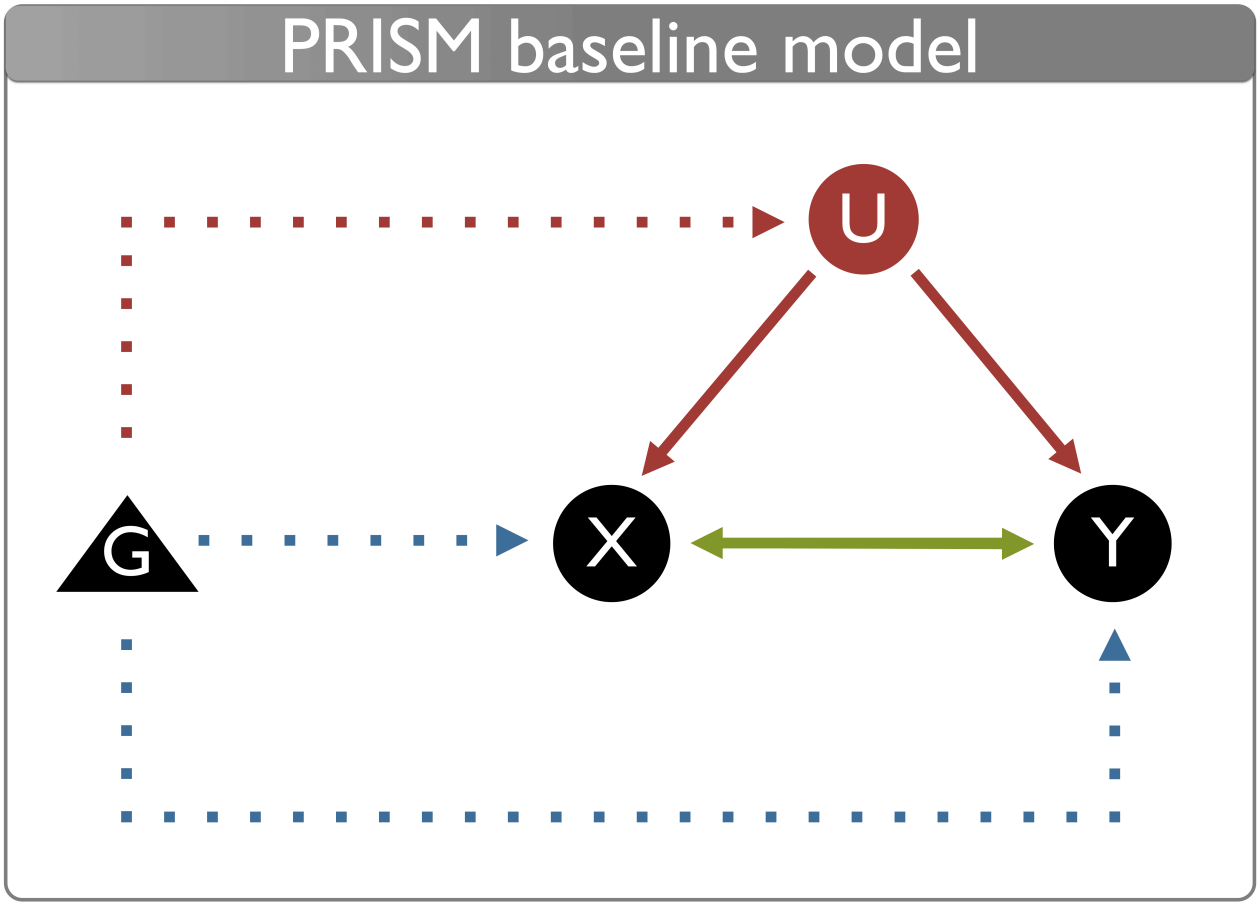
Supplementary material: Inferring genetic variant causal networks by leveraging pleiotropy

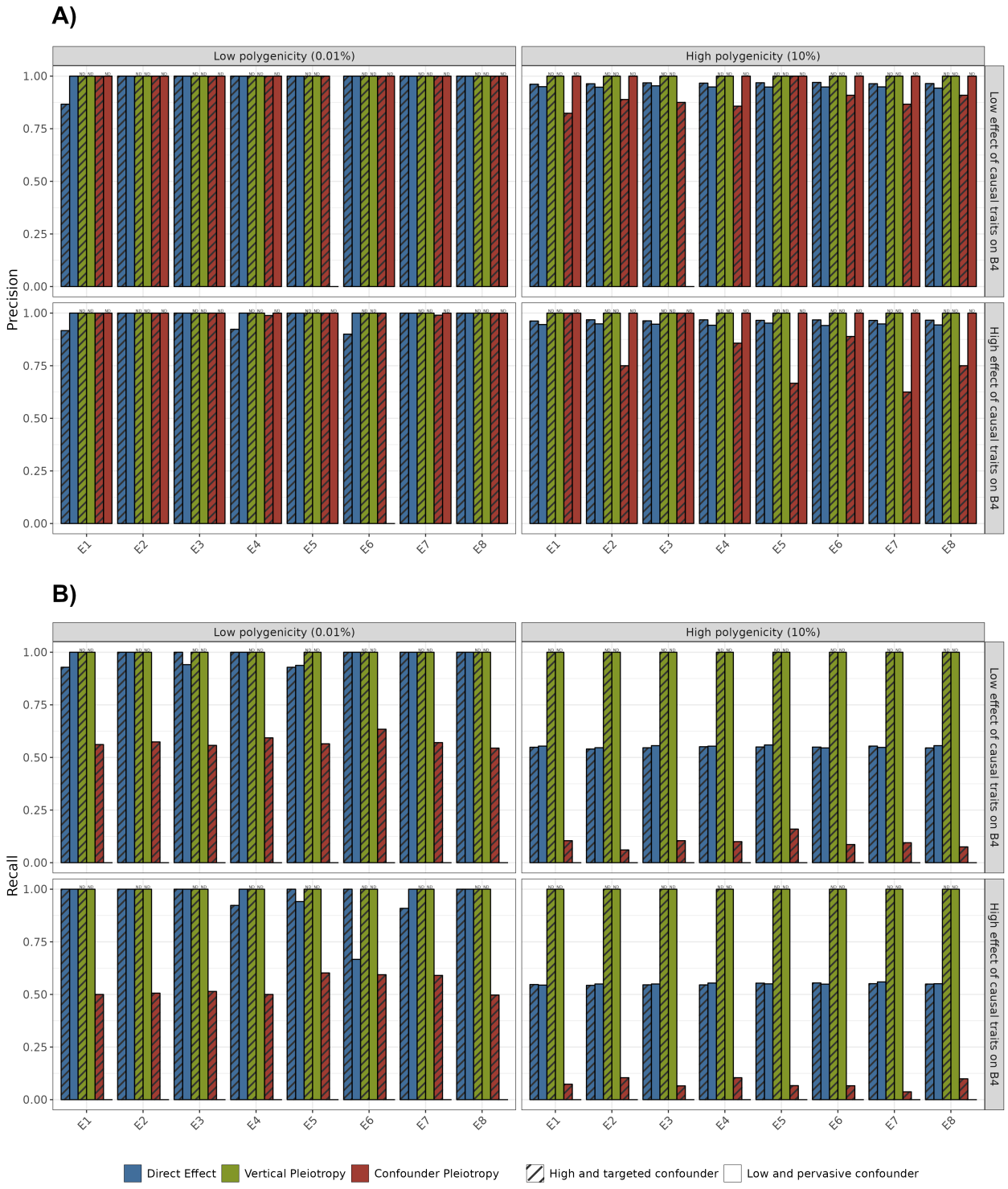
Martin Tournaire[1], Asma Nouira[1], Mario Favre Moiron[1], Yves Rozenholc[1], Marie Verbanck[1,2]

[1] *UR 7537 BioSTM, Biostatistique, Traitement et Modélisation des données biologiques, Université Paris Cité.* [2] *Inserm U900, Institut Curie, PSL Research University, 75248 Paris, France.*

Supplementary Figures

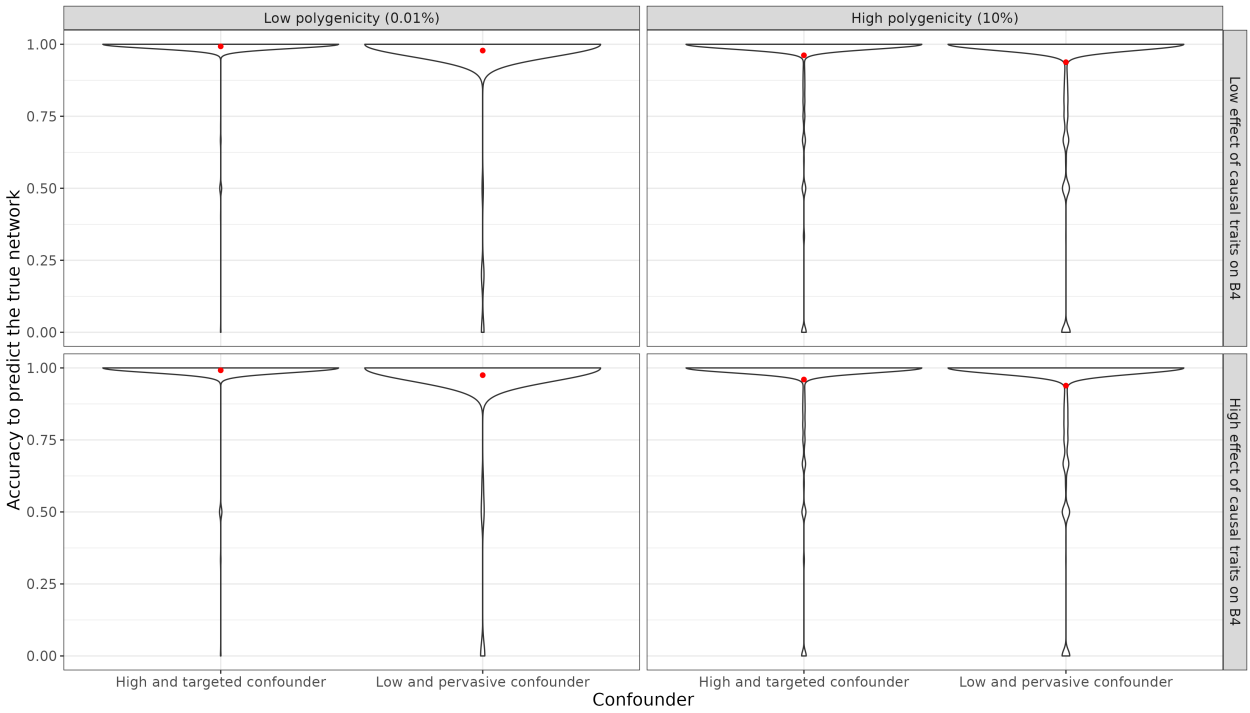


Supplementary Figure 1. *X and Y are two complex traits, and U a latent confounder with causal effects on each other. G represents a genetic variant, with putative effects on X, Y and U.*

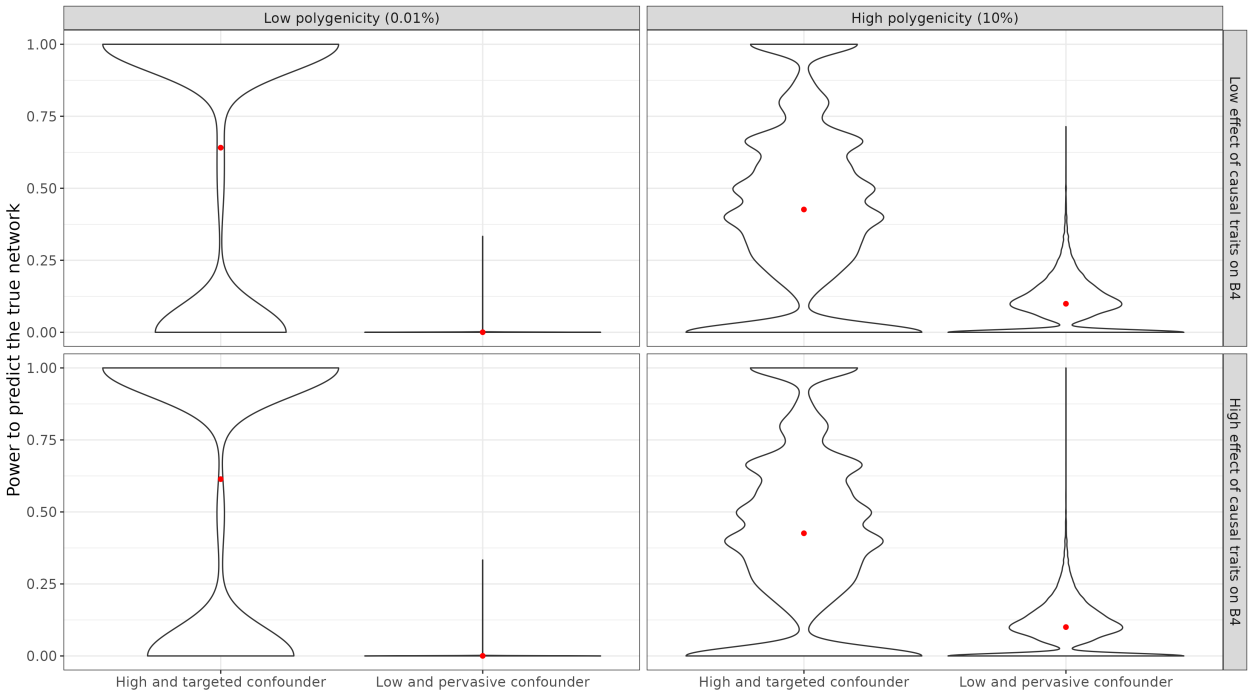


Supplementary Figure 2. A) *Precision and* ***B)*** *Recall of PRISM predictions for significant variant-trait effects, on simulations. The x-axis represents the simulated traits in the network (See Fig. 6 and Online Methods). Significant effects are defined with* $P < 5 \times 10^{-8}/17$*, the recommended threshold from PRISM (See Online Methods). Bars are colored according to predicted direct and pleiotropic labels. By convention, when variants neither exist nor were predicted for a given label, the accuracy or the recall equals 1, such cases are indicated with ND. Eight scenarios are represented across facets, with varying parameters. Polygenicity represents the proportion of variants with a direct effect on each trait. Effect on* $B4$*, the most pivotal trait, represents the proportion of effect passed to* $B4$*, for all traits with a non-zero vertical effect on* $B4$*. High targeted confounder means that few variants (0.01%) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounder means that a large proportion (5%) of variants have an effect on the confounder, but with low magnitude.* ***A)*** *The y-axis represents the*

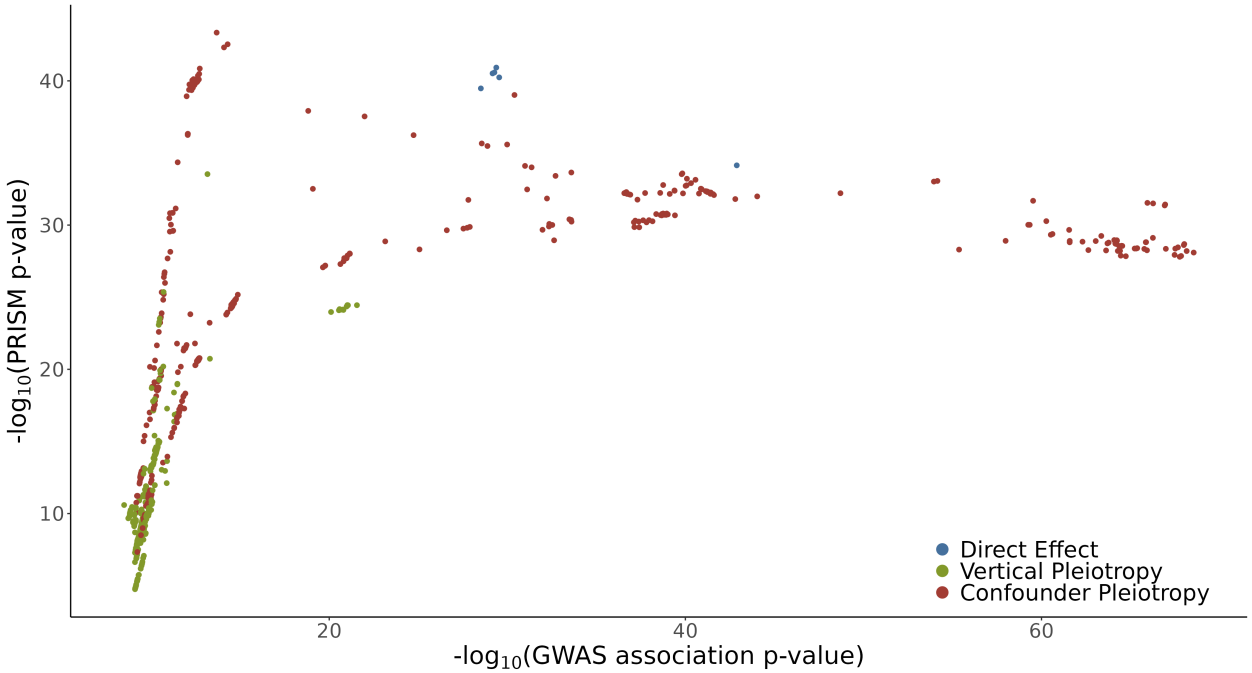
precision which is the proportion of well-predicted variants among all predicted variants for a given label. ***B)*** *The y-axis represents the recall, which is the proportion of well-predicted variants among all true variants for a given label. NB: results for traits* A-D *can be found in Fig. 2*



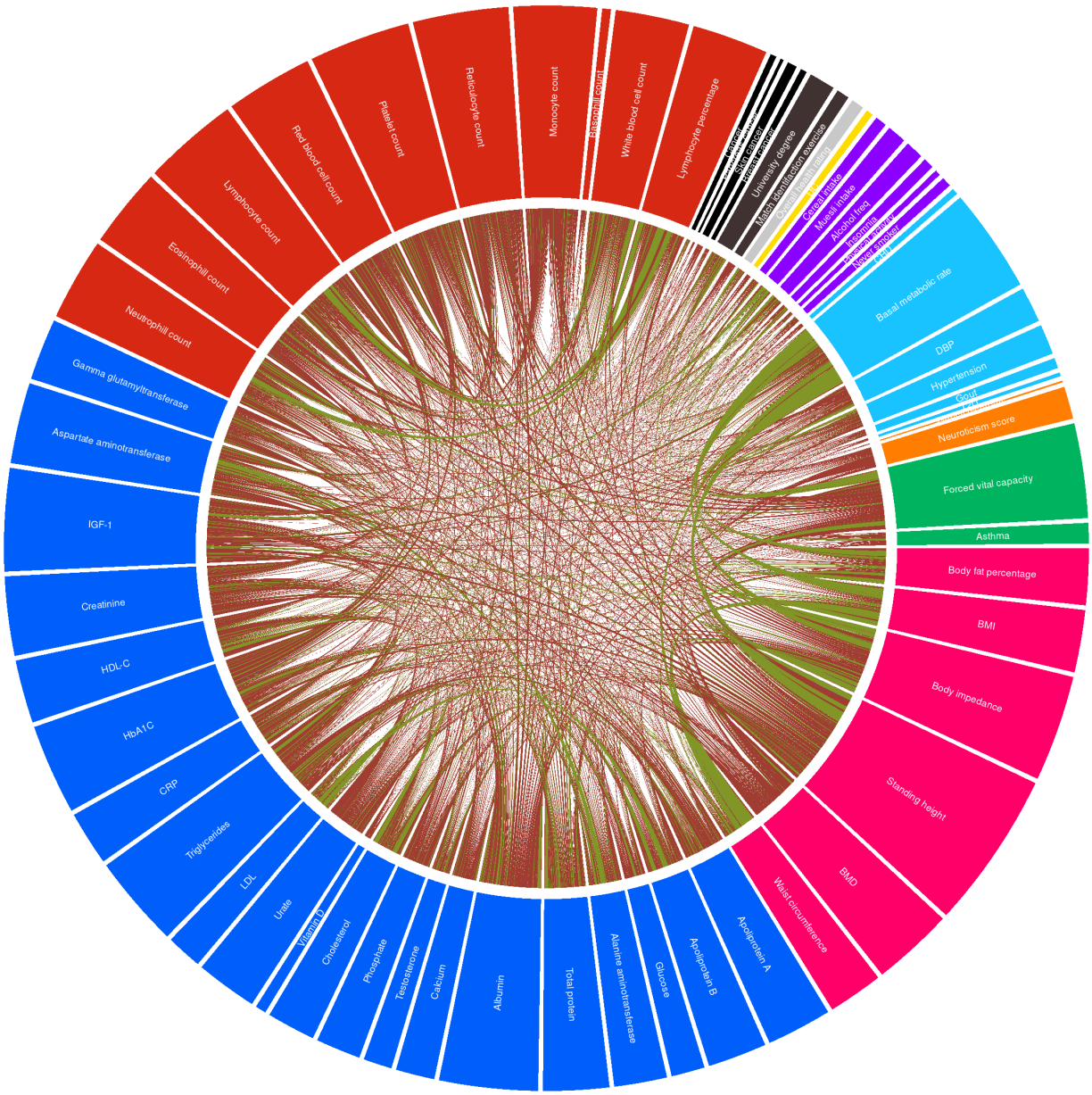
Supplementary Figure 3. *Violin plot of the similarity between PRISM-predicted variant networks and true simulated networks, evaluated using the Simple Matching Coefficient (SMC), represented on the y-axis. The SMC measures the accuracy of the predicted networks by matching their edges with those of the true networks. A SMC of 1 indicates that all predicted edges are true edges, while a SMC of 0 indicates that none of the predicted edges correspond to the true edges. The x-axis displays the eight simulated scenarios involving 18 traits with high heritability, as detailed in Supplementary Table 1. The red dot represents the average SMC.*



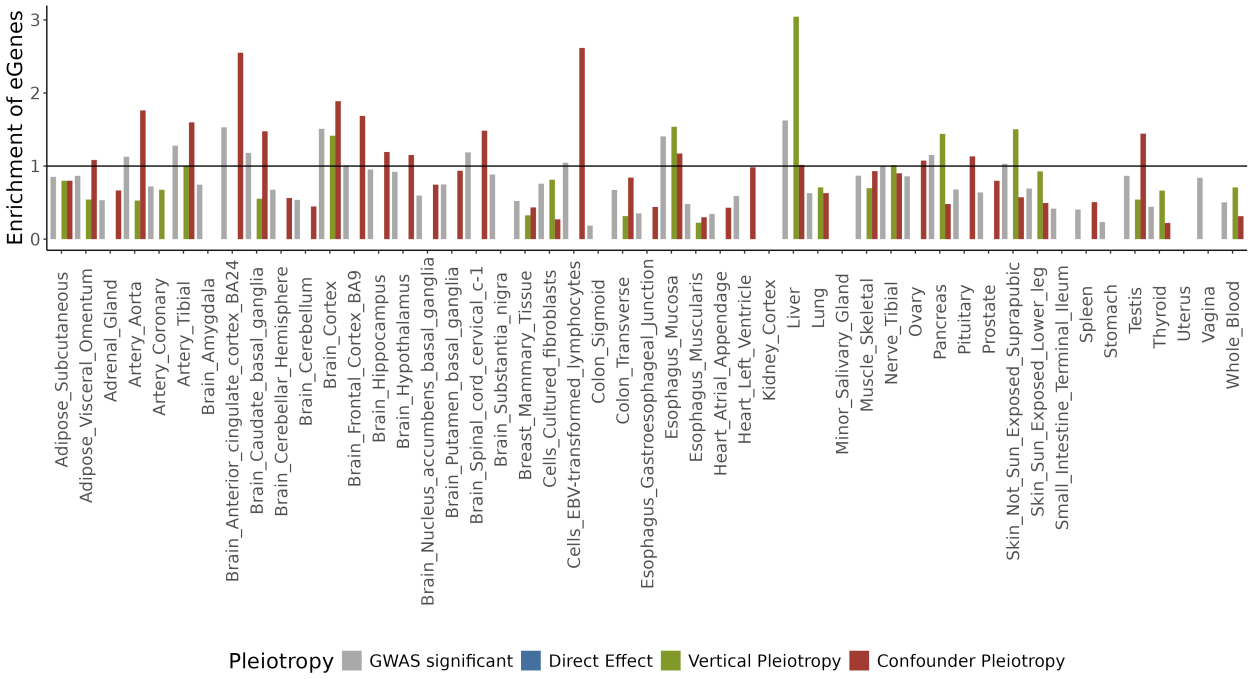
Supplementary Figure 4. *Violin plot of the similarity between true simulated networks and PRISM-predicted variant networks, evaluated using the Simple Matching Coefficient (SMC), represented on the y-axis. The SMC measures the power of detecting true networks by matching their edges with those of the predicted networks. A SMC of 1 indicates that all true edges are predicted, while a SMC of 0 indicates that none of the true edges are predicted. The x-axis displays the eight simulated scenarios involving 18 traits with high heritability, as detailed in Supplementary Table 1. The red dot represents the average SMC.*



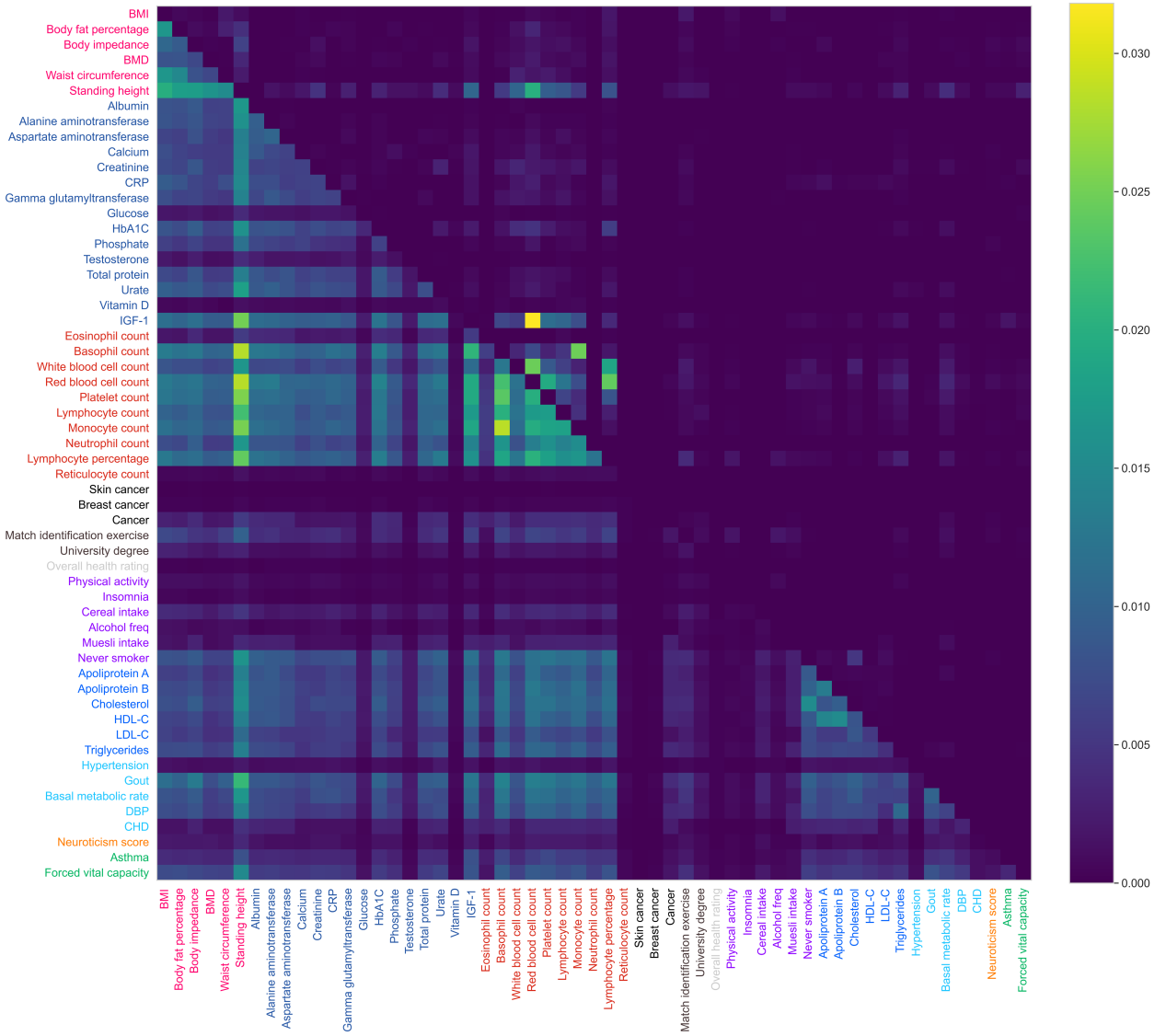
Supplementary Figure 5. *P-values ($-log_{10}$) from the UK Biobank GWAS on coronary heart disease (CHD) (x-axis) and PRISM (y-axis) for 947 genetic variants. Each dot represents a significant variant according to GWAS or PRISM, colored according to PRISM predicted labels: direct (blue), vertical (green) and confounder (red).*



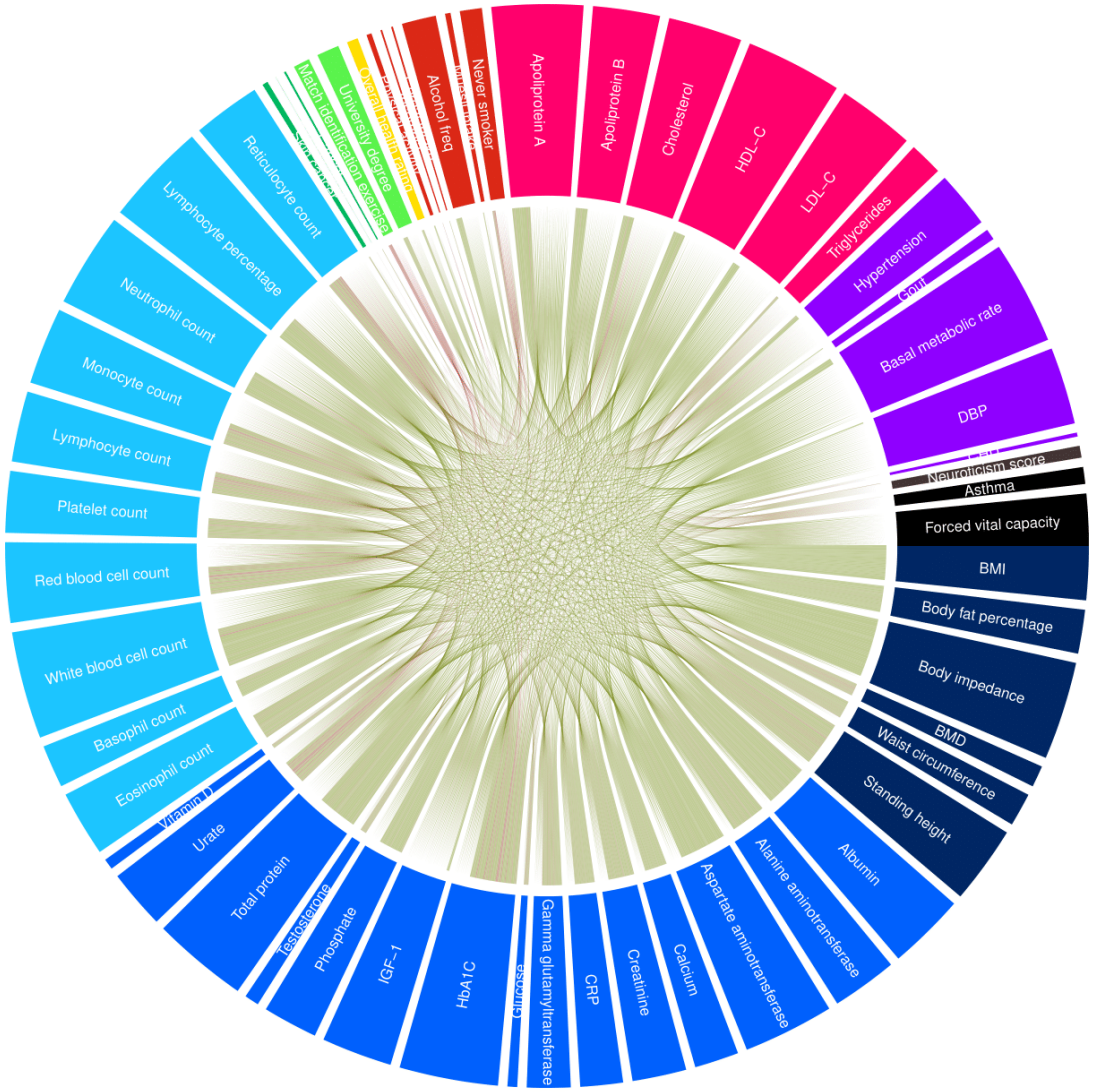
Supplementary Figure 6. *Circos plot of shared variants, for all traits processed by PRISM. This represents genetic pleiotropy obtained from PRISM, with each line representing a significant variant with direct horizontal or pleiotropic effect on two traits according to PRISM. Blue lines represent horizontal pleiotropy, green lines represent vertical pleiotropy, red lines represent confounder pleiotropy.*



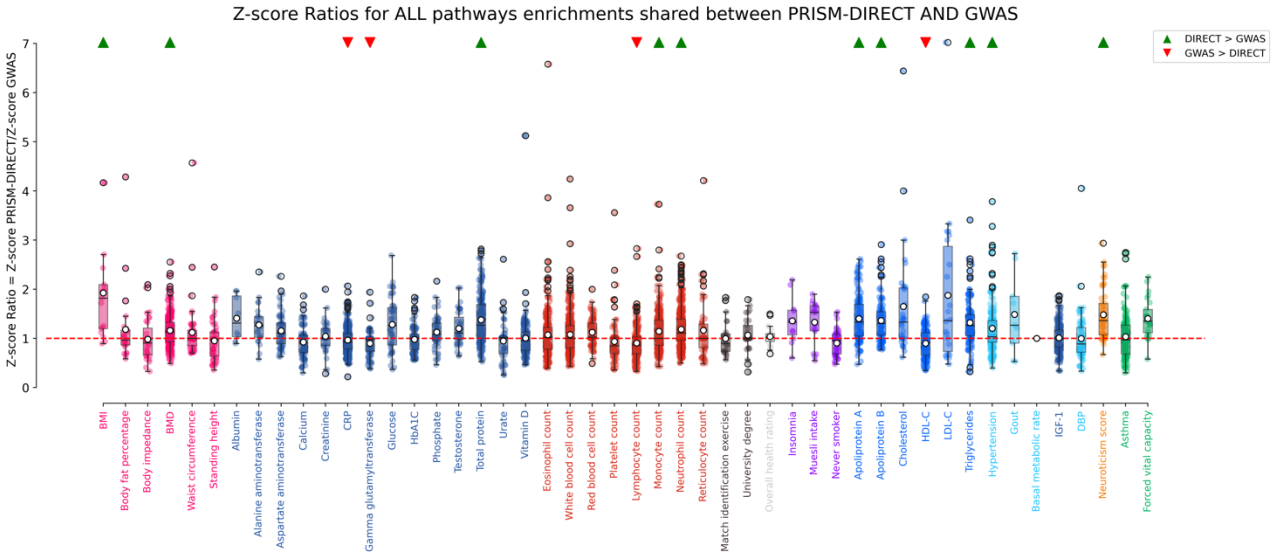
Supplementary Figure 7. *Enrichment of eGenes, from different tissues, mapped to genetic variants, according to variants labels. The x-axis represents all tissues. The y-axis represents the enrichment of eGenes. The studied trait is I25 from UK Biobank, chronic ischaemic heart disease (CHD). The eGenes were retrieved from GTEx v8.*



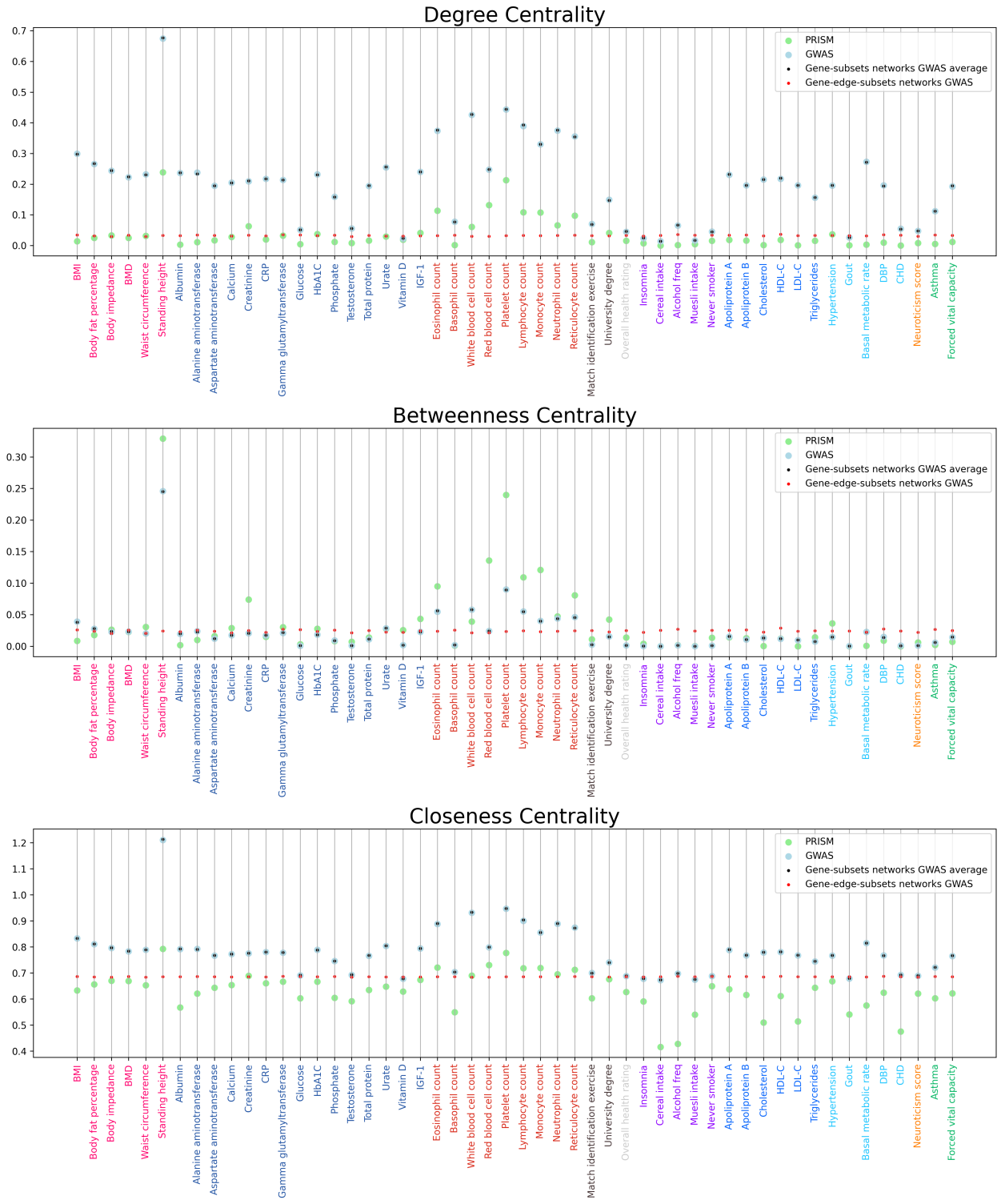
Supplementary Figure 8. *Heatmap of shared genes between our 61 traits. The bottom-left triangle represents common genes mapped from GWAS variants. The top-right triangle represents common genes mapped from PRISM direct-effect variants. Each tile is the intersection between two traits, and the color of the tile represents the number of common genes between those two traits divided by the total number of genes (PRISM or GWAS).*



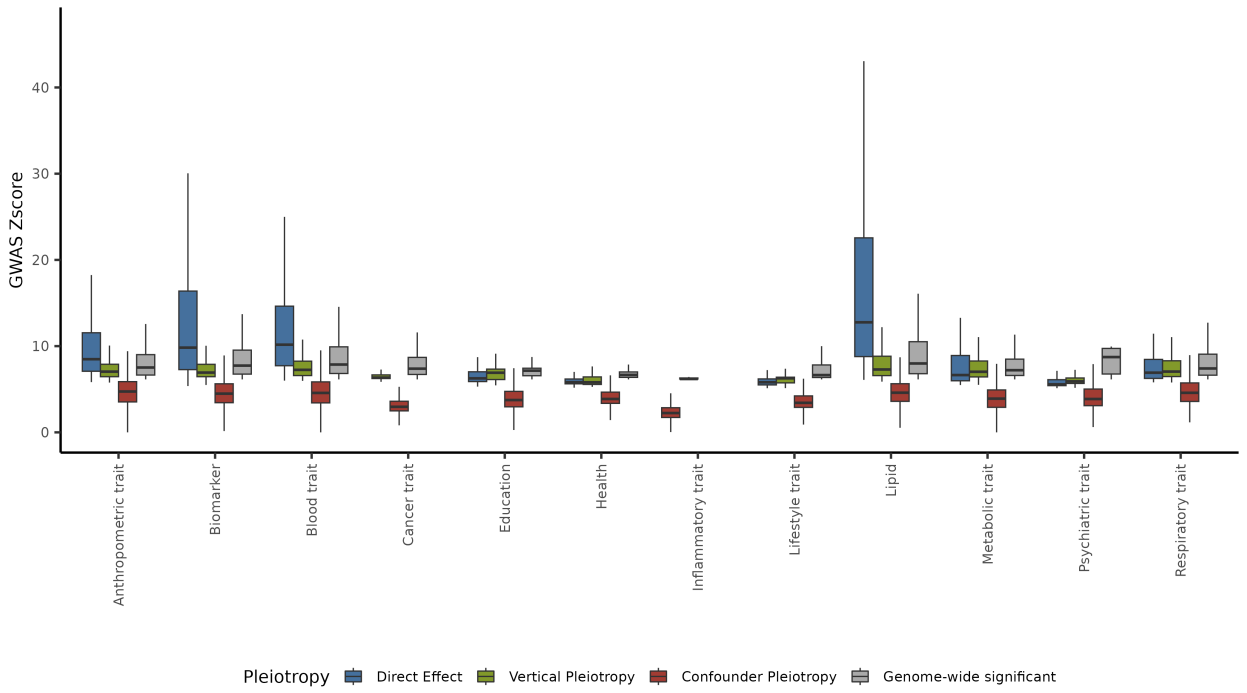
Supplementary Figure 9. *Circos plot of genes mapped to shared variants, for all traits processed by PRISM. This represents genetic pleiotropy obtained from PRISM, with each line representing a gene mapped to a significant variant with direct horizontal or pleiotropic effect on two traits according to PRISM. Blue lines represent horizontal pleiotropy, green lines represent vertical pleiotropy, red lines represent confounder pleiotropy. To map a gene to a genetic variant, we prioritize first horizontal pleiotropy, then vertical pleiotropy, then confounder pleiotropy.*



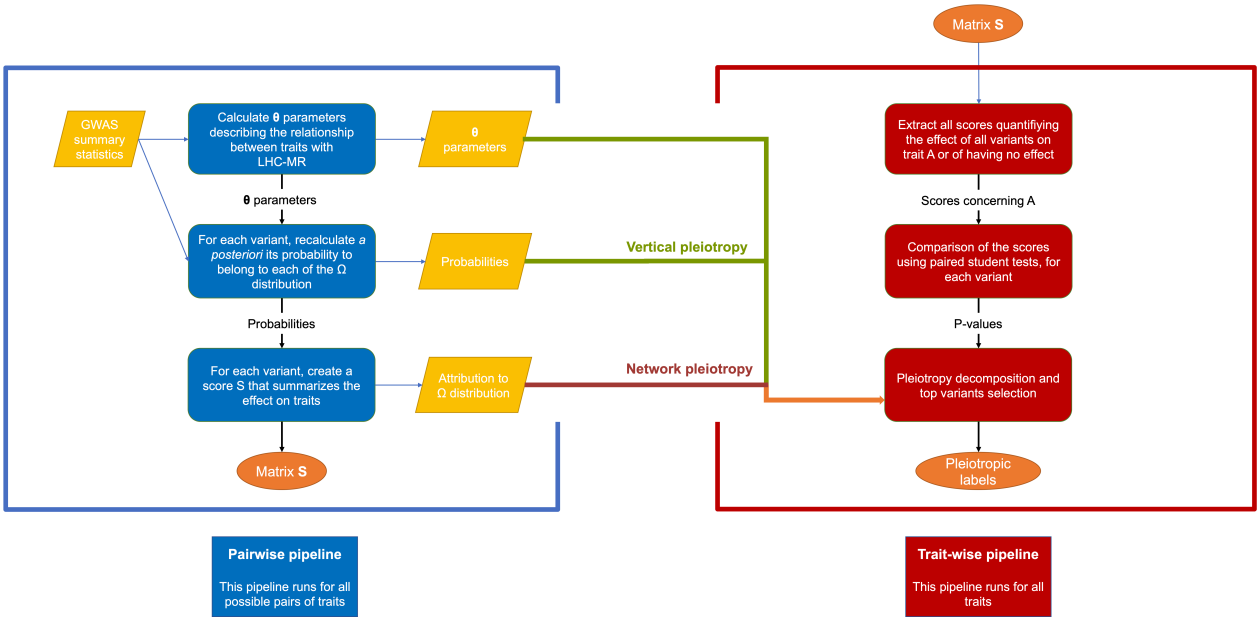
Supplementary Figure 10. *Z-score ratios for shared pathways enrichment between PRISM direct mapped genes and GWAS mapped genes. The x-axis represents all traits, colored by category. The y-axis represents a boxplot of Z-score ratios between PRISM direct and GWAS, from all pathways. A green triangle above a boxplot means that the enrichment is significantly higher than 1 for this specific trait. A red triangle above a boxplot means that the enrichment is significantly lower than 1, for this specific trait.*



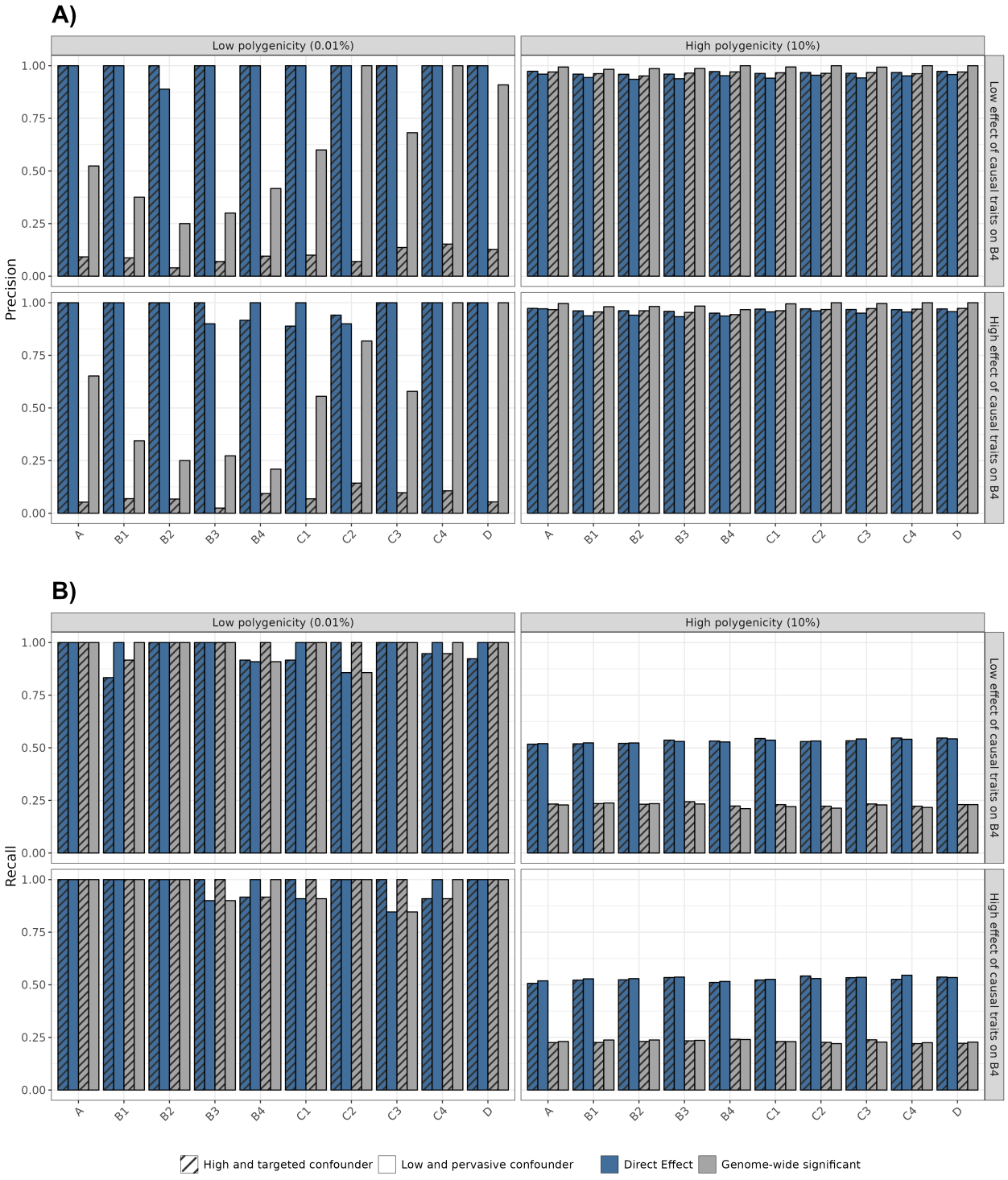
Supplementary Figure 11. *Centrality measures of PRISM direct bipartite gene-trait network and GWAS bipartite gene-trait network. The three sub-plots show respectively the degree, the betweeness, and the closeness metrics, represented on the y-axis. The x-axis represents the traits, colored by category, as the metrics are specific to a trait in the network. Green dots correspond to the PRISM direct network. Grey dots correspond to the GWAS network. Blue dots correspond to the average of multiple networks with randomly removed genes, to have the same number of genes as PRISM. Red dots correspond to the average of multiple networks with randomly removed genes and edges, to have the same number of genes and edges as PRISM.*



Supplementary Figure 12. *Initial GWAS Z-scores of labeled genetic variants, grouped by labels and trait categories. The y-axis represents the initial Z-scores in GWAS summary statistics from UK Biobank. The x-axis represents traits grouped by broad categories. Boxplots are colored according to PRISM predicted labels, and GWAS significance at threshold* $P < 5 \times 10^{-8}/60$.

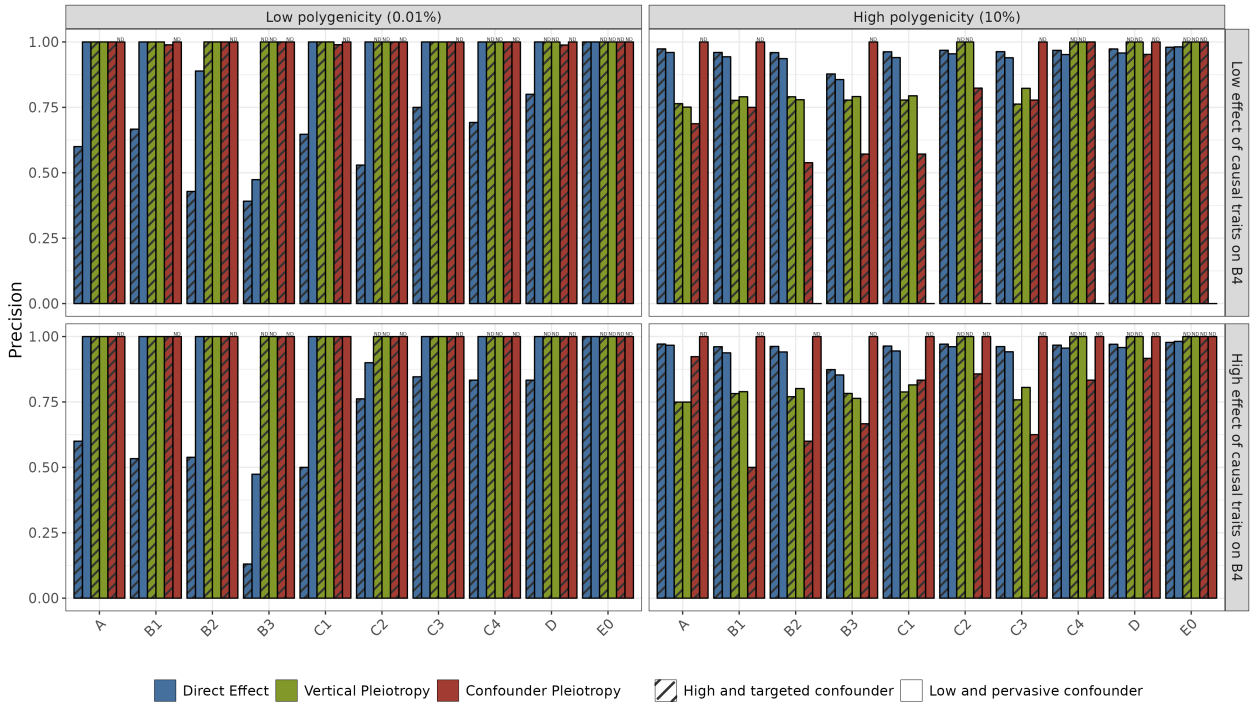


Supplementary Figure 13. *The PRISM pipeline is divided in two main steps, pairwise and traitwise pipeline. The left side represents the pairwise pipeline, whereas the right side represents the traitwise pipeline.*

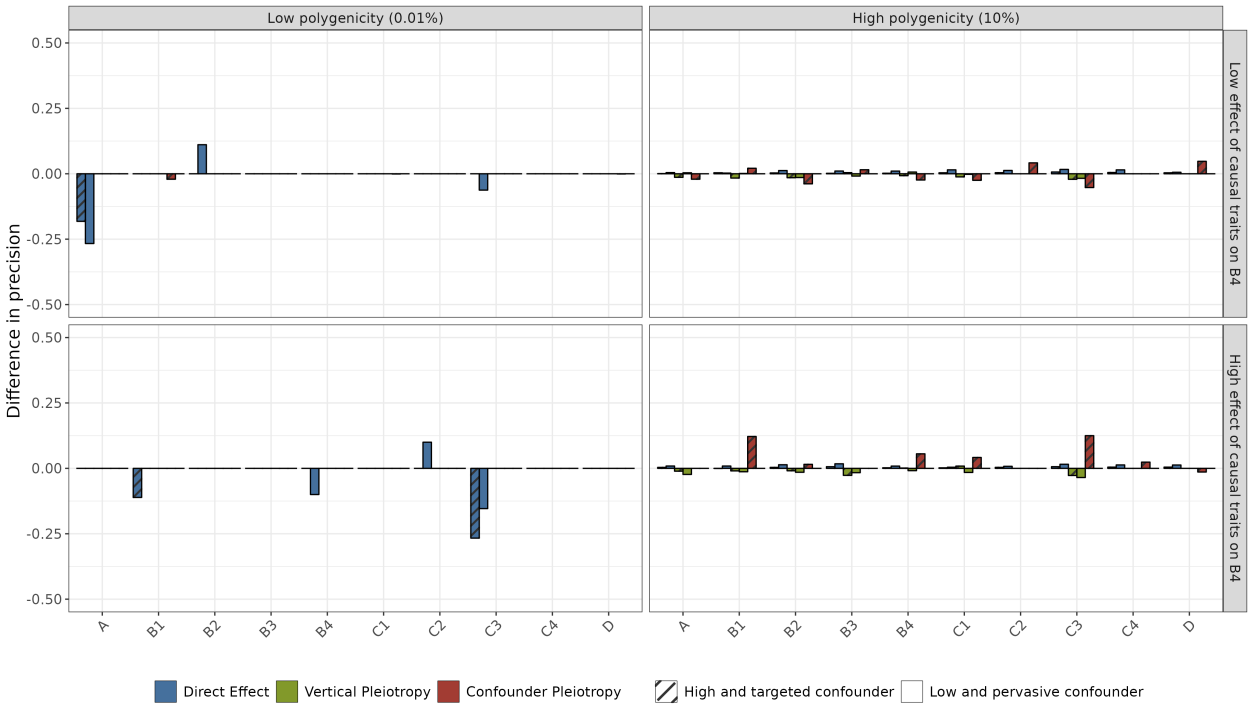


Supplementary Figure 14. A) *Precision and* ***B)*** *Recall of PRISM and GWAS predictions for significant variant-trait effects, on simulations. The x-axis represents the simulated traits in the network (See Fig. 6 and Online Methods). Significant effects of PRISM and GWAS are defined with* $P < 5 \times 10^{-8}/17$*, the recommended threshold from PRISM (See Online Methods). Bars are colored according to predicted direct effects from PRISM or significant associations from GWAS. Eight scenarios are represented across facets, with varying parameters. Polygenicity represents the proportion of variants with a direct effect on each trait. Effect on* $B4$*, the most pivotal trait, represents the proportion of effect passed to* $B4$*, for all traits with a non-zero vertical effect on* $B4$*. High targeted confounder means that few variants (0.01%) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounder means that a large proportion (5%) of variants have an effect on the confounder, but with low magnitude.* ***A)*** *The y-axis represents the precision which is the proportion of well-predicted variants among all predicted variants*

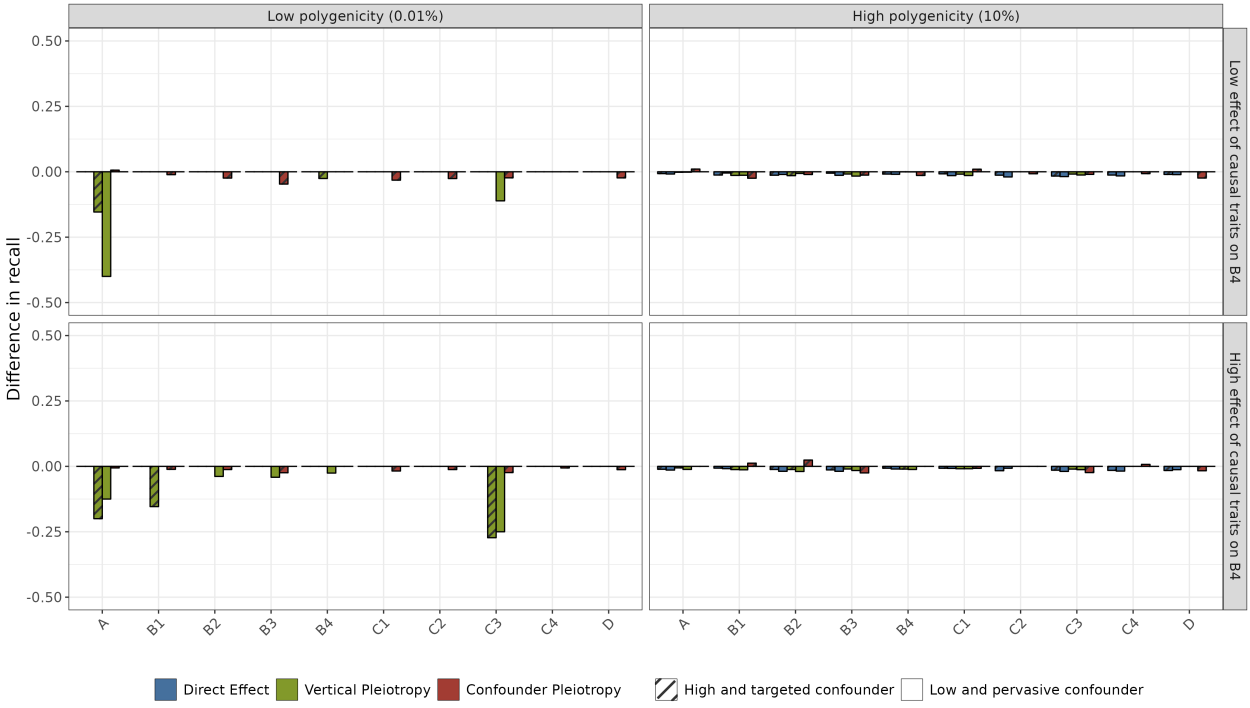
for a given label. ***B)*** *The y-axis represents the recall, which is the proportion of well-predicted variants among all true variants for a given label.*



Supplementary Figure 15. *Precision of PRISM predictions for significant variant-trait effects on simulations, without* $B4$*. The y-axis represents the precision which is the proportion of well-predicted variants among all predicted variants for a given label. The x-axis represents the simulated traits in the network (See Online Methods). Trait* $B4$ *was used to simulate the network of traits, but was not included in PRISM. To maintain the total number of traits, E0, a trait with no pleiotropy, was processed. Significant effects are defined with* $P < 5 \times 10^{-8}/17$*, the recommended threshold from PRISM (See Online Methods). Bars are colored according to predicted direct and pleiotropic labels. By convention, when no variants were predicted for a given label, the accuracy equals 1, such cases are indicated with ND. Eight scenarios are represented across facets, with varying parameters. Polygenicity represents the proportion of variants with a direct effect on each trait. Effect on* $B4$*, the most pivotal trait, represents the proportion of effect passed to* $B4$*, for all traits with a non-zero vertical effect on* $B4$*. High targeted confounder means that few variants (0.01%) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounder means that a large proportion (5%) of variants have an effect on the confounder, but with low magnitude.*

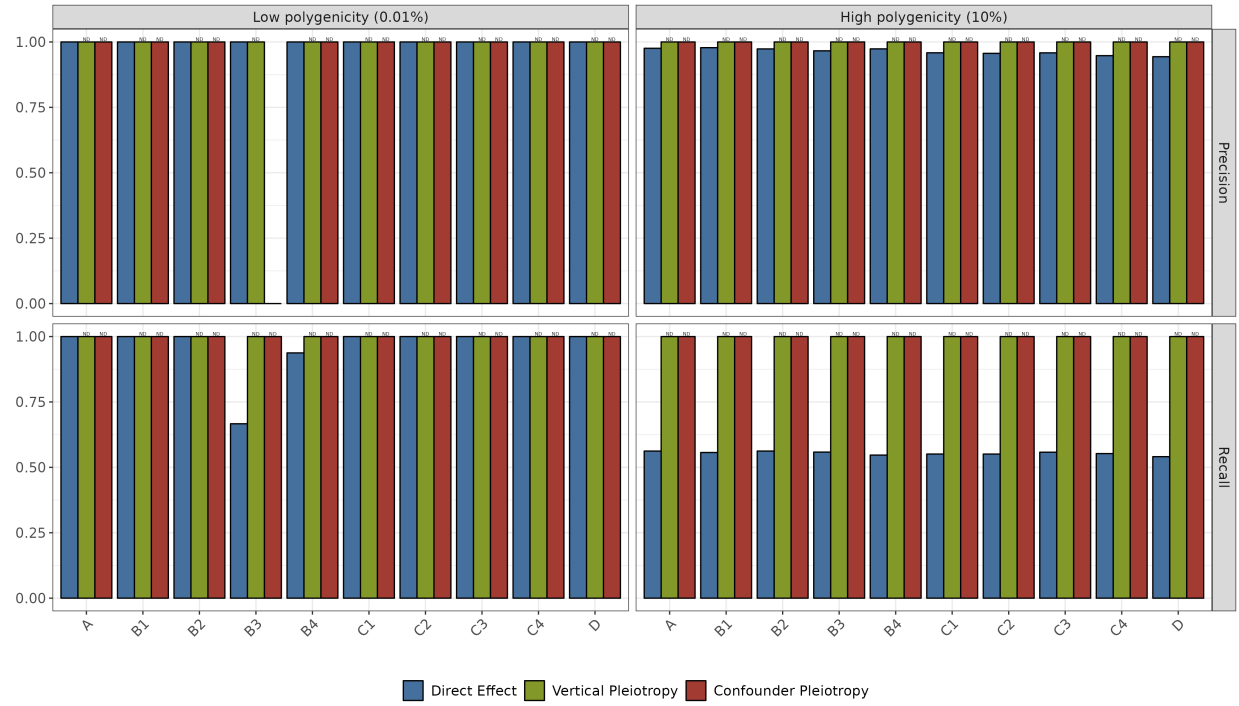


Supplementary Figure 16. *Difference of precision between PRISM predictions on simulations with or without noised parameters. The y-axis represents the difference in precision, in different conditions. The x-axis represents the simulated traits in the network (See Online Methods). Bars are colored according to predicted direct and pleiotropic labels. Eight scenarios are represented across facets, with varying parameters. Polygenicity represents the proportion of variants with a direct effect on each trait. Effect on $B4$, the most pivotal trait, represents the proportion of effect passed to $B4$, for all traits with a non-zero vertical effect on $B4$. High targeted confounder means that few variants (0.01%) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounder means that a large proportion (5%) of variants have an effect on the confounder, but with low magnitude.*

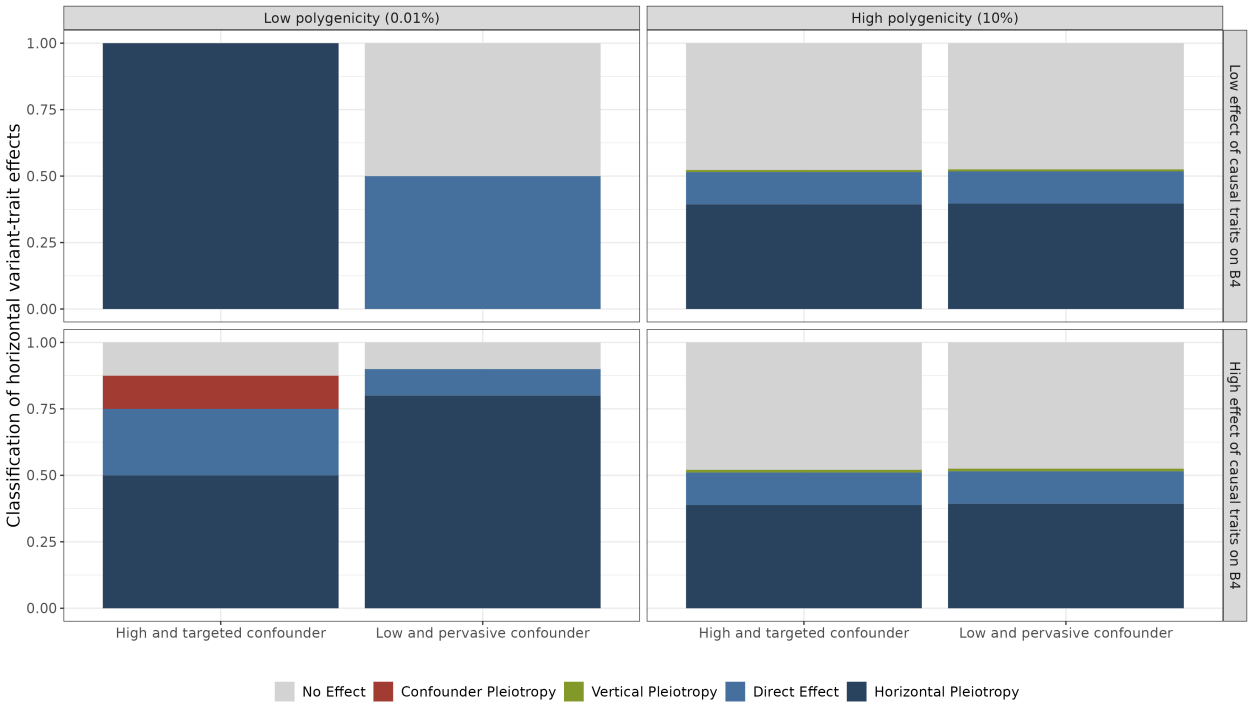


Supplementary Figure 17. *Difference of recall between PRISM predictions on simulations with or without noised parameters. The y-axis represents the difference in recall, in different conditions. The x-axis represents the simulated traits in the network (See Online Methods). Bars are colored according to predicted direct and pleiotropic labels. Eight*

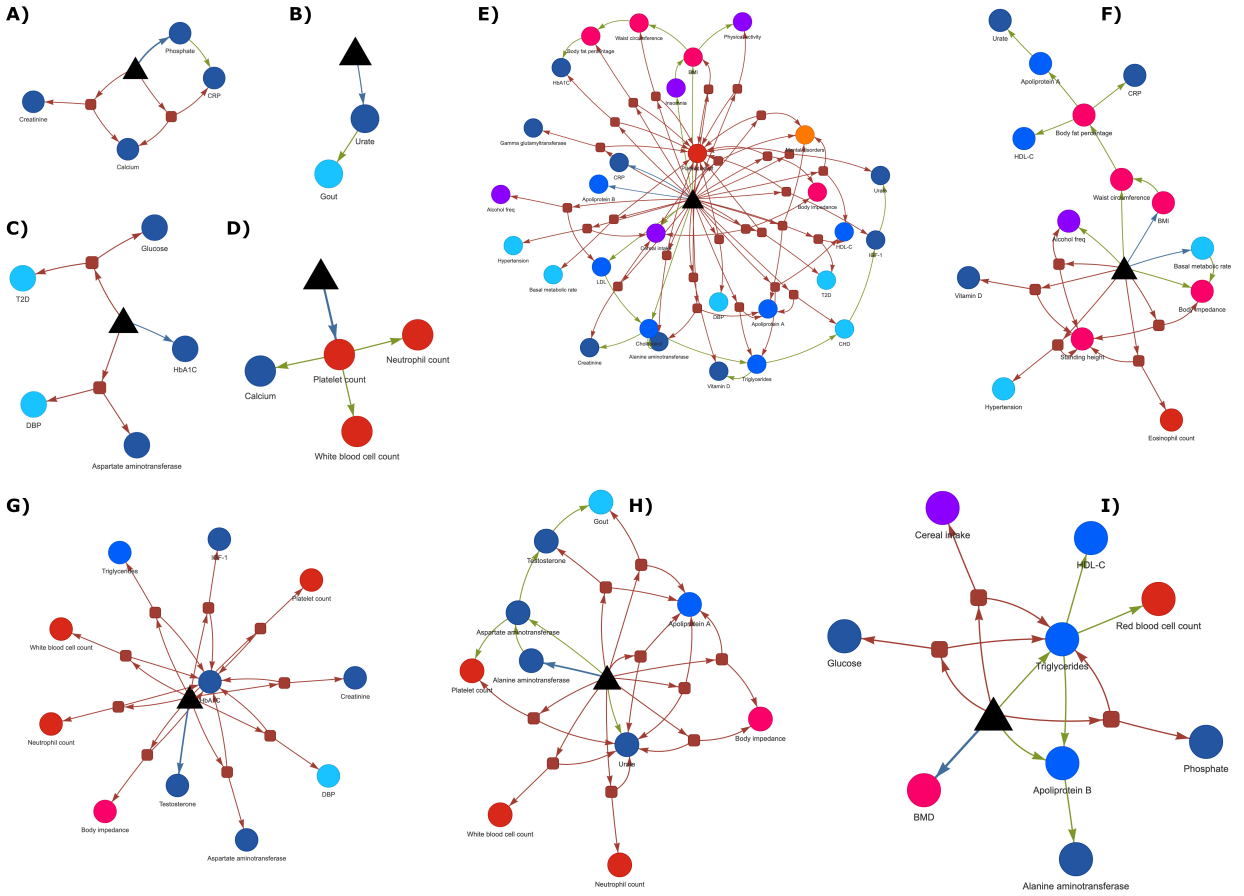
scenarios are represented across facets, with varying parameters. Polygenicity represents the proportion of variants with a direct effect on each trait. Effect on $B4$, the most pivotal trait, represents the proportion of effect passed to $B4$, for all traits with a non-zero vertical effect on $B4$. High targeted confounder means that few variants (0.01%) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounder means that a large proportion (5%) of variants have an effect on the confounder, but with low magnitude.



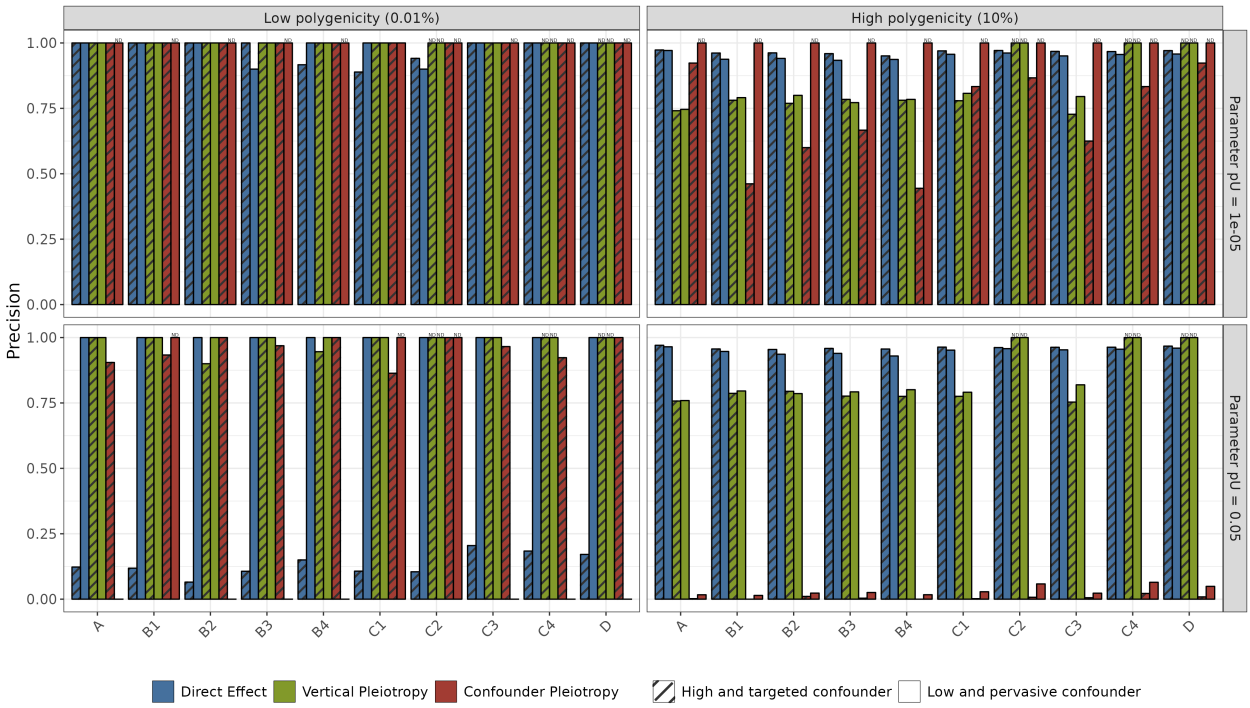
Supplementary Figure 18. *Precision and recall of PRISM predictions for significant variant-trait effects, on simulations without pleiotropy. The y-axis represents precision or recall in different conditions. The x-axis represents the simulated traits in the network (See Online Methods). Significant effects are defined with $P < 5 \times 10^{-8}/17$, the recommended threshold from PRISM (See Online Methods). Bars are colored according to predicted direct and pleiotropic labels. By convention, when variants neither exist nor were predicted for a given label, the accuracy or the recall equals 1, such cases are indicated with ND. Two scenarios are represented across facets, with varying polygenicity (the proportion of variants with a direct effect on each trait).*



Supplementary Figure 19. *PRISM predictions for true horizontal variant-trait effects, on simulations. The y-axis represents the proportion of prediction in each group, in different conditions. The x-axis represents the confounder effects. Each bar represents a scenario, with variant-trait predictions merged for each trait. Significant effects are defined with* $P < 5 \times 10^{-8}/17$*, the recommended threshold from PRISM (See Online Methods). Bars are colored according to the group the horizontal effect was predicted as. Eight scenarios are represented across facets, with varying parameters. Polygenicity represents the proportion of variants with a direct effect on each trait. Effect on* $B4$*, the most pivotal trait, represents the proportion of effect passed to* $B4$*, for all traits with a non-zero vertical effect on* $B4$*. High targeted confounder means that few variants (0.01%) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounder means that a large proportion (5%) of variants have an effect on the confounder, but with low magnitude.*



Supplementary Figure 20. *PRISM inferred causal network for a panel of validated genetic variants. Genetic variants are represented as black triangles. Red arrows are effects of the variant through a confounder meaning confounder pleiotropy. Green arrows are effects of variant through a causal trait meaning vertical pleiotropy. Blue arrows are direct causal effects from the variant to traits. Confounders are represented as red squares and traits are represented as circles colored according to trait categories.*



Supplementary Figure 21. *Precision of PRISM predictions for significant variant-trait effects, on simulations. The y-axis represents the precision which is the proportion of well-predicted variants among all predicted variants for a given label. The x-axis represents the simulated traits in the network (See Online Methods). Significant effects are defined with* $P < 5 \times 10^{-8}/17$*, the recommended threshold from PRISM (See Online Methods). Bars are colored according to predicted direct and pleiotropic labels. By convention, when no variants were predicted for a given label, the accuracy equals 1, such cases are indicated with ND. Eight scenarios are represented across facets, with varying parameters. Polygenicity represents the proportion of variants with a direct effect on each trait. Parameter pU represents PRISM chosen value for latent parameter* π_u*. High targeted confounder means that few variants (0.01%) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounder means that a large proportion (5%) of variants have an effect on the confounder, but with low magnitude.*

Supplementary Tables

Supplementary Table 1. *Parameters of all 32 scenarios, used to simulate GWAS summary statistics.*

h^2 AB	h^2 CDE	Trait Polygenicity	Confounder Polygenicity	Causal effect (except B4)	Causal effect (to B4)	Scenario
0.001	0.001	0.0001	0.05	0.3	0.3	1
0.001	0.6	0.0001	0.05	0.3	0.3	2
0.6	0.001	0.0001	0.05	0.3	0.3	3
0.6	0.6	0.0001	0.05	0.3	0.3	4
0.001	0.001	0.0001	0.0001	0.3	0.3	5
0.001	0.6	0.0001	0.0001	0.3	0.3	6
0.6	0.001	0.0001	0.0001	0.3	0.3	7
0.6	0.6	0.0001	0.0001	0.3	0.3	8
0.001	0.001	0.0001	0.05	0.3	0.05	9
0.001	0.6	0.0001	0.05	0.3	0.05	10
0.6	0.001	0.0001	0.05	0.3	0.05	11
0.6	0.6	0.0001	0.05	0.3	0.05	12

0.001	0.001	0.0001	0.0001	0.3	0.05	13
0.001	0.6	0.0001	0.0001	0.3	0.05	14
0.6	0.001	0.0001	0.0001	0.3	0.05	15
0.6	0.6	0.0001	0.0001	0.3	0.05	16
0.001	0.001	0.1	0.05	0.3	0.3	17
0.001	0.6	0.1	0.05	0.3	0.3	18
0.6	0.001	0.1	0.05	0.3	0.3	19
0.6	0.6	0.1	0.05	0.3	0.3	20
0.001	0.001	0.1	0.0001	0.3	0.3	21
0.001	0.6	0.1	0.0001	0.3	0.3	22
0.6	0.001	0.1	0.0001	0.3	0.3	23
0.6	0.6	0.1	0.0001	0.3	0.3	24
0.001	0.001	0.1	0.05	0.3	0.05	25
0.001	0.6	0.1	0.05	0.3	0.05	26
0.6	0.001	0.1	0.05	0.3	0.05	27
0.6	0.6	0.1	0.05	0.3	0.05	28
0.001	0.001	0.1	0.0001	0.3	0.05	29
0.001	0.6	0.1	0.0001	0.3	0.05	30
0.6	0.001	0.1	0.0001	0.3	0.05	31
0.6	0.6	0.1	0.0001	0.3	0.05	32

Supplementary Table 2. *61 heritable traits from UK Biobank. The first column is a trait category, the second column corresponds to the Trait code in UK Biobank, the third column corresponds to the description of the trait in UK Biobank.*

Category	Trait Code in UK Biobank	Description of the trait in UK Biobank
Anthropometric trait	23099_irnt	Body fat percentage
Anthropometric trait	21001_irnt	Body mass index (BMI)
Anthropometric trait	23106_irnt	Impedance of whole body
Anthropometric trait	50_irnt	Standing height
Anthropometric trait	3148_irnt	Heel bone mineral density (BMD)
Anthropometric trait	48_irnt	Waist circumference
Biomarker	30740_irnt	Glucose (mmol/L)
Biomarker	30620_irnt	Alanine aminotransferase (U/L)
Biomarker	30860_irnt	Total protein (g/L)
Biomarker	30600_irnt	Albumin (g/L)
Biomarker	30680_irnt	Calcium (mmol/L)
Biomarker	30850_irnt	Testosterone (nmol/L)
Biomarker	30810_irnt	Phosphate (mmol/L)
Biomarker	30890_irnt	Vitamin D (nmol/L)
Biomarker	30880_irnt	Urate (μmol/L)
Biomarker	30710_irnt	C-reactive protein (mg/L)
Biomarker	30750_irnt	Glycated hemoglobin (mmol/mol)

Biomarker	30700_irnt	Creatinine (μmol/L)
Biomarker	30770_irnt	IGF-1 (nmol/L)
Biomarker	30650_irnt	Aspartate aminotransferase (U/L)
Biomarker	30730_irnt	Gamma glutamyltransferase (U/L)
Blood trait	30140_irnt	Neutrophil count
Blood trait	30150	Eosinophil count
Blood trait	30120_irnt	Lymphocyte count
Blood trait	30010_irnt	Red blood cell (erythrocyte) count
Blood trait	30080_irnt	Platelet count
Blood trait	30250_irnt	Reticulocyte count
Blood trait	30130_irnt	Monocyte count
Blood trait	30160	Basophil count
Blood trait	30000_irnt	White blood cell (leukocyte) count
Blood trait	30180_irnt	Lymphocyte percentage
Cancer trait	II_NEOPLASM	Neoplasms
Cancer trait	C_PANCREAS	Malignant neoplasm of pancreas
Cancer trait	C3_SKIN	Malignant neoplasm of skin
Cancer trait	C50	Diagnoses - main ICD10: C50 Malignant neoplasm of breast
Education	6138_1	Qualifications: College or University degree
Education	20023_irnt	Mean time to correctly identify matches
Health	2178	Overall health rating
Inflammatory trait	K51	Diagnoses - main ICD10: K51 Ulcerative colitis
Lifestyle trait	1458	Cereal intake
Lifestyle trait	1468_4	Cereal type: Muesli
Lifestyle trait	1558	Alcohol intake frequency.
Lifestyle trait	1200	Sleeplessness / insomnia
Lifestyle trait	884	Number of days/week of moderate physical activity 10+ minu
Lifestyle trait	20116_0	Smoking status: Never
Lipid	30630_irnt	Apoliprotein A (g/L)
Lipid	30640_irnt	Apoliprotein B (g/L)
Lipid	30690_irnt	Cholesterol (mmol/L)
Lipid	30780_irnt	LDL direct (mmol/L)
Lipid	30870_irnt	Triglycerides (mmol/L)
Lipid	30760_irnt	HDL cholesterol (mmol/L)
Metabolic trait	I25	Diagnoses - main ICD10: I25 Chronic ischaemic heart disease
Metabolic trait	23105_irnt	Basal metabolic rate
Metabolic trait	4079_irnt	Diastolic blood pressure, automated reading
Metabolic trait	20002_1065	Non-cancer illness code, self-reported: hypertension
Metabolic trait	20002_1466	Non-cancer illness code, self-reported: gout
Metabolic trait	20002_1223	Non-cancer illness code, self-reported: type 2 diabetes
Psychiatric trait	V_MENTAL_BEHAV	Mental and behavioral disorders
Psychiatric trait	20127_irnt	Neuroticism score

Respiratory trait	3062_irnt	Forced vital capacity (FVC)
Respiratory trait	20002_1111	Non-cancer illness code, self-reported: asthma

Supplementary Results

PRISM vastly improves the interpretation of variant-trait associations detected in GWAS.

Genome-wide association studies (GWASs) identify significant associations between genetic variants and traits, while PRISM re-examines these associations through the prism of other traits. We aimed to assess the added value of PRISM compared to traditional GWAS results. To do so, in our simulations, we calculated GWAS precision and recall for all traits by comparing genome-wide significant associations with true direct effects. GWAS accuracy and recall were then compared with those from PRISM (Supplementary Fig. 14). First, in scenarios with low polygenicity, GWAS exhibited high recall but low precision to predict direct effects. Precision was particularly reduced in cases of strong pleiotropy, *i.e.* high confounder effects, and traits affected by vertical pleiotropy (A, $B1$-$B4$, $C1$, $C3$). In contrast, PRISM consistently demonstrated high recall and high precision across all traits, regardless of pleiotropic complexity. Second, in scenarios with high polygenicity, as individual genetic variant effect sizes were small, GWAS exhibited high precision but low recall. GWAS only detected the strongest effect sizes, *i.e.* predominantly direct effects. On the other hand, PRISM largely outperformed GWAS in recall while maintaining comparable precision. We conclude that PRISM, using GWAS results, can vastly improve the biological interpretation of significant associations detected in GWAS.

PRISM results are dependent on the set of processed traits.

In our simulations, we applied PRISM to the complex network of traits (See Fig. 6), but purposefully omitted the most pivotal trait ($B4$, See Online Methods) from input data. To keep the total number of traits processed by PRISM constant, we included a trait $E0$ with no causal interaction nor confounder with any other trait. As shown in Supplementary Fig. 15, the precision to predict direct effects decreased across all traits in scenarios with high confounder pleiotropy, since confounder variant effects induced by $B4$ were misclassified as direct for the other traits. Additionally, the precision to predict direct effects in $B3$ decreased in all scenarios, since the vertical pleiotropy from $B4$ to $B3$ was not included in the analysis. In this case, PRISM was unable to establish the mediation link with the omitted pivotal trait $B4$.

PRISM is robust to reasonable parameter estimation errors.

PRISM relies on global parameters from LHC-MR to estimate the relationships between traits. Therefore, we tested the robustness of PRISM to inaccurate estimations of the global parameters in the inference of causal variant networks. We added random noise within 20% of their original values to the global parameters. Supplementary Fig. 16 and 17 show that PRISM precision and recall are robust to these reasonable parameter

estimation errors. As an example, let us examine trait A under two scenarios, both characterized by low polygenicity and low vertical effects on trait B4. In the high confounder scenario, the causal effect of trait D on trait A was reduced by nearly 10%, whereas in the low confounder scenario, this effect was increased by more than 8%. Both scenarios exhibited reduced precision in predicting direct effects and reduced recall in predicting vertical effects on trait A. Overall, most classification errors were due to fluctuations in specific parameters estimating causal effects between traits, leading to misclassifications of vertical effects as direct effects. However, since estimating causal effects is the primary objective of LHC-MR, the estimation errors of the causal effect parameters are minimized in real applications.

PRISM accurately predicts direct variant-trait effects in the absence of pleiotropy.

In our simulations, we applied PRISM to the complex network of traits (See Fig. 6), but purposefully removed pleiotropy. Specifically, we set all causal effects between traits to zero (no vertical pleiotropy), set all confounders polygenicity to zero (no confounder pleiotropy), and forbid variants to exert direct effects on more than one trait (no horizontal pleiotropy). As shown in Supplementary Fig. 18, PRISM exhibited excellent precision and high recall in these very simple scenarios. In the highly polygenic scenario, a recall of approximately 50% may appear underwhelming. However, given the fixed total heritability and high polygenicity, there were thousands of direct variants, many of which had extremely small effect sizes that PRISM could not reliably distinguish from non-causal variants. We sought to determine the smallest effect sizes that PRISM could reliably detect. For a fixed level of total heritability, the scenario with high polygenicity involved many variants with small effects, while low polygenicity involved fewer variants with larger effects. Analyzing the distribution of effect sizes, we found that the smallest detectable effects had median magnitudes of ~1.9 and ~5.0 standard deviations for high and low polygenicity, respectively.

PRISM reliably detects horizontal pleiotropy in simulations.

In the real application, horizontal pleiotropy is surprisingly rare, accounting for only 0.2% of the observed pleiotropy in GWAS. In simulations, we investigated whether this scarcity could result from a bias in PRISM against horizontal pleiotropy. A horizontal variant-trait effect is defined as a direct effect from a variant with at least one other direct effect. Given the low probability of randomly generating horizontal pleiotropy in scenarios with low polygenicity, we introduced four true horizontal variants (See Supplementary Methods for details). In our simulations, we assessed PRISM ability to predict all true horizontal variant-trait effects (Supplementary Fig. 19). Our findings indicate that PRISM is adequately capable of detecting horizontal pleiotropy and does not exhibit bias against it.

PRISM direct variants show no distinctive features in large-scale genetic studies.

One of our key results is the enriched per-variant heritability observed in PRISM direct variants. Consequently, we investigated whether direct variants exhibit distinctive characteristics in other large-scale genetic studies. First, we compared CADD scores[1], that assess the deleteriousness of genetic variants, between direct, pleiotropic and GWAS variants. Second, we compared minor allele frequencies from GNOMAD[2] v4.0 between

direct, pleiotropic and GWAS variants, across African, American, and European ancestries. Third, we compared predicted pathogenicity scores of variants from AlphaMissense[3] between direct, pleiotropic and GWAS variants. In all three analyses, we found no significant differences between direct, pleiotropic, and GWAS variants. We concluded that direct variants were typical genetic variants that are neither less common, nor significantly more pathogenic or deleterious. The distinction between direct and pleiotropic variants lies in what they affect rather than the nature of their impact.

PRISM produces coherent results on a panel of literature-validated variants.

As mentioned in the Results section, we examined whether the variant-trait effects identified and labeled by PRISM had already been previously validated in the literature for a panel of variants. Here are some additional examples where PRISM inferred a causal network aligned with the literature.

Variant rs1697421 (Supplementary Fig. 20A) mapping to gene ALPL which encodes for Alkaline Phosphatase has been validated for its association with phosphorus levels. Additionally, ALPL has been studied as a potential biomarker for C-reactive protein (CRP)[4]. These findings highlight the significance of ALPL in metabolic and inflammatory processes.

The genetic variant rs1471633 (Supplementary Fig. 20B) is located in the PDZK1 gene, which encodes a protein involved in the regulation of urate transport[5]. Additionally, the variant rs1967017, which is in complete linkage disequilibrium with rs1471633, has been validated as impacting PDZK1 expression and altering the excretion of urate[6]. This is consistent with PRISM predictions, which indicate that rs1471633 directly affects urate levels, subsequently inducing a vertical effect on gout disease.

Variant rs560887 (Supplementary Fig. 20C) is located in the G6PC2 gene, coding for a protein involved in the gluconeogenic and glycogenolytic pathways[7]. Polymorphisms in G6PC2 have been validated in mice to alter fasting blood glucose levels, which are associated with type 2 diabetes (T2D)[8]. Interestingly, PRISM predicts a direct effect of this variant on HbA1c, a marker of blood glucose levels. It has been showed that variant rs560887 is linked to higher HbA1c in individuals with glucokinase mutations in MODY (Maturity Onset Diabetes of the Young) patients. Specifically, individuals who are GG homozygotes are more likely to meet the diagnostic criteria for diabetes based on their HbA1c levels[9]. Unfortunately, PRISM does not establish a vertical link between HbA1c and T2D, but does identify a confounder effect on glucose levels and T2D.

Variant rs6993770 (Supplementary Fig. 20D) maps to the ZFPM2 gene, which encodes a transcription factor of the FOG family, known for regulating hematopoiesis. This variant has been validated as causal for plateletcrit, which is the proportion of blood volume occupied by platelets. Although associated with multiple blood traits in GWAS, PRISM only finds a direct effect on platelet count and only vertical effects on the additional blood traits.

Variant rs429358 (Supplementary Fig. 20E) make allele APOE-$\varepsilon 4$ of apolipoprotein E. The allele is well known for its involvement in congnitive impairment, specifically Alzheimer's disease. But the allele has also been linked to the metabolic syndrome[10,11] which is a cluster of conditions that occur together, increasing the risk of heart disease, stroke, and type 2 diabetes, and includes high blood pressure, high blood sugar, excess

body fat around the waist, and abnormal cholesterol or triglyceride levels in full agreement with PRISM findings.

The rs12970134 variant (Supplementary Fig. 20F) is located in the MC4R (Melanocortin 4 Receptor) gene, coding for the melanocortin 4 receptor protein, a key regulator of energy homeostasis. This variant has been significantly associated with body mass index (BMI) in a case-control study[12]. Interestingly, the MC4R gene is heavily studied due to its critical role in obesity[13,14,15,16,17]. Notably, a recent article[18] have highlighted that loss-of-function variants in the MC4R gene have substantial impacts on BMI, weight, fat mass, and lean mass. These findings align with PRISM, which identifies a direct effect of this variant on BMI and basal metabolic rate. Consequently, a myriad of pleiotropic ripple effects is predicted, for example vertical effects on body impedance, waist circumference, body fat percentage, HDL-cholesterol, and Apolipoprotein A. Pleiotropic effects on inflammation markers, urate, and hypertension further link this variant to obesity-related complications.

The rs12150660 variant (Supplementary Fig. 20G) is located in the SHBG gene, coding for the Sex Hormone-Binding Globulin protein. This protein binds and regulates androgens and estrogens, exhibiting a particularly high affinity for testosterone[19]. PRISM does predict that this variant has a direct impact on testosterone levels. Additionally, PRISM forecasts potential confounder effects on various traits, such as diastolic blood pressure, inflammation markers, and body impedance. A review of the scientific literature suggests that these confounder effects might be linked to metabolic syndrome, as testosterone deficiency is implicated in its development[20].

Variant rs738408 (Supplementary Fig. 20H) maps to the PNPLA3 gene, coding for a triacylglycerol lipase involved in the hydrolysis of triacylglycerol in adipocytes. This variant is associated with NAFLD[21] (non-alcoholic fatty liver disease), and is in tight LD with rs738409, which accounts for large fraction of liver disease heritability[22]. However, the precise role of PNPLA3 in liver lipid metabolism remains unclear[23]. In any case, the link between these variants and liver diseases accounts for the numerous confounder effects on inflammation markers, body impedance, and even gout[24] predicted by PRISM. Furthermore, variant rs738409 is linked to variations in alanine aminotransferase levels[25], aligning with PRISM direct effect, though further details are lacking.

Genetic variant rs7741021 (Supplementary Fig. 20I) is located in the RSPO3 gene, coding for the R-spondin-3 protein. This variant has been significantly associated with bone mass[26]. Additionally, rs7741021 has been shown to affect circulating levels of R-spondin-3 protein, where elevated levels are linked to increased bone mineral density (BMD) and a substantially reduced risk of distal forearm fractures[27]. These results align with the direct effect of this variant on BMD as predicted by PRISM. Furthermore, PRISM predicts pleiotropic effects on lipid traits, including ApoB, triglycerides, and HDL-C levels. This can be explained by the role of RSPO3 in influencing peripheral adipose tissue storage capacity[28], providing insight into these additional effects.

Supplementary Methods

PRISM - Pleiotropic Relationships to Infer the SNP Model.

The PRISM method relies on the Mendelian Randomization (MR) model from LHC-MR[29,30]. LHC-MR is an integrative MR method that aims to infer the causal relationship of

a pair of traits from GWAS summary statistics, while taking into account a latent confounder. The principle of PRISM is to reroute the trait-level MR model from LHC-MR to infer information at the level of genetic variants. Thus, to introduce our PRISM model, we recapitulate some elements of the basic model from LHC-MR.

Part 1: Basic structural Equation Model

The structural equation model is defined by the following equations:

$$X = q_x \cdot U + \alpha_{y\to x} Y + \boldsymbol{G} \cdot \vec{\gamma}_x + \epsilon_X \qquad (A)$$

$$Y = q_y \cdot U + \alpha_{x\to y} Y + \boldsymbol{G} \cdot \vec{\gamma}_y + \epsilon_Y \qquad (B)$$

$$U = \boldsymbol{G} \cdot \vec{\gamma}_u + \epsilon_U \qquad (C)$$

In this model (See Supplementary Fig. 1):

- X and Y are continuous random variables representing two complex traits.
- U is a continuous random variable representing a latent heritable confounder, with causal effects q_x and q_y on X and Y, respectively.
- To simplify the notations, we assume that $E(X) = E(Y) = E(U) = 0$ and $Var(X) = Var(Y) = Var(U) = 1$.
- X and Y have causal effects on each other, denoted $\alpha_{x\to y}$ and $\alpha_{y\to x}$.
- The genome-wide genotype data for m genetic variants and n individuals are denoted by $\mathbf{G} \in \mathbb{R}^{n\times m}$. The vectors $\vec{\gamma}_x, \vec{\gamma}_y, \vec{\gamma}_u \in \mathbb{R}^m$ represent the true multivariate direct effects of all m genetic variants on X, Y and U, respectively.
- $\epsilon_X \sim \mathcal{N}(0, \sigma_x^2)$, $\epsilon_Y \sim \mathcal{N}(0, \sigma_y^2)$, and $\epsilon_U \sim \mathcal{N}(0, \sigma_u^2)$ are normally distributed and mutually independent error terms.

We assume that only a proportion $0 \leq \lambda_x, \lambda_y, \lambda_u \leq 1$ of the genome has a direct effect on X, Y and U, respectively. The vectors $\vec{\gamma}_x, \vec{\gamma}_y, \vec{\gamma}_u$ are modeled using a spike-and-slab distribution:

$$\vec{\gamma}_a = \vec{\xi}_a \odot \vec{\kappa}_a \qquad (D)$$

where $\xi_a \sim \mathcal{B}(m, \lambda_a)$ and $\kappa_a \sim \mathcal{N}\left(0, \frac{h_a^2}{m\lambda_a}\right)$ for $a = x, y, u$.

$\vec{\xi}_a, \vec{\kappa}_a \in \mathbb{R}^m$ are vectors of dimension m. Let A denote either trait X, Y, or U. The vector $\vec{\xi}_a$ consists of binary elements (0 or 1) indicating whether the genetic variant has an effect on A. The effect $\vec{\kappa}_a$ follows a Gaussian distribution. h_x^2, h_y^2, h_u^2 represent the heritability of X, Y et U, respectively. The symbol $\odot$ denotes the element-wise product. We assume zero covariance between the direct effects of a genetic variant on X, Y and U, thus $\mathbf{cov}(\vec{\gamma}_x, \vec{\gamma}_y) = \mathbf{cov}(\vec{\gamma}_x, \vec{\gamma}_y) = \mathbf{cov}(\vec{\gamma}_x, \vec{\gamma}_y) = 0$.

Part 2: Reparametrization of the summary statistics distributions.

We observe the univariate association summary statistics on traits X and Y from finite samples N_x and N_y, of size n_x and n_y respectively. The realizations of X, Y and U are denoted by $\vec{x}, \vec{y}, \vec{u} \in \mathbb{R}^{n_x}$. The genome-wide genetic data for n_x individuals with m variants

are represented by $\mathbf{G}_x \in \mathbb{R}^{n_x \times m}$. The genetic data for a variant k which is tested for association across n_x individuals are denoted $\vec{g}_k \in \mathbb{R}^{n_x}$. All variant genotypes are standardized to a standard normal distribution.

$$\mathbf{G} = \begin{pmatrix} g_1^1 & g_2^1 & \dots & \overrightarrow{g_k} & \dots & g_m^1 \\ \vdots & \vdots & & \vdots & & \vdots \\ g_1^i & g_2^i & \dots & g_k^i & \dots & g_m^i \\ \vdots & \vdots & & \vdots & & \vdots \\ g_1^{n_x} & g_2^{n_x} & \dots & g_k^{n_x} & \dots & g_m^{n_x} \end{pmatrix} \in \mathbb{R}^{n_x \times m}$$

The association summary statistics β_k^x for variant k on trait X can be written:

$$\beta_k^x = \frac{\langle \vec{g}_k, \vec{x} \rangle}{n_x}$$

Substituting $\vec{x}$ with its expression from equation (A), we obtain:

$$\beta_k^x = \frac{q_x \cdot \langle \vec{g}_k, \vec{u} \rangle + \alpha_{y \to x} \cdot \langle \vec{g}_k, \vec{y} \rangle + \langle \vec{g}_k, \mathbf{G}_x \cdot \vec{\gamma}_x \rangle + g_k \cdot \epsilon_x}{n_x}$$

Further substituting $\vec{y}$ and $\vec{u}$ with their expression from equations (B) and (C), we obtain:

$$\beta_k^x = \frac{q_x \cdot \langle \vec{g}_k, \mathbf{G}_x \cdot \vec{\gamma}_u + \epsilon_u \rangle}{n_x} + \frac{\alpha_{y \to x} \cdot \langle \vec{g}_k, \left(q_y \cdot \mathbf{G}_x \cdot \vec{\gamma}_u + \epsilon_u\right) + \alpha_{x \to y} \vec{x} + \left(\mathbf{G}_x \cdot \vec{\gamma}_y + \epsilon_y\right) \rangle}{n_x} + \frac{\langle \vec{g}_k, \mathbf{G}_x \cdot \vec{\gamma}_x \rangle + g_k \cdot \epsilon_x}{n_x}$$

$$\beta_k^x = \frac{q_x \cdot \langle \vec{g}_k, \mathbf{G}_x \cdot \vec{\gamma}_u \rangle}{n_x} + q_x \cdot \epsilon_k^u + \frac{\alpha_{y \to x} q_y \cdot \langle \vec{g}_k, \mathbf{G}_x \cdot \vec{\gamma}_u \rangle}{n_x} + \alpha_{y \to x} q_y \cdot \epsilon_k^u + \alpha_{y \to x} \alpha_{x \to y} \beta_k^x + \frac{\alpha_{y \to x} \langle \vec{g}_k, \mathbf{G}_x \cdot \vec{\gamma}_y \rangle}{n_x} + \alpha_{y \to x} \epsilon_k^y + \frac{\langle \vec{g}_k, \mathbf{G}_x \cdot \vec{\gamma}_x \rangle}{n_x} + \epsilon_k^x$$

because $\beta_k^x = \frac{\langle \vec{g}_k, \vec{x} \rangle}{n_x}$ and with: $\epsilon_k^x = \frac{\langle \vec{g}_k \cdot \epsilon_x \rangle}{n_x} \sim \mathcal{N}\left(0, \frac{\sigma_x^2}{n_x}\right)$, $\epsilon_k^u = \frac{\langle \vec{g}_k \cdot \epsilon_u \rangle}{n_x} \sim \mathcal{N}\left(0, \frac{\sigma_u^2}{n_x}\right)$, $\epsilon_k^y = \frac{\langle \vec{g}_k \cdot \epsilon_y \rangle}{n_x} \sim \mathcal{N}\left(0, \frac{\sigma_y^2}{n_x}\right)$

To simplify, we denote $\vec{\rho}_k = \frac{\mathbf{G}_x^\top \cdot \vec{g}_k}{n_x}$, as the Linkage Disequilibrium (LD) coefficients between variant k and all markers in the genome. So:

$$\beta_k^x = q_x \cdot \langle \vec{\rho}_k, \vec{\gamma}_u \rangle + q_x \cdot \epsilon_k^u + \alpha_{y \to x} q_y \cdot \langle \vec{\rho}_k, \vec{\gamma}_u \rangle + \alpha_{y \to x} q_y \cdot \epsilon_k^u + \alpha_{y \to x} \alpha_{x \to y} \beta_k^x + \alpha_{y \to x} \langle \vec{\rho}_k, \vec{\gamma}_y \rangle + \alpha_{y \to x} \epsilon_k^y + \langle \vec{\rho}_k, \vec{\gamma}_x \rangle + \epsilon_k^x$$

We denote $\eta_k^x = \left(q_x + \alpha_{y\to x} \cdot q_y\right) \cdot \epsilon_u^k + \alpha_{y\to x}\epsilon_y^k + \epsilon_x^k \sim \mathcal{N}\left(0, \frac{i_x}{n_x}\right)$, with $i_x = \left(q_x + \alpha_{y\to x} \cdot q_y\right)^2 \sigma_u^2 + \alpha_{y\to x}\sigma_y^2 + \sigma_x^2$. i_x is equivalent to the LD score regression intercept[31]. So:

$$\beta_k^x = \left(\alpha_{y\to x} q_y + q_x\right) \cdot \langle \vec{\rho}_k, \vec{\gamma}_u \rangle + \alpha_{y\to x} \langle \vec{\rho}_k, \vec{\gamma}_y \rangle + \langle \vec{\rho}_k, \vec{\gamma}_x \rangle + \alpha_{y\to x}\alpha_{x\to y}\beta_k^x + \eta_k^x$$

We hypothesize that $\alpha_{y\to x}\alpha_{x\to y} \approx 0$ because in real traits, either one of the two causal effects is zero, or the product of the bidirectional causal effects is very small. In addition, as mentioned in the LHC-MR paper, in case of high bidirectional effects, the interpretation of all parameters becomes delicate. So, regrouping all β_k^x on the left and factorizing:

$$\beta_k^x \approx \left(1 - \alpha_{y\to x}\alpha_{x\to y}\right)\beta_k^x = \left(\alpha_{y\to x} q_y + q_x\right) \cdot \langle \vec{\rho}_k, \vec{\gamma}_u \rangle + \alpha_{y\to x} \langle \vec{\rho}_k, \vec{\gamma}_y \rangle + \langle \vec{\rho}_k, \vec{\gamma}_x \rangle + \eta_k^x$$

Then, we substitute equation (D):

$$\beta_k^x = \left(\alpha_{y\to x} q_y + q_x\right) \cdot \langle \vec{\rho}_k, \vec{\xi}_u \odot \vec{\kappa}_u \rangle + \alpha_{y\to x} \langle \vec{\rho}_k, \vec{\xi}_x \odot \vec{\kappa}_x \rangle + \langle \vec{\rho}_k, \vec{\xi}_y \odot \vec{\kappa}_y \rangle + \eta_k^x$$

$$\beta_k^x = \left(\alpha_{y\to x} q_y + q_x\right) \cdot \langle \vec{\rho}_k \odot \vec{\xi}_u, \vec{\kappa}_u \rangle + \alpha_{y\to x} \langle \vec{\rho}_k \odot \vec{\xi}_y, \vec{\kappa}_y \rangle + \langle \vec{\rho}_k \odot \vec{\xi}_x, \vec{\kappa}_x \rangle + \eta_k^x$$

Assuming similar LD structures ($\vec{\rho}_k$), we apply the same reasoning for β_k^y:

$$\beta_k^y = \left(\alpha_{x\to y} q_x + q_y\right) \cdot \langle \vec{\rho}_k \odot \vec{\xi}_u, \vec{\kappa}_u \rangle + \alpha_{x\to y} \langle \vec{\rho}_k \odot \vec{\xi}_x, \vec{\kappa}_x \rangle + \langle \vec{\rho}_k \odot \vec{\xi}_y, \vec{\kappa}_y \rangle + \eta_k^y$$

with $\eta_k^x \sim \mathcal{N}\left(0, \frac{i_y}{n_y}\right)$ and $i_y = \left(q_y + \alpha_{x\to y} \cdot q_x\right)^2 \sigma_u^2 + \alpha_{x\to y}\sigma_x^2 + \sigma_y^2$

Therefore, the joint distribution of can be written as:

$$\begin{pmatrix} \beta_k^x \\ \beta_k^y \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ \alpha_{x\to y} \end{pmatrix} \cdot \langle \vec{\rho}_k \odot \vec{\xi}_x, \vec{\kappa}_x \rangle + \begin{pmatrix} \alpha_{y\to x} \\ 1 \end{pmatrix} \cdot \langle \vec{\rho}_k \odot \vec{\xi}_y, \vec{\kappa}_y \rangle + \begin{pmatrix} \alpha_{y\to x} \cdot t_y + t_x \\ \alpha_{x\to y} \cdot t_x + t_y \end{pmatrix} \cdot \langle \vec{\rho}_k \odot \vec{\xi}_u, \vec{\kappa}_u \rangle + \begin{pmatrix} \eta_x^k \\ \eta_y^k \end{pmatrix}$$

Following the same reasoning as the cross-trait LD-score regression[31], the noise term distribution is[29]:

$$\begin{pmatrix} \eta_x^k \\ \eta_y^k \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} i_x/n_x & \frac{corr(X,Y) \cdot n_{x\cap y}}{n_x \cdot n_y} \\ \frac{corr(X,Y) \cdot n_{x\cap y}}{n_x \cdot n_y} & i_y/n_y \end{pmatrix} \right)$$

Here, $n_{x\cap y}$ represents the sample overlap between the two studies used to calculate $\hat{\beta}_k^x$ and $\hat{\beta}_k^y$, and $corr(X,Y)$ is the phenotypic correlation between variables X and Y. $\frac{corr(X,Y) \cdot n_{x\cap y}}{\sqrt{n_x \cdot n_y}}$ is akin to the cross-trait LD Score regression intercept[32]. We set $corr(X,Y) = 0$. Our objective with PRISM is to identify pleiotropic variant-trait effects. We consider a variant-trait effect induced by the correlation between two traits not as noise, but as vertical pleiotropy. Furthermore, in the case of large sample size, the bias caused by sample overlap is expected to be small[33].

Part 3: Derivation of the likelihood function.

Let π_x, π_u, π_y denote the fractions of variants for which $\vec{\rho}_k \odot \vec{\xi}_x$, $\vec{\rho}_k \odot \vec{\xi}_u$, $\vec{\rho}_k \odot \vec{\xi}_y$ are non-zero respectively. These are the fractions of observed variants that are in non-zero LD with at least one causal variant for X, U, Y respectively. We assume that causal markers are sufficiently tagged by GWAS variants. Then $\pi_x > \lambda_x, \pi_u > \lambda_u, \pi_y > \lambda_y$, since LD propagates association signals beyond strictly causal markers. Next, we calculate the variance of the building blocks. For $a = x, u, y$:

$$Var\left(\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle\right) = E\left[\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle^2\right] - E\left[\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle\right]^2$$

Given that ρ_k, ξ_a, κ_a are independent:

$$Var\left(\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle\right) = E\left[\left(\vec{\rho}_k \odot \vec{\xi}_a\right)^2\right] \cdot E[\vec{\kappa}_a^2] - E\left[\vec{\rho}_k \odot \vec{\xi}_a\right]^2 \cdot E[\vec{\kappa}_a]^2$$

Since $E[\vec{\kappa}_a] = 0$:

$$Var\left(\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle\right) = E\left[\left(\vec{\rho}_k \odot \vec{\xi}_a\right)^2\right] \cdot E[\vec{\kappa}_a^2] = E[\vec{\rho}_k^2] \cdot E\left[\vec{\xi}_a^2\right] \cdot E[\vec{\kappa}_a^2] = l_k \cdot \lambda_a \cdot \frac{h_x^2}{\lambda_a m} = l_k \cdot \frac{h_x^2}{m}$$

where $l_k = \langle \vec{\rho}_k, \vec{\rho}_k \rangle$ corresponds to the LD score.

Since $\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle$ is modeled as a spike-and-slab Gaussian mixture model, with mixing proportion π_a, the non-zero component variance is the total variance divided by the mixing proportion. Thus, the distribution of the component can be expressed as:

$$\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle \sim \pi_a \cdot \mathcal{N}\left(0, l_k \cdot \frac{h_a^2}{m \cdot \pi_a}\right) + (1 - \pi_a) \cdot \mathcal{N}(0,0)$$

Here, π_a represents the proportion of genetic variants with a true causal effect on A or in non-zero LD with such variants, and $(1 - \pi_a)$ represents the variants with no effect and in zero LD with causal variants.

- Next, we partition the genomic variants into eight disjoint components based on their associations with X, Y and U:
  (0) No association
  (1) Associated with X
  (2) Associated with U
  (3) Associated with Y
  (4) Associated with X and U
  (5) Associated with X and Y
  (6) Associated with U and Y
  (7) Associated with X, Y and U
- We assume that the components are independent, so the proportion of variants in each component can be expressed as:
  (0) $\omega_0 = (1 - \pi_u)(1 - \pi_x)\left(1 - \pi_y\right)$

(1) $\omega_1 = \pi_x\,(1-\pi_u)(1-\pi_y)$

(2) $\omega_2 = \pi_u\,(1-\pi_x)(1-\pi_y)$

(3) $\omega_3 = \pi_y\,(1-\pi_u)(1-\pi_x)$

(4) $\omega_4 = \pi_x\,\pi_u\,(1-\pi_y)$

(5) $\omega_5 = \pi_x\,\pi_y\,(1-\pi_u)$

(6) $\omega_6 = \pi_y\,\pi_u\,(1-\pi_x)$

(7) $\omega_7 = \pi_x\,\pi_y\,\pi_u$

The joint distribution of $\begin{pmatrix}\beta_k^x\\ \beta_k^y\end{pmatrix}$ depends on the component from which variant k is drawn. For example, if variant k is drawn from component (1):

$$\begin{pmatrix}\beta_k^x\\ \beta_k^y\end{pmatrix} = \begin{pmatrix}1\\ \alpha_{x\to y}\end{pmatrix}\cdot\langle\vec{\rho}_k\odot\vec{\xi}_x,\vec{\kappa}_x\rangle + \begin{pmatrix}\eta_x^k\\ \eta_y^k\end{pmatrix}$$

because $\vec{\xi}_y = \vec{0}$ and $\vec{\xi}_u = \vec{0}$.

$$\begin{pmatrix}\beta_k^x\\ \beta_k^y\end{pmatrix} \sim \begin{pmatrix}1\\ \alpha_{x\to y}\end{pmatrix}\cdot\pi_x\cdot\mathcal{N}\left(0, l_k\cdot\frac{h_x^2}{m\cdot\pi_x}\right) + \begin{pmatrix}\eta_x^k\\ \eta_y^k\end{pmatrix}$$
$$= \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, l_k\cdot\frac{h_x^2}{m\cdot\pi_x}\begin{bmatrix}1 & \alpha_{x\to y}\\ \alpha_{x\to y} & {\alpha_{x\to y}}^2\end{bmatrix}\right) + \begin{pmatrix}\eta_x^k\\ \eta_y^k\end{pmatrix}$$

Following the same reasoning for all 8 components, and summing them according to their respective proportion ω_j:

$$\begin{pmatrix}\beta_k^x\\ \beta_k^y\end{pmatrix} \sim \sum_{i=0}^{7}\omega_j\cdot\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \boldsymbol{\Omega}_{ki}\right)$$

- where:
  - (0) $\boldsymbol{\Omega}_{k0} = \boldsymbol{\Sigma}_0$
  - (1) $\boldsymbol{\Omega}_{k1} = l_k\cdot\boldsymbol{\Sigma}_x + \boldsymbol{\Omega}_0$
  - (2) $\boldsymbol{\Omega}_{k2} = l_k\cdot\boldsymbol{\Sigma}_u + \boldsymbol{\Omega}_0$
  - (3) $\boldsymbol{\Omega}_{k3} = l_k\cdot\boldsymbol{\Sigma}_y + \boldsymbol{\Omega}_0$
  - (4) $\boldsymbol{\Omega}_{k4} = l_k\cdot(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_u) + \boldsymbol{\Omega}_0$
  - (5) $\boldsymbol{\Omega}_{k5} = l_k\cdot(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y) + \boldsymbol{\Omega}_0$
  - (6) $\boldsymbol{\Omega}_{k6} = l_k\cdot(\boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_y) + \boldsymbol{\Omega}_0$
  - (7) $\boldsymbol{\Omega}_{k7} = l_k\cdot(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_y) + \boldsymbol{\Omega}_0$

and:

$$\boldsymbol{\Sigma}_u = \frac{h_u^2}{m \cdot \pi_u}\begin{bmatrix} (\alpha_{y\to x} \cdot q_y + q_x)^2 & (\alpha_{y\to x} \cdot q_y + q_x) \cdot (\alpha_{x\to y} \cdot q_x + q_y) \\ (\alpha_{y\to x} \cdot q_y + q_x) \cdot (\alpha_{x\to y} \cdot q_x + q_y) & (\alpha_{x\to y} \cdot q_x + q_y)^2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_x = \frac{h_x^2}{m \cdot \pi_x}\begin{bmatrix} 1 & \alpha_{x\to y} \\ \alpha_{x\to y} & {\alpha_{x\to y}}^2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_y = \frac{h_y^2}{m \cdot \pi_y}\begin{bmatrix} {\alpha_{y\to x}}^2 & \alpha_{y\to x} \\ \alpha_{y\to x} & 1 \end{bmatrix}$$

$$\Sigma_0 = \begin{bmatrix} \frac{i_x}{n_x} & \rho_{x\cap y} \\ \rho_{x\cap y} & \frac{i_y}{n_y} \end{bmatrix}$$

Since the sum of two independent normally distributed random variables is Gaussian, when calculating the distribution of $\begin{pmatrix} \beta_k^x \\ \beta_k^y \end{pmatrix}$ for each component, the terms $\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle$ are either $\vec{0}$ or independent Gaussian distributions (as affecting X, Y or U is independent). Therefore, the sum of $\langle \vec{\rho}_k \odot \vec{\xi}_a, \vec{\kappa}_a \rangle$ terms is a Gaussian distribution, making each component modeled with a Gaussian distribution. It is worth mentioning that, although each component is modeled with a Gaussian distribution, the global model is not Gaussian. This is not an issue since likelihoods and posterior probabilities are calculated independently for each component.

We chose to use a model with 8 nested components, whereas it would be possible to use a simpler model with just 4 components (O, X, Y, and U). However, *U* is a latent variable and π_u is an unknown prior. We do not want to risk masking effects on X and Y because of U. For example, let us consider a genetic variant v with causal effects on X and U. With a 4-components model, v would have a high probability to belong to component X and to belong to component U. However, we would have no way to interpret these probabilities. We would not be able to compare probabilities to conclude if v affects X, U, or both, nor to conclude on a type of pleiotropy. In the 8-component model, v will have a high probability of belonging to model (4), $X\&U$, that we choose to classify as “direct pleiotropy”. In others words, this is an ingenious way of performing a fuzzy classification where categories are not strictly defined. This is why we preferred the more complex model, which can handle more subtle effects.

To account for the added complexity of the multi-effect components, we penalize each component with the prior probability ω_j for a variant to belong to the component. The more complex the component, the lower its prior, resulting in a greater penalty.

Part 4: Calculating posterior probabilities and scores for variants.

Parameters $\theta = i_x, i_y, \pi_x, \pi_y, h_x^2, h_y^2, q_x, q_y, \alpha_{x\to y}, \alpha_{y\to x}, \rho_{x\cap y}$ are estimated from the observed association summary statistics, using LHC-MR. Parameters n_x, n_y, m are known, h_u^2 is a latent parameter fixed to 1, π_u is a latent parameter fixed to 10^{-5}.

We chose $\pi_u = 10^{-5}$ based on simulations indicating that using a very low value for π_u had no adverse effects. As shown in Supplementary Fig. 21, setting a very low value for

π_u significantly improved the precision in detecting direct variants, regardless of the actual polygenicity of the confounder.

Applying Bayes' theorem, the probability that a given variant k belongs to the Gaussian component j is determined by the following posterior probability:

$$\hat{P}\left(k \in j | \hat{\beta}_k^x, \hat{\beta}_k^y\right) = \frac{\phi\left(\hat{\beta}_k^x, \hat{\beta}_k^y | 0, \widehat{\mathbf{\Omega}}_\mathbf{j}\right)\hat{P}(k \in j)}{\sum_{i=0}^{7} \phi\left(\hat{\beta}_k^x, \hat{\beta}_k^y | 0, \widehat{\mathbf{\Omega}}_\mathbf{j}\right)\hat{P}(k \in j)}$$

Here, $\widehat{\mathbf{\Omega}}_\mathbf{j}$ denotes the variance-covariance matrix of Gaussian component j. Function ϕ represents the joint probability density function of the bivariate normal distribution, and $\hat{P}(k \in j) = \omega_j$ is the prior probability of variant k to belong to component j, conditional on parameters θ. These probabilities will be use to differentiate between direct and pleiotropic effects in Part 6.

The probabilities for variant k to belong to the 8 Gaussian components are converted into scores, indicating whether k has no effect (denoted O), or an effect on X, or on Y.

No Effect: the score indicating no effect of variant k on neither X nor Y corresponds to the probability of belonging to component (0): $\hat{S}_k^O = \hat{P}\left(k \in \{\text{component } 0\} | \hat{\beta}_k^x, \hat{\beta}_k^y\right)$

Effect on X: the score indicating an effect of variant k on X corresponds to the highest probability among the components that include an effect on X, specifically components (1), (4), (5) and (7): $\hat{S}_k^X = \hat{P}\left(k \in \{\text{component } 1,4,5,7\} | \hat{\beta}_k^x, \hat{\beta}_k^y\right)$

Effect on Y: the score indicating an effect of variant k on Y corresponds to the highest probability among components indicating an effect on Y, specifically components (3), (5), (6) and (7): $\hat{S}_k^Y = \hat{P}\left(k \in \{\text{component } 3,5,6,7\} | \hat{\beta}_k^x, \hat{\beta}_k^y\right)$

$\vec{S}_k = (\hat{S}_k^O \quad \hat{S}_k^X \quad \hat{S}_k^Y) \in \mathbb{R}^3$ is a vector of dimension 3 that contains these scores, for each variant k.

$$\hat{\mathbf{S}} = \begin{pmatrix} \hat{S}_1^O & \hat{S}_1^X & \hat{S}_1^Y \\ \hat{S}_2^O & \hat{S}_2^X & \hat{S}_2^Y \\ \vdots & \vdots & \vdots \\ \hat{S}_m^O & \hat{S}_m^X & \hat{S}_m^Y \end{pmatrix} \in \mathbb{R}^{m \times 3}$$

$\hat{\mathbf{S}}$ is a matrix of dimensions $m \times 3$ stemming from the concatenation of all m $\vec{S}_k$ vectors.

Part 5: Trait-wise workflow of PRISM and classification of the genetic variants.

We detailed the process to obtain $\hat{\mathbf{S}}$ for two complex traits X and Y, using their GWAS summary statistics. The idea of PRISM is to apply this process to a large number of traits T pairwise. Consequently, $\frac{T(T-1)}{2}$ different $\hat{\mathbf{S}}$ matrices are obtained from the pairwise pipeline.

Then, for each variant k, we extract the scores $\hat{S}$ corresponding to each trait. For example, let us consider the variant-trait association of variant k on a new trait A, included in PRISM workflow. We previously observed the effect of variant k on trait A in $T-1$ contexts, trait A paired with all the other traits. So, we aggregate all previously calculated scores $\hat{S}_k^A$ and

$\hat{S}_k^O$ from all $\hat{\mathbf{S}}$ matrices containing A. As a result, we obtain $\vec{E}_k^A = \left(\hat{S}_k^{A_1} \quad \hat{S}_k^{A_2} \quad \cdots \quad \hat{S}_k^{A_{T-1}}\right) \in \mathbb{R}^{T-1}$ and $\vec{E}_k^O = \left(\hat{S}_k^{O_1} \quad \hat{S}_k^{O_2} \quad \cdots \quad \hat{S}_k^{O_{T-1}}\right) \in \mathbb{R}^{T-1}$. Vector $\vec{E}_k^A$ is the collection of $T-1$ observations of the score of the very same variant k to have a causal effect on trait A. Vector $\vec{E}_k^O$ is the collection of $T-1$ observations of the score of the very same variant k to have no effect on trait A.

Part 6: Statistical test of the variant-trait effect consistency.

Next, we compare these values using a paired Student test. We define μ_k^A as the average of all S_k^A and μ_k^O as the average of all S_k^O. The hypotheses of the test are $H_0: \mu_k^A = \mu_k^O$ and $H_1: \mu_k^A > \mu_k^O$.

By applying this test to all k variants, we obtain one p-value per variant. Variant k is considered significant if p-value$_k < \frac{5\times10^{-8}}{T-1}$. This threshold corresponds to a Bonferroni correction in addition to the usual GWAS significance threshold of $p < 5 \times 10^{-8}$.

Then, to label significant variants for trait A according to whether their effect is direct or pleiotropic, we follow a three-step procedure.

1) A significant variant is flagged with vertical pleiotropy on trait A if any trait B is causal to trait A, and $\hat{P}\left(k \in \{\text{component } 3\}|\hat{\beta}_k^x, \hat{\beta}_k^y\right) > \hat{P}\left(k \in \{\text{component } 0,1,2,4,5,6,7\}|\hat{\beta}_k^x, \hat{\beta}_k^y\right)$ in the causal inference model for pair (A, B). This means that the probability that the genetic variant k has an effect only on trait B is higher than all other possibilities.

2) A significant variant is flagged with confounder pleiotropy on trait A if, in at least one causal inference model involving trait A, $\hat{P}\left(k \in \{\text{component } 2\}|\hat{\beta}_k^x, \hat{\beta}_k^y\right) > \hat{P}\left(k \in \{\text{component } 0,1,3,4,5,6,7\}|\hat{\beta}_k^x, \hat{\beta}_k^y\right)$. This means that the probability that the genetic variant k has an effect on A only through a confounder is higher than all other possibilities.

3) All other significant variants are considered to have a direct effect on trait A.

The same procedure is applied to each trait. Finally, for each variant, we construct the causal network by creating a graph where nodes represent the variant and traits, and edges represent the relationships inferred from PRISM (direct, vertical, or confounder effects). We refine the causal network for each variant by removing vertical edges between the variant and specific traits, conditioned on other traits involved in vertical relationships. This process ensures that the causal pathways are accurately represented. Additionally, when a variant shows vertical effects on multiple traits through a common causal trait, we average the effect on the causal trait and eliminate redundant edges.

GWAS simulation

We started by creating an intricate network of 18 traits (See Online Methods, Fig. 6). We tested 32 different scenarios, varying parameters for heritability, polygenicity, and causal relationships, as detailed in Supplementary Table 1. To approximate a genome-wide model while maintaining computational feasibility, we opted for sets of $m = 100{,}000$ simulated variants. The standardized effects of all these variants on all the traits were

simulated, taking into account the network relationships between the traits. First, for each trait, we randomly selected genetic variants to have a true direct effect on trait X and on the confounders between trait X and other traits. We also randomly selected 4 genetic variants to have a true direct effect on two traits, specifically $B3\&C2$, $C3\&E4$, $B4\&E5$ and $E2\&E3$ to test for horizontal pleiotropy. Then, for the selected genetic variants, the true effects were drawn from a Gaussian distribution with parameters depending on the trait and the scenario. Specifically, true direct effects of all m variants on trait X were computed as:

$$\vec{\gamma}_x = \vec{\xi}_x \odot \vec{\kappa}_x$$

Here, $\xi_x \sim \mathcal{B}(m, \pi_x)$ consisted of binary elements (0 or 1) indicating the presence or absence of a genetic variant effect on trait X. The effect $\kappa_x \sim \mathcal{N}\left(0, \frac{h_x^2}{m\lambda_x}\right)$ followed a Gaussian distribution with variance depending on the heritability of X, h_x^2, and on the total number of variants with effect on X, $m\lambda_x$. Similarly, the true confounder effects of all m variants on all confounders U were computed separately as:

$$\vec{\gamma}_u = \vec{\xi}_u \odot \vec{\kappa}_u$$

Similarly, $\xi_u \sim \mathcal{B}(m, \pi_u)$ consisted of binary elements (0 or 1) indicating the presence or absence of a genetic variant effect on U. The effect $\kappa_x \sim \mathcal{N}\left(0, \frac{h_x^2}{m\lambda_x}\right)$ followed a Gaussian distribution with variance depending on the heritability of U, $h_u^2 = 1$, and on the total number of variants with effect on U, $m\lambda_u$.

Then, the true effects were propagated to the other T traits through vertical and confounder pleiotropy. Specifically,

$$\vec{\beta}_x = \vec{\gamma}_x + \sum_{y=1}^{T} \left(\left(q_x + b q_y \right) \odot \vec{\gamma}_{u_{xy}} \right) + b\vec{\gamma}_y$$

Additionally, the effects were propagated according to the LD structure of each variant. For example, for variant k in LD with all J variants with coefficients denoted r_j^2:

$$B_k^x = \beta_k^x + \sum_{j=1}^{J} \beta_j^x \, r_j^2$$

Finally, an error term was added:

$$\overrightarrow{\widehat{\beta_x}} = \vec{B}_x + \mathcal{N}\left(\vec{O}_m, med\left(\frac{\sqrt{(1 - B_x^2)}}{\sqrt{N}} \right) I_m \right)$$

This gives us, for all genetic variants, standardized effects $\overrightarrow{\widehat{\beta_x}}$ on all traits, and LD scores. The LD score of a variant is the sum of its LD with all variants (including itself). The LD structure that we used was derived from 1000 Genomes[34] LD data from chromosome 1. We chose to simulate small independent LD blocks, rather large LD blocks deemed computationally prohibitive for extensive simulations.

References

1. Schubach, M., Maass, T., Nazaretyan, L., Röner, S. & Kircher, M. CADD v1.7: Using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic Acids Research* **52**, D1143–D1154 (2024).

2. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

3. Tordai, H. *et al.* Analysis of AlphaMissense data in different protein groups and structural context. *Scientific Data* **11**, 495 (2024).

4. Webber, M., Krishnan, A., Thomas, N. G. & Cheung, B. M. Y. Association between serum alkaline phosphatase and C-reactive protein in the United States National Health and Nutrition Examination Survey 2005-2006. *Clinical Chemistry and Laboratory Medicine* **48**, 167–173 (2010).

5. Anzai, N. *et al.* The Multivalent PDZ Domain-containing Protein PDZK1 Regulates Transport Activity of Renal Urate-Anion Exchanger URAT1 via Its C Terminus. *Journal of Biological Chemistry* **279**, 45942–45950 (2004).

6. Ketharnathan, S. *et al.* A non-coding genetic variant maximally associated with serum urate levels is functionally linked to HNF4A-dependent PDZK1 expression. *Human Molecular Genetics* (2018) doi:10.1093/hmg/ddy295.

7. Ebert, D. H. *et al.* Structure and promoter activity of an islet-specific glucose-6-phosphatase catalytic subunit-related gene. *Diabetes* **48**, 543–551 (1999).

8. Boortz, K. A. *et al.* Functional Analysis of Mouse G6pc1 Mutations Using a Novel In Situ Assay for Glucose-6-Phosphatase Activity and the Effect of Mutations in Conserved Human G6PC1/G6PC2 Amino Acids on G6PC2 Protein Expression. *PLOS ONE* **11**, e0162439 (2016).

9. Borowiec, M. *et al.* HbA1c-based diabetes diagnosis among patients with glucokinase mutation (GCK-MODY) is affected by a genetic variant of glucose-6-phosphatase (G6PC2). *Diabetic Medicine: A Journal of the British Diabetic Association* **29**, 1465–1469 (2012).

10. Frisardi, V. Apolipoprotein E Genotype: The Innocent Bystander or Active Bridge Between Metabolic Syndrome and Cognitive Impairment? *Journal of Alzheimer's Disease* **30**, S283–S304 (2012).

11. Morris, J. K. *et al.* Effect of APOE $E4$ Genotype on Metabolic Biomarkers in Aging and Alzheimer's Disease. *Journal of Alzheimer's Disease* **58**, 1129–1135 (2017).

12. Batarfi, A. A. *et al.* MC4R variants Rs12970134 and Rs17782313 are associated with obese polycystic ovary syndrome patients in the Western region of Saudi Arabia. *BMC Medical Genetics* **20**, 144 (2019).

13. Clément, K. *et al.* Efficacy and safety of setmelanotide, an MC4R agonist, in individuals with severe obesity due to LEPR or POMC deficiency: Single-arm, open-label, multicentre, phase 3 trials. *The Lancet Diabetes & Endocrinology* **8**, 960–970 (2020).

14. Brouwers, B. *et al.* Human MC4R variants affect endocytosis, trafficking and dimerization revealing multiple cellular mechanisms involved in weight regulation. *Cell Reports* **34**, 108862 (2021).

15. Lotta, L. A. *et al.* Human Gain-of-Function MC4R Variants Show Signaling Bias and Protect against Obesity. *Cell* **177**, 597–607.e9 (2019).

16. Hatoum, I. J. *et al.* Melanocortin-4 Receptor Signaling Is Required for Weight Loss after Gastric Bypass Surgery. *The Journal of Clinical Endocrinology & Metabolism* **97**, E1023–E1031 (2012).

17. Farooqi, I. S. *et al.* Clinical Spectrum of Obesity and Mutations in the Melanocortin 4 Receptor Gene. *New England Journal of Medicine* **348**, 1085–1095 (2003).

18. Wade, K. H. *et al.* Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. *Nature Medicine* **27**, 1088–1096 (2021).

19. Dunn, J. F., Nisula, B. C. & Rodbard, D. Transport of Steroid Hormones: Binding of 21 Endogenous Steroids to Both Testosterone-Binding Globulin and Corticosteroid-Binding Globulin in Human Plasma. *The Journal of Clinical Endocrinology & Metabolism* **53**, 58–68 (1981).

20. Zitzmann, M. Testosterone deficiency, insulin resistance and the metabolic syndrome. *Nature Reviews Endocrinology* **5**, 673–681 (2009).

21. Najafi, M., Rafiei, A., Ghaemi, A. & Hosseini, V. Association between Rs738408, Rs738409 and Rs139051polymorphisms in PNPLA3 gene and non-alcoholic fatty liver disease. *Gene Reports* **26**, 101472 (2022).

22. Cherubini, A., Casirati, E., Tomasi, M. & Valenti, L. PNPLA3 as a therapeutic target for fatty liver disease: The evidence to date. *Expert Opinion on Therapeutic Targets* **25**, 1033–1043 (2021).

23. Pingitore, P. & Romeo, S. The role of PNPLA3 in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1864**, 900–906 (2019).

24. Khrais, A., Kahlam, A., Tahir, A., Shaikh, A. & Ahlawat, S. Outcomes of gout in patients with cirrhosis: A national inpatient sample-based study. *World Journal of Hepatology* **15**, 303–310 (2023).

25. Zhang, L. *et al.* *PNPLA3* polymorphisms (Rs738409) and non-alcoholic fatty liver disease risk and related phenotypes: A meta-analysis. *Journal of Gastroenterology and Hepatology* **30**, 821–829 (2015).

26. Correa-Rodríguez, M., Schmidt Rio-Valle, J. & Rueda-Medina, B. The *RSPO3* gene as genetic markers for bone mass assessed by quantitative ultrasound in a population of young adults. *Annals of Human Genetics* **82**, 143–149 (2018).

27. Nilsson, K. H. *et al.* RSPO3 is important for trabecular bone and fracture risk in mice and humans. *Nature Communications* **12**, 4923 (2021).

28. Loh, N. Y. *et al.* RSPO3 impacts body fat distribution and regulates adipose cell biology in vitro. *Nature Communications* **11**, 2797 (2020).

29. Darrous, L., Mounier, N. & Kutalik, Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *medRxiv* 2020.01.27.20018929 (2020) doi:10.1101/2020.01.27.20018929.

30. Darrous, L., Mounier, N. & Kutalik, Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *Nature Communications* **12**, 7274 (2021).

31. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291 (2015).

32. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236–1241 (2015).

33. Mounier, N. & Kutalik, Z. Bias correction for inverse variance weighting Mendelian randomization. *Genetic Epidemiology* **47**, 314–331 (2023).

34. Devuyst, O. The 1000 Genomes Project: Welcome to a New World. *Peritoneal Dialysis International: Journal of the International Society for Peritoneal Dialysis* **35**, 676–677 (2015).