A phenome-wide comparative analysis of genetic discordance between obesity and type 2 diabetes

Coral, Daniel E.; Fernandez-Tajes, Juan; Tsereteli, Neli; Pomares-Millan, Hugo; Fitipaldi, Hugo; Mutie, Pascal M.

Published in:
Nature Metabolism

DOI:
10.1038/s42255-022-00731-5

Publication date:
2023

nature metabolism

Article

https://doi.org/10.1038/s42255-022-00731-5

A phenome-wide comparative analysis of genetic discordance between obesity and type 2 diabetes

Received: 14 December 2021

Accepted: 20 December 2022

Published online: 26 January 2023

Check for updates

Daniel E. Coral[1] ✉, Juan Fernandez-Tajes[1], Neli Tsereteli[1], Hugo Pomares-Millan[1], Hugo Fitipaldi[1], Pascal M. Mutie[1], Naeimeh Atabaki-Pasdar[1,2], Sebastian Kalamajski[1], Alaitz Poveda[1], Tyne W. Miller-Fleming[3], Xue Zhong[3], Giuseppe N. Giordano[1], Ewan R. Pearson[1,4], Nancy J. Cox[2] & Paul W. Franks[1,5] ✉

Obesity and type 2 diabetes are causally related, yet there is considerable heterogeneity in the consequences of both conditions and the mechanisms of action are poorly defined. Here we show a genetic-driven approach defining two obesity profiles that convey highly concordant and discordant diabetogenic effects. We annotate and then compare association signals for these profiles across clinical and molecular phenotypic layers. Key differences are identified in a wide range of traits, including cardiovascular mortality, fat distribution, liver metabolism, blood pressure, specific lipid fractions and blood levels of proteins involved in extracellular matrix remodelling. We find marginal differences in abundance of Bacteroidetes and Firmicutes bacteria in the gut. Instrumental analyses reveal prominent causal roles for waist-to-hip ratio, blood pressure and cholesterol content of high-density lipoprotein particles in the development of diabetes in obesity. We prioritize 17 genes from the discordant signature that convey protection against type 2 diabetes in obesity, which may represent logical targets for precision medicine approaches.

Cardiometabolic diseases are the leading cause of death globally, with obesity and type 2 diabetes mellitus (T2D) accounting for a large proportion of this burden[1]. The prevalences of obesity and T2D have risen sharply over the past decades worldwide[2], corresponding with a shift to sedentary lifestyles and poor diet[3]. Even though obesity and T2D often coincide, their relationship is complex and remains incompletely understood. Indeed, while more than 80% of people with T2D also have obesity, 10–30% of people with obesity appear metabolically healthy[4–6]. Conversely, metabolic abnormalities occur in ~30% of normal-weight individuals[7–9]. Likewise, despite weight loss improving glycaemic control in people with T2D[10], when T2D occurs in people with normal weight, mortality rates are higher than those in people with overweight or obesity[11,12]. Here, we refer to these divergent features as 'discordant diabesity'. We focus on this unusual phenotype because it helps leverage the independent roles of excess adiposity and T2D in life-threatening disease.

To some extent, this discordance can be attributed to the imprecision with which body mass index (BMI), the conventional metric used to define obesity, characterizes adiposity[13,14]. For instance, even when BMI is comparable, lean and fat mass distributions often vary from one

[1]Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Department of Clinical Science, Lund University, Skåne University Hospital, Malmö, Sweden. [2]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [3]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. [4]Population Health and Genomics, University of Dundee, Dundee, UK. [5]Harvard T.H. Chan School of Public Health, Boston, MA, USA. ✉e-mail: daniel.coral@med.lu.se; paul.franks@med.lu.se

https://doi.org/10.1038/s42255-022-00731-5

person to the next[15]. Genetics has helped provide pathophysiological explanations for discordant diabesity, whereby, collectively, common variants affecting adipose distribution mimic monogenic syndromes such as familial lipodystrophies[16–21]. Expanding our knowledge of the phenotypic signature of discordant diabesity using the quantitative framework of genetics may help elucidate the mechanisms by which the broader health consequences of excess adiposity varies from one person to the next.

Here, we characterize genetically determined discordant diabesity through a comparative analysis with its concordant counterpart (that is, where higher genetic risk of obesity and T2D coincide). We used a range of machine learning methods to undertake phenome-wide scans to identify traits other than T2D that distinctively characterize these profiles. We concluded by undertaking robust causal inference analyses to determine the causal relationships underlying discordant diabesity with other features of health and disease.

Results

Assembly of concordant and discordant diabesity profiles

An analysis flowchart is presented in Extended Data Fig. 1a. We first identified genetic instruments for BMI[22] and T2D[23] by cross-referencing publicly accessible genome-wide association study (GWAS) summary statistics, finding 67 relatively independent single nucleotide polymorphisms (SNPs) strongly associated with both conditions ($P < 5 \times 10^{-8}$). After alignment to the BMI-increasing allele, these variants were labelled as 'concordant' (48 SNPs) or 'discordant' (19 SNPs) according to the positive or negative sign of their coefficients for T2D, respectively (Extended Data Fig. 1b and Supplementary Table 1; replication shown in Supplementary Table 2). Visual inspection of correlation patterns between BMI and T2D signals at each locus was undertaken using regional association plots (Supplementary Figs. 1 and 2).

Phenome-wide scans

Among the clinical phenotypes, we found that concordant and discordant diabesity profiles differed predominantly in cardiometabolic features including high-density lipoprotein (HDL) cholesterol, waist-to-hip ratio (WHR), waist circumference, and blood pressure (Fig. 1 and Supplementary Table 3). Generally, the discordant profile was associated with a favourable phenotypic signature compared to the concordant profile. For example, systolic blood pressure (SBP) was lower in the discordant compared to the concordant profile (SBP: $\beta_C = 0.002$ s.d. units per allele (95% confidence interval (CI): −0.001, 0.004), $\beta_D = -0.008$ s.d. units per allele (95% CI: −0.012, −0.004), $p\delta = 1.39 \times 10^{-4}$). We also found differences in risk of coronary heart disease (CHD) and stroke, which were lower in the discordant compared to the concordant profile (for example, CHD: odds ratio $(OR)_c = 1.01$ per allele (95% CI: 1.01, 1.02), $OR_D = 0.98$ per allele (95% CI: 0.97, 0.99), $p\delta = 1.3 \times 10^{-6}$). The levels of biomarkers of liver function such as gamma-glutamyl transferase (GGT) and alanine aminotransferase (ALT) enzymes were lower in the discordant relative to the concordant profile (for example, ALT: $\beta_C = 0.008$ s.d. units per allele (95% CI: 0.006, 0.011), $\beta_D = -0.011$ (95% CI: −0.019, −0.003), $p\delta = 2.07 \times 10^{-6}$). SHBG, a protein also produced in the liver, was higher in the discordant as opposed to the concordant profile ($\beta_C = -0.008$ s.d. units per allele (95% CI: −0.012, −0.004), $\beta_D = 0.013$ s.d. units per allele (95% CI: 0.007, 0.019), $p\delta = 1.94 \times 10^{-8}$). Additionally, the discordant profile was associated with higher mean corpuscular volume ($\beta_C = -0.002$ s.d. units per allele (95% CI: −0.005, 0), $\beta_D = 0.006$ s.d. units per allele (95% CI: 0.002, 0.01), $p\delta = 8.76 \times 10^{-4}$) and lower levels of urate ($\beta_C = 0.007$ s.d. units per allele (95% CI: 0.004, 0.01), $\beta_D = -0.005$ s.d. units per allele (95% CI: −0.01, −0.001), $p\delta = 3 \times 10^{-6}$) compared to the concordant profile. The odds of receiving treatment with alendronate was higher in the discordant than in the concordant profile, a drug indicated for osteoporosis ($OR_C = 0.99$ per allele (95% CI: 0.99, 0.99), $OR_D = 1.001$ per allele (95% CI: 1.001, 1.001), $p\delta = 3.26 \times 10^{-6}$).

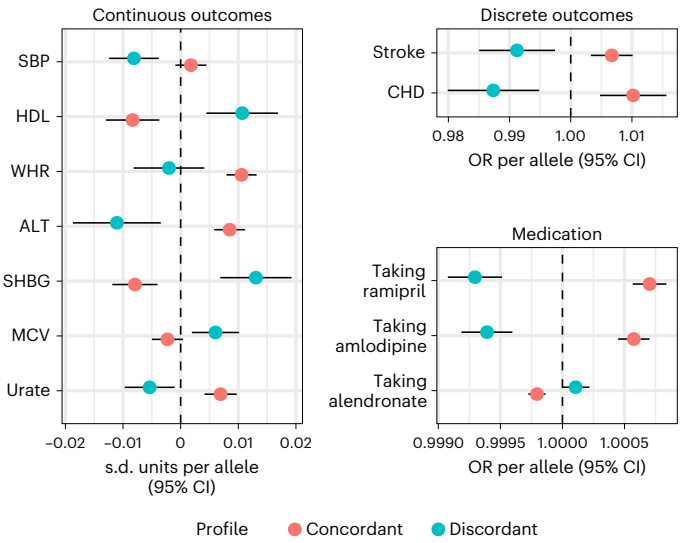


Fig. 1 | Summary-based comparison of concordant and discordant profiles. Concordant and discordant GRS coefficients for traits where we found differences between profiles using GWAS summary data. All are per-allele effect sizes, in s.d. units for continuous outcomes and ORs for binary traits (diseases and self-reported medication). Traits shown had at least one estimate significant after 5% FDR correction and the difference between profiles was also significant after 5% FDR. Statistical tests were based on a *z*-distribution and were two-sided. Bars show 95% CIs. Sample sizes vary for every trait ($N > 100{,}000$ for all traits).

Profile decomposition

Further exploration of the molecular features of the discordant and concordant profiles revealed that some variants used to characterize these profiles deviated from the overall pattern of trait association for their respective SNP set. Using single-linkage clustering on the SNP–trait matrix (Extended Data Fig. 1c), we identified two outliers in the concordant profile, one near *GCKR*, associated with higher SHBG and lower liver enzymes (SHBG: 0.07 (95% CI: 0.07, 0.08), $P = 7.5 \times 10^{-199}$) and a second near *TOMM40* associated with higher HDL (0.07 s.d. units per allele (95% CI: 0.06, 0.08), $P = 3.7 \times 10^{-107}$). In the discordant profile, the last variant to be aggregated to the clustering tree (that is, the SNP most distal from the other SNPs within its set) is located near *SLC2A2* and, in contrast to the overall discordant estimates, was associated with higher levels of AST and GGT (GGT: 0.02 s.d. units per allele (95% CI: 0.02, 0.03), $P = 2.7 \times 10^{-22}$).

External validation in BioVU

We sought replication of the discoveries outlined above in an independent European-ancestry cohort from BioVU, a de-identified collection of electronic health records and a linked biobank including inpatient and outpatient data from Vanderbilt University Medical Center (VUMC), a tertiary-care centre in Nashville, Tennessee, USA[24–26]. We constructed separate genetic risk score (GRS) coefficients for concordant and discordant profiles and assessed their association with multiple phenotypes (Fig. 2 and Supplementary Table 4). We first confirmed that the concordant and discordant GRSs were associated with higher obesity risk, respectively ($OR_C = 1.03$ per allele (95% CI: 1.03, 1.04), $OR_D = 1.02$ per allele (95% CI: 1.01, 1.02), $p\delta = 1.6 \times 10^{-3}$) and that the concordant and discordant profiles were positively and negatively associated with diabetes diagnosis, respectively ($OR_C = 1.03$ per allele, (95% CI: 1.02, 1.04), $OR_D = 0.95$ per allele (95% CI: 0.94, 0.96), $p\delta = 3.2 \times 10^{-49}$). Both scores were associated with increased odds of bariatric surgery ($OR_C = 1.05$ per allele (95% CI: 1.03, 1.06), $OR_D = 1.03$ per allele (95% CI: 1.008, 1.06), $p\delta = 0.24$). We found divergent associations in multiple diseases directly related to the main traits (for example, essential hypertension

https://doi.org/10.1038/s42255-022-00731-5

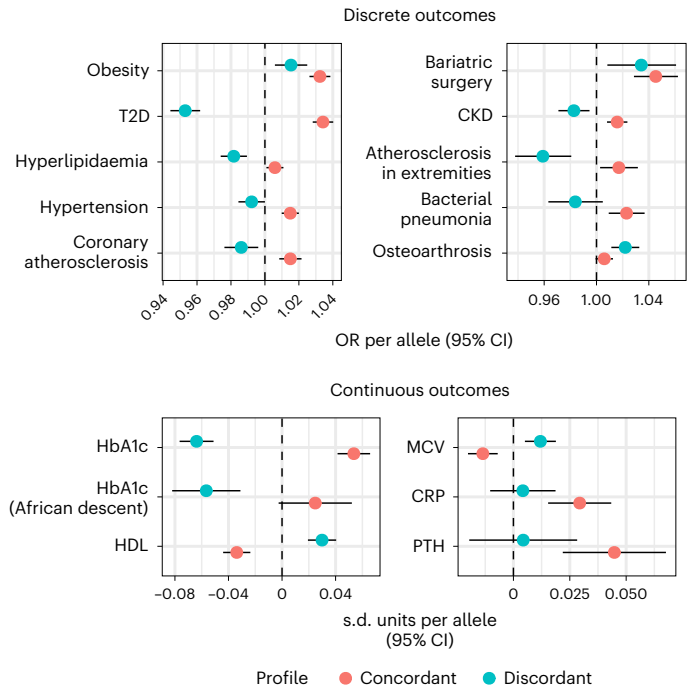


Fig. 2 | Comparison of concordant and discordant profiles in BioVU. Concordant and discordant GRS coefficients for traits where we found differences between profiles in BioVU. Analyses of disease endpoints included data for up to 48,544 individuals. Continuous outcomes included data for up to 68,724 and 13,661 individuals of European and African descent, respectively. All are per-allele effect sizes, in s.d. units for continuous outcomes and ORs for disease endpoints. Traits shown had at least one estimate significant after 5% FDR correction and the difference between profiles was also significant after 5% FDR. Statistical tests were based on a *z*-distribution and were two-sided. Bars show 95% CIs. CRP, C-reactive protein.

(HT): $OR_C = 1.014$ per allele (95% CI: 1.009, 1.019), $OR_D = 0.99$ per allele (95% CI: 0.98, 0.99), $p\delta = 1.2 \times 10^{-6}$). We also observed differences for other disease outcomes such as chronic kidney disease ($OR_C = 1.02$ per allele (95% CI: 1.01, 1.02), $OR_D = 0.98$ per allele (95% CI: 0.97, 0.99), $p\delta = 2.9 \times 10^{-6}$) and osteoarthrosis ($OR_C = 1.01$ per allele (95% CI: 1, 1.01), $OR_D = 1.02$ per allele (95% CI: 1.01, 1.03), $p\delta = 0.012$). Because both scores were also strongly associated with type 1 diabetes (T1D; $OR_C = 1.05$, (95% CI: 1.03, 1.05), $OR_D = 0.96$, (95% CI: 0.94, 0.97), $p\delta = 4.8 \times 10^{-17}$), we repeated the analyses excluding individuals with T1D. This attenuated the differences between concordant and discordant profiles for a small subset of traits including diabetic retinopathy and end-stage chronic kidney disease.

We also assessed the association of each GRS to multiple laboratory measurements in individuals of European ($n > 68{,}000$) and African American ($n > 14{,}000$) descent (Supplementary Table 5). The value per individual was computed as the median value over all measurements after a quality-control pipeline described in detail elsewhere[27]. Significant differences were found for several glycaemic traits consistent with the diabetes risk profiles (for example, HbA1c: s.d. difference per concordant allele: 0.05 (95% CI: 0.04, 0.06), s.d. difference per discordant allele: −0.06 (95% CI: −0.08, −0.05), $p\delta = 1.5 \times 10^{-39}$). We confirmed the difference between the two profiles in HDL and observed differences in the other two main lipid fractions (for example, triglycerides: s.d. difference per concordant allele: 0.02 (95% CI: 0.01, 0.03), s.d. difference per discordant allele: −0.04 (95% CI: −0.05, −0.03), $p\delta = 1.5 \times 10^{-14}$). The findings for red blood cell phenotypes were also replicated, and additional differences were found in leucocyte count, urea, creatinine, phosphate, C-reactive protein and parathyroid hormone (PTH), all of which were higher in carriers of concordant SNPs. Of the liver enzymes, only ALT values were available for comparison, whose levels were weakly associated with the concordant but not the discordant GRS (s.d. difference per concordant allele: 0.14 (95% CI: 0.02, 0.26), s.d. difference per discordant allele: 0.06 (95% CI: −0.08, 0.2), $p\delta = 0.2$). In individuals of African American descent, significant differences were found in HbA1c, glucose and urea levels in urine.

Differences in mortality in UK Biobank

We examined the relationship of GRSs to mortality owing to cardiovascular events in >337,000 participants of European descent from the UK Biobank (mean follow-up of 11.8 years). Around 35,000 deaths were reported, of which approximately 20% were related to cardiovascular events. The concordant GRS was associated with higher mortality (hazard ratio (HR) per allele: 1.01 (95% CI: 1.01, 1.02)), whereas the discordant GRS was not (HR per allele: 0.99 (95% CI: 0.98, 1.01), $p\delta = 0.02$). However, when assessing each SNP separately, we observed that the concordant variant near *TOMM40* was associated with lower incidence of cardiovascular mortality (HR per allele: 0.85 (95% CI: 0.81, 0.90), $P = 4.54 \times 10^{-9}$ and Supplementary Table 6).

Differences in serum metabolites

Of the metabolites available, those related to lipid subfractions were the strongest discriminators of concordant and discordant profiles (Fig. 3 and Supplementary Table 7). Discordant diabesity was associated with higher cholesterol in lipoprotein particles of all densities, while lower triglyceride content in lipoprotein particles of low densities, as opposed to concordant diabesity (for example, free cholesterol in HDL: $\beta_C = -0.008$ s.d. units per allele (95% CI: −0.01, −0.005), $\beta_D = 0.008$ s.d. units per allele (95% CI: 0.004, 0.013), $p\delta = 3.09 \times 10^{-10}$). Discordant diabesity also correlated with lower levels of branched-chain amino acids and aromatic amino acids, whereas in concordant diabesity they tended to be higher (total concentration of branched-chain amino acids: $\beta_C = 0.004$ s.d. units per allele (95% CI: 0.002, 0.008), $\beta_D = -0.008$ s.d. units per allele (95% CI: −0.012, −0.003), $p\delta = 1.46 \times 10^{-6}$).

Differences in gut microbiota

There were no differences between pooled concordant and discordant estimates for bacterial abundance in the gut that were statistically significant after false discovery rate (FDR) correction. Across ten taxa, several were nominally associated ($P < 0.05$) within either the concordant or the discordant profiles (Fig. 3 and Supplementary Table 8). Four of these belonged to the phylum Bacteroidetes (family Bacteroidaceae and geni *Bacteroides*, *Parabacteroides* and *Butyricimonas*), all of which were less abundant in discordant relative to concordant diabesity (for example, family Bacteroidaceae: $\beta_C = 0.005$ s.d. units per allele (95% CI: 0.001, 0.008), $\beta_D = -0.004$ s.d. units per allele (95% CI: −0.004, −0.01), $p\delta = 0.004$). The remaining taxa belonged to the phylum Firmicutes, most of them members of the obligately anaerobic class Clostridia, which tended to be more abundant in the discordant profile compared to the concordant profile (for example, genus *Subdoligranulum*: $\beta_C = -0.003$ s.d. units per allele (95% CI: −0.006, 0.001), $\beta_D = 0.006$ s.d. units per allele (95% CI: 0.007, 0.011), $p\delta = 0.006$). The family Lactobacillaceae was also lower in the discordant compared to the concordant profile ($\beta_C = 0.006$ s.d. units per allele (95% CI: 0.001, 0.01), $\beta_D = -0.006$ s.d. units per allele (95% CI: −0.014, 0.003), $p\delta = 0.02$).

Differences in serum protein levels

We found a significant difference between concordant and discordant estimates after FDR correction in a single protein: heparan sulfate 6-*O*-sulfotransferase 2 (HS6ST2), which was higher in discordant relative to concordant diabesity ($\beta_C = -0.01$ s.d. units per allele (95% CI: −0.017, 0), $\beta_D = 0.03$ s.d. units per allele (95% CI: 0.02, 0.04), $p\delta = 7.52 \times 10^{-7}$; Fig. 3). These analyses may be underpowered given that the effect of variants in *trans* is likely to be weaker than that of

https://doi.org/10.1038/s42255-022-00731-5

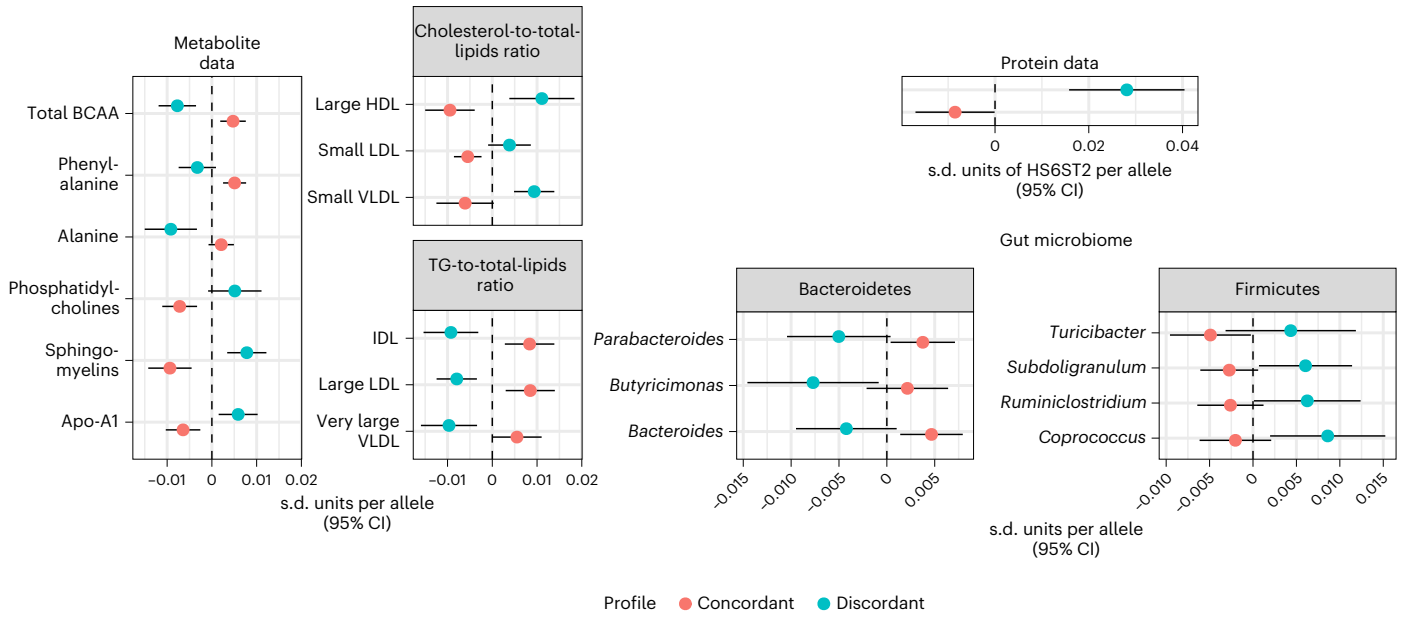


Fig. 3 | Comparison of concordant and discordant profiles in molecular phenotypes. Concordant and discordant GRS coefficients for traits where we found differences between profiles in molecular phenotypes. All are per-allele effect sizes, in s.d. units. Findings in metabolites shown here are derived from TwinsUK + KORA F4 (N = 7,824) and the UK Biobank (N = 115,078). Protein data were derived from the INTERVAL study (N = 3,301). Traits shown in these two domains had at least one estimate significant after 5% FDR correction, and the difference between profiles was also significant after 5% FDR. Statistical tests were based on a z-distribution and were two-sided. Bars show 95% CIs. Microbiome data came from the MiBioGen consortium (N = 18,340); the genii shown here had at least one estimate nominally significant, and the difference between estimates was also nominally significant (two-sided P < 0.05).

those in the gene encoding the protein. Thus, we also searched for strong *cis* effects ($P < 5 \times 10^{-8}$) in the discordant profile. We found one association between the discordant variant near *PPARG* and metalloproteinase inhibitor 4 (TIMP4; $\beta = -0.28$ s.d. units per allele (95% CI: −0.35, −0.2), $P = 5 \times 10^{-14}$).

Functional annotation using DEPICT

We used the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)[28] tool to compare the enrichment for tissues and biological pathways in each profile (Supplementary Figs. 3–5). The most notable difference was the significant enrichment ($P < 0.05$) for adipose tissue in the discordant profile, which was not found in the concordant profile. We also found significant enrichment for adrenal glands, ileum and kidney in the discordant but not in the concordant profile. Conversely, there was significant enrichment for endocrine tissue and retina in the concordant but not the discordant profile. Tissues for which there was significant enrichment in both profiles included pancreas and myometrium.

Gene expression and splicing in discordant diabesity

We found 506 genes whose expression/splicing was significantly influenced by concordant SNPs and 76 which were influenced by discordant SNPs across multiple tissues in GTEx[29]. In eQTLGen[30], we found significant associations of concordant SNPs with expression of 493 genes. Discordant SNPs were associated with 94 genes. Around 46% of all the associations found in GTEx were replicated in eQTLGen (47% of the genes associated with concordant SNPs; 39% of the genes associated with discordant SNPs).

To identify genes most likely involved in the molecular mechanisms leading to discordant diabesity, we chose genetic instruments for the 76 genes whose expression was influenced by discordant SNPs in the corresponding tissues in GTEx, and for the 94 genes in eQTLGen. A prerequisite for these instruments was that they are strongly associated with BMI ($P < 5 \times 10^{-8}$). We followed the SMR & HEIDI approach[31], which utilizes the strongest instrument for gene expression/splicing within the *cis* region of the corresponding gene (±500 Mb from the transcription start site) to calculate an estimate of the pleiotropic association across gene expression, BMI and T2D risk. This approach then determines if the association found reflects true pleiotropy rather than mere linkage by testing for heterogeneity of the estimates of SNPs in linkage disequilibrium (LD) with the lead SNP. We found 17 genes with robust expression signals for obesity and T2D whose directions of effect were in contrast (FDR-corrected $P < 5\%$, $p_{HEIDI} > 0.01$; Fig. 4 and Supplementary Table 9). To locate the most likely tissue of action for these genes, we followed a scoring procedure[32] through which a tissue specificity score is derived for each gene. This is calculated as (i) the proportion of median expression (in transcripts per million) across specific tissue types catalogued in GTEx and (ii) evidence of promoter/enhancer histone marks surrounding the genetic instruments, derived from multiple cell lines classified anatomically by the RoadMap Epigenomics Project[33] that could be mapped to tissue samples in GTEx. For each gene, we sorted tissues where we found pleiotropic links according to its specificity score and presence of promoter/enhancer signals for the genetic instrument. This allowed us to prioritize potential main action sites of relevance to discordant diabesity, for example, *LYPLAL1* in adipose tissue and *JAZF1* in vasculature and pancreas, while confirming the widespread effects of *SLC22A3* across multiple organs.

Discordant diabesity genes as therapeutic targets

We performed a lookup of the genes identified previously in the comprehensive public access databases DGIdb[34] and PHAROS[35]. Notably, there was evidence of interaction between three of the genes with strong pleiotropic associations (*SLC2A2*, *SLC22A3* and *KCNJ11*) and metformin in both databases. *SLC22A3* interacted with various quinoline derivatives (decynium-22, disprocynium-24, found in both databases), SarCNU (an antineoplastic drug in phase 2 clinical trials), derivatives of the alpha blocker phenoxybenzamine, corticosterone and colchicine. There is also evidence of potent inhibition of GLUT2, the protein product of *SLC2A2*, by a specific class of pyrazolopyrimidines. *SLC38A11*, *MAU2* and *FBXO46* are classified in the 'Tdark' level of target

https://doi.org/10.1038/s42255-022-00731-5

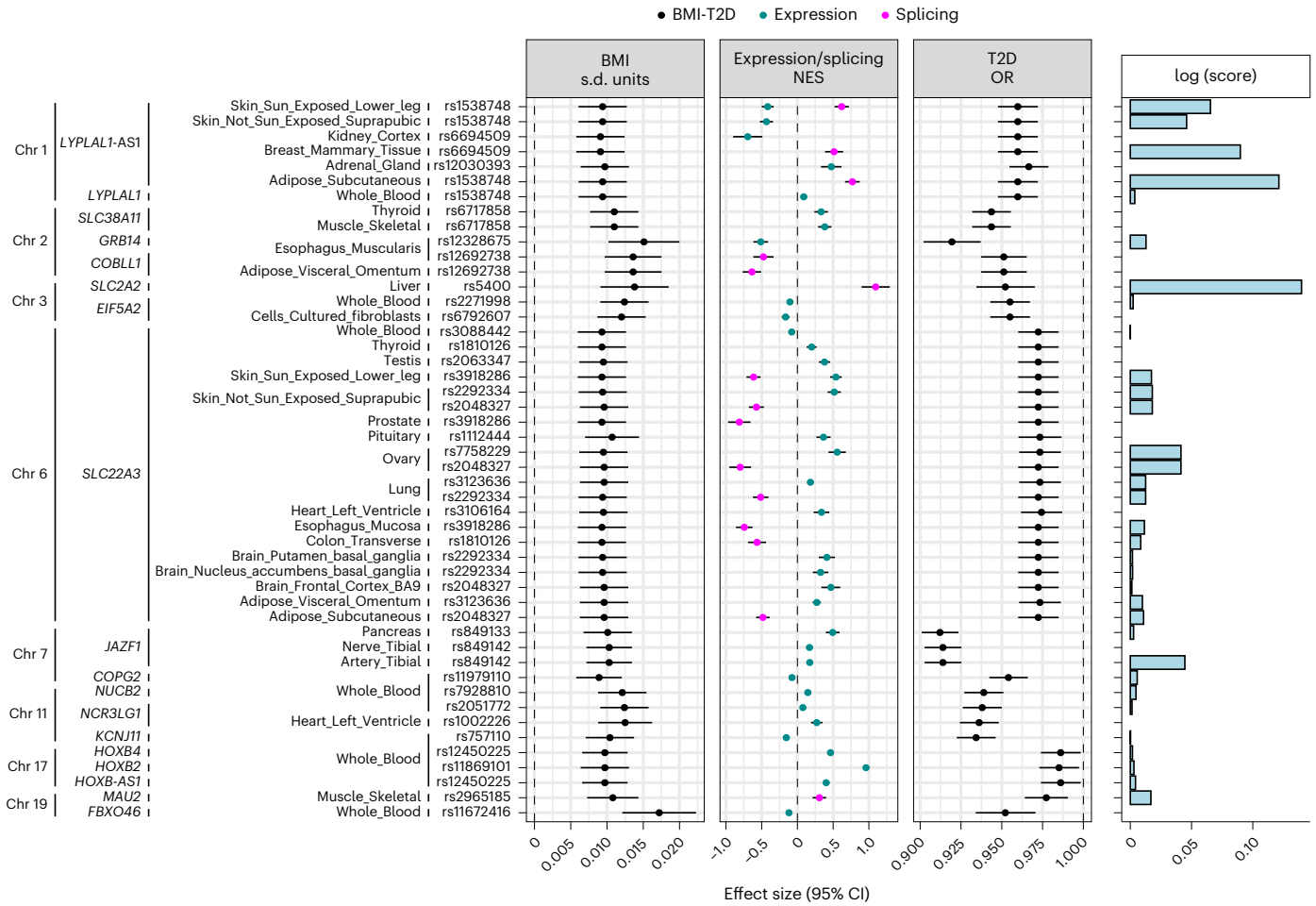


Fig. 4 | Genes with likely discordant pleiotropic effects on BMI and T2D. Genes with likely pleiotropic, yet discordant, effects on BMI and T2D, as found in the SMR & HEIDI analysis. Genes were sorted by their chromosome location and tissue where the pleiotropic association was found, as well as the lead expression quantitative trait loci (eQTL). The first three panels comprise the effect sizes of the lead eQTL on BMI (s.d. units), gene expression/splicing (normalized effect size) and T2D risk (OR), respectively. Bars represent 95% CIs. The right panel represents the logarithm of the tissue-of-action score. BMI data were derived from the GIANT + UK Biobank meta-analysis (N = 681,275). Gene expression data came from the GTEx (N = 838) and eQTLGen consortia (N = 31,684). T2D data came from the DIAGRAM meta-analysis dataset (N = 158,186).

development in the PHAROS database, composed of understudied targets, while the remaining genes fall under the 'Tbio' level, which includes targets with no known interactions yet satisfying other conditions, such as having functional annotations based on experimental evidence, repeated mentions in publications indexed in PubMed, and 50 or more available commercial antibodies.

Instrumental variable analyses

To quantify the potential impact of traits that emerged from the previous steps on offsetting the diabetogenic effect of obesity, we derived genetic instruments for each of these traits using SNPs that were also robustly associated with BMI ($P < 5 \times 10^{-8}$) and decomposed these instruments into two groups based on their direction of effect on the trait of interest after alignment to the BMI-increasing allele. We then constructed two GRSs, one for each group of variants, and calculated the T2D risk conferred by each GRS using summary data from the DIAGRAM consortium; we focused on GRSs that confer protection from T2D.

From the clinical phenotypes, the GRSs that conveyed higher BMI but lower WHR and SBP were significantly associated with lower T2D risk (Extended Data Fig. 2a and Supplementary Table 10). For example, the estimate for the GRS associated with higher BMI but lower WHR had an OR of 0.96 per allele (95% CI: 0.94–0.98, $P = 6.71 \times 10^{-5}$). Some traits in the clinical phenotypes required instruments to be in *cis* with the gene encoding the corresponding protein (for example, SHBG), to prevent confounding due to pleiotropy. We found two such instruments for ApoA1 and SHBG, respectively, which were not associated with T2D risk ($P = 0.17$ and 0.84, respectively; Supplementary Table 11) despite their strong association with higher BMI. From the analysis of metabolites, we found two GRS coefficients associated with higher BMI and lower T2D risk. The strongest protection was found for the GRS conferring higher total concentration of lipoprotein particles (OR: 0.98, 95% CI: 0.96, 0.99, $P = 0.006$; Supplementary Table 12), consistent with our findings in the phenome scans.

To test for the potential causal effect on diabesity discordance of HS6ST2 and TIMP4, the two proteins identified in the previous analysis, we searched for valid instruments (P value for both protein levels and BMI $< 5 \times 10^{-8}$) in the *cis* region of the corresponding genes. We could only derive a valid instrument for TIMP4. Using the SMR & HEIDI method, we found a significant pleiotropic effect ($P = 3.8 \times 10^{-7}$, $p_{HEIDI} = 0.4$; Extended Data Fig. 2b). However, we noted that the lead instrument and its closest proxies were located within *PPARG*, which is proximal to *TIMP4*.

No instruments for the microbial taxa where we found nominally significant differences reached the significance threshold required for BMI. Extending the exploration to other taxa revealed a single significant association ($P < 5 \times 10^{-8}$) of the A allele of rs1530559 (a variant

within the lactase persistence haploblock[36] in LD with the lactase functional variant rs4988235 ($r^2 = 0.4$)) with higher BMI and lower abundance of the order Bifidobacteriales. This variant was not associated with T2D risk ($P = 0.76$).

Individuals within the top decile for each profile

To determine the relevance of concordant and discordant profiles in people with obesity (≥30 kg/m^2), we focused on this subgroup in UK Biobank who localized to the top decile of one of the two profile GRSs[37]. Consistent with a binomial distribution, 18% of individuals with obesity were present in the two groups of extreme GRSs. The health characteristics of these individuals differed from all others with obesity (Supplementary Table 13) in several ways: for example, HbA1c levels in individuals with obesity and the extreme concordant GRS were higher compared to all other individuals with obesity (Kruskal–Wallis $P = 5.94 \times 10^{-12}$). Conversely, individuals with obesity and an extreme discordant GRS had significantly lower HbA1c compared to the rest of individuals with obesity ($P = 2.71 \times 10^{-42}$). Persons with obesity at both extreme GRSs are also distinguished from the rest by the main clinical features identified in our previous analyses such as SBP, HDL and ALT. WHR did not adequately separate the concordant or discordant extreme GRS from the wider group of people with obesity. However, because the initial phenome scan revealed a gender-specific difference in WHR between concordant and discordant profiles, (Extended Data Fig. 1c), we performed an additional analysis for WHR stratified by sex, where we found that women with extreme discordant GRS had significantly lower WHR compared to other women with obesity ($P = 6.05 \times 10^{-10}$).

Comparison with previous studies of discordant diabesity

We compared our results to those obtained in three previous investigations of discordant variants. For instance, Mahajan et al.[23] calculated the change in estimates of SNPs associated with T2D before and after adjustment for BMI. They found 15 loci where signals were enhanced after adjustment, which was attributed to discordance. Consistently, the SNP effects in the diabesity discordant profile derived here were enhanced, while those of the concordant profile were attenuated after adjustment (as described in ref. [38]; Supplementary Table 14). The change in SNP effect estimates was consistently associated with SNP effects on BMI ($R^2 = 0.8$; Supplementary Fig. 6). However, we observed that for four of the 19 SNPs (20%) from the discordant set near *PPARG*, *JAZF1*, *KCNJ11* and *LYPLAL1*, the change in SNP effect estimates was less than predicted. These discordant variants are those most likely to directly alter the relationship between BMI and T2D. This is consistent with the known effect of *PPARG* on adipocyte differentiation, and with our findings linking adipose tissue-specific gene expression at the *LYPLAL1* locus with higher BMI but lower T2D risk. Similarly, we found that *KCNJ11* and *JAZF1* had discordant effects on BMI and T2D, which is related to tissue-specific expression in heart and arteries; variants at both loci are known to influence insulin secretion.

We also sought replication of a finding from Pigeyre et al.[27] linking discordance to levels of the protein IGFBP-3 in blood. We were not able to replicate this finding (Supplementary Table 15), possibly due to differences in the characteristics of the cohorts where this relationship was found. For our analysis, we used summary data from the INTERVAL study[39], which includes predominantly healthy blood donors of European ancestry. In contrast, Pigeyre et al. used data from the ORIGIN trial, a cohort composed of individuals of European (47%) and Latin American (53%) ancestries, enriched for T2D cases (>80% had a prior diagnosis).

Finally, we searched for the SNPs comprising the concordant and discordant profiles described above in the cluster analysis of discordant SNPs performed by Huang et al.[21]. Fourteen of the 19 discordant SNPs identified in our analysis (78%) are among or in LD with the 62 SNPs identified by Huang et al ($r^2 > 0.1$ within a 1-Mb window, as specified in the publication; Supplementary Table 16). Two of the subclusters described by the authors were significantly overrepresented by these 14 SNPs: 5 (*ARAP1*, *ADCY5*, *PPARG*, *TCF7L2*, *KCNJ11-NCR3LG1*) were in the subcluster characterized mainly by higher BMI and lower fasting glucose and risk of T2D (enrichment $P = 1.6 \times 10^{-3}$) and 4 (*GRB14*, *LYPLAL1*, *ADAMTS9* and *VEGFA*) in the subcluster that conveyed an apparent protective effect on multiple cardiometabolic traits via peripheral adipose distribution (higher BMI and body fat percentage, and lower WHR; enrichment $P = 0.04$). Four concordant variants (at *GCKR*, *TOMM40*, *AKAP6* and *PPP1R3B-TNKS-MSRA*) were also among the 62 SNPs described by Huang et al. As opposed to other variants in the concordant set, the variant in *AKAP6* was associated with lower SBP (in ICBP GWAS: $\beta = -0.25$ mm Hg (95% CI: −0.38, −0.12), $P = 1.25 \times 10^{-4}$) and the variant near *PPP1R3B-TNKS-MSRA* was associated with higher HDL ($\beta = 0.02$ s.d. units (95% CI: 0.012, 0.027), $P = 1.72 \times 10^{-6}$). As we found and discussed in our analyses, *TOMM40* and *GCKR* deviate from the concordant set owing to their favourable associations with lipids and liver enzymes that resemble the discordant set, a pattern that was also reported by Huang et al.

Discussion

Obesity conveys heterogenous effects in cardiometabolic health, making disease prevention and management challenging. Here we used genetics to deconstruct the obesity phenotype into concordant and discordant diabesity, with strikingly different health characteristics beyond diabetes and obesity. Through transcriptomic, metabolomic and metagenomic analyses, we identified biomarkers that shed light on mechanisms of action and may aid risk stratification. Further analyses identified potential targets for drug development and drug repurposing.

Obesity and T2D often coalesce, owing largely to the mediating effect of peripheral insulin resistance caused by excess adiposity. The trait discordances described here reflect mechanisms involved in uncoupling obesity risk from T2D risk, thereby exposing diabetes-independent pathways through which obesity affects disease risk, for example, through adipose distribution. It is likely that both a higher capacity to expand adipose tissue in the gluteo-femoral compartment[40,41] and lower abdominal region around organs such as the liver, which might underlie the difference seen in biomarkers of liver failure[42], play important and independent roles in genetically determined diabesity discordance.

Another key phenotypic distinction between concordant and discordant profiles concerns blood pressure. Although T2D often causes vascular dysfunction, changes in the vascular bed may also precede metabolic perturbations through nutrient and hormonal flux[43,44], affecting pancreas, muscle and adipose tissue[45]. For instance, capillary recruitment and permeability are key determinants of whole-body glucose uptake and glycaemic variation[46].

Our findings relating to lipid metabolites support the use of more refined profiling of lipid subfractions to help determine risk in people with obesity. The cholesterol content of HDL particles and BCAA levels appear especially informative biomarkers[47], possibly because they enhance glucose homeostasis in obesity by improving cross-talk between peripheral tissue and the liver[48].

Despite the contrasting health consequences of the two diabesity profiles, bariatric surgery was equally likely, which may predispose one group to health benefits following surgery, whereas the other may not benefit in this way.

We found a significant difference between concordant and discordant profiles in levels of HS6ST2, a protein expressed in brain, kidney and ovaries, which in animal knock-out models shows a strong association with increased body weight and insulin resistance, possibly owing to enhanced adipocyte differentiation[49,50]. We found only one robust pleiotropic effect for discordant diabesity at *TIMP4*, which is proximal to *PPARG*, the likely causal gene. Moreover, PPARG activator medication inhibits matrix metalloproteinases[51,52]. *TIMP4* has been associated with adipogenesis, possibly through its effect on the adipose tissue extracellular matrix in obesity[53].

The colocalization analyses underscore the importance of tissue pleiotropy and tissue cross-talk in the molecular mechanisms of diabesity discordance. This is especially evident for *SLC22A3*, but also for other potential targets such as *LYPLAL1*, whose differential expression in both adipose tissue and adrenal glands appears linked to discordant diabesity. Moreover, three of the genes with pleiotropic links to T2D risk (*SLC2A2*, *SLC22A3* and *KCNJ11*) interact with metformin. This suggests a potential effect of metformin in shifting individuals with obesity from a concordant to a discordant phenotype. *SLC2A2* encodes GLUT2, which is part of the glucose sensor apparatus in pancreas and liver and is involved in intestinal glucose absorption in the gut[54]. Variants in this gene have been associated with preference for sugary foods[55] and modified response to metformin[56,57]. *SLC22A3* encodes OCT3, a protein widely expressed across tissues that aids adipocyte beiging[58] and perivascular adipose tissue remodelling[59]. *KCNJ11* encodes the Kir6.2 subunit of the ATP-sensitive potassium channel. As this is the target of sulphonylureas, this group of drugs may also harbor potential candidates for the phenotype shift to discordance in diabesity. The other ligands identified in the lookup may also constitute potential therapeutic agents to prevent cardiometabolic complications in obesity. For the rest of the genes, especially those in the 'Tdark' level in PHAROS, follow-up functional experiments in the tissues indicated by the lead genetic instruments and its corresponding epigenetic annotations are warranted.

Certain SNPs deviate from the overall association pattern of the profile within which they reside. In the concordant profile, the BMI-increasing allele of the variant near *TOMM40* increases T2D risk but, unlike other SNPs in the same profile, is associated with a better lipid profile and lower cardiovascular disease mortality. This gene and others in its proximity (*APOC1* and *APOE*) have been consistently implicated in lipid metabolism[60]. In the discordant profile, the variant *SLC2A2* conveys protection against T2D risk despite being associated with heavier weight and higher blood pressure, and worse liver function and dyslipidaemia. The opposite pattern was observed in the concordant variant in *GCKR*, which encodes a regulatory protein that inhibits glucokinase. This reflects disparate phenotypic effects of modulating the glucose sensor apparatus at different levels[54]. Deeper characterization of these mechanisms can further improve obesity stratification.

Although no statistically robust differences were observed in gut microbiota between the two diabesity profiles, possibly owing to low statistical power, nominal differences emerged in taxa belonging to the Bacteroidetes and Firmicutes phyla, which together constitute 90% of the human intestinal flora[61]. Our results indicate higher Firmicutes and lower Bacteroidetes abundance in discordant diabesity, which may result in enhanced production of short-chain fatty acid species such as butyrate, which is involved in glucose-lowering and anti-inflammatory mechanisms[62].

Previous strategies to characterize the discordance between BMI and metabolic risk have been based on predefined sets of phenotypes traditionally linked with metabolic status[19,21]. Our phenome-wide approach consisted of leveraging the wealth of genetic associations harvested to date to dissect the phenotypic structure relevant for discordant diabesity, having three main advantages: (1) variables defining the differential phenotypic structure of each profile are selected a in data-driven manner across many phenotypic layers; (2) leveraging genetic data across multiple datasets enhances power and minimizes cohort-specific biases that would be anticipated if analyses were performed in a single cohort; and (3) although concordant and discordant diabesity profiles may be driven by molecular mechanisms that are independent of DNA variation, using germline DNA variants helps mitigate reverse causality and other sources of confounding that hamper the interpretation of associations for most other types of biological variation and phenotypes. An example of this is the analysis of epigenetic factors, which has led to identification of obesity sub-phenotypes even in the context of genetic homogeneity, as found in monozygotic twins that are discordant for adiposity traits[63]. However, these findings might be driven by variations in environmental exposures and behaviours that exist within and between twin pairs, as well as confounded by factors such as age, which differed between twin pairs in the reported analyses.

In conclusion, obesity profiles with either diabetogenic or anti-diabetogenic proclivities reveal distinctive aetiological subtypes, with key differences in fat distribution, blood pressure and cholesterol content in HDL particles. We identified 1 protein (TIMP4) and 17 genes potentially involved in the molecular mechanisms leading to diabesity discordance, involving pleiotropic effects across multiple tissues.

Methods

Study populations

BioVU. Collection of electronic health records in BioVU was established in 1990 and includes data on billing codes from the International Classification of Diseases, 9th and 10th editions (ICD-9 and ICD-10). Disease phenotypes ('phecodes') are derived from these billing codes as described previously[24] and case, control and exclusion criteria are defined. Two codes on different visit days were required to instantiate a case for each phecode. The biobank was launched in 2007 and comprises excess blood samples that their donors had consented for use in biomedical research. Details of programme operations, ethical considerations, continuing oversight and patient engagement are published elsewhere[25]. DNA samples were analysed using genome-wide genotyping platforms including Illumina multi-ethnic genotyping array. After quality assessment, the genotype data were then imputed to the Haplotype Reference Consortium reference panel at the Michigan imputation server. Populations of African American and European descent were identified by projecting individuals onto the major principal-component space derived from 1000 Genomes reference panel.

UK Biobank. The UK Biobank is an ongoing prospective study of approximately 500,000 adults. Initial enrolment took place from 2006 to 2010 and included individuals aged 40–69 years across the United Kingdom[64]. It has collected comprehensive genetic and phenotypic information, biochemical assays and longitudinal health outcomes through health records, such as hospitalization and mortality. The genotypes were assayed using the UK Biobank Lung Exome Variant Evaluation and the Applied Biosystems UK Biobank Axiom Array, and imputed to the Haplotype Reference Consortium panel. Population structure was also assessed using principal-component analysis. We excluded individuals with inconsistency between their reported and genetic sex, had sex chromosome aneuploidy or were outliers for heterozygosity or missingness. Only individuals who were included in the calculation of genetic principal components were included, which ensures minimal genetic kinship with other participants.

Single-nucleotide polymorphism selection to construct concordant and discordant genetic profiles

We cross-referenced the largest GWAS for BMI and T2D and extracted common biallelic SNPs (minor allele frequency (MAF) > 1%). Insertions, deletions and potentially ambiguous palindromic SNPs (A/T or C/G alleles with MAF > 30%) were excluded. Because both GWAS were conducted predominantly in populations of European descent, we used 1000 Genomes EUR reference panel for clumping ($r^2 < 0.01$ over a 500-kb window) to identify nearly independent SNPs that were strongly associated with both conditions ($P < 5 \times 10^{-8}$). The directions of the effect of these SNPs on T2D were consistent in a second independent set of GWAS summary statistics extracted from the FinnGen database[65] (Supplementary Table 2).

Phenome-wide scans

We extracted association data for concordant and discordant SNPs from a variety of sources. From the curated repository Open GWAS[66], which we queried using the 'ieugwasr' package in R, we gathered data

for >3,500 traits derived from UK Biobank and other consortia for a variety of traits; these traits are termed 'clinical phenotypes'. In cases where the effect of a SNP on a trait was not found, we looked for the effect of the nearest proxy SNP up to an r^2 of 0.5 over a 500-kb window. We only kept estimates obtained from European ancestry populations in order to be consistent with the GWAS used to identify the genetic profiles. To prevent inclusion of inflated signals due to low sample size, we only included studies of more than 500 individuals; for binary traits, we required at least 25 minor alleles in the smallest group[67]. We calculated *z*-scores by dividing the β coefficients by their corresponding standard errors, and then we computed standardized effect sizes as a function of MAF and sample size n using equations (1) and (2) (ref. [31]):

$$\mathrm{SE} = \frac{1}{\sqrt{2 \times \mathrm{MAF} \times (1 - \mathrm{MAF}) \times (n + z^2)}} \quad (1)$$

$$\beta = z \times \mathrm{SE} \quad (2)$$

We aligned all the estimates from these scans to the BMI-increasing alleles, so that they represent phenotypic variations associated with higher BMI in both profiles.

We also obtained data for 657 blood metabolites[68,69] and 3,282 proteins in plasma[39]. Associations with several bacterial taxa in the gut were obtained from the MiBioGen consortium[70]. Association with expression and splicing of nearby genes in multiple tissue samples and in whole blood were obtained from data generated by GTEx and eQTLGen consortia, respectively.

Profile comparison

We then compared the effects of discordant versus concordant SNPs for every trait in two stages: we first obtained the combined effect of concordant (β_C) and discordant (β_D) SNPs separately using a random-effects meta-analysis with the Paule–Mandel estimator of between-SNP variance τ^2 (refs. [71,72]). We then calculated their difference $\delta = |\beta_C - \beta_D|$ and computed its standard error as $SE_\delta^2 = SE_C^2 + SE_D^2$[73]. We excluded from these analyses T2D traits (Supplementary Table 17). Traits in which any of the combined estimates β_C or β_D and δ were statistically significant after 5% FDR correction were taken forward to stage two, where we converted the effect estimates for each SNP and the selected traits to *z*-scores and placed them in a SNP–trait matrix, with SNPs coded as '0' if concordant and '1' if discordant. We then trained several Random Forest classifiers (1,000 iterations) to this matrix, which attempted to classify SNPs in their correct category, and used the Boruta algorithm[74] to identify which traits were relevant to distinguish discordant from concordant SNPs. Briefly, this algorithm creates randomly shuffled copies of all traits in the SNP–trait matrix, and then evaluates for each trait if its contribution to the accuracy of decision trees in the Random Forest is higher than its corresponding random set.

Genetic risk score analyses

Concordant and discordant GRSs for an individual i were calculated as:

$$\mathrm{PRS}_{Pi} = \sum_{j \in P}^{M_P} G_{ij} \quad (3)$$

Where P is the set of M_P SNPs belonging to the concordant or discordant profiles and G_{ij} is the genotype for SNP j in individual i. In BioVU, association analyses were carried out for each GRS using R package 'PheWAS' (v0.99.5-5)[24]. We kept phecodes with at least 200 cases[67] and identified those associated with either of the GRS coefficients and a significant difference between the estimated effects after a 5% FDR correction.

In the UK Biobank, we examined the relationship of GRSs to mortality due to cardiovascular events in individuals followed up to the latest censor date (30th September 2021) using Cox regression. Primary cause of death was ascertained using ICD-10 codes reported in death certificates (Supplementary Table 18). All association models were adjusted for age, sex and first ten genetic principal components.

SMR & HEIDI

The SMR method consists of identifying for a protein or gene the strongest association signal, which is used as a genetic instrument to test for its pleiotropic effect on an outcome. The HEIDI method consists of calculating the heterogeneity in the estimates of SNPs in LD with the lead SNP used in SMR. A higher p_{HEIDI} value means heterogeneity is less likely, which supports true pleiotropy across the gene/protein and outcome signal, while a lower p_{HEIDI} value means there is heterogeneity in the estimates, and the SMR signal is probably due to linkage. We consider an association to be true pleiotropy if $p_{HEIDI} > 0.01$ (ref. [75]). We retained signals where we found evidence of true pleiotropy for both BMI and T2D.

Scoring method using epigenetic annotation

The scoring method to identify the most likely tissue of action assumes that if a genetic instrument for the expression of a gene in a certain tissue where it is highly expressed (that is, high tissue specificity) is in or close (in LD) to a promoter/enhancer region in the same tissue, and this genetic instrument is also associated with an outcome, then it is likely that the pleiotropic association on the outcome is due to perturbation of gene activity in that tissue. Promoter/enhancer signals were obtained by querying the RoadMap Epigenomics Project through the 'haploR' package in R.

Genes as therapeutic targets

The lookups in DGIdb and PHAROS were performed using the web-based tool. DGIdb assigns an interaction score to the drug–gene interactions, which is the result of combining publication count, source count, relative drug specificity and relative gene specificity. The PHAROS database classifies targets into four 'Target Development Levels', according to the evidence of drug interactions available: 'Tdark' contains understudied targets, 'Tbio' contains highly studied targets but without interaction with compounds, 'Tchem' includes targets that bind to small molecules, and 'Tclin' interact with approved drugs.

All analyses were done using packages within the R environment (v4.1.2)[76].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The GWAS summary data analysed in this study are available from the GIANT (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium) and DIAGRAM (https://diagram-consortium.org/) consortia websites, the Open GWAS database (https://gwas.mrcieu.ac.uk/), the GTEx consortium website (https://gtexportal.org/home/ and the MiBioGen repository (https://mibiogen.gcc.rug.nl/). UK Biobank data are available through a procedure described at http://www.ukbiobank.ac.uk/using-the-resource/. Individual-level genetic and clinical data from BioVU cannot be shared publicly due to patient confidentiality. However, summary statistics can be viewed in tabular form at https://phewascatalog.org/labwas/. The DGIdb and the PHAROS databases can be accessed online at https://www.dgidb.org/ and https://pharos.nih.gov/, respectively.

Code availability

The codes used for our analyses are available at https://github.com/danielcoral/DVA_codes/.

References

1. World Health Organization. Cardiovascular diseases (CVDs). Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed October 2022.
2. Magliano, D. J. et al. Trends in incidence of total or type 2 diabetes: systematic review. *BMJ.* **366**, l5003 (2019).
3. Blüher, M. Obesity: global epidemiology and pathogenesis. *Nat. Rev. Endocrinol.* **15**, 288–298 (2019).
4. van Vliet-Ostaptchouk, J. V. et al. The prevalence of metabolic syndrome and metabolically healthy obesity in Europe: a collaborative analysis of ten large cohort studies. *BMC Endocr. Disord.* **14**, 9 (2014).
5. Blüher, M. Metabolically healthy obesity. *Endocr. Rev.* **41**, 405–420 (2020).
6. Vecchié, A. et al. Obesity phenotypes and their paradoxical association with cardiovascular diseases. *Eur. J. Intern Med* **48**, 6–17 (2018).
7. Ding, C., Chan, Z. & Magkos, F. Lean, but not healthy: the 'metabolically obese, normal-weight' phenotype. *Curr. Opin. Clin. Nutr. Metab. Care* **19**, 408–417 (2016).
8. Wang, B. et al. Prevalence of metabolically healthy obese and metabolically obese but normal weight in adults worldwide: a meta-analysis. *Horm. Metab. Res* **47**, 839–845 (2015).
9. Loos, R. J. F. & Kilpeläinen, T. O. Genes that make you fat, but keep you healthy. *J. Intern Med.* **284**, 450–463 (2018).
10. Schwartz, S. S. & Kohl, B. A. Glycemic control and weight reduction without causing hypoglycemia: the case for continued safe aggressive care of patients with type 2 diabetes mellitus and avoidance of therapeutic inertia. *Mayo Clin. Proc.* **85**, S15–S26 (2010).
11. Carnethon, M. R. et al. Association of weight status with mortality in adults with incident diabetes. *JAMA* **308**, 581–590 (2012).
12. Badrick, E., Sperrin, M., Buchan, I. E. & Renehan, A. G. Obesity paradox and mortality in adults with and without incident type 2 diabetes: a matched population-level cohort study. *BMJ Open Diabetes Res. Care* **5**, e000369 (2017).
13. Shah, N. R. & Braverman, E. R. Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat and leptin. *PLoS ONE* **7**, e33308 (2012).
14. Müller, M. J. et al. The case of GWAS of obesity: does body weight control play by the rules? *Int. J. Obes.* **42**, 1395–1405 (2018).
15. Goossens, G. H. The metabolic phenotype in obesity: fat mass, body fat distribution, and adipose tissue function. *Obes. Facts* **10**, 207–215 (2017).
16. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
17. Rask-Andersen, M., Karlsson, T., Ek, W. E. & Johansson, Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat. Commun.* **10**, 339 (2019).
18. Lotta, L. A. et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).
19. Ji, Y. et al. Genome-wide and abdominal MRI data provide evidence that a genetically determined favorable adiposity phenotype is characterized by lower ectopic liver fat and lower risk of type 2 diabetes, heart disease, and hypertension. *Diabetes* **68**, 207–219 (2019).
20. Kilpeläinen, T. O. et al. Genetic variation near *IRS1* associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.* **43**, 753–760 (2011).
21. Huang, L. O. et al. Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. *Nat. Metab.* **3**, 228–243 (2021).
22. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
23. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
24. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
25. Roden, D. et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* **84**, 362–369 (2008).
26. Dennis, J. K. et al. Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. *Genome Med.* **13**, 6 (2021).
27. Pigeyre, M. et al. Identification of novel causal blood biomarkers linking metabolically favorable adiposity with type 2 diabetes risk. *Diabetes Care* **42**, 1800–1808 (2019).
28. Pers, T. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
29. The GTEX Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
30. Võsa, 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).
31. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
32. Torres, J. M. et al. A multi-omic integrative scheme characterizes tissues of action at loci associated with type 2 diabetes. *Am. J. Hum. Genet.* **107**, 1011–1028 (2020).
33. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
34. Freshour, S. L. et al. Integration of the drug–gene interaction database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.* **49**, D1144–D1151 (2021).
35. Sheils, T. K. et al. TCRD and Pharos 2021: mining the human proteome for disease biology. *Nucleic Acids Res.* **49**, D1334–D1346 (2021).
36. Enattah, N. et al. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002).
37. Udler, M. S. et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med.* **15**, e1002654 (2018).
38. Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96**, 329–339 (2015).
39. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
40. Emdin, C. A. et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA* **317**, 626–634 (2017).
41. Lotta, L. A. et al. Association of genetic variants related to gluteofemoral vs abdominal fat distribution with type 2 diabetes, coronary disease, and cardiovascular risk factors. *JAMA* **320**, 2553–2563 (2018).
42. Silva, N. M. G. D. et al. Liver function and risk of type 2 diabetes: bidirectional Mendelian randomization study. *Diabetes* **68**, 1681–1691 (2019).

43. Karaca, Ü., Schram, M. T., Houben, A. J. H. M., Muris, D. M. J. & Stehouwer, C. D. A. Microvascular dysfunction as a link between obesity, insulin resistance and hypertension. *Diabetes Res. Clin. Pract.* **103**, 382–387 (2014).
44. Stehouwer, C. D. A. Microvascular dysfunction and hyperglycemia: a vicious cycle with widespread consequences. *Diabetes* **67**, 1729–1741 (2018).
45. Barrett, E. J. et al. The vascular actions of insulin control its delivery to muscle and regulate the rate-limiting step in skeletal muscle insulin action. *Diabetologia* **52**, 752–764 (2009).
46. Meijer, R. I. et al. Insulin-induced microvascular recruitment in skin and muscle are related and both are associated with whole-body glucose uptake. *Microcirculation* **19**, 494–500 (2012).
47. Holeček, M. Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. *Nutr. Metab.* **15**, 33 (2018).
48. Guasch-Ferré, M. et al. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes Care* **39**, 833–846 (2016).
49. Nagai, N. et al. Involvement of heparan sulfate 6-*O*-sulfation in the regulation of energy metabolism and the alteration of thyroid hormone levels in male mice. *Glycobiology* **23**, 980–992 (2013).
50. Matsuzawa, T. et al. Heparan sulfate promotes differentiation of white adipocytes to maintain insulin sensitivity and glucose homeostasis. *J. Biol. Chem.* **297**, 101006 (2021).
51. Sakamuri, S. S. V. P. et al. Absence of tissue inhibitor of metalloproteinase-4 (TIMP4) ameliorates high fat diet-induced obesity in mice due to defective lipid absorption. *Sci. Rep.* **7**, 6210 (2017).
52. Marx, N. et al. Antidiabetic PPAR gamma-activator rosiglitazone reduces MMP-9 serum levels in type 2 diabetic patients with coronary artery disease. *Arterioscler Thromb. Vasc. Biol.* **23**, 283–288 (2003).
53. Maquoi, E., Munaut, C., Colige, A., Collen, D. & Lijnen, H. R. Modulation of adipose tissue expression of murine matrix metalloproteinases and their tissue inhibitors with obesity. *Diabetes* **51**, 1093–1101 (2002).
54. Thorens, B. GLUT2, glucose sensing and glucose homeostasis. *Diabetologia* **58**, 221–232 (2015).
55. Eny, K. M., Wolever, T. M. S., Fontaine-Bisson, B. & El-Sohemy, A. Genetic variant in the glucose transporter type 2 is associated with higher intakes of sugars in two distinct populations. *Physiol. Genomics* **33**, 355–360 (2008).
56. Zhou, K. et al. Variation in the glucose transporter gene *SLC2A2* is associated with glycemic response to metformin. *Nat. Genet.* **48**, 1055–1059 (2016).
57. Rathmann, W. et al. A variant of the glucose transporter gene SLC2A2 modifies the glycaemic response to metformin therapy in recently diagnosed type 2 diabetes. *Diabetologia* **62**, 286–291 (2019).
58. Song, W. et al. Organic cation transporter 3 (Oct3) is a distinct catecholamines clearance route in adipocytes mediating the beiging of white adipose tissue. *PLoS Biol.* **17**, e2006571 (2019).
59. Saxton, S. N. et al. Restoring perivascular adipose tissue function in obesity using exercise. *Cardiovasc. Drugs Ther.* **35**, 1291–1304 (2021).
60. Middelberg, R. P. et al. Genetic variants in *LPL*, *OASL* and *TOMM40/APOE-C1-C2-C4* genes are associated with multiple cardiovascular-related traits. *BMC Med. Genet.* **12**, 123 (2011).
61. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
62. den Besten, G. et al. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.* **54**, 2325–2340 (2013).
63. Yang, C. H. et al. Independent phenotypic plasticity axes define distinct obesity sub-types. *Nat. Metab.* **4**, 1150–1165 (2022).
64. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
65. Kurki, M. I. et al. FinnGen: unique genetic insights from combining isolated population and national health register data. Preprint at https://www.medrxiv.org/content/10.1101/2022.03.03.22271360v1 (2022).
66. Elsworth B, et al. The MRC IEU OpenGWAS data infrastructure. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.10.244293 (2020).
67. Verma, A. et al. PheWAS and beyond: the landscape of associations with medical diagnoses and clinical measures across 38,662 individuals from Geisinger. *Am. J. Hum. Genet* **102**, 592–608 (2018).
68. Shin, S.-Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
69. Kettunen, J. et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 11122 (2016).
70. Kurilshikov, A. et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* **53**, 156–165 (2021).
71. Dastani, Z. et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**, e1002607 (2012).
72. Veroniki, A. A. et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Methods* **7**, 55–79 (2016).
73. Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. Multiple outcomes or time-points within a study. In *Introduction to Meta-Analysis*, 225–238 (John Wiley & Sons, 2009).
74. Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
75. Wu, Y. et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).
76. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed at https://www.R-project.org/ (2022).

Acknowledgements

This work was undertaken as part of the Stratification of Obesity Phenotypes to Optimize Future Therapy (SOPHIA) and the Novel Approach to Systematically Characterize Exercise- and Nutrient-responsive genes in Type 2 diabetes and cardiovascular disease (NASCENT) projects. SOPHIA has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 875534. This Joint Undertaking is supported by the European Union's Horizon 2020 research and innovation programme and EFPIA and T1D Exchange, JDRF and Obesity Action Coalition. NASCENT was funded by a grant from the European Commission (ERC-2015-CoG–681742 NASCENT). Additonal funding was provided by the Swedish Research Council (Distinguished Young Researcher Award) and the Swedish Foundation for Strategic Research (LUDC-IRC, 15-0067). All awards to PWF.

Author contributions

P.W.F. and D.E.C. conceived and designed and analyses, and wrote the manuscript. D.E.C., T.W.M.-F. and X.Z. performed the analyses. All other co-authors contributed materials/analysis tools and reviewed the manuscript.

Funding

Open access funding provided by Lund University.

Ethics statements

Ethics approval for the UK Biobank was obtained from the North West Centre for Research Ethics Committee. Analysis of individual-level data from UK Biobank participants in Lund University was approved by the Swedish Ethical Review Authority (2021-0317). The BioVU project was approved by the VUMC Institutional Review Board. The analysis of individual-level data was performed in VUMC, and only summary results were shared with researchers at Lund University. Both studies conformed to the ethical principles for medical research involving human participants outlined in the Declaration of Helsinki. All participants provided written informed consent at enrolment.

Inclusion & ethics statement

All collaborators of this study have fulfilled the criteria for authorship required by Nature Portfolio journals have been included as authors, as their participation was essential for the design and implementation of the study. Roles and responsibilities were agreed among collaborators ahead of the research. This work includes findings that are locally relevant, which have been determined in collaboration with local partners. This research was not severely restricted or prohibited in the setting of the researchers, and does not result in stigmatization, incrimination, discrimination or personal risk to participants. Local and regional research relevant to our study was taken into account in citations.

Competing interests

P.W.F. has received research grants from numerous diabetes drug companies and fees as consultant from Novo Nordisk, Lilly and Zoe. He is currently the Head of the Department of Translational Medicine at the Novo Nordisk Foundation. All other authors declare non-competing interests.

Additional information

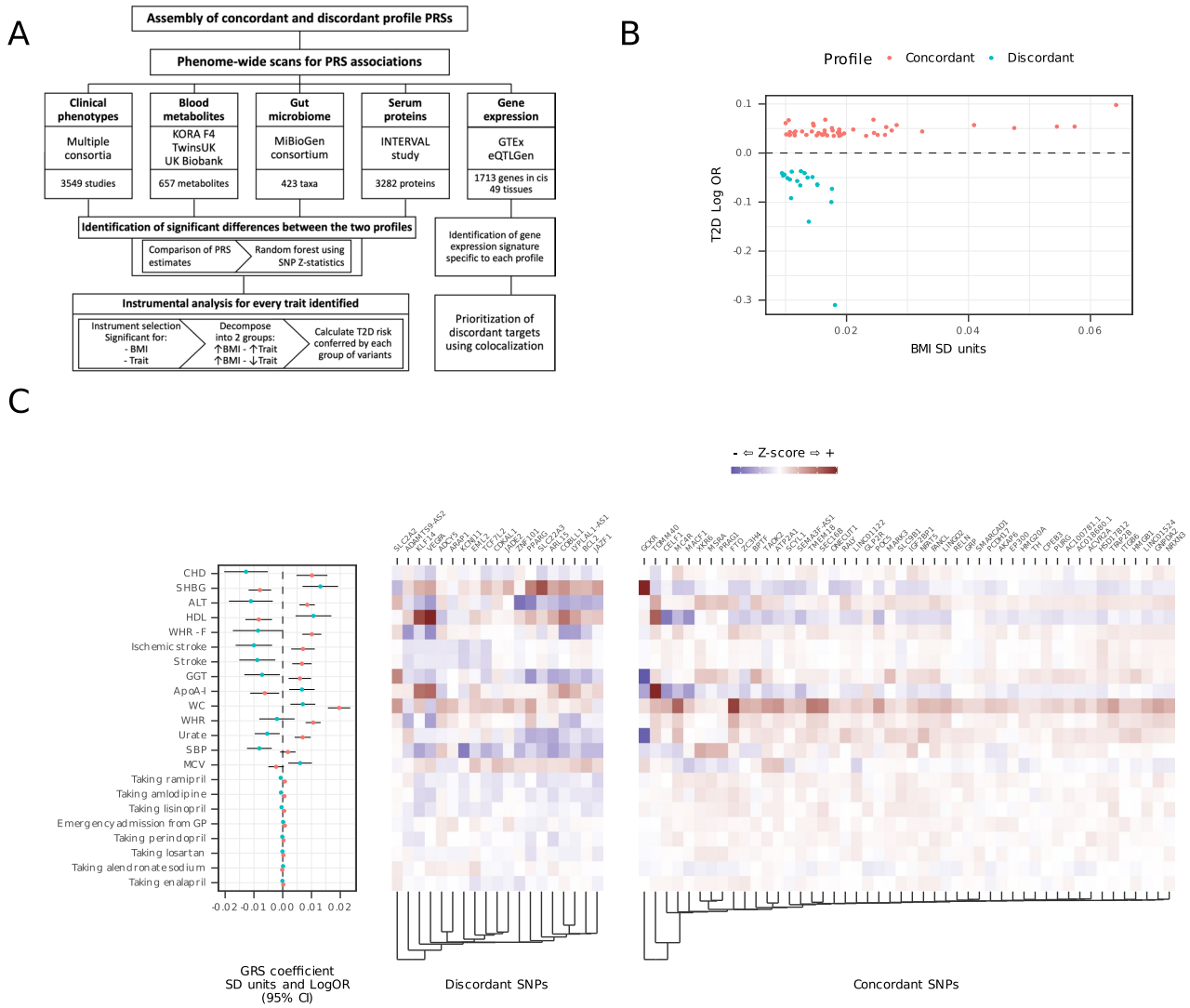
Extended data is available for this paper at https://doi.org/10.1038/s42255-022-00731-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42255-022-00731-5.

Correspondence and requests for materials should be addressed to Daniel E. Coral or Paul W. Franks.

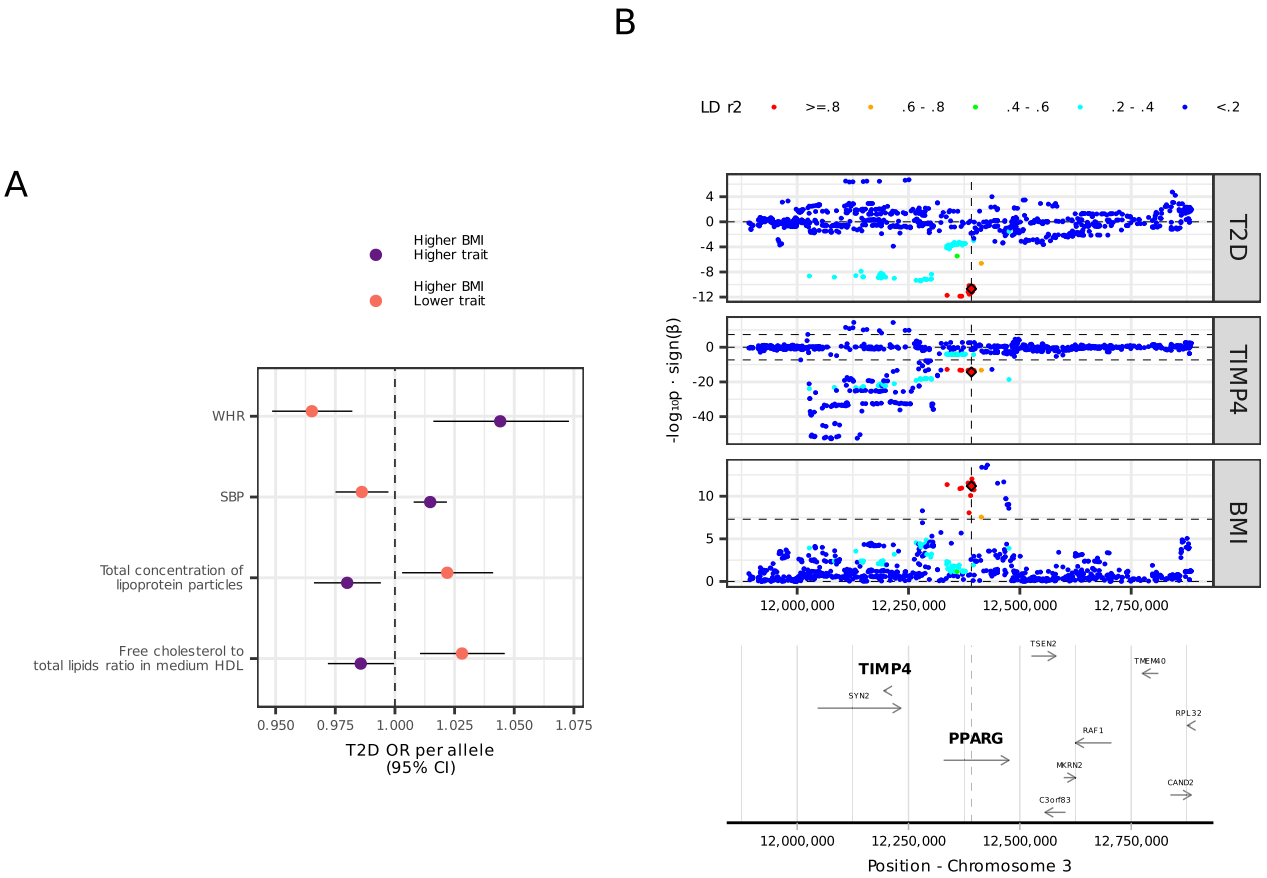
Peer review information *Nature Metabolism* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Isabella Samuelson, in collaboration with the *Nature Metabolism* team.

https://doi.org/10.1038/s42255-022-00731-5



Extended Data Fig. 1 | Analysis flowchart and profile identification. Panel A: Analysis flowchart. Panel B: BMI and T2D risk estimates of concordant and discordant SNPs after alignment to the BMI increasing allele. Panel C: summary-based concordant and discordant GRS coefficients (standard deviation units for continuous traits, log OR for binary traits). Traits shown have at least 1 estimate significant after 5% FDR correction and the difference between profiles is also significant after 5% FDR. Statistical tests were based on a Z-distribution and were two-sided. Bars show 95% confidence intervals. Sample sizes vary for every trait (>100.000 for all traits). The heatmap shows the Z-scores of the SNPs in every trait, with the single-linkage tree at the bottom, separately for concordant and discordant SNPs.

https://doi.org/10.1038/s42255-022-00731-5



Extended Data Fig. 2 | Traits with potential causal effect on diabesity discordance. Panel A: Traits where a difference was found in the comparison of profiles and one of the two direction-specific GRS associated with BMI was associated with lower risk of T2D (two-sided Z-statistic P < 0.05). To derive the GRS we used BMI data from the GIANT + UK Biobank meta-analysis (N = 681,275). WHR data came from the GIANT consortium (N = 212,244). SBP data came from the meta-analysis performed by the ICBP (N = 757,601). Metabolite data came from the UK Biobank (N = 115,078). Estimates represent T2D OR, bars represent 95% confidence intervals, which are derived from the DIAGRAM meta-analysis (N = 158,186). Panel C: Regional association plot showing the pleiotropic effect of genetic instruments for blood levels of TIMP4 protein and high BMI and lower T2D risk. Protein data was derived from the INTERVAL study (N = 3,301).

nature portfolio

Corresponding author(s): Paul W. Franks

Last updated by author(s): Dec 15, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We collected GWAS summary statistics for BMI and T2D from the GIANT and the DIAGRAM consortium websites, respectively, using the command line interface. The links are provided in the computer codes. We downloaded data from multiple large consortia from the publicly available repository Open GWAS, hosted by the University of Bristol Medical Research Council Integrative Epidemiology Unit, which we queried through the R interface using the "ieugwasr" package. Summary statistics for tissue specific gene expression and microbiome data were downloaded from the GTEx and MiBioGen online repositories, respectively, using the command line interface, whose links are also provided in the computer code. Epigenetic data generated by the RoadMap Epigenomics Project were collected through the R interface using the "haploR" package. Finally, we queried the genes and proteins with likely pleiotropic effects on BMI and T2D in oppositve directions in the publicly available drug target repositories DGIdb and PHAROS. The R version used was 4.1.2. The codes used for our analyses are available at https://github.com/danielcoral/DVA_codes.
Data analysis	Data manipulation was done using mainly packages from the "tidyverse" collection. Pooled concordant and discordant estimates were calculated using the package "meta". To select traits where we find the most relevant differences between genetically determined obesity profiles we used the package "Boruta". For the comparison in BioVU we used the 'PheWAS' package in R. For the SMR & HEIDI analysis, we adapted code from the original publication by Zhu et al. 2016 to be used within the R environment. For clumping genetic variants we used PLINK version 1.9. The R version used was 4.1.2. The codes used for our analyses are available at https://github.com/danielcoral/DVA_codes.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The GWAS summary data analysed in this study are available from the GIANT (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium) and DIAGRAM (https://diagram-consortium.org/) consortia websites, the Open GWAS database (https://gwas.mrcieu.ac.uk/), the GTEx consortium website (https://gtexportal.org/home/ and the MiBioGen repository (https://mibiogen.gcc.rug.nl/). UK Biobank data are available through a procedure described at http://www.ukbiobank.ac.uk/using-the-resource/. Individual level genetic and clinical data from BioVU cannot be shared publicly due to patient confidentiality. However, summary statistics can be viewed in tabular form at: https://phewascatalog.org/labwas. The DGIdb and the PHAROS databases can be accessed online at https://www.dgidb.org/ and https://pharos.nih.gov/, respectively.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We found that concordant and discordant genetic profiles differ in waist-to-hip ratio, a measure of central to peripheral obesity, predominantly in women. Therefore, when assessing whether individuals with extreme concordant and discordant GRS conveyed obesity profiles that are different from other obesity, we included an additional analysis were we stratified by sex. Sex was determined by genotyping analysis; individuals whose genetic sex did not match reported sex were excluded, in order to have results relevant to biological sex and guard against distortion of estimates due to possible sex chromosome aneuploidies. These and other analyses that we perfomed in UK Biobank only included individuals from European descent. This is because our initial step to find concordant and discordant variants was done using GWAS summary statistics that were done in European populations. European descent was determined using genotyping data. The association of concordant and discordant GRS with laboratory measures were tested on individuals of African descent in BioVU. This was also determined by genotyping data.
Population characteristics	The mean age of individuals in UK Biobank when they attended the first assessment centre was 56 years, ranging between 37 to 85 years. 54% of participants are females. To analyse the association between concordant and discordant GRS to diagnoses in BioVU we included up to 48,544 individuals of European descent. In the analyses of laboratory measures, we included 68,724 participants of European descent and 13,661 participants of African descent. In both the proportion of females is around 51%.
Recruitment	UK Biobank participants were assessed between 2006 and 2010 in 22 assessment centres throughout the UK, covering a variety of different settings to provide socioeconomic and ethnic heterogeneity and urban–rural mix. Invitations to participate were sent via mail to potential participants identified through the National Health Service. Participants that were included are not representative of the sampling population, as there is evidence for healthy volunteer selection bias. Recruitment in BioVU consists of an opt-out clinical collection of patients from the Vanderbilt University Medical Center (VUMC) in an outpatient setting. DNA is extracted from discarded blood drawn for routine clinical care. This is also not representative of the general population of Tennessee and the United States, due to its dependence on clinical registry.
Ethics oversight	Ethics approval for the UK Biobank was obtained from the North West Centre for Research Ethics Committee. Analysis of individual level data from UK Biobank participants in Lund University was approved by the Swedish Ethical Review Authority (2021-0317). The BioVU project was approved by the VUMC Institutional Review Board. The analysis of individual level data was performed in VUMC, and only summary results were shared with researchers at Lund University. Both studies conformed to the ethical principles for medical research involving human participants outlined in the Declaration of Helsinki. All participants provided written informed consent at enrolment.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	To obtain genetic variants with significant effects on BMI and T2D, we selected studies including >700,000 participants, which were the studies with the largest sample sizes to date for both traits. For the phenome-wide scan, we selected studies with at least 500 individuals since for variants with MAF higher than 10% (the minimum value for concordant and discordant SNP sets, see Supplementary Table 1) and expecting an effect of at least 0.1 standard deviation units the power to detect an association under a nominal alpha value of 0.05 and an additive genetic mode is higher than 80%. In BioVU, we selected diagnoses with at least 200 cases, as recommended in previous simulations of phenome-wide association studies (citation 74 in the manuscript). To assess the association with mortality, we calculated a statistical power of 41% to detect an association between a SNP with a MAF higher than 10% and cardiovascular mortality, expecting a event rate of 1 in every 1000 person-years (using the "survSNP package" in R).
Data exclusions	In the phenome-wide comparison between concordant and discordant SNPs, we excluded studies that were performed in population other than Europeans, to be able to compare effects across phenotypes. To test the association of concordant and discordant GRSs with diseases in BioVU we only included population from European descent. We tested the association of these GRS with laboratory values separately for European and African descent individuals. To test the association with mortality in UK Biobank, we only included individuals of European ancestry. In both BioVU and UK Biobank, we excluded individuals with inconsistency between their reported and genetic sex, had sex chromosome aneuploidy or were outliers for heterozygosity or missingness.
Replication	To verify the sign of the association of concordant and discordant SNPs on T2D we used GWAS summary data from FinnGen. To verify our results of the comparison between concordant and discordant obesity profiles in GWAS summary data, we constructed genetic risk scores with each SNP set and assessed their associations with diseases and laboratory values in an independent dataset (BioVU). After correcting for multiple testing, we checked if the associations reflected the findings from the summary data. We successfully replicated our results. These replication procedures were performed only once.
Randomization	This study does not involve randomization. However, in our analyses we constructed two genetically determined obesity profiles using genetic instruments that are randomly allocated at birth and remain invariant throughout life, which helps prevent confounding and reverse causality. All genetic associations gathered included age and sex as covariates.
Blinding	This study does not involve blinding. GWAS are observational studies and therefore participants know at least the exposure or the outcome, and often both.However, by using genetic instruments randomly allocated we prevent confounding and reverse causality in our analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging