

# Supplementary material: Inferring genetic variant causal network by leveraging pleiotropy

Martin Tournaire<sup>1</sup>, Asma Nouira<sup>1</sup>, Yves Rozenholc<sup>1</sup>, Marie Verbanck<sup>1</sup>

<sup>1</sup>UR 7537 BioSTM, Biostatistique, Traitement et Modélisation des données biologiques, Université Paris Cité.

## Supplementary Figures

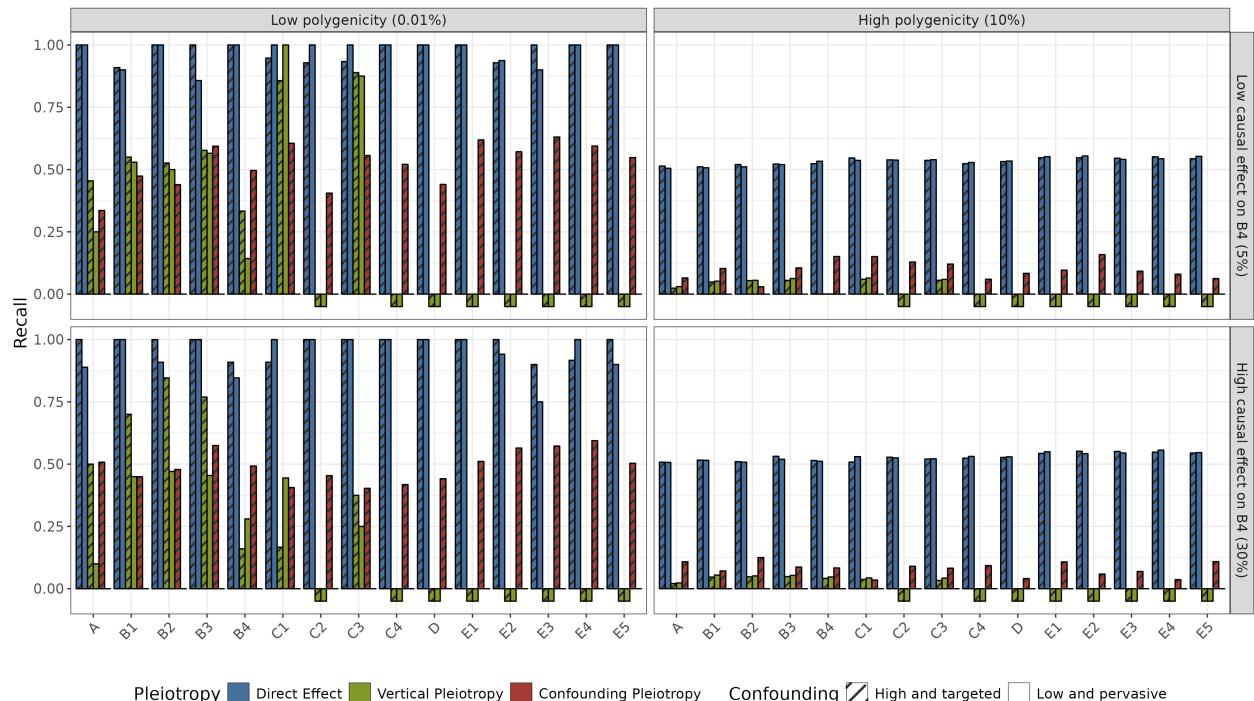
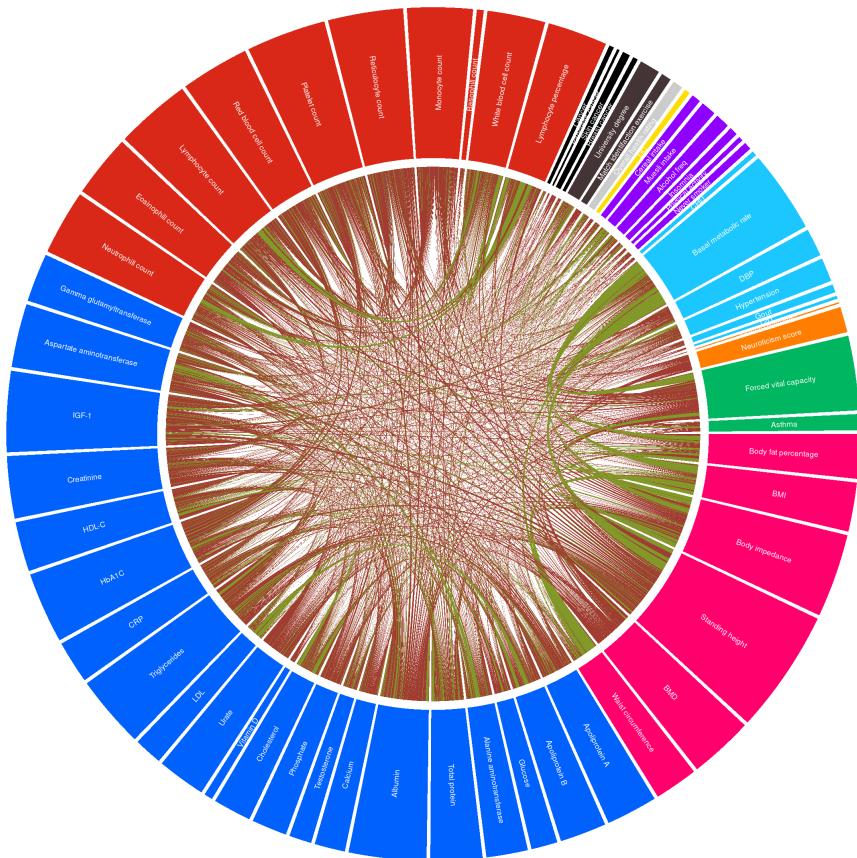


Figure 1: Recall of PRISM predictions for significant variant-trait effects, on simulations. The y-axis represents recall in different conditions. the x-axis represents the set of 15 simulated traits (See Online Methods and Fig. ??). Significant effects are defined with  $P < 5 \times 10^{-9}$ , which is slightly less strict than the PRISM recommended threshold. Bars are colored according to predicted direct and pleiotropic labels. A bar below zero means that PRISM predicted 0 effect in this category. 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 represents the proportion of effect passed to B4, for all traits with a non-zero vertical effect on B4. High targeted confounding means that few variants (0.01) have an effect on the confounder U, but with magnitude of effect rivaling direct effects. Low pervasive confounding means that a large proportion (5) of variants have an effect on the confounder, but with low magnitude.



**Figure 2: Circos plots of shared variants, on all PRISM processed traits. 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 confounding pleiotropy.**

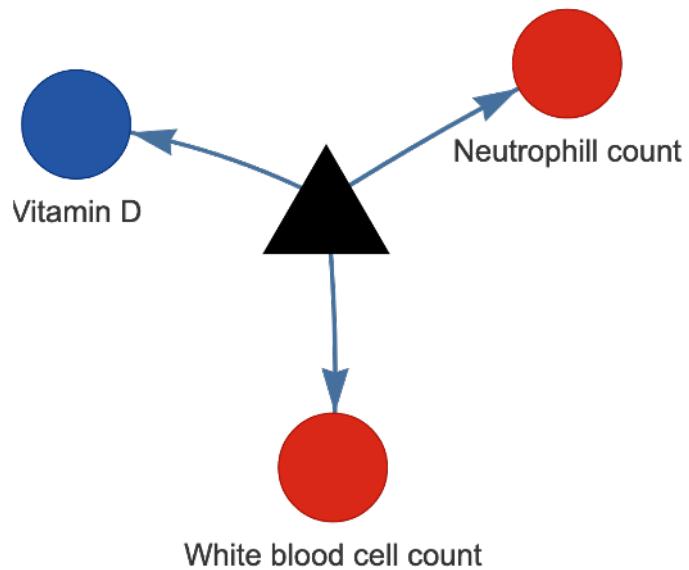
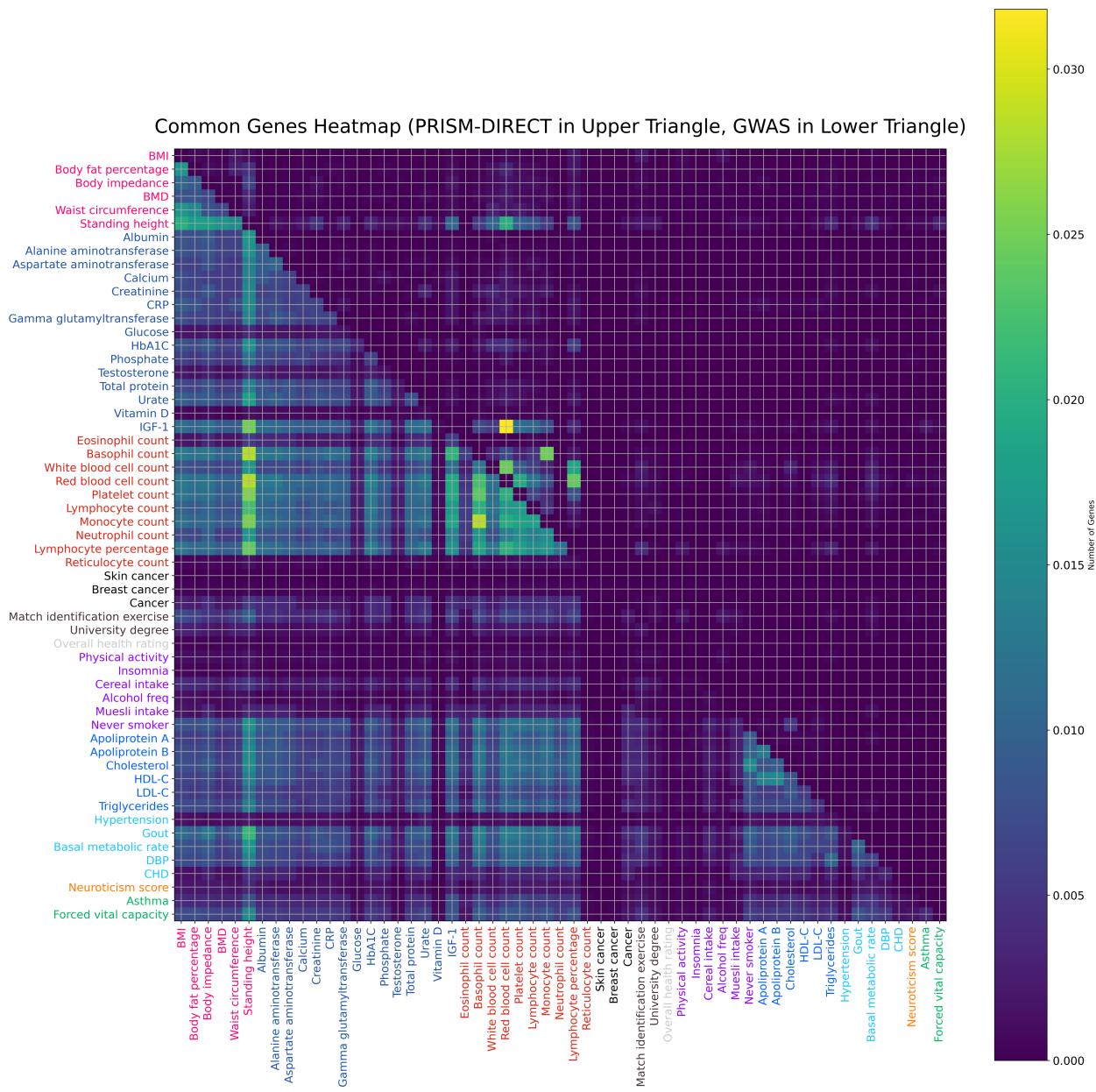
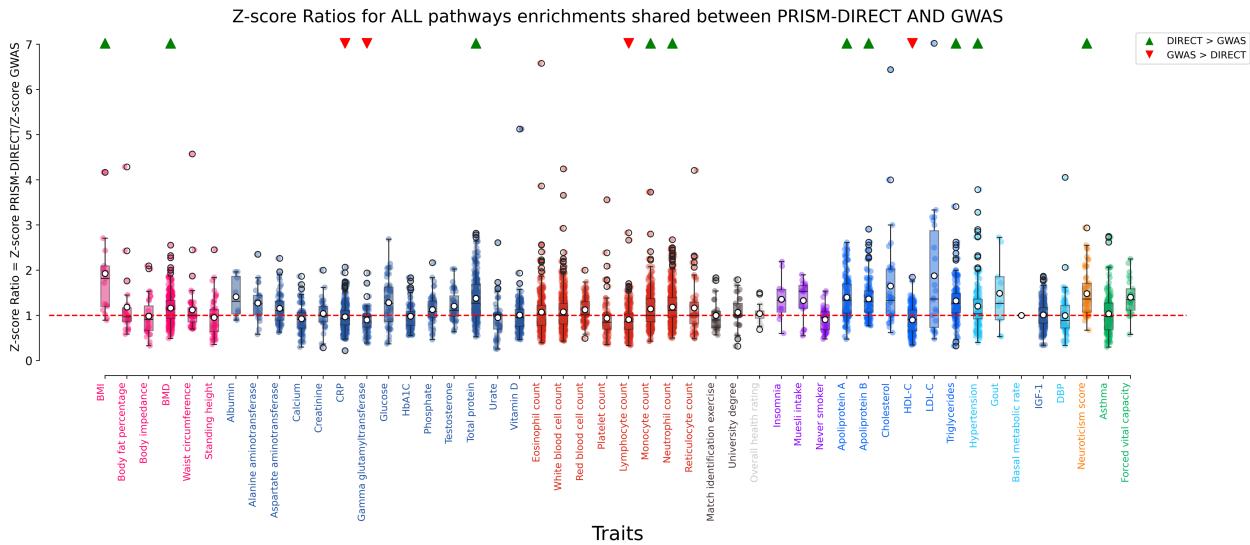


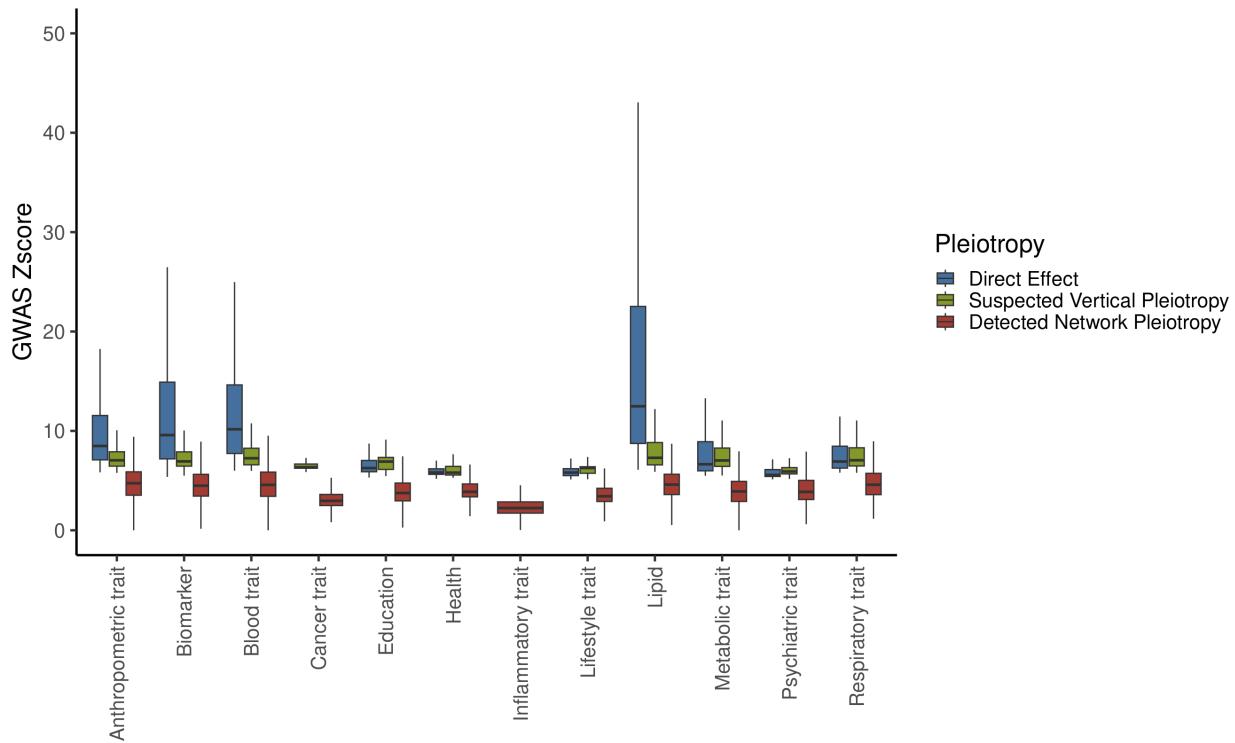
Figure 3: PRISM causal network of variants rs2282679/rs2298850. These variants are represented as a black triangle. Arrows represent causal effects. Red arrows are effects of variants through a confounder, represented as a red square, meaning confounding pleiotropy. Green arrows are effects of variants through traits, represented as circled colored by general category, meaning vertical pleiotropy. Blue arrows are direct causal effects from variants to traits. Variants rs2282679/rs2298850 have direct effects on vitamin D levels, white blood cells accumulation, and neutrophill accumulation.



**Figure 4:** Heatmap of genes shared between traits. The bottom-left triangle represents common GWAS mapped genes between 61 traits. The top-right triangle represents common PRISM direct mapped genes between the same traits. Each square is the intersection between two traits, and the color of this square represents the proportion of common genes between those two traits.



**Figure 5:** 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.



**Figure 6:** Initial GWAS Z-scores of labeled genetic variants, groupes by pleiotropic labels and traits categories. Y-axis represents a boxplot of the initial z-scores in GWAS summary statistics from UK Biobank, for robust genetic associations according to PRISM. X-axis represents traits grouped by categories.

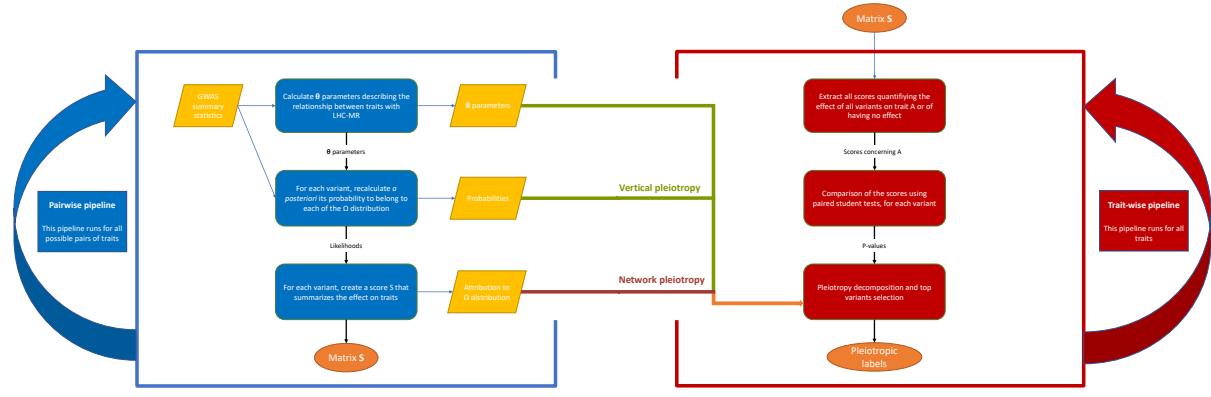


Figure 7: The PRISM pipeline is divided in two main steps, pairwise and traitwise pipeline. The left side represents the pairwise pipeline, whereas the right side represents the traitwise pipeline.

## Supplementary Tables

<b><math>h^2_{AB}</math></b>	<b><math>h^2_{CDE}</math></b>	<b>Polygenicity</b>	<b>Confounder Polygenicity</b>	<b>Causal except B4</b>	<b>Causal to B4</b>	<b>LD</b>	<b>pU Computation</b>
0.001	0.001	0.0001	0.05	0.3	0.3	1000G	0.00001
0.001	0.6	0.0001	0.05	0.3	0.3	1000G	0.00001
0.6	0.001	0.0001	0.05	0.3	0.3	1000G	0.00001
0.6	0.6	0.0001	0.05	0.3	0.3	1000G	0.00001
0.001	0.001	0.0001	0.0001	0.3	0.3	1000G	0.00001
0.001	0.6	0.0001	0.0001	0.3	0.3	1000G	0.00001
0.6	0.001	0.0001	0.0001	0.3	0.3	1000G	0.00001
0.6	0.6	0.0001	0.0001	0.3	0.3	1000G	0.00001
0.001	0.001	0.0001	0.05	0.3	0.05	1000G	0.00001
0.001	0.6	0.0001	0.05	0.3	0.05	1000G	0.00001
0.6	0.001	0.0001	0.05	0.3	0.05	1000G	0.00001
0.6	0.6	0.0001	0.05	0.3	0.05	1000G	0.00001
0.001	0.001	0.0001	0.0001	0.3	0.05	1000G	0.00001
0.001	0.6	0.0001	0.0001	0.3	0.05	1000G	0.00001
0.6	0.001	0.0001	0.0001	0.3	0.05	1000G	0.00001
0.6	0.6	0.0001	0.0001	0.3	0.05	1000G	0.00001
0.001	0.001	0.1	0.05	0.3	0.3	1000G	0.00001
0.001	0.6	0.1	0.05	0.3	0.3	1000G	0.00001
0.6	0.001	0.1	0.05	0.3	0.3	1000G	0.00001
0.6	0.6	0.1	0.05	0.3	0.3	1000G	0.00001

0.001	0.001	0.1	0.0001	0.3	0.3	1000G	0.00001
0.001	0.6	0.1	0.0001	0.3	0.3	1000G	0.00001
0.6	0.001	0.1	0.0001	0.3	0.3	1000G	0.00001
0.6	0.6	0.1	0.0001	0.3	0.3	1000G	0.00001
0.001	0.001	0.1	0.05	0.3	0.05	1000G	0.00001
0.001	0.6	0.1	0.05	0.3	0.05	1000G	0.00001
0.6	0.001	0.1	0.05	0.3	0.05	1000G	0.00001
0.6	0.6	0.1	0.05	0.3	0.05	1000G	0.00001
0.001	0.001	0.1	0.0001	0.3	0.05	1000G	0.00001
0.001	0.6	0.1	0.0001	0.3	0.05	1000G	0.00001
0.6	0.001	0.1	0.0001	0.3	0.05	1000G	0.00001
0.6	0.6	0.1	0.0001	0.3	0.05	1000G	0.00001

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

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 (umol/L)
Biomarker	30710_irnt	C-reactive protein (mg/L)
Biomarker	30750_irnt	Glycated haemoglobin (mmol/mol)
Biomarker	30700_irnt	Creatinine (umol/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	Neutrophill 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+ minutes
Lifestyle trait	20116_0	Smoking status: Never
Lipid	30630_irnt	Apolipoprotein A (g/L)
Lipid	30640_irnt	Apolipoprotein 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 behavioural disorders
Psychiatric trait	20127_irnt	Neuroticism score

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

---

*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*