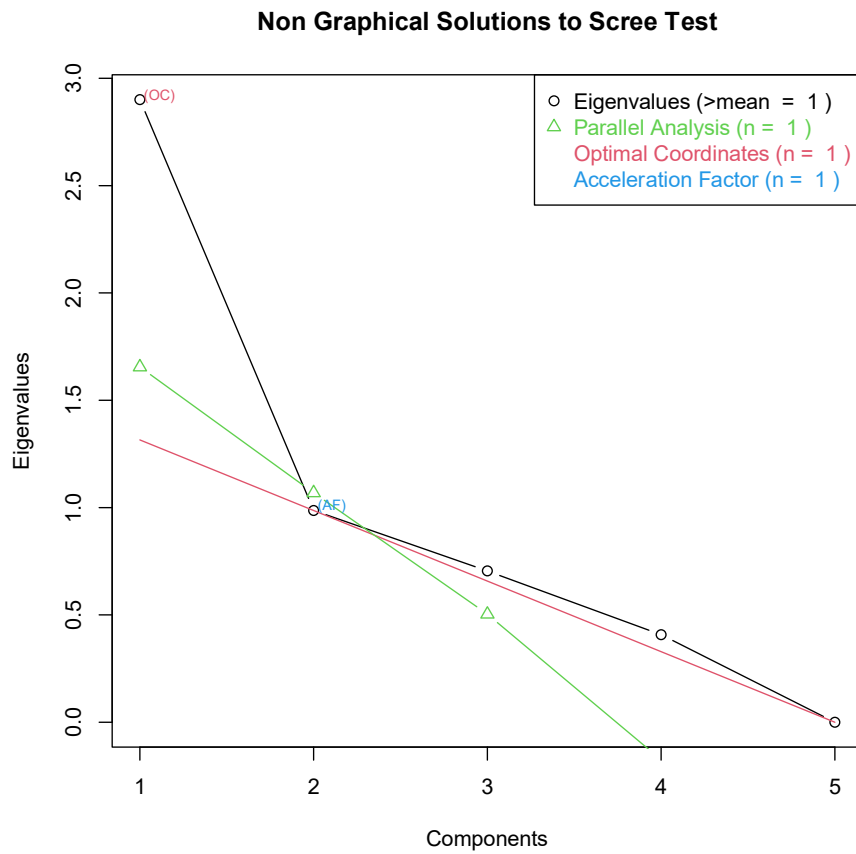
Running a common factor GWAS without WC resulted in 35 genomic risk loci (21 (60%) overlap with genomic risk loci significant for MetS factor), 112 significantly mapped genes by MAGMA (57 (51%) overlap with MetS factor), 26 significantly associated gene sets (8 (31%) overlap with MetS factor), three enriched tissues (none overlap with MetS factor), two enriched mouse cell types (none overlap with MetS factor) and four enriched human brain cell types (one (25%) overlaps with MetS factor) (Supplementary Table 20-25). LDSC derived SNP heritability was 0.12 (0.01) and genetic correlation (r_G) with MetS factor GWAS was 0.74 (SE 0.013, P 0).

3B. Additional drug gene set (repurposing) results

There is significant enrichment of plain lipid modifying agents (ATC code C10A; $P = 5.28\text{e-}10$), and antimycotics for systemic use (J02A; $P = 0.0002$) (Supplementary Table 30). Furthermore, drugs clinically indicated for hypercholesterolemia ($P = 0.0001$) show enrichment (Supplementary Table 31), as well as the mechanism of action categories of HMG-CoA reductase inhibitors (i.e., statins; $P = 2.97\text{e-}06$), and peroxisome proliferator-activated receptor agonists ($P = 7.18\text{e-}06$) (Supplementary Table 32).

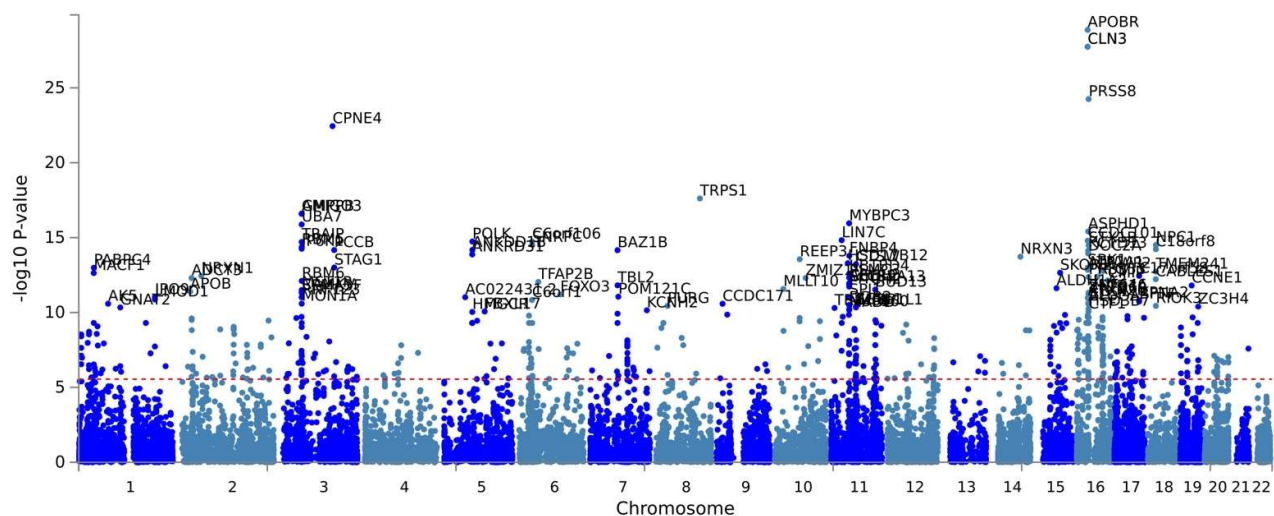
4. Supplementary figures

4A. Exploratory factor analysis scree plot (R nFactors package)



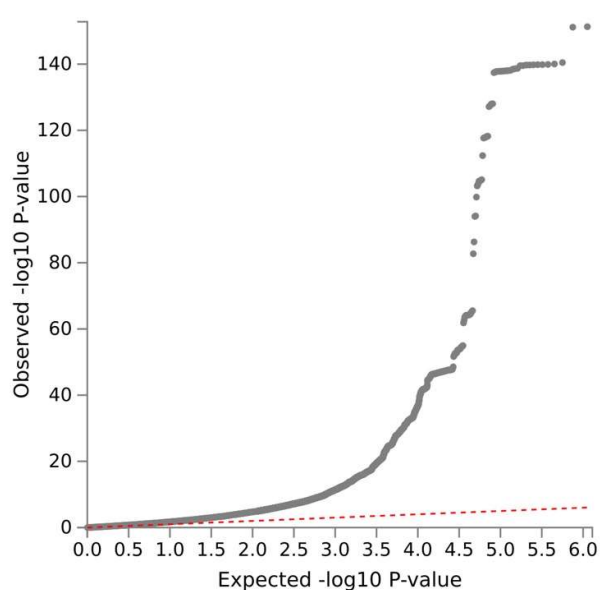
The suggested number of factors to retain was one with each of the methods (see legend).

4B. MAGMA gene-based Manhattan plot.

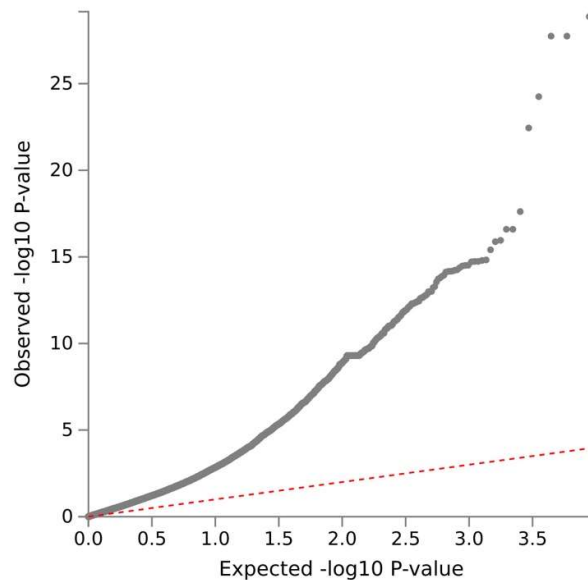


Genes were significant at $P < 0.05/17,706$ (number of tested genes) = $2.824e-6$. The red line indicates Bonferroni-corrected significance, shown here as the negative \log_{10} -transformed P value on the y axis. The top 100 most significant genes are labelled. This plot is generated by FUMA v.1.3.7.²⁸

4C and D. SNP-based (4C) and gene-based (4D) QQ plots.



4C



4D

QQ plots display the expected $-\log_{10}$ transformed P -values on the x-axis and the observed $-\log_{10}$ transformed P -values on the y-axis.

References

1. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
2. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
3. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
4. Hartmann, K., Krois, J. & Waske, B. R factanal package. (2018).
5. R Core Team. R: A language and environment for statistical computing. (2019).
6. David Kaplan. Structural Equation Modeling: Foundations and Extension (2018).
7. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
8. Gibbs, R. A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
9. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. (2018).
10. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
11. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
12. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, (2010).
13. Ernst, J. & Kellis, M. ChromHMM: Automating chromatin-state discovery and characterization. *Nature Methods* vol. 9 215–216 (2012).
14. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
15. GTEx Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–5 (2013).
16. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* 2021 539 **53**, 1290–1299 (2021).
17. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for

- the human brain. *Science*. **362**, (2018).
18. Van Der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies cell type-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493 (2018).
 19. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, (2018).
 20. Vösa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
 21. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
 22. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
 23. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
 24. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
 25. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
 26. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418 (2014).
 27. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
 28. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
 29. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22 (2015).
 30. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
 31. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
 32. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018).
 33. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-

- Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
34. Purcell, S. PLINK 2.0. (2019).
 35. Alberti, K. G. M. M. *et al.* Harmonizing the metabolic syndrome: A joint interim statement of the international diabetes federation task force on epidemiology and prevention; National heart, lung, and blood institute; American heart association; World heart federation; International . *Circulation* vol. 120 1640–1645 (2009).
 36. Bell, N., Uffelmann, E. & Posthuma, D. Using DRUG Gene SEt Analysis (DRUGSEA) to identify drug repurposing candidates for psychiatric and non-psychiatric phenotypes. *Prep.*
 37. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
 38. Corsello, S. M. *et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat. Med.* 2017 234 **23**, 405–408 (2017).
 39. Cotto, K. C. *et al.* DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* **46**, D1068–D1073 (2018).