Genome-wide association study for circulating metabolic traits in 619,372 individuals

Ralf Tambets[1], Jaanika Kronberg[2], Adriaan van der Graaf[3], Mihkel Jesse[1], Erik Abner[2], Urmo Võsa[2], Ida Rahu[1], Nele Taba[2], Anastassia Kolde[2,4], Dzvenymyra Yarish[1], Estonian Biobank Research Team[2], Krista Fischer[2,3], Zoltán Kutalik[3,5], Tõnu Esko[2,*], Kaur Alasoo[1,*], Priit Palta[2,*]

[1]Institute of Computer Science, University of Tartu, Tartu, Estonia
[2]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
[3]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
[4]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia
[5]University Center for Primary Care and Public Health, Unisanté, University of Lausanne, Lausanne, Switzerland
*These authors jointly supervised this work.
Correspondence: kaur.alasoo@ut.ee, priit.palta@ut.ee

Abstract

Interpreting genetic associations with complex traits can be greatly improved by detailed understanding of the molecular consequences of these variants. However, although genome-wide association studies (GWAS) for common complex diseases routinely profile 1M+ individuals, studies of molecular phenotypes have lagged behind. We performed a GWAS meta-analysis for 249 circulating metabolic traits in the Estonian Biobank and the UK Biobank in up to 619,372 individuals, identifying 88,604 significant locus-metabolite associations and 8,774 independent lead variants, including 987 lead variants with a minor allele frequency less than 1%. We demonstrate how common and low-frequency associations converge on shared genes and pathways, bridging the gap between rare-variant burden testing and common-variant GWAS. We used Mendelian randomisation (MR) to explore putative causal links between metabolic traits, coronary artery disease and type 2 diabetes (T2D). Surprisingly, up to 85% of the tested metabolite-disease pairs had statistically significant genome-wide MR estimates, likely reflecting complex indirect effects driven by horisontal pleiotropy. To avoid these pleiotropic effects, we used *cis*-MR to test the phenotypic impact of inhibiting specific drug targets. We found that although plasma levels of branched-chain amino acids (BCAAs) have been associated with T2D in both observational and genome-wide MR studies, inhibiting the BCAA catabolism pathway to lower BCAA levels is unlikely to reduce T2D risk. Our publicly available results provide a valuable novel resource for GWAS interpretation and drug target prioritisation.

Introduction

Systematic mapping and interpretation of the heritable determinants underlying complex traits and disease predisposition can be greatly improved by detailed understanding of molecular consequences of genetic variation. Studying metabolic traits is crucial as they serve as key indicators of various biological processes and disease states. Metabolite studies can reveal the complex interactions between genes and metabolic pathways, providing a more comprehensive understanding of human molecular biology and the potential for novel therapeutic hypotheses. This understanding can lead to the identification of new biomarkers for disease diagnosis, prognosis, and treatment, as well as the development of improved personalised medical interventions.

Although genome-wide association studies (GWAS) for several traits and diseases now exceed the sample size of 1 million individuals (*1*–*6*), studies of molecular traits such as gene expression (*7*), plasma proteins (*8*) or circulating metabolites have lagged behind. Notable exceptions are five blood lipid traits with the latest meta-analysis including data from 1.65 million individuals (*9*). Therefore, as expected, recent large-scale GWAS studies of metabolic traits continue to uncover novel associations and biological insights (*10*–*15*). However, for more than half of the metabolites captured by NMR, the proportion of heritability explained by genome-wide significant variants remains below 50% (*13*), indicating that much larger sample sizes are needed to discover the remaining genetic effects. Furthermore, existing GWAS studies using the Nightingale Health NMR platform have been limited to common variants (MAF > 1%) due to limited sample sizes as well as lower imputation accuracy of low frequency variants (*10*–*13*). Therefore, limited attention has been paid to low-frequency and rare variation, which, while explaining less heritability overall, can still provide novel biological insights (*14*, *16*).

At the same time, larger sample sizes and increased statistical power of metabolic GWAS studies also brings new challenges for effectively integrating these results with GWAS findings for other complex traits and diseases. Owing to their relative simplicity, genome-wide Mendelian randomisation (MR) studies have recently gained popularity (*17*–*19*). Similarly to the established causal role of circulating LDL cholesterol on coronary artery disease (*20*), these studies seek to determine metabolic traits (exposures) that have a causal effect on any number of outcomes (often complex diseases). However, inferences from MR studies are only valid if certain assumptions are met (*21*, *22*). A key assumption of MR is that the genetic variants are associated with the outcome only via the exposure of interest (*23*). In practice, this assumption can be challenging to satisfy, because genetic variants can have pleiotropic effects on multiple metabolic traits (*10*–*12*, *24*). As a result, genome-wide MR studies often identify hundreds or even thousands of associations that can be challenging to interpret causally (*10*, *14*, *25*). A promising alternative to genome-wide MR is *cis*-MR that focuses the analysis to a specific *cis* region around the target gene of interest (*11*, *26*). While *cis*-MR is less susceptible to horizontal pleiotropy (*15*, *21*), it is limited by the number of independent association signals that can be identified at any gene region, thus requiring very well powered GWAS studies.

Here, we present a genome-wide association study for 249 circulating metabolic traits quantified by nuclear magnetic resonance spectroscopy (NMR) across the complete set of UK Biobank participants (n = 434,020) from diverse genetic ancestry groups, and European-ancestry individuals from the Estonian Biobank (n = 185,352) (**Figure 1**). We also

performed two separate meta-analyses across the predominantly European ancestry samples (n = 599,249) and across all samples (n = 619,372), resulting in 3-5× larger sample size compared to previous studies (*10*–*13*) and 40% larger sample size compared to a parallel study (*14*). To illustrate the value of our resource for interpreting complex disease associations, we performed systematic genetic colocalisation between GWAS signals for coronary artery disease (CAD) and type 2 diabetes (T2D), all metabolic trait associations from this study, and gene expression and splicing QTLs from the eQTL Catalogue (*27*, *28*). Finally, we demonstrate how leveraging our data in the *cis*-MR framework can be used to test specific therapeutic hypotheses. We find that although plasma levels of branched-chain amino acids (BCAAs) have been associated with T2D in both observational and genome-wide Mendelian randomisation studies (*29*–*32*), lowering plasma levels of BCAAs by targeting the BCAA catabolism pathway is unlikely to reduce T2D risk. Our publicly available results will provide a valuable resource for GWAS interpretation and target prioritisation studies.

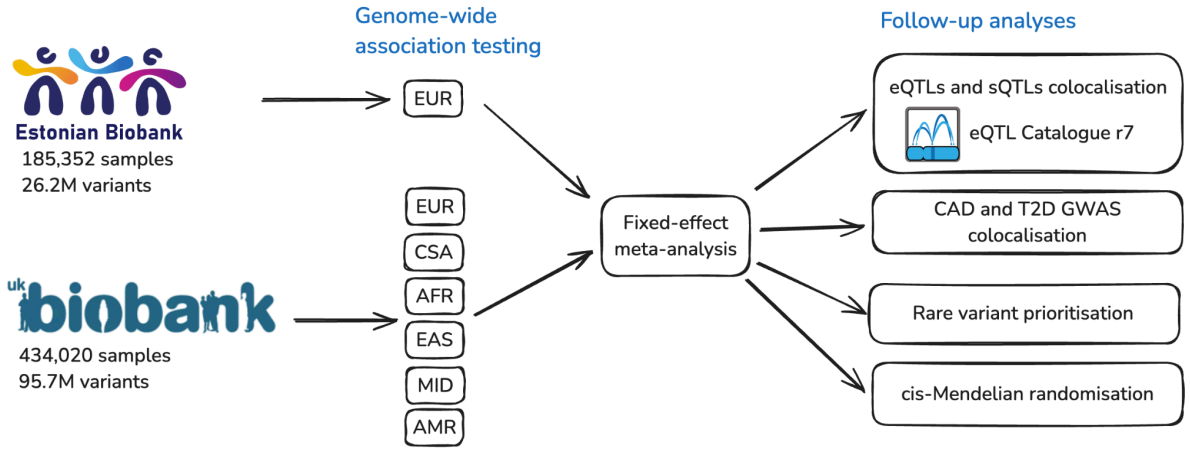

Figure 1. **Outline of the study.** First, a separate GWAS was performed for each metabolite in the Estonian Biobank and six genetic ancestry groups of the UK Biobank: EUR (European), AFR (African), AMR (Admixed American), MID (Middle Eastern), EAS (East Asian), CSA (Central/South Asian) as defined by Pan-UKBB (*33*). These ancestry-specific GWAS results were then combined in a fixed-effect meta-analysis. Finally, several follow-up analyses were performed to demonstrate the value of the resource. eQTL - gene expression quantitative trait locus; sQTL - splicing quantitative trait locus; CAD - coronary artery disease; T2D - type 2 diabetes.

Results

Association testing and meta-analysis

We performed GWAS for 249 metabolic traits (**Table S1**) in the Estonian Biobank (EstBB) and six genetic ancestry groups from the UK Biobank (UKBB) (**Figure 1**). The UKBB genetic ancestry groups were defined previously by the Pan-UKBB project (*33*). Relying on the population-specific genotype imputation panel for the EstBB (*34*) and the Genomics England (*35*) and the TopMed (*36*) imputation panels for the UKBB allowed us to test 10-96 million variants across genetic ancestry groups (up to 9× more than previous studies using the same NMR platform (*10*, *13*)). The number of genome-wide significant ($p < 5\times10^{-8}$) locus-trait

pairs ranged from 37 (UKBB_AMR) to 63,147 (UKBB_EUR) and the number of independent lead variants ($r^2 < 0.8$) ranged from 24 to 6,410, with most associations detected in the UKBB_EUR and EstBB subsets (**Table 1**). Using LD score regression (*37*), we observed high genetic correlation for matched metabolic traits between the EstBB (n = 185,352) and the UKBB_EUR (n = 413,897) subsets (median rg = 0.91, mean rg = 0.89), indicating that genetic effects are largely shared between the two biobanks (**Table S2**).

Table 1. Number of significant locus-metabolic trait pairs ($p < 5\times10^{-8}$) and unique lead variants ($r^2 > 0.8$) detected in each genetic ancestry group and the two meta-analyses.

Biobank / genetic ancestry group	**Locus-trait pairs ($p < 5\times10^{-8}$)**	**Unique lead variants**	**Sample size**
EstBB	27,002	2,356	185,352
UKBB_EUR (European)	63,147	6,410	413,897
UKBB_CSA (Central/South Asian)	1,070	112	8,652
UKBB_AFR (African)	916	143	6,439
UKBB_EAS (East Asian)	358	40	2,604
UKBB_MID (Middle Eastern)	92	44	1,500
UKBB_AMR (Admixed American)	37	24	928
meta_EUR	87,373	8,641	599,249
meta_ALL	88,604	8,774	619,372

Motivated by the high genetic correlation between the two biobanks, we proceeded with the meta-analyses. In the meta-analysis of EstBB and UKBB_EUR (meta_EUR, n = 599,249), we identified 87,373 locus-trait pairs, corresponding to 8,641 independent lead variants ($r^2 < 0.8$). This represented an approximately 10-fold increase compared to the previous study by Karjalainen *et al* (*10*) (n = 136,016, 8,578 locus-trait pairs) and a 66% increase compared to a parallel study performed by Zoodsma *et al* (*14*) on the overlapping set of UKBB samples (n = 450,016, 52,662 locus-trait pairs). The estimated heritability of individual metabolic traits ranged from 2.8% for Acetoacetate to 19.5% for HDL_size (median 10.2%) and we observed a clear linear relationship between heritability and the number of loci associated with each metabolic trait (**Figure S1**, **Table S3**). On average, 95% of the lead variant associations detected by Karjalainen *et al* (*10*) also replicated in our meta_EUR analysis with a highly concordant direction of effect (**Figure S2**). We also detected many novel associations for all tested metabolites. The fraction of novel associations ranged from 27% for

3-Hydroxybutyrate (bOHbutyrate) to 85% for Lactate (**Figure 2A**). Altogether, we identified 4,085 novel independent lead variants ($r^2 < 0.8$) not previously reported by Karjalainen *et al* (*10*), including 248 lead variants on the X chromosome.

In addition to the EUR genetic ancestry group, we also performed GWAS in five smaller genetic ancestry groups of the UK Biobank (AFR, AMR, CSA, EAS, MID) (**Table 1**). Including these summary statistics into our meta-analysis (meta_ALL) increased the number of independent lead variants from 8,641 to 8,774 (**Table 1**), 41 of which were not tested in the EstBB and UKBB_EUR cohorts due to low allele frequency (allele count < 20). This highlights the need to substantially increase the sample sizes for under-represented genetic ancestry groups to enable the discovery of ancestry-specific associations.

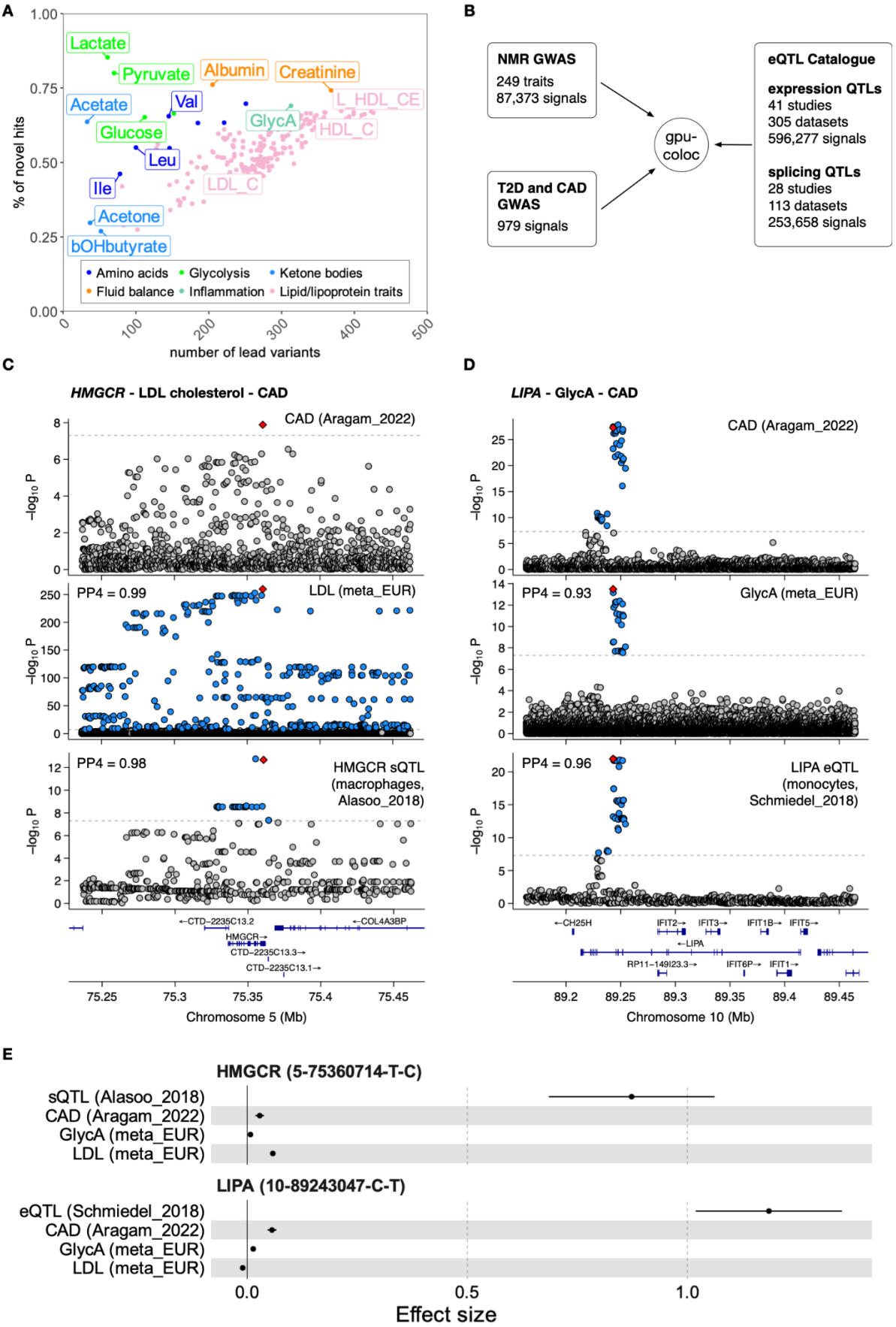
A
% of novel hits
number of lead variants
Lactate
Pyruvate
Albumin
Creatinine
Acetate
Val
L_HDL_CE
GlycA
HDL_C
Glucose
Leu
Ile
LDL_C
Acetone
bOHbutyrate
Amino acids
Glycolysis
Ketone bodies
Fluid balance
Inflammation
Lipid/lipoprotein traits
B
NMR GWAS
249 traits
87,373 signals
T2D and CAD GWAS
979 signals
gpu-coloc
eQTL Catalogue
expression QTLs
41 studies
305 datasets
596,277 signals
splicing QTLs
28 studies
113 datasets
253,658 signals
C
HMGCR - LDL cholesterol - CAD
CAD (Aragam_2022)
PP4 = 0.99
LDL (meta_EUR)
PP4 = 0.98
HMGCR sQTL (macrophages, Alasoo_2018)
Chromosome 5 (Mb)
D
LIPA - GlycA - CAD
CAD (Aragam_2022)
PP4 = 0.93
GlycA (meta_EUR)
PP4 = 0.96
LIPA eQTL (monocytes, Schmiedel_2018)
Chromosome 10 (Mb)
E
HMGCR (5-75360714-T-C)
sQTL (Alasoo_2018)
CAD (Aragam_2022)
GlycA (meta_EUR)
LDL (meta_EUR)
LIPA (10-89243047-C-T)
eQTL (Schmiedel_2018)
CAD (Aragam_2022)
GlycA (meta_EUR)
LDL (meta_EUR)
Effect size

Figure 2. Known and novel genetic associations with metabolic traits. (**A**) Number of genome-wide significant loci ($p < 5\times10^{-8}$) detected for each metabolic trait and the proportion of those associations that were not detected by Karjalainen *et al*. (**B**) Overview of the datasets included in the colocalisation analysis. (**C**) Regional association plots for CAD, GlycA, LDL cholesterol and *HMGCR* exon 13 skipping sQTL at the *HMGCR* locus. PP4 values show pairwise colocalisation posterior probabilities between CAD and the other three traits. (**D**) Regional association plots for CAD, GlycA, LDL cholesterol and *LIPA* gene expression at the *LIPA* locus. (**E**) Effect sizes and 95% confidence intervals of the *HMGCR* and *LIPA* lead variants on CAD risk, LDL_C, GlycA and transcriptomic traits.

Systematic colocalisation across molecular layers to interpret disease associations

To demonstrate how genetic associations with metabolic traits can help interpret disease associations, we colocalised all 87,373 signals from our meta_EUR analysis with quantitative trait loci (QTL) detected for gene expression (eQTLs, n = 596,277 signals) and splicing (sQTLs, n = 253,658 signals) in the eQTL Catalogue release 7 (*27*, *28*), as well as GWAS signals for coronary artery disease (CAD) and type 2 diabetes (T2D) from two recent well-powered GWAS studies (n = 979 associations) (*1*, *5*) (**Figure 2B**). This analysis was made feasible by our GPU-accelerated re-implementation of the coloc algorithm (*38*), leading to up to 1000-fold speed-up compared to the original R implementation (see **Methods**). Colocalisation with eQTLs and sQTLs linked 32,115 metabolic trait signals to 3,238 unique genes. Restricting the analysis to metabolic trait signals that also colocalised with T2D and CAD revealed 270 and 126 unique genes, respectively.

As expected, we detected several known gene-metabolite-disease relationships. For example, a splicing QTL affecting the inclusion of exon 13 of 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGCR*) colocalised with both CAD and LDL cholesterol as well as 127 other metabolic traits from our analysis (**Figure 2C, Table S4**). *HMGCR* is a known target for statin therapy to lower circulating LDL levels and reduce CAD risk (*39*). The strongest sQTL signal for *HMGCR* was detected in the Alasoo_2018 (*40*) macrophage dataset (**Figure S3),** where statistical fine-mapping prioritised 5-75355259-A-G as the most likely causal variant (posterior inclusion probability (PIP) = 0.64), a finding that has also been validated experimentally (*41*). Consistent with a recent report, we did not detect colocalisation between the *HMGCR* sQTL signal and T2D, suggesting that T2D association at this locus involves additional genetic mechanisms (*42*). Similarly, a known fine-mapped liver-specific eQTL for *SORT1* (1-109274968-G-T, PIP = 0.99) (*43*) colocalised with LDL cholesterol and CAD risk (**Figure S4**). Also, a known pancreatic islet-specific enhancer variant 3-123346931-A-G (*44*) was specifically detected as an eQTL in the PISA pancreatic islet dataset (*45*) for *ADCY5* and colocalised with both glucose and T2D (**Figure S4**).

While focusing on novel associations, we detected a colocalisation between an eQTL signal for *LIPA* (10-89243047-C-T), CAD and 11 metabolic traits (**Table S5**). The strongest metabolic association was detected for an inflammatory biomarker GlycA (**Figure 2D**). Consistent with a recent report (*46*), the *LIPA* eQTL signal was strongest in CD14+ monocytes and replicated across multiple independent monocyte datasets (**Table S5**). Furthermore, the GWAS effect size for GlycA was 4.3x times smaller than for CAD (**Figure**

2E), indicating that GlycA is unlikely to be a causal mediator between *LIPA* expression and CAD risk. Instead, GlycA probably represents a biomarker of an underlying inflammatory process (*47*) that drives CAD risk. This is consistent with existing literature that *LIPA* is expressed in atherosclerotic plaque macrophages, where it might contribute to atherosclerotic inflammation (*46*, *48*). Thus, our results demonstrate how systematic colocalisation leveraging well-powered studies across molecular layers (RNA splicing, gene expression, metabolic traits) can aid disease GWAS interpretation and causal hypothesis generation.

Associations with low-frequency and rare variants

While previous GWAS studies of metabolic traits profiled with NMR have primarily focused on common variation (MAF > 1%) (*10*, *13*, *14*), we tested all variants with minor allele count greater than 20. Thus, in our meta_EUR meta-analysis, 11.4% of the independent lead variants (n = 987 variants) had MAF < 1% (**Figure S5**). As expected, these low-frequency and rare variants also had larger effect sizes than those at higher allele frequencies (**Figure S5**). This estimate is a lower bound as many loci with common lead variants are likely to harbour secondary causal variants with lower allele frequencies.

Low-frequency variants have less extensive LD with neighbouring variants. We hypothesised that this should simplify causal gene prioritisation, because the lead variants should be more enriched for causal variants. To test this, we used computational variant effect prediction tools (*49–51*) to prioritise 63 low-frequency (MAF < 1%) missense or splice-altering variants (see **Methods**). Reassuringly, 31/63 genes with low-frequency missense or splice variants in our analysis had already been annotated in the GWAS Catalog to be associated with the same or a related trait (**Table S6**). These included low-frequency variants in *PCSK9* and *APOC3* genes associated with lipid/lipoprotein traits, a missense variant in phenylalanine hydroxylase (*PAH*) associated with phenylalanine, and a missense variant in histidine ammonia lyase (*HAL*) associated with histidine, among many others (**Table S6**). However, 32/63 of our prioritised variants were in novel genes. For example, we detected an association between the X-24503382-A-G (rs138321172) missense variant in pyruvate dehydrogenase kinase 3 (*PDK3*) and pyruvate. The *PDK3* association had likely been missed by previous studies, because the X chromosome is often excluded from GWAS studies.

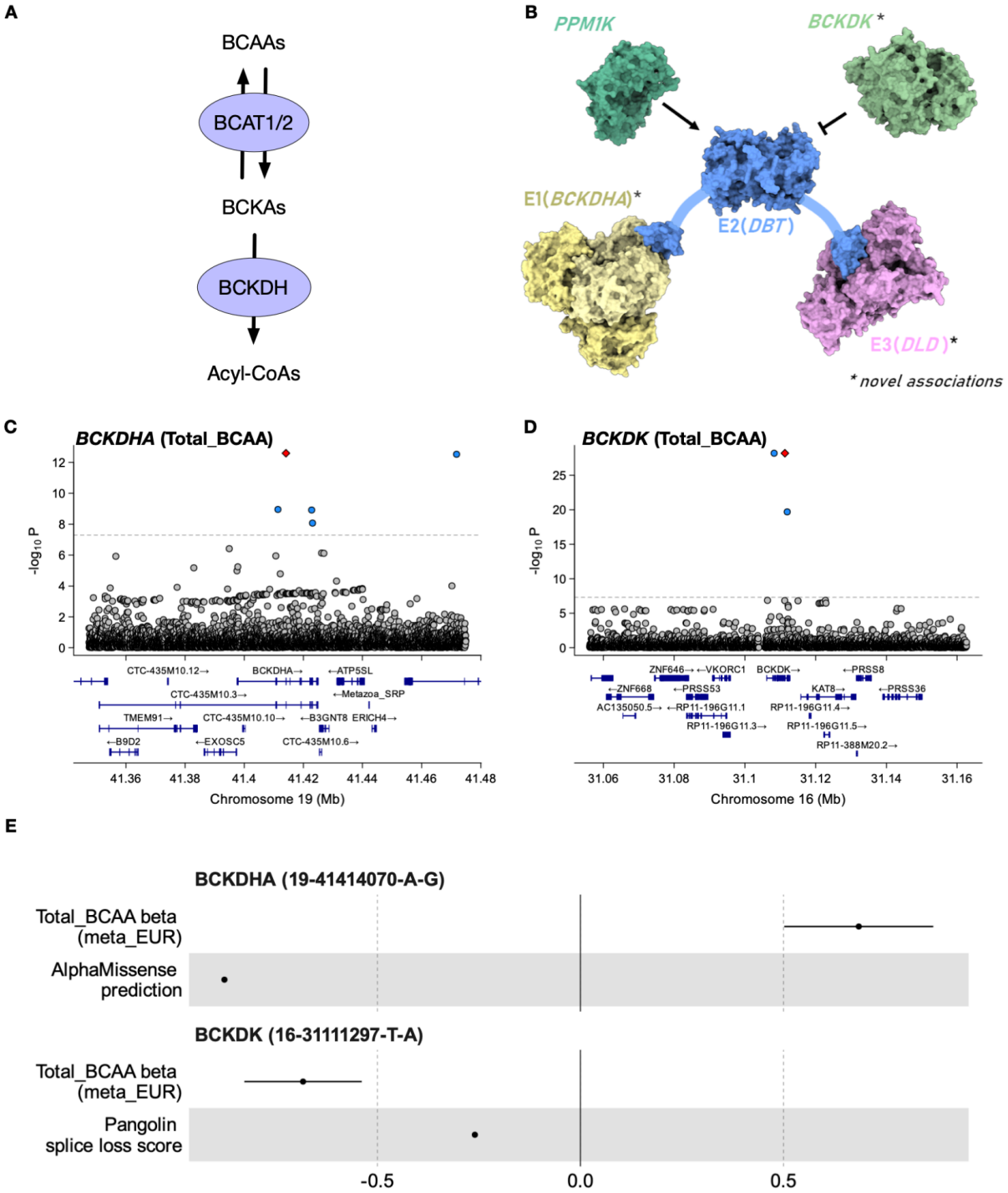

Figure 3. Convergence of common and low-frequency associations at the branched-chain amino acid catabolism pathway. (**A**) Branched-chain amino acids (BCAAs) are converted to branched chain keto acids (BCKAs) by branched-chain amino acid aminotransferase (encoded by *BCAT1* and *BCAT2*). This process is reversible. BCKAs can be further catabolised by the branched-chain α-keto acid dehydrogenase (BCKDH) complex to acyl-CoAs. (**B**) BCKDH complex is made up of subunits E1 (encoded by *BCKDHA* and *BCKDHB)*, E2 encoded by *DBT* and E3 encoded by *DLD*. The activity of the BCKDH complex is controlled by a kinase (*BCKDK*) that inhibits its function, and phosphatase (*PPM1K*) that reactivates it (*52*). (**C**) GWAS association signal for Total_BCAA

in the *BCKDHA* gene region. The *BCKDHA* missense variant 19-41414070-A-G is highlighted in red. (**D**) GWAS association signal for Total_BCAA in the *BCKDK* gene region. The predicted *BCKDK* splice loss variant 16-31111297-T-A is highlighted in red. (**E**) The effect of the *BCKDHA* missense variant and *BCKDK* splice loss variant on Total_BCAA (beta + 95% confidence interval) together with variant effect predictions on protein function from AlphaMissense (*53*) and Pangolin (*51*) models (arbitrary units).

As another example of novel genes not previously reported in the GWAS Catalog, we observed convergence of common and low-frequency variants on the BCAA catabolism pathway. The first two steps of BCAA catabolism are transamination of valine, leucine and isoleucine catalyzed by branched-chain aminotransferase (encoded by *BCAT1/2* genes) followed by oxidative decarboxylation catalyzed by the branched-chain α-keto acid dehydrogenase (BCKDH) complex (**Figure 3A**) (*52*). The BCKDH complex is made up of three proteins: E1 subunit encoded by the branched chain keto acid dehydrogenase E1 subunit alpha and beta genes (*BCKDHA, BCKDHB*), E2 subunit encoded by dihydrolipoyl transacylase (*DBT*) and E3 subunit encoded by dihydrolipoamide dehydrogenase (*DLD*) (*52*) (**Figure 3B**). The activity of the *BCKDH* complex is further regulated by the branched chain keto acid dehydrogenase kinase (*BCKDK*) that inhibits its activity, and protein phosphatase 2Cm (encoded by *PPM1K)* that reactivates it (**Figure 3B**).

Common-variant associations for three of the six genes (*BCAT2*, *DBT*, and *PPM1K*) have been reported in previous GWAS studies for BCAAs. We additionally detected a low frequency (MAF = 0.012%, $p = 2.6\times10^{-13}$) missense variant 19-41414070-A-G (rs771686663) in *BCKDHA* (**Figure 3C)** and a low frequency (MAF = 0.047%, $p = 6.5\times10^{-29}$) splice region variant 16-31111297-T-A (rs118042732) in *BCKDK* (**Figure 3D)**. In both cases, the predicted variant effects were directionally consistent with the sign of the GWAS associations (**Figure 3E**). The *BCKDHA* missense variant was predicted by CADD (*54*) and AlphaMissense (*53*) to be deleterious and was associated with increased BCAA levels. In contrast, the rs118042732 splice region variant was predicted by Pangolin deep learning splice site prediction model (*51*) to lead to splice acceptor loss and was associated with decreased BCAA levels (consistent with BCKDK being a negative regulator of the BCKDH complex) (**Figure 3B**). Reassuringly, both *BCKDK* ($p < 1\times10^{-30}$) and *BCKDHA* ($p < 1\times10^{-13}$) were also found to be associated with BCAAs in a parallel effort that performed rare-variant burden testing and exome-wide association testing using overlapping UK Biobank NMR samples (*14*), confirming that our rare variant imputation is reliable.

Finally, we detected a novel common-variant (7-107837919-T-A, MAF ~50%) association at the *DLD* locus (beta = -0.01, $p = 9.8\times10^{-13}$). Thus, we identified GWAS hits for all six key enzymes involved in the catabolism of BCAAs. This illustrates that very large sample sizes are needed to saturate the discovery of key regulators of biological processes due to either very small effects of some common variants on the target genes (*DLD*) or very low allele frequency of the genetic variants that affect those genes (*BCKDHA*, *BCKDK*). Interestingly, while the GWAS and burden testing analysis by Zoodsma *et al (14)* identified largely divergent set of genes in this pathway (*BCAT2*, *BCKDK* and *BCKDHA* from burden testing and *BCAT2*, *DLD*, *DBT* and *PPM1K* from GWAS), our well-powered GWAS using imputed low frequency variants was able to discover all six genes in a single analysis. This is consistent with recent reports that the differences between GWAS and burden testing results can be largely explained by differential statistical power (*55*).

Extent of horizontal pleiotropy across metabolic traits

To understand the shared genetic control of various classes of metabolic traits, we explored genetic correlations between all 249 metabolic traits (*37*). Although the median genetic correlation across all traits was low (rg = 0.16), there were high genetic correlations between various lipoprotein traits (median rg = 0.52) as well as between other closely regulated metabolites, such as BCAAs (rg = 0.97) (**Figure S6**, **Table S7**). To better understand the molecular mechanisms behind genetic correlations, we identified clusters of lead variants that were shared ($r^2 > 0.8$) between metabolic traits. Among the 249 metabolic traits, most lead variants were significantly associated ($p < 5\times10^{-8}$) with multiple metabolites (mean = 10; median = 2). Most prominently, a common missense variant (MAF = 40%) in the glucokinase regulatory protein (*GCKR*) (2-27508073-T-C, *GCKR*:p.Leu446Pro) was significantly associated ($p < 5\times10^{-8}$) with 231 (out of 249) metabolites (**Figure 4A**). However, in many other cases, pleiotropy was restricted to the same class of metabolites such as the 5-75355259-A-G variant at the *HMGCR* locus with multiple lipid traits (**Figure S7**).

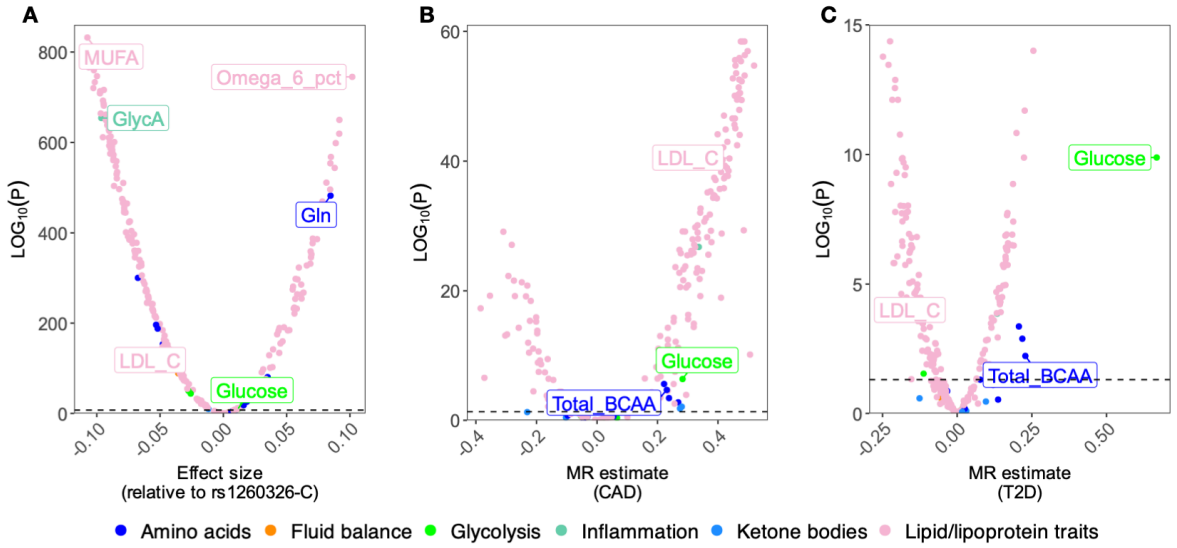

Figure 4. **Extent of pleiotropic associations across metabolic traits.** (**A**) Pleiotropic effects of the *GCKR* missense variant 2-27508073-T-C (rs1260326) on 249 metabolites. (**B**) Genome-wide MR estimates using all 249 metabolic traits as exposures and CAD as outcome. (**C**) Genome-wide MR estimates using all 249 metabolic traits as exposures and T2D as outcome.

To characterise the impact of pleiotropic genetic effects on interpreting disease associations, we performed genome-wide Mendelian randomisation (MR) using all 249 metabolic traits as exposures and either coronary artery disease (CAD) (*5*) or type 2 diabetes (T2D) (*1*) as outcomes (see Methods). For CAD, 211 of the 249 (85%) tested metabolic traits yielded significant MR estimates (FDR < 5%, **Figure 4B**) while for T2D (**Figure 4C**) the number of significant associations was 157 (63% of tested traits) (**Table S8**). Reassuringly, we recapitulated known causal effects between genetically regulated LDL cholesterol and CAD (beta = 0.43, p-value = 6.02×10^{-42}) and between glucose and T2D (beta = 0.67, p-value = 6.3×10^{-12}). We also detected a known negative association between genetically lower LDL cholesterol and T2D (beta = -0.11, p-value = 2.37×10^{-4}) (*56*). Finally, we detected genome-wide significant MR estimates between BCAA levels and both CAD and T2D

(**Figure 4B-C**). However, these genome-wide MR estimates, beyond the known causal effects of LDL and glucose, can be tricky to interpret due to extensive genetic correlation between the metabolic traits (**Figure S6, Table S7**), widespread horizontal pleiotropy, and large heterogeneity between the effect estimates from individual variants (**Table S8**).

Evaluating drug targets with *cis*-Mendelian randomisation

To limit the impact of horizontal pleiotropy, we interrogated a subset of the genome-wide MR associations using a more conservative *cis*-MR (drug-target MR) approach (*21*, *26*). Instead of capturing average genome-wide effects of circulating metabolic traits, *cis*-MR uses genetic variation in *cis* of a drug target gene to estimate the impact of perturbing gene function on disease risk (*26*). If these genes have a direct biological effect on metabolic traits, then we can use the variant effect on those traits as a proxy read-out for the (unmeasured) effect of these variants on gene function (*15*). First, we focussed on three genes with direct effect on regulating plasma LDL cholesterol levels: low density lipoprotein receptor (*LDLR*), *HMGCR*, and proprotein convertase subtilisin/kexin type 9 (*PCSK9*) (**Figure 5A**). In all three cases, we observed robust causal effects of perturbing these genes on CAD risk as previously reported (*11*, *56*). We then estimated the causal effect of lowering LDL cholesterol via these mechanisms on T2D. At the *HMGCR* locus, we detected a negative association between genetically regulated LDL cholesterol and T2D, which is consistent with previous MR studies as well as large clinical trials demonstrating that statin use is associated with increased T2D risk (*42*, *56*). Notably, while the effect of genetically regulated LDL cholesterol on CAD risk was even higher at the *LDLR* and *PCSK9* loci, the effect on T2D was strongly attenuated relative to *HMGCR* (*56*). This is consistent with clinical trials of *PCSK9* inhibitors not detecting increased risk of T2D as a side effect (*57*). Our results illustrate how a carefully conducted *cis*-MR analysis with well-powered proxy exposures can be used to evaluate drug targets and quantify their potential side effects.

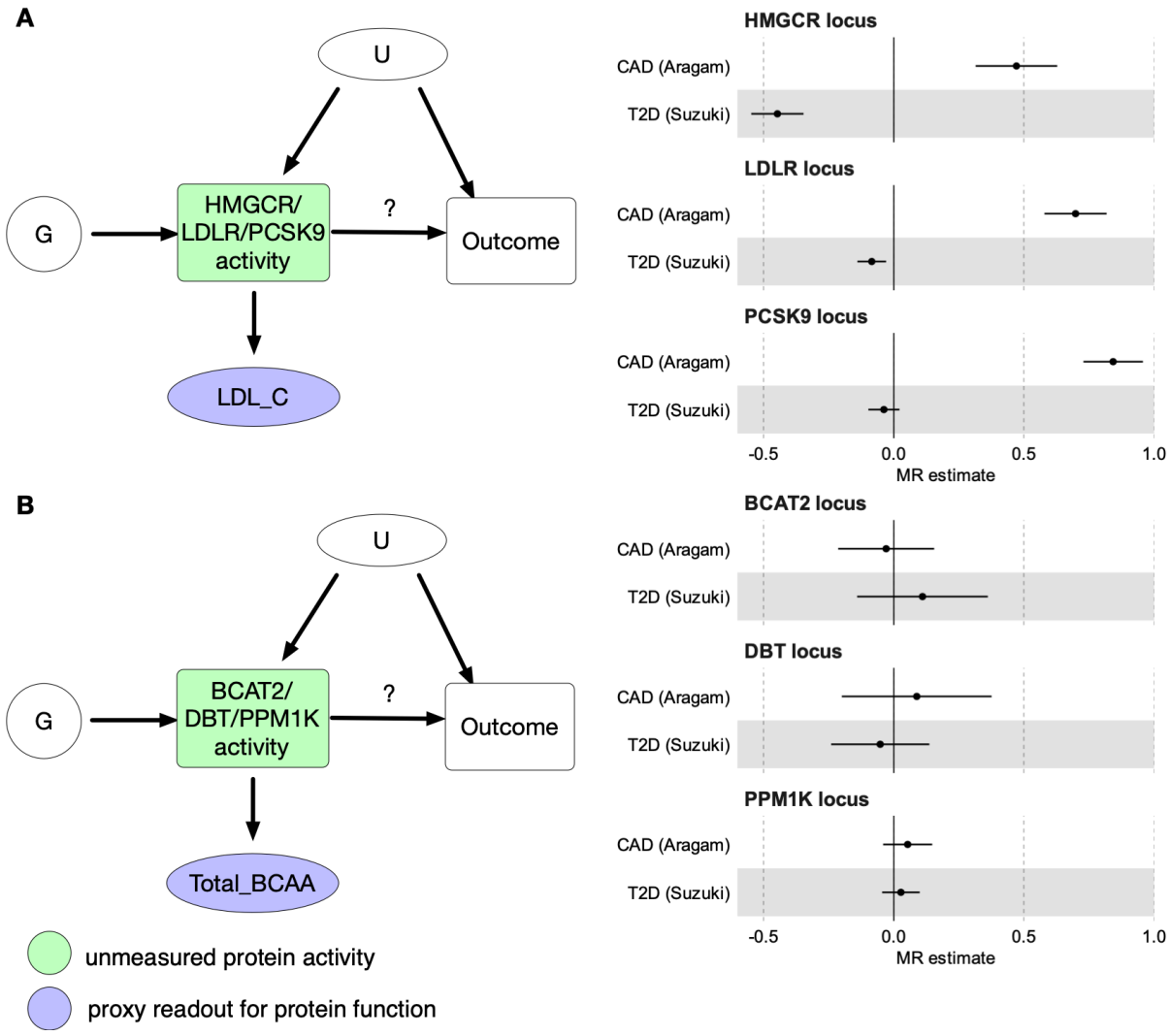

Figure 5. Drug target evaluation with *cis*-Mendelian randomisation. (**A**) Performing *cis*-MR using genetic variants G from the *cis* regions of *HMGCR*, *LDLR* and *PCSK9* genes to estimate the causal effect of inhibiting the corresponding gene function on T2D and CAD risk. LDL cholesterol (LDL_C) is used as a proxy readout for variant effects on *HMGCR*, *LDLR* and *PCSK9* function. (**B**) Performing *cis*-MR using genetic variants G from the *cis* regions of *BCAT2*, *DBT* and *PPM1K* genes to estimate the causal effect of inhibiting the corresponding protein function on T2D and CAD risk. Total branched-chain amino acid levels (Total_BCAA) is used as a proxy readout for variant effect on *BCAT2*, *DBT* and *PPM1K* function. U - unmeasured confounders.

Reassured by the ability of *cis*-MR to rediscover known associations with lipid-lowering drug targets, we next followed up the significant genome-wide MR estimates between BCAAs and both CAD and T2D (**Figure 4B-C**). Although the association between genetically regulated BCAAs and T2D has been reported before (*29*), genome-wide MR necessarily averages effects across multiple distinct mechanisms, only some of which might influence T2D risk. Furthermore, recent studies have suggested that the genome-wide MR signal between BCAAs and T2D first observed by (*29*) might be confounded by horizontal pleiotropy (*14*, *58*). As discussed above, a prominent mechanism regulating plasma BCAA levels is BCAA catabolism controlled by BCAT2 and the BCKDH complex (**Figure 3A**). To clarify the contradictory results obtained from genome-wide MR analysis and motivated by the recent

discovery of a clinical candidate of the BCKDK kinase inhibitor (*59*), we sought to assess if lowering plasma BCAA levels via specific inhibition of the BCKDK kinase could reduce T2D and CAD risk.

The most direct way to assess this would be to perform *cis*-MR with genetic variants in the *cis* region of *BCKDK* as genetic instruments, plasma BCAA levels as a proxy exposure and T2D as outcome. However, the *BCKDK* region lacks strong common variant associations and the low-frequency splice donor variant that we detected (**Figure 3D**) is too rare to have sufficient power for *cis*-MR. Instead, we focussed on *cis* variation near *DBT* and *PPM1K*, two other members of the BCKDH complex (**Figure 3B**) that have robust common variant associations. We also included *cis* variation near *BCAT2*, an enzyme directly upstream of BCKDH complex in the BCAA catabolism pathway (**Figure 3A**). In all three gene regions, the results were broadly consistent with a null effect of BCKDK inhibition on T2D risk (**Figure 5B**). None of these loci had genome-wide significant hits for T2D (**Figures S8-S9**). Although some MR method and outcome GWAS combinations (see **Methods**) did yield non-null causal effect estimates, these were not consistent across the three *cis* regions (**Tables S9-S11**). Thus, current genetic evidence does not support the hypothesis that *BCKDK* inhibition would have a large beneficial effect on reducing T2D risk.

One limitation of *cis*-MR is that due to the focus on a narrow *cis*-region for the selection of genetic instruments, it can be sensitive to linkage disequilibrium (LD) between independent causal variants (*60*, *61*). For example, although the association signals between LDL cholesterol and T2D at the *HMGCR* locus do not colocalise with each other (*42*), the two lead variants are in moderate LD ($r^2 = 0.53$), which could likely bias our *cis*-MR estimates for T2D (**Figure 2A**) at this locus. This issue of LD is likely to become more pronounced as the sample size and power of GWAS studies increase. For example, even though the lead variant at the *BCAT2* locus (19-48800958-C-T) has a genome-wide significant ($p < 5\times10^{-8}$) association with 30 lipid traits, this seems to be entirely driven by low LD ($r^2 = 0.08$) with a neighbouring *FUT2* locus lead variant (19-48703417-G-A) (**Figure S8**).

Discussion

We have created a comprehensive resource of both common and low-frequency genetic variants associated with 249 metabolic traits in up to 619,372 individuals across multiple ancestry groups. We have demonstrated the utility of the resource for GWAS interpretation via systematic phenome-wide colocalisation, low-frequency variant prioritisation and *cis*-MR analysis. To ensure that our results can be used as widely as possible, we have publicly released all summary statistics via the GWAS Catalog (*62*) and we have also made the results easy to query via an online browser (https://nmrmeta.gi.ut.ee/). Thus, our study provides the most comprehensive catalogue of genetic associations with these metabolites yet. All our analyses were performed on the latest human reference genome (GRCh38), greatly simplifying integration with other large-scale human genetics resources such as FinnGen (*63*), Million Veterans Program (*64*) and the eQTL Catalogue (*28*) without the risk of losing genetic variants through the lift-over process (*65*).

Although plasma BCAA levels have been robustly associated with T2D diabetes in observational studies (*30*–*32*), this effect seems to be primarily driven by reverse causality whereby variants associated with insulin resistance (*66*, *67*) and T2D (*58*) increase plasma

BCAA levels. Genome-wide MR studies using BCAA levels as exposure and T2D as outcome have yielded contradictory results, with our genome-wide MR and other early studies suggesting that BCAA levels increase T2D risk (*29*), but more recent studies finding that this association is significantly attenuated when pleiotropic genetic variants are excluded (*14*, *58*). However, even when care is taken to exclude pleiotropic variants, genome-wide MR still averages over multiple distinct mechanisms that could alter plasma BCAA levels, only some of which could potentially have a causal effect on T2D. Here, we performed focussed *cis*-MR analysis with genetic variants from *PPM1K*, *DBT* and *BCAT2* gene regions as instruments to directly assess the potential therapeutic effect of inhibiting BCKDK kinase on reducing T2D risk. Our results were consistent with a null effect, suggesting that the recently published clinical candidate of the BCKDK kinase inhibitor (*59*) is unlikely to reduce T2D risk.

Our work is highly complementary to a parallel effort by Zoodsma *et al* (*14*) that used an overlapping set of ~450,000 UK Biobank samples (*14*). First, including the Estonian Biobank into the meta-analysis increased the sample size by ~40% compared to UK Biobank alone and increased the number of genome-wide significant discoveries by approximately 64%. Secondly, instead of relying on old genotype imputation from the UK Biobank for GWAS analysis (*68*), we used the latest available set of imputed genotypes from both biobanks (*34–36*). This allowed us to test ~10x more variants in the UK Biobank and 2.5x more variants in the Estonian Biobank compared to Zoodsma *et al.* As a result, 11.4% of the independent lead variants detected in our analysis had a minor allele frequency (MAF) of less than 1% (**Figure S5**). Most notably, our analysis detected genome-wide significant hits for six members of the BCAA catabolism pathway, two of which (*BCKDHA* and *BCKDK*) were missed by the GWAS performed by Zoodsma *et al.,* due to lack of common variant associations at these loci. Reassuringly, both genes were detected by rare variant burden testing and exome-wide association analysis (*14*), confirming that our findings are reliable even at low allele frequencies. Thus, our results illustrate how well-powered GWAS studies that test low-frequency and rare variants can help to reconcile the apparent differences in gene discovery from common-variant GWAS and rare variant burden testing studies (*55*, *69*).

Our study also has several limitations. First, 97% of the samples included in the analysis were of predominantly European genetic ancestries. This skew limited our ability to detect genome-wide significant signals in other genetic ancestry groups and may influence the generalizability of our findings across genetic ancestry groups. As a result, the number of genome-wide significant signals increased by only 1.4% (**Table 1**) when samples from other UKBB genetic ancestry groups (AFR, AMR, CSA, EAS, MID) were included in the meta-analysis. Secondly, due to significant methodological challenges (*70*), we did not perform statistical fine mapping of the identified loci. As a result, we are likely missing many secondary signals at the genome-wide significant loci. Lastly, we applied a global genome-wide significance threshold ($p < 5\times10^{-8}$) tailored for common variants. To control for false-positive associations, this threshold may need adjustment to account for the large number of metabolic traits tested and the low allele frequency threshold (allele count > 20) utilised in our study.

While biobank-scale datasets provide unprecedented power for genetic discovery, they also introduce complexities in interpreting genetic associations due to pervasive pleiotropy. Our

results reinforce previous reports of extensive pleiotropy across metabolic trait GWASs (*10*–*12*). Some of this pleiotropy is readily interpretable, such as co-regulation between various lipid traits (*11*) or opposing effects between substrates and products of enzymatic reactions (*12*). However, given our large sample size, we also detected more cryptic pleiotropic effects such as the *GCKR* missense variant that was associated with 231/249 tested metabolic traits (**Figure 4A**). Such extensive overlaps exemplify that as cohorts grow larger, the detection of pleiotropic signals becomes more pronounced, making it harder to disentangle direct and indirect genetic effects. As a result, we caution against interpreting genome-wide MR results as evidence of direct causal effect of tested metabolic traits on the outcomes of interest (*23*, *24*). *Cis*-MR analyses are potentially less susceptible to these pleiotropic effects, but still require careful consideration of LD and detailed understanding of metabolic pathways to ensure that MR assumptions are met (*15*, *21*, *26*).

Methods

Cohorts

Estonian Biobank. The Estonian Biobank (EstBB) is a population-based biobank at the Institute of Genomics, University of Tartu (*71*). The current EstBB data freeze consists of 212,955 adult (age≥18y) participants, reflecting the age, sex and geographical distribution of the adult Estonian population, for whom biological samples as well a variety of health-related and demographic information have been collected. All biobank participants have signed a broad informed consent form and their blood sample collection was undertaken across the country between 2002 and 2021 (*71*, *72*). The activities of EstBB are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of EstBB. Nightingale Health NMR platform was used to generate plasma metabolite profiles for all individual samples in the biobank. The assay covers 249 metabolic traits ranging from low molecular weight compounds to lipids and lipoproteins. Individual level data analysis in EstBB was carried out under ethical approval 1.1-12/624 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application 6-7/GI/8988 from the EstBB.

UK Biobank. The UK Biobank is a longitudinal biomedical study of approximately half a million participants between 38-71 years old from the United Kingdom (*68*). Participant recruitment was conducted on a volunteer basis and took place between 2006 and 2010. Initial data were collected in 22 different assessment centers throughout Scotland, England, and Wales. Data collection includes elaborate genotype, environmental and lifestyle data. Blood samples were drawn at baseline for all participants, with an average of four hours since the last meal, i.e. generally non-fasting. NMR metabolic traits (Nightingale Health, quantification library 2020) were measured from EDTA plasma samples (aliquot 3) during 2019–2024 from the entire cohort. Details on the NMR metabolomic measurements in UK Biobank have been described previously for the first tranche of ~120,000 samples (*73*). The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee. This research was conducted using the UK Biobank Resource under application numbers 91233 and 30418.

Genotype imputation

Estonian Biobank. All EstBB participants have been genotyped at the Core Genotyping Lab of the Institute of Genomics, University of Tartu, using Illumina Global Screening Array v1.0, v2.0 and v3.0. Samples were genotyped and PLINK format files were created using Illumina GenomeStudio v2.0.4. Individuals were excluded from the analysis if their call-rate was < 95%, if they were outliers of the absolute value of heterozygosity (> 3SD from the mean) or if sex defined based on heterozygosity of X chromosome did not match sex in phenotype data (*34*). Before imputation, variants were filtered by call-rate < 95%, HWE p-value < 1e-4 (autosomal variants only), and minor allele frequency < 1%. Genotyped variant positions were lifted over from GRCh37 to GRCh38 with Picard. Phasing was performed using the Beagle v5.4 software (*74*). Imputation was performed with Beagle v5.4 (beagle.22Jul22.46e.jar) using default settings. Dataset was split into batches of 5,000 variants. A population specific reference panel consisting of 2,695 WGS samples (*34*) was

utilised for imputation and standard Beagle hg38 recombination maps were used. Based on principal component analysis, samples who were not of European ancestry were removed. Duplicate and monozygous twin detection was performed with KING 2.2.7 (*75*), and one sample was removed out of the pair of duplicates.

UK Biobank autosomes. Genotype imputation for the UK Biobank (UKBB) autosomal data was conducted using a high-coverage whole genome sequencing reference panel (342 million autosomal variants) from 78,195 individuals from the Genomics England (GEL) project. Reference panel construction and UK Biobank imputations have been described previously (UKBB data field 21008) (*35*).

Briefly, the UKBB SNP array data consisted of 784,256 autosomal variants. Initially, 113,515 sites identified by previous centralised UK Biobank analysis as failing quality control were removed, along with an additional 39,165 sites failing a Hardy–Weinberg equilibrium test on 409,703 GBR samples, with a p-value threshold of 1^{-10}. The resulting SNP array data were mapped from the GRCh37 to GRCh38 genome build using the GATK Picard LiftOver tool. Alleles with mismatching strands but matching alleles were flipped. A further 495 sites were removed due to incompatibility between the two reference genomes, resulting in a final SNP array incorporating 631,081 autosomal variants used for phasing and imputation.

Haplotype estimation of the SNP array data, a prerequisite for imputation, was carried out one chromosome at a time using SHAPEIT4 v4.2.2 (*76*) without a reference panel, utilising the full set of UK Biobank samples. SHAPEIT4 was run with its default 15 Markov chain Monte Carlo iterations and 30 threads. Autosomal imputation using the GEL reference panel was conducted with IMPUTE5 (*77*) (v.1.1.4). The SNP array data were divided into 408 consecutive and overlapping chunks of approximately 5 megabases (Mb) each, with a 2.5 Mb buffer across the genome using the Chunker program in IMPUTE5. Each chunk was further divided into 24 sample batches, each containing 20,349 samples. IMPUTE5 was run on each of the 9,792 subsets using a single thread and default settings. The resulting imputed genotype dosages are stored in BGEN format, and phasing information is stored in VCF format.

UK Biobank X chromosome. As the UKBB genotypes imputed by Genomics England did not include the X chromosome, we used the TOPMed r2 imputation for the X chromosome (UKBB data field 21007). Imputation was performed using the TopMed Imputation Server (*78*). The data were divided into 10 Mb chunks, and each chunk underwent several checks to ensure validity. These checks included verifying the inclusion of variants in the reference panel, ensuring a sufficient overlap with the reference panel, and maintaining an adequate sample call rate. Chunks that did not meet these criteria were excluded from further analysis. Overall, quality control methods employed by the TopMed Imputation Server were slightly more conservative than those employed by GEL and thus the sample size for each sub-population decreased by roughly 0.5% (final sample sizes: AFR - 6,411; AMR - 925; CSA - 8,627; EAS - 2,595; EUR - 412,523; MID - 1,491). Genotype phasing was performed with Eagle2 (*79*) and imputation was conducted with mimimac4 (*78*). After imputation, all chunks of each chromosome were merged into a single file. For chromosome X, additional checks were performed to verify ploidy and ensure the accuracy of mixed genotypes. The chromosome was split into three regions (PAR1, non-PAR, PAR2) for phasing and imputation, and these regions were later merged into a complete chromosome X file.

NMR metabolite data quality control and normalisation

NMR data generation in the EstBB and UKBB has been previously described (*80*). During the quality control of the Nuclear Magnetic Resonance spectroscopy (NMR) metabolomics data, we detected a difference between distributions of several metabolites (notably Ala and His) driven primarily by spectrometer and batch effect. We removed this unwanted technical variation using the R package 'ukbnmr' in both EstBB and UKBB data (*81*). We excluded individuals with more than 5 missing metabolite measurements from the cohort, confirmed that none of the 249 metabolites had a significant number of missing measurements (8000 for EstBB, 24000 for UKBB), and applied a metabolite-wise inverse normal transformation to obtain the final dataset.

Association testing and meta-analysis

We conducted genome-wide association tests for each of the seven genetic ancestry groups separately using regenie v3.1.1 (*82*), with sex, age, age squared and the top principal components of the genotype data used as covariates (PC1-PC10 for EstBB, PC1-PC20 for UKBB). For step 1 (whole genome model), we used genotype calls for UKBB and genotyping data for EstBB and included variants with a minor allele frequency (MAF) of at least 1%, a minor allele count (MAC) of at least 20, Hardy-Weinberg equilibrium exact test p-values of 10^{-15} or less, and maximum per-variant and per-sample missing genotype rates of 0.1. For step 2 (association testing using a linear regression model), we used imputed genotypes and selected variants with a MAC of at least 20 and an imputation INFO score of at least 0.6.

We performed two different inverse-variance weighted fixed-effect meta-analyses: meta_EUR on individuals of predominantly European genetic ancestry (EstBB cohort and EUR genetic ancestry group of UKBB), and meta_ALL which encompasses all seven genetic ancestry groups from UKBB and EstBB.

Genetic correlations

We employed LD score regression (LDSC) (*83*) to obtain pairwise genetic correlations for all 249 NMR metabolites. Correlations were calculated between biobanks for each metabolite and between all metabolites in three of the largest datasets (EstBB, UKBB_EUR and meta_EUR) using the European reference panel LD scores.

Lead variant and locus definition

We obtained the set of dataset-metabolite-variant triplets by iterating over variants that met the genome-wide significance threshold of 5×10^{-8}. The variant with the lowest p-value was designated as the lead variant within a 2 Mb locus. In each dataset, neighbouring loci were merged into one if their lead variants were in LD with an r^2 of at least 0.05. To better evaluate the independence of lead variants, we utilised PLINK v1.90b6.26 to calculate pairwise LD between all lead variants in a single genetic ancestry group, assigning them into shared cross-metabolite clusters if r^2 was at least 0.8. The variant with the smallest p-value was assigned as the lead variant for each cluster.

Colocalisation

We performed colocalisation between all genome-wide significant locus-trait pairs from the meta_EUR meta-analysis, GWAS hits for T2D and CAD from two large meta-analyses (*1*, *5*), and all gene expression and splicing QTLs from the eQTL Catalogue release 7 (*27*, *28*). For metabolite and disease GWAS loci, we extracted summary statistics from the +/- 1Mb window around each lead variant and converted those to approximate Bayes factors using the formula defined in the process.dataset() function of the coloc R package (*84*). For the eQTL Catalogue datasets, we used the fine-mapped (*85*) log Bayes factors available from the eQTL Catalogue FTP Server (*28*). For eQTLs, we included all datasets from the eQTL Catalogue dataset metadata table where the quantification method was either 'ge' or 'microarray'. For sQTLs, we included all datasets where the quantification method was 'leafcutter' (*86*). We then used the algorithm defined in the bf_bf() function of the coloc R package (*38*) to test for colocalisation between the GWAS approximate Bayes factors and eQTL/sQTL log Bayes factors. For prior probabilities, we used p1 = 1e-4, p2 = 1e-4, p12 = 1e-6. We re-implemented all computations in Python and used PyTorch (*87*) to add support for GPU acceleration. Parallelisation of colocalisation testing on GPUs combined with efficient storage of summary statistics in Parquet files allowed us to achieve ~1000-fold speed-up compared to the original R implementation (*38*). Our gpu-coloc source code is available at https://github.com/mjesse-github/gpu-coloc.

Prioritisation of low-frequency missense and splice-altering variants

We first identified all independent lead variants in the meta_EUR analysis that had MAF < 1%. We then narrowed the set down by only including SNPs identified as missense or splice regions variants by the Ensembl Variant Effect Predictor (VEP) (*49*) or were predicted to alter splice junction usage by SpliceAI (*50*) or Pangolin (*51*). This approach identified 63 variants that we were able to assign to putative effector genes (Table S6). We then downloaded all associations from the GWAS Catalog (*62*) (https://www.ebi.ac.uk/gwas/docs/file-downloads, accessed 8 February 2025) where the 'Mapped gene' column matched one of our 63 prioritised genes. Finally, we assessed if the traits reported in the GWAS Catalog were related to the metabolic traits detected in our analysis (Table S6).

Genome-wide Mendelian randomisation

We performed genome-wide MR between all 249 metabolic traits and two diseases, CAD (*5*) and T2D (*1*), resulting in a total of 498 analyses. For each metabolic trait, we identified instrumental variables using a greedy LD pruning approach applied to its lead variants with MAF > 1%. This involved (A) assigning the lead with the lowest p-value in the initial set to the instrument set, (B) discarding that variant and all variants in LD with it ($r^2 < 0.01$) from the initial set, and (C) repeating steps A and B until no variants remained in the initial set. For the MR analysis itself, we used multiplicative random-effects inverse-variance weighted MR (IVW-MR) (implemented in MendelianRandomization R package (*88*)) as recommended by recent best practices (*21*).

cis-Mendelian randomisation

For the primary *cis*-MR analysis, we included only variants from the +/- 200kb region around the gene body of the target gene that had MAF > 1%. For instrument selection, we used the

LD information from the UK Biobank Genomics England imputation (Pan-UKBB EUR subset, n = 413,897) and used greedy pruning strategy to only retain variants with $p < 5\times10^{-8}$ and $r^2 < 0.01$. Although previous studies have used more relaxed r^2 thresholds for LD pruning (*11*), we found that our high statistical power required a more stringent filtering to avoid including many variants with low residual LD. We performed the primary *cis*-MR analysis using two pleiotropy-robust methods that we found to perform well in the *cis*-MR context in our previous benchmark study (*89*): multiplicative random-effects IVW-MR implemented in the MendelianRandomization R package (*88*) and MRLocus (*90*). For the IVW-MR method, we also specified 'weights = delta'. We also repeated the same analysis using MR-Egger (*91*),

To further assess the robustness of our MR results, we also tested two other *cis*-MR methods: MR-link-2 (*92*) and MR-PCA (*93*). The advantage of these methods is that instead of requiring instrument selection via LD pruning, they explicitly model the LD between all associated variants in the *cis* region. For MR-link-2 and MR-PCA, all association summary statistics were harmonized to the UK10K genotype reference (*94*) and both methods were using the default parameters provided by the 'mr_link_2_standalone.py function': MAF > 0.005, regional definition +/- 250kb, and instrument selection threshold $p < 5\times10^{-8}$ and variance explained of the LD matrix = 99%.

Structural modelling

Models of the three subunits of the BCKD complex in Figure 3b were generated using AlphaFold 3 via the AlphaFold Server (*95*). Structures of the regulatory proteins BCKDK and PPM1K were retrieved from the AlphaFold Protein Structure Database (*96*, *97*). For clarity in visualization, disordered or unstructured regions at the N- and C-termini were manually removed. Molecular graphics were performed with UCSF ChimeraX (*98*).

Data and code availability

Complete genetic ancestry group-specific and meta-analysis association summary statistics from this study can be downloaded from the GWAS Catalog (*62*) (accessions GCST90449363 - GCST90451603, Table S12). GWAS lead variants are available from Zenodo (https://dx.doi.org/10.5281/zenodo.13937265). The meta_EUR meta-analysis results can also be viewed in our PheWeb browser (https://nmrmeta.gi.ut.ee/). Meta-analysis code is available from https://github.com/ralf-tambets/EstBB-UKBB-metaanalysis/. The individual-level UK Biobank data are available for approved researchers through the UK Biobank data-access protocol (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access). The individual-level data from Estonia Biobank can be accessed through a research application to the Institute of Genomics of the University of Tartu (https://genomics.ut.ee/en/content/estonian-biobank).

Author contributions

R.T. performed GWAS analysis on the EstBB and UKBB data. I.R. developed the initial GWAS workflow. N.T., A.K., K.F. developed quality control criteria for the EstBB metabolite data. R.T. and K.A. designed the *cis*-MR and genome-wide MR analyses and interpreted the results. AvdG performed the MR-link-2 and MR-PCA analyses. M.J. performed genetic

colocalisation between molecular QTLs, metabolic traits and diseases. E.A. performed structural modelling of the BCKDH complex. D.Y. obtained SpliceAI and Pangolin predictions for all lead variants. T.E. established the NMR dataset within the EstBB. Z.K., T.E., K.A. and P.P. supervised the research. K.A., P.P., R.T, U.V., E.A. and J.K. wrote the manuscript with feedback from all authors.

Acknowledgements

We want to acknowledge the participants of the UK Biobank and Estonian Biobank for their contributions. The Estonian Genome Center analyses were partially carried out in the High Performance Computing Center, University of Tartu. We thank Eric B Fauman for feedback on an earlier version of the manuscript. This research has been conducted using the UK Biobank Resource under application numbers 91233 and 30418. Nightingale Health Plc is acknowledged for early access to the UK Biobank NMR metabolite data. UCSF ChimeraX was developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. The Estonian Biobank Research Team consists of Mari Nelis, Georgi Hudjasov, Reedik Mägi, Andres Metspalu, and Lili Milani.

Funding

K.A. and M.J. were supported by the Estonian Research Council (grant no PSG415 and MOB3ERC115). P.P., R.T., E.A, U.V., J.K. and T.E. were supported by the Estonian Research Council (grant no PRG1291). I.R. was supported by the Estonian Research Council (grant no PSG415). E.A was supported by the European Union through Horizon 2020 and Horizon Europe research and innovation programs under grants no. 894987, 101137201 and 101137154. K.F and A.K. were supported by a grant from the Estonian Research Council no PRG1197. N.T. was supported by the Estonian Research Council grant no PRG1414. Z.K. was supported by the Department of Computational Biology of the University of Lausanne.

Supplementary Figures

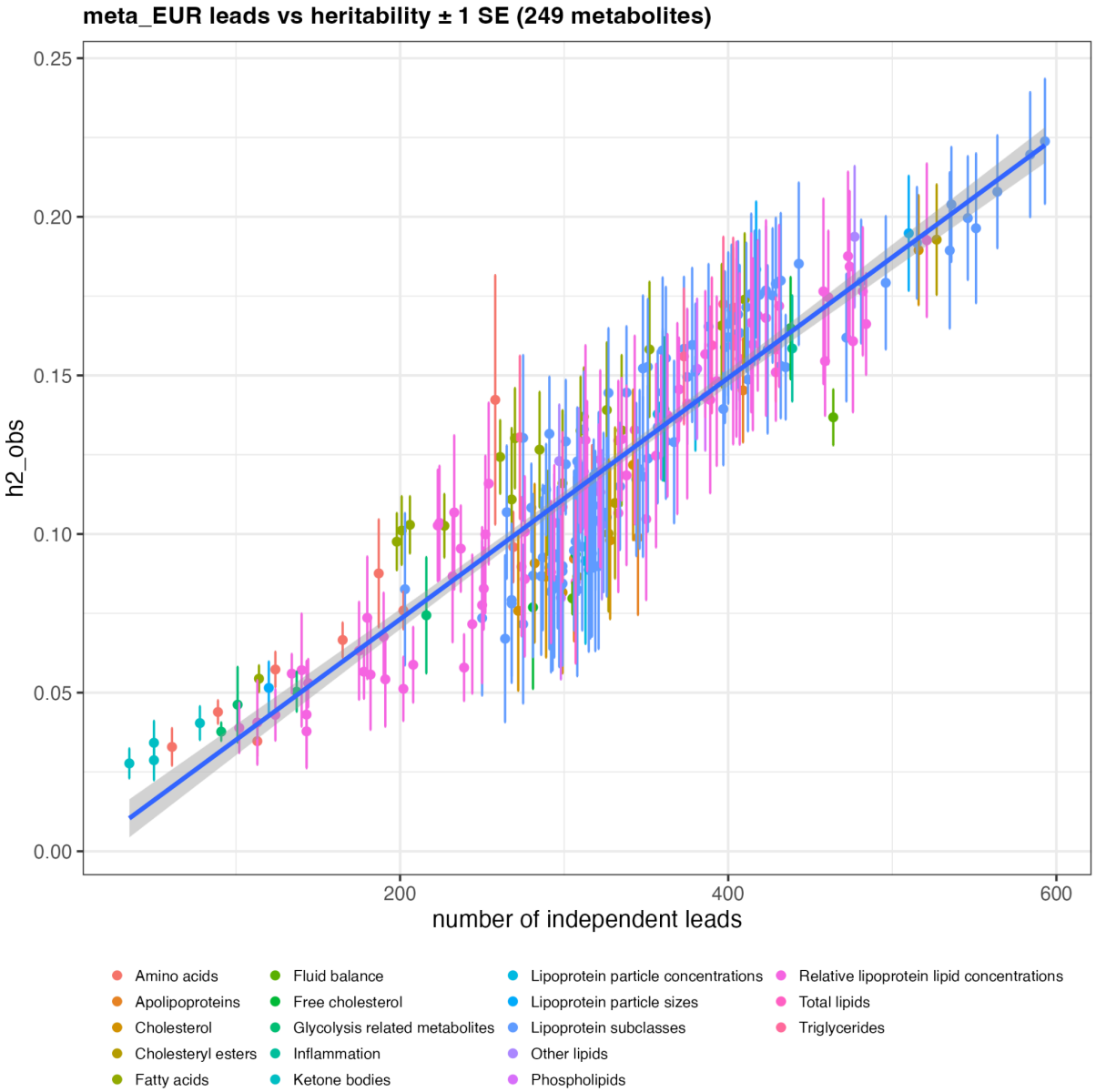

Figure S1. Relationship between heritability (h2_obs) and the number of genome-wide significant hits detected for each of the 249 metabolic traits in the meta_EUR meta-analysis.

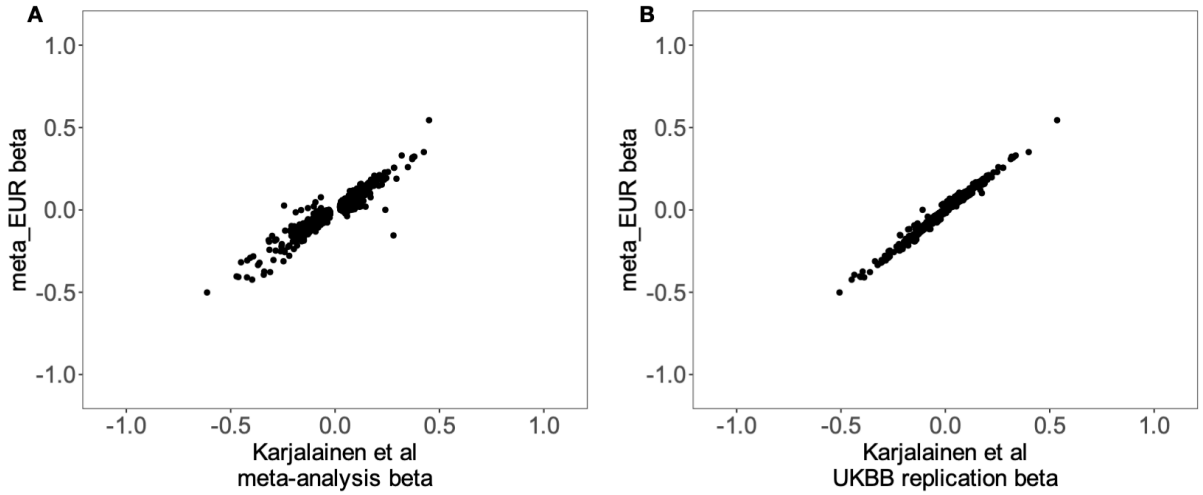

Figure S2. Comparison of shared lead variant betas in our European-ancestry meta-analysis (meta_EUR) and results presented by Karjalainen *et al*. (**A**) Scatter plot of GWAS lead variant effect sizes from Karjalainen *et al* main analysis (n = 137k, 33 cohorts) and our meta_EUR. (**B**) GWAS lead variant effect sizes from Karjalainen *et al* UK biobank replication (n = 100k) and our meta_EUR meta-analysis. Even though Karjalainen *et al* included 3,701 samples from the Estonian Biobank, these were older samples profiled in 2011- 2012 that were excluded from our meta-analysis due to significant batch effects. Thus, there is no sample overlap between our meta-analysis and the primary analysis conducted by Karjalainen *et al* (panel **A**). The ~100,000 UK Biobank samples used for replication by Karjalainen *et al* were also part of our meta-analysis, explaining the extremely high concordance in GWAS effect sizes.

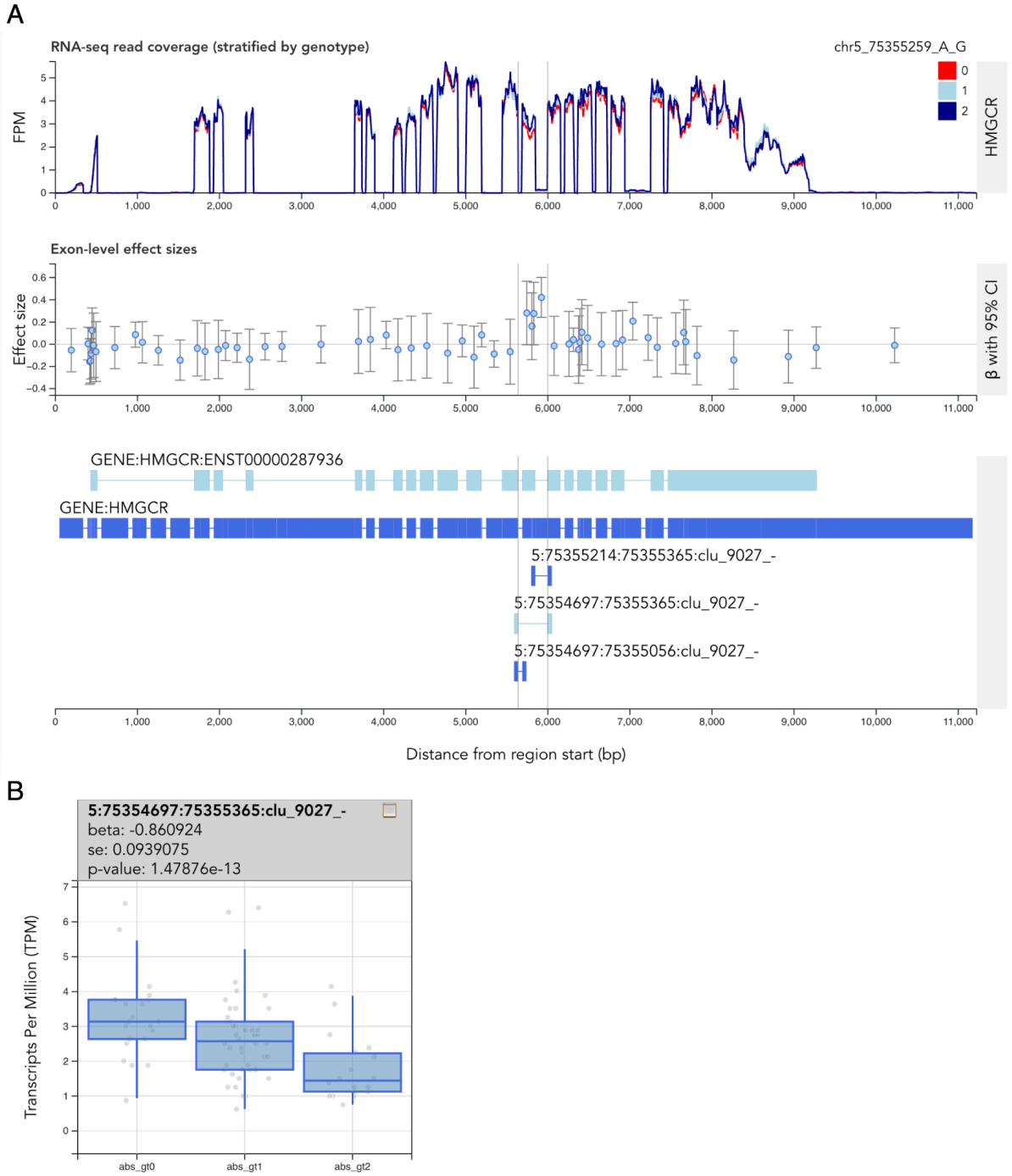

Figure S3. ***HMGCR* sQTL signal in the Alasoo_2018 dataset.** (**A**) RNA-seq read coverage across the HMGCR gene stratified by the genotype of the lead sQTL variant (5-75355259-A-G). (**B**) Usage of the exon13-skipping splice junction stratified by the genotype of the lead sQTL variant. Interactive visualisation available from here.

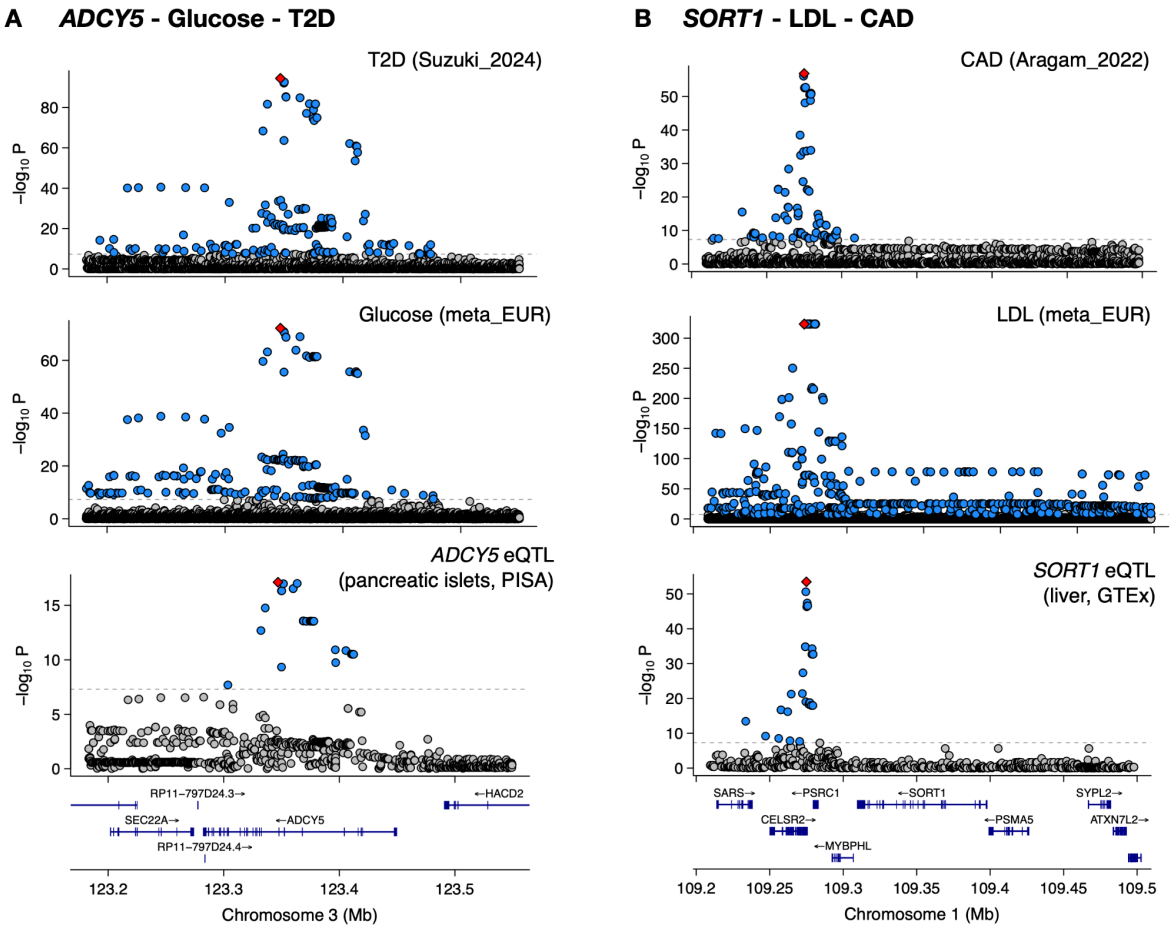

Figure S4. Examples of colocalising eQTL - metabolic trait - disease triplets. (**A**) Colocalisation between *ADCY5* eQTL in pancreatic islets, plasma glucose and T2D GWAS. (**B**) Colocalisation between SORT1 eQTL in the liver, plasma LDL cholesterol and CAD.

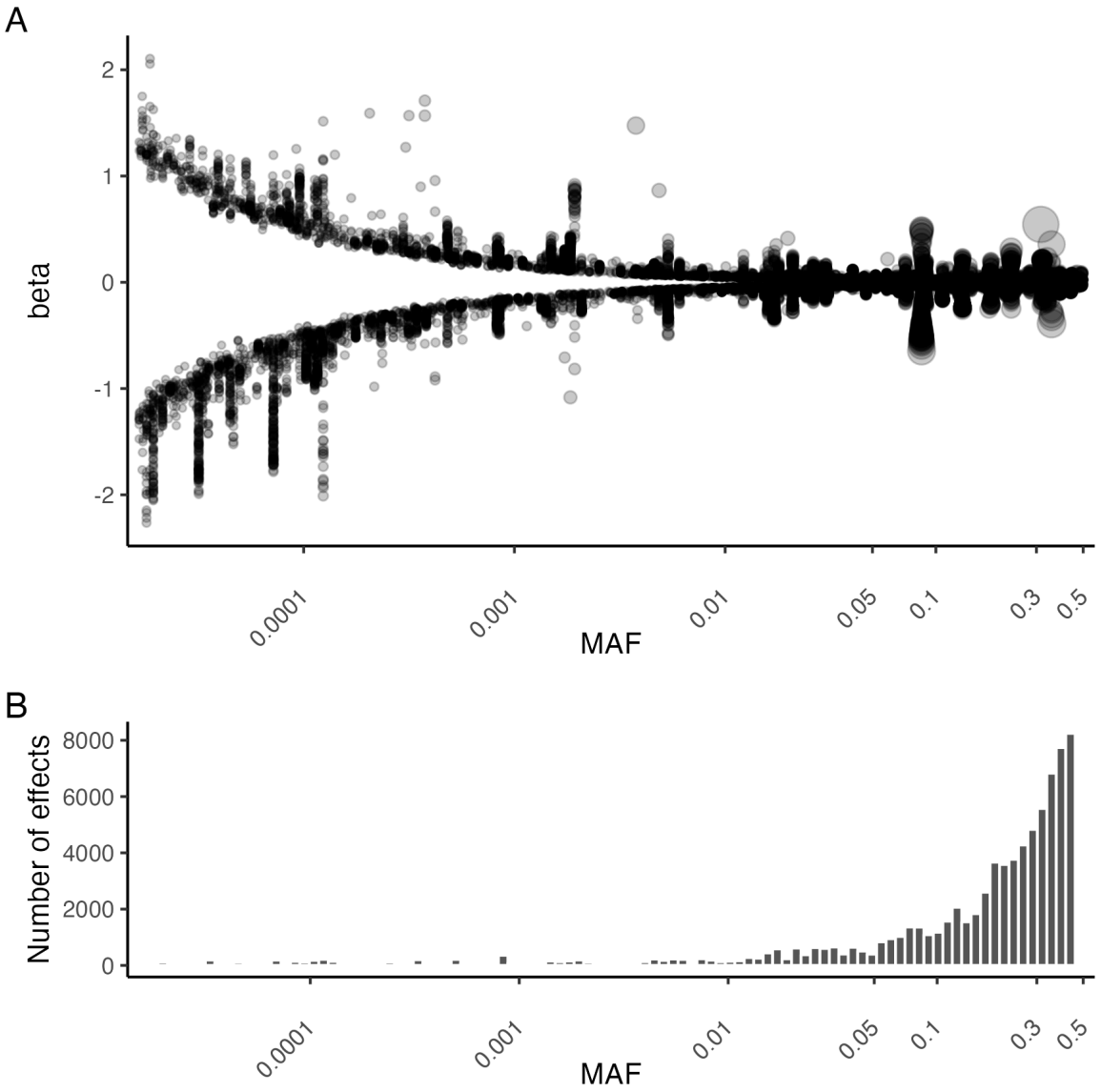

Figure S5. **Detected metabolite trait associations with low-frequency variants.** (**A**) Relationship between the lead variant minor allele frequency (MAF) and effect size (beta). Each dot signifies the lead variant (+/- 1Mb window) from each locus-trait pair (meta_EUR). The size of each dot has been scaled by $-\log_{10}$ p-value. (**B**) Number of detected significant associations in relation to the lead variant MAF in meta_EUR analysis.

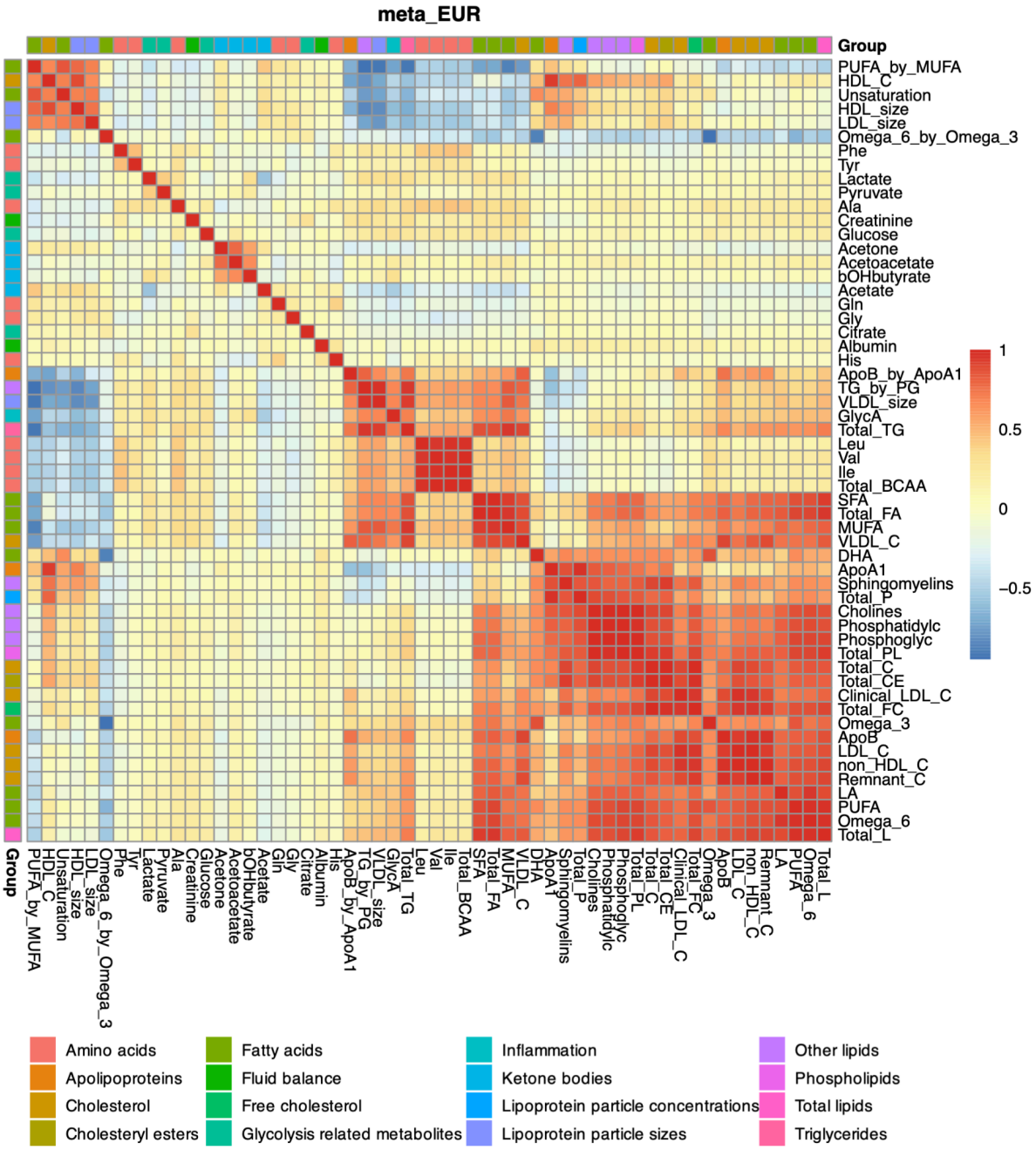

Figure S6. Heatmap of pairwise genetic correlations between metabolic traits in the meta_EUR dataset. The heatmap shows a representative subset of 56 metabolic traits from the main metabolic classes. The complete genetic correlation matrix for all 249 metabolic traits is presented in Table S7.

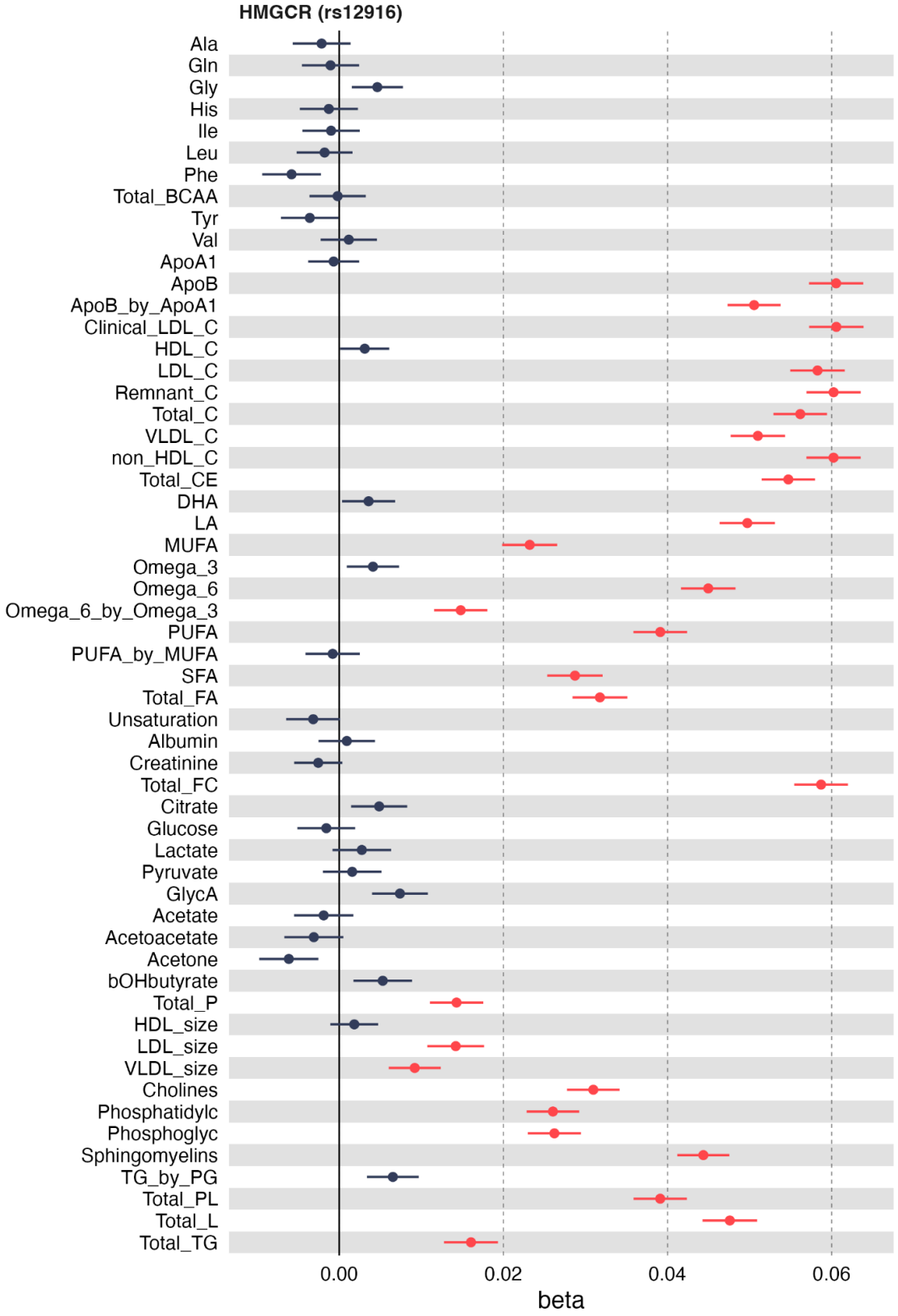

Figure S7. Pleiotropic effects of the *HMGCR* locus lead variant rs12916 on many lipid-related metabolites. The forest plot shows a representative subset of 56 metabolic traits from the main metabolic classes.

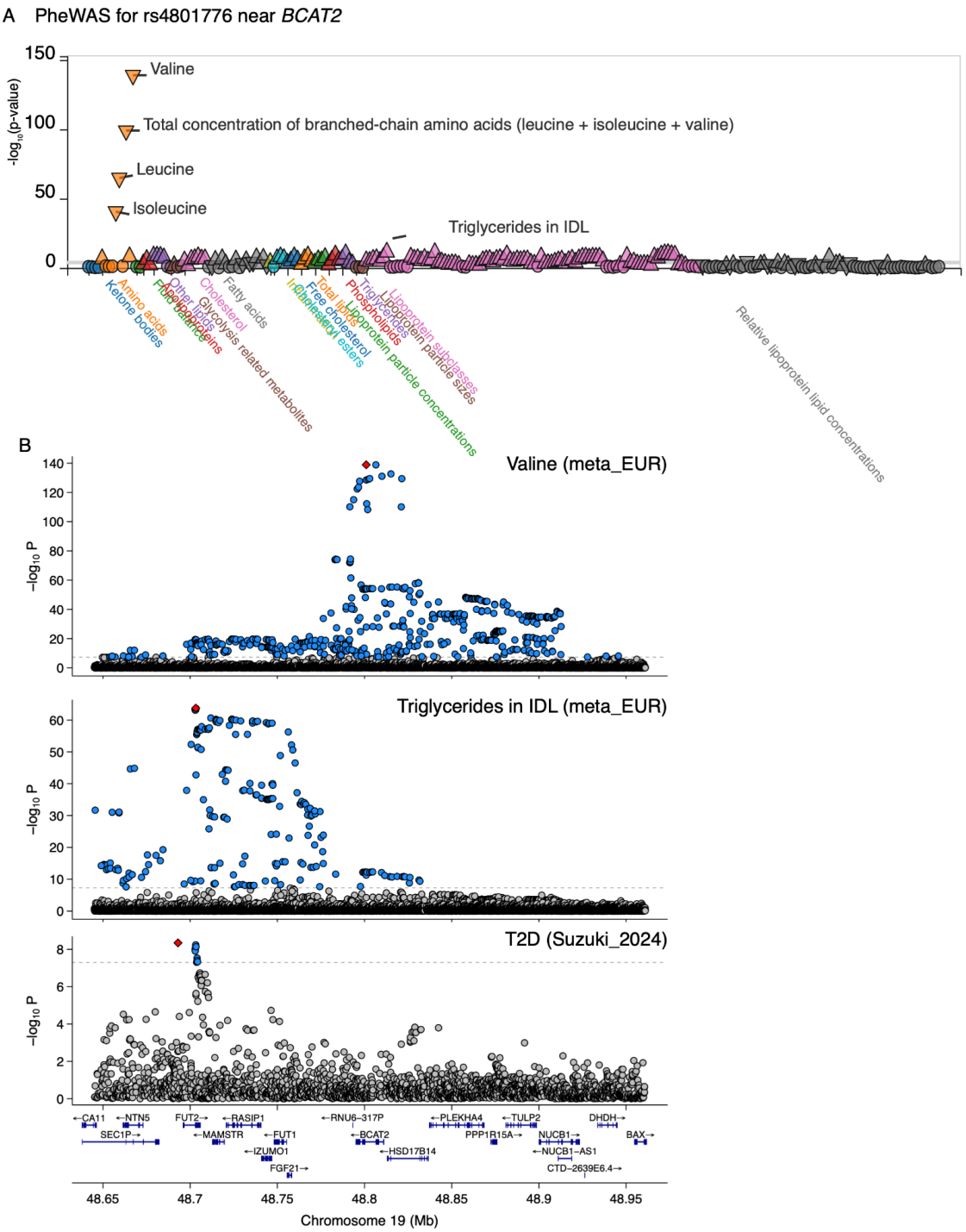

Figure S8. Association signals for BCAAs, lipoprotein traits and T2D near *BCAT2*. (**A**) PheWAS plot for the valine lead variant (rs4801776) in the intron of *BCAT2*. In addition to very strong associations with all three branched-chain amino acids, we also see genome-wide significant associations with various lipid traits. (**B**) Regional association plots for Valine, Triglycerides in IDL and T2D in the *BCAT2* region. The association between rs4801776 and lipid traits seems to be driven by an independent lipid signal near the *FUT2* gene that has low LD (r^2 = 0.08) with the valine lead variant.

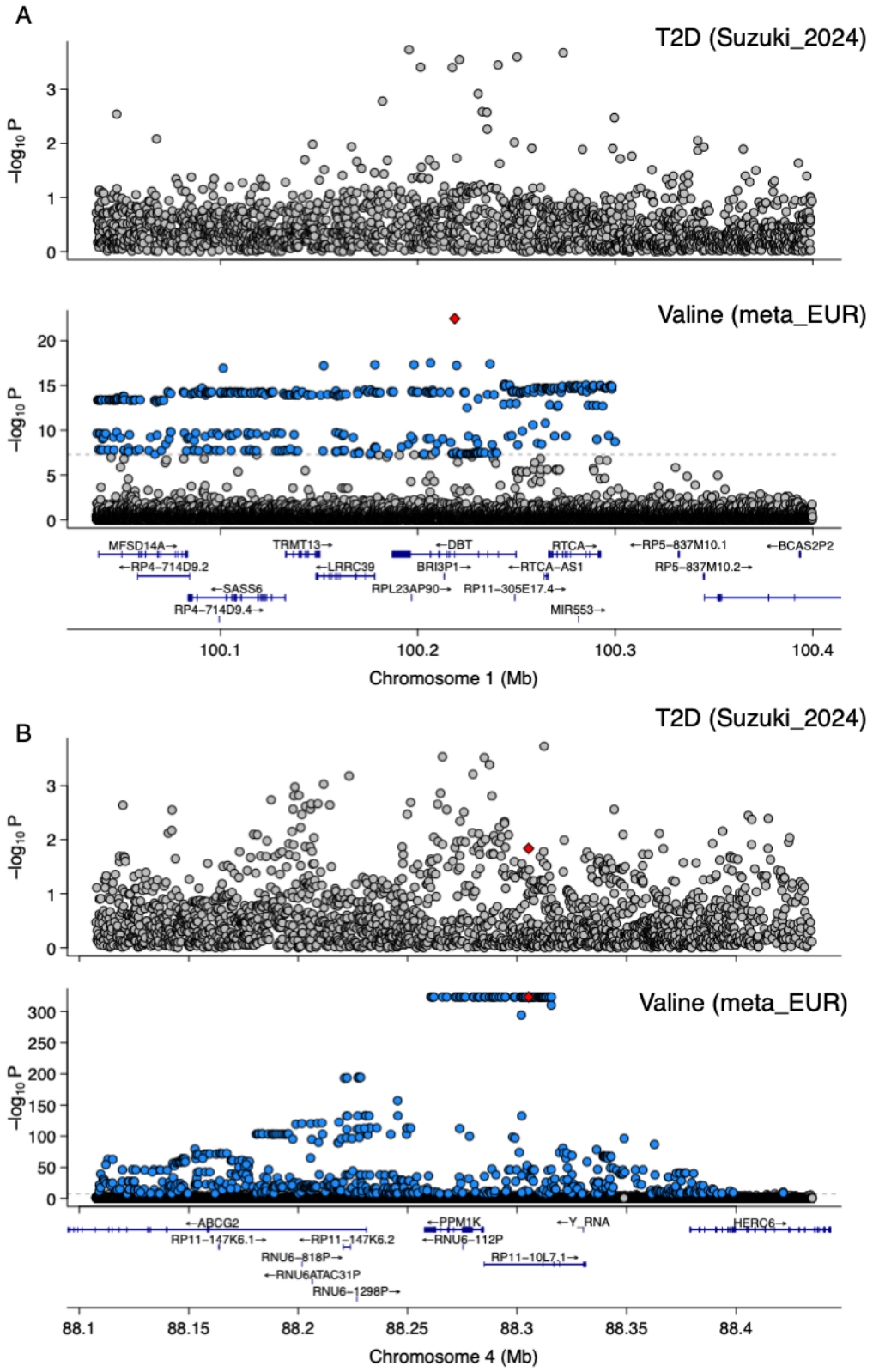

Figure S9. Regional association plots for T2D and valine in the *cis* regions of (**A**) *DBT* and (**B**) *PPM1K*.

References

1. K. Suzuki, K. Hatzikotoulas, L. Southam, H. J. Taylor, X. Yin, K. M. Lorenz, R. Mandla, A. Huerta-Chagoya, G. E. M. Melloni, S. Kanoni, N. W. Rayner, O. Bocher, A. L. Arruda, K. Sonehara, S. Namba, S. S. K. Lee, M. H. Preuss, L. E. Petty, P. Schroeder, B. Vanderwerff, M. Kals, F. Bragg, K. Lin, X. Guo, W. Zhang, J. Yao, Y. J. Kim, M. Graff, F. Takeuchi, J. Nano, A. Lamri, M. Nakatochi, S. Moon, R. A. Scott, J. P. Cook, J.-J. Lee, I. Pan, D. Taliun, E. J. Parra, J.-F. Chai, L. F. Bielak, Y. Tabara, Y. Hai, G. Thorleifsson, N. Grarup, T. Sofer, M. Wuttke, C. Sarnowski, C. Gieger, D. Nousome, S. Trompet, S.-H. Kwak, J. Long, M. Sun, L. Tong, W.-M. Chen, S. S. Nongmaithem, R. Noordam, V. J. Y. Lim, C. H. T. Tam, Y. Y. Joo, C.-H. Chen, L. M. Raffield, B. P. Prins, A. Nicolas, L. R. Yanek, G. Chen, J. A. Brody, E. Kabagambe, P. An, A. H. Xiang, H. S. Choi, B. E. Cade, J. Tan, K. A. Broadaway, A. Williamson, Z. Kamali, J. Cui, M. Thangam, L. S. Adair, A. Adeyemo, C. A. Aguilar-Salinas, T. S. Ahluwalia, S. S. Anand, A. Bertoni, J. Bork-Jensen, I. Brandslund, T. A. Buchanan, C. F. Burant, A. S. Butterworth, M. Canouil, J. C. N. Chan, L.-C. Chang, M.-L. Chee, J. Chen, S.-H. Chen, Y.-T. Chen, Z. Chen, L.-M. Chuang, M. Cushman, J. Danesh, S. K. Das, H. J. de Silva, G. Dedoussis, L. Dimitrov, A. P. Doumatey, S. Du, Q. Duan, K.-U. Eckardt, L. S. Emery, D. S. Evans, M. K. Evans, K. Fischer, J. S. Floyd, I. Ford, O. H. Franco, T. M. Frayling, B. I. Freedman, P. Genter, H. C. Gerstein, V. Giedraitis, C. González-Villalpando, M. E. González-Villalpando, P. Gordon-Larsen, M. Gross, L. A. Guare, S. Hackinger, L. Hakaste, S. Han, A. T. Hattersley, C. Herder, M. Horikoshi, A.-G. Howard, W. Hsueh, M. Huang, W. Huang, Y.-J. Hung, M. Y. Hwang, C.-M. Hwu, S. Ichihara, M. A. Ikram, M. Ingelsson, M. T. Islam, M. Isono, H.-M. Jang, F. Jasmine, G. Jiang, J. B. Jonas, T. Jørgensen, F. K. Kamanu, F. R. Kandeel, A. Kasturiratne, T. Katsuya, V. Kaur, T. Kawaguchi, J. M. Keaton, A. N. Kho, C.-C. Khor, M. G. Kibriya, D.-H. Kim, F. Kronenberg, J. Kuusisto, K. Läll, L. A. Lange, K. M. Lee, M.-S. Lee, N. R. Lee, A. Leong, L. Li, Y. Li, R. Li-Gao, S. Ligthart, C. M. Lindgren, A. Linneberg, C.-T. Liu, J. Liu, A. E. Locke, T. Louie, J. ’an Luan, A. O. Luk, X. Luo, J. Lv, J. A. Lynch, V. Lyssenko, S. Maeda, V. Mamakou, S. R. Mansuri, K. Matsuda, T. Meitinger, O. Melander, A. Metspalu, H. Mo, A. D. Morris, F. A. Moura, J. L. Nadler, M. A. Nalls, U. Nayak, I. Ntalla, Y. Okada, L. Orozco, S. R. Patel, S. Patil, P. Pei, M. A. Pereira, A. Peters, F. J. Pirie, H. G. Polikowsky, B. Porneala, G. Prasad, L. J. Rasmussen-Torvik, A. P. Reiner, M. Roden, R. Rohde, K. Roll, C. Sabanayagam, K. Sandow, A. Sankareswaran, N. Sattar, S. Schönherr, M. Shahriar, B. Shen, J. Shi, D. M. Shin, N. Shojima, J. A. Smith, W. Y. So, A. Stančáková, V. Steinthorsdottir, A. M. Stilp, K. Strauch, K. D. Taylor, B. Thorand, U. Thorsteinsdottir, B. Tomlinson, T. C. Tran, F.-J. Tsai, J. Tuomilehto, T. Tusie-Luna, M. S. Udler, A. Valladares-Salgado, R. M. van Dam, J. B. van Klinken, R. Varma, N. Wacher-Rodarte, E. Wheeler, A. R. Wickremasinghe, K. W. van Dijk, D. R. Witte, C. S. Yajnik, K. Yamamoto, K. Yamamoto, K. Yoon, C. Yu, J.-M. Yuan, S. Yusuf, M. Zawistowski, L. Zhang, W. Zheng, VA Million Veteran Program, L. J. Raffel, M. Igase, E. Ipp, S. Redline, Y. S. Cho, L. Lind, M. A. Province, M. Fornage, C. L. Hanis, E. Ingelsson, A. B. Zonderman, B. M. Psaty, Y.-X. Wang, C. N. Rotimi, D. M. Becker, F. Matsuda, Y. Liu, M. Yokota, S. L. R. Kardia, P. A. Peyser, J. S. Pankow, J. C. Engert, A. Bonnefond, P. Froguel, J. G. Wilson, W. H. H. Sheu, J.-Y. Wu, M. G. Hayes, R. C. W. Ma, T.-Y. Wong, D. O. Mook-Kanamori, T. Tuomi, G. R. Chandak, F. S. Collins, D. Bharadwaj, G. Paré, M. M. Sale, H. Ahsan, A. A. Motala, X.-O. Shu, K.-S. Park, J. W. Jukema, M. Cruz, Y.-D. I. Chen, S. S. Rich, R. McKean-Cowdin, H. Grallert, C.-Y. Cheng, M. Ghanbari, E.-S. Tai, J. Dupuis, N. Kato, M. Laakso, A. Köttgen, W.-P. Koh, D. W. Bowden, C. N. A. Palmer, J. S. Kooner, C. Kooperberg, S. Liu, K. E. North, D. Saleheen, T. Hansen, O. Pedersen, N. J. Wareham, J. Lee, B.-J. Kim, I. Y. Millwood, R. G. Walters, K. Stefansson, E. Ahlqvist, M. O. Goodarzi, K. L. Mohlke, C. Langenberg, C. A. Haiman, R. J. F. Loos, J. C. Florez, D. J. Rader, M. D. Ritchie, S. Zöllner, R. Mägi, N. A. Marston, C. T. Ruff, D. A. van Heel, S. Finer, J. C. Denny, T. Yamauchi, T. Kadowaki, J. C. Chambers, M. C. Y. Ng, X. Sim, J. E. Below, P. S. Tsao, K.-M. Chang, M. I.

McCarthy, J. B. Meigs, A. Mahajan, C. N. Spracklen, J. M. Mercader, M. Boehnke, J. I. Rotter, M. Vujkovic, B. F. Voight, A. P. Morris, E. Zeggini, Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).

2. W. Zhou, M. Kanai, K.-H. H. Wu, H. Rasheed, K. Tsuo, J. B. Hirbo, Y. Wang, A. Bhattacharya, H. Zhao, S. Namba, I. Surakka, B. N. Wolford, V. Lo Faro, E. A. Lopera-Maya, K. Läll, M.-J. Favé, J. J. Partanen, S. B. Chapman, J. Karjalainen, M. Kurki, M. Maasha, B. M. Brumpton, S. Chavan, T.-T. Chen, M. Daya, Y. Ding, Y.-C. A. Feng, L. A. Guare, C. R. Gignoux, S. E. Graham, W. E. Hornsby, N. Ingold, S. I. Ismail, R. Johnson, T. Laisk, K. Lin, J. Lv, I. Y. Millwood, S. Moreno-Grau, K. Nam, P. Palta, A. Pandit, M. H. Preuss, C. Saad, S. Setia-Verma, U. Thorsteinsdottir, J. Uzunovic, A. Verma, M. Zawistowski, X. Zhong, N. Afifi, K. M. Al-Dabhani, A. Al Thani, Y. Bradford, A. Campbell, K. Crooks, G. H. de Bock, S. M. Damrauer, N. J. Douville, S. Finer, L. G. Fritsche, E. Fthenou, G. Gonzalez-Arroyo, C. J. Griffiths, Y. Guo, K. A. Hunt, A. Ioannidis, N. M. Jansonius, T. Konuma, M. T. M. Lee, A. Lopez-Pineda, Y. Matsuda, R. E. Marioni, B. Moatamed, M. A. Nava-Aguilar, K. Numakura, S. Patil, N. Rafaels, A. Richmond, A. Rojas-Muñoz, J. A. Shortt, P. Straub, R. Tao, B. Vanderwerff, M. Vernekar, Y. Veturi, K. C. Barnes, M. Boezen, Z. Chen, C.-Y. Chen, J. Cho, G. D. Smith, H. K. Finucane, L. Franke, E. R. Gamazon, A. Ganna, T. R. Gaunt, T. Ge, H. Huang, J. Huffman, N. Katsanis, J. T. Koskela, C. Lajonchere, M. H. Law, L. Li, C. M. Lindgren, R. J. F. Loos, S. MacGregor, K. Matsuda, C. M. Olsen, D. J. Porteous, J. A. Shavit, H. Snieder, T. Takano, R. C. Trembath, J. M. Vonk, D. C. Whiteman, S. J. Wicks, C. Wijmenga, J. Wright, J. Zheng, X. Zhou, P. Awadalla, M. Boehnke, C. D. Bustamante, N. J. Cox, S. Fatumo, D. H. Geschwind, C. Hayward, K. Hveem, E. E. Kenny, S. Lee, Y.-F. Lin, H. Mbarek, R. Mägi, H. C. Martin, S. E. Medland, Y. Okada, A. V. Palotie, B. Pasaniuc, D. J. Rader, M. D. Ritchie, S. Sanna, J. W. Smoller, K. Stefansson, D. A. van Heel, R. G. Walters, S. Zöllner, Biobank of the Americas, Biobank Japan Project, BioMe, BioVU, CanPath - Ontario Health Study, China Kadoorie Biobank Collaborative Group, Colorado Center for Personalized Medicine, deCODE Genetics, Estonian Biobank, FinnGen, Generation Scotland, Genes & Health Research Team, LifeLines, Mass General Brigham Biobank, Michigan Genomics Initiative, National Biobank of Korea, Penn Medicine BioBank, Qatar Biobank, QSkin Sun and Health Study, Taiwan Biobank, HUNT Study, UCLA ATLAS Community Health Initiative, Uganda Genome Resource, UK Biobank, A. R. Martin, C. J. Willer, M. J. Daly, B. M. Neale, Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genom.* **2**, 100192 (2022).

3. L. Yengo, S. Vedantam, E. Marouli, J. Sidorenko, E. Bartell, S. Sakaue, M. Graff, A. U. Eliasen, Y. Jiang, S. Raghavan, J. Miao, J. D. Arias, S. E. Graham, R. E. Mukamel, C. N. Spracklen, X. Yin, S.-H. Chen, T. Ferreira, H. H. Highland, Y. Ji, T. Karaderi, K. Lin, K. Lüll, D. E. Malden, C. Medina-Gomez, M. Machado, A. Moore, S. Rüeger, X. Sim, S. Vrieze, T. S. Ahluwalia, M. Akiyama, M. A. Allison, M. Alvarez, M. K. Andersen, A. Ani, V. Appadurai, L. Arbeeva, S. Bhaskar, L. F. Bielak, S. Bollepalli, L. L. Bonnycastle, J. Bork-Jensen, J. P. Bradfield, Y. Bradford, P. S. Braund, J. A. Brody, K. S. Burgdorf, B. E. Cade, H. Cai, Q. Cai, A. Campbell, M. Cañadas-Garre, E. Catamo, J.-F. Chai, X. Chai, L.-C. Chang, Y.-C. Chang, C.-H. Chen, A. Chesi, S. H. Choi, R.-H. Chung, M. Cocca, M. P. Concas, C. Couture, G. Cuellar-Partida, R. Danning, E. W. Daw, F. Degenhard, G. E. Delgado, A. Delitala, A. Demirkan, X. Deng, P. Devineni, A. Dietl, M. Dimitriou, L. Dimitrov, R. Dorajoo, A. B. Ekici, J. E. Engmann, Z. Fairhurst-Hunter, A.-E. Farmaki, J. D. Faul, J.-C. Fernandez-Lopez, L. Forer, M. Francescatto, S. Freitag-Wolf, C. Fuchsberger, T. E. Galesloot, Y. Gao, Z. Gao, F. Geller, O. Giannakopoulou, F. Giulianini, A. P. Gjesing, A. Goel, S. D. Gordon, M. Gorski, J. Grove, X. Guo, S. Gustafsson, J. Haessler, T. F. Hansen, A. S. Havulinna, S. J. Haworth, J. He, N. Heard-Costa, P. Hebbar, G. Hindy, Y.-L. A. Ho, E. Hofer, E. Holliday, K. Horn, W. E. Hornsby, J.-J. Hottenga, H. Huang, J. Huang, A. Huerta-Chagoya, J. E. Huffman, Y.-J.

Hung, S. Huo, M. Y. Hwang, H. Iha, D. D. Ikeda, M. Isono, A. U. Jackson, S. Jäger, I. E. Jansen, I. Johansson, J. B. Jonas, A. Jonsson, T. Jørgensen, I.-P. Kalafati, M. Kanai, S. Kanoni, L. L. Kårhus, A. Kasturiratne, T. Katsuya, T. Kawaguchi, R. L. Kember, K. A. Kentistou, H.-N. Kim, Y. J. Kim, M. E. Kleber, M. J. Knol, A. Kurbasic, M. Lauzon, P. Le, R. Lea, J.-Y. Lee, H. L. Leonard, S. A. Li, X. Li, X. Li, J. Liang, H. Lin, S.-Y. Lin, J. Liu, X. Liu, K. S. Lo, J. Long, L. Lores-Motta, J. 'an Luan, V. Lyssenko, L.-P. Lyytikäinen, A. Mahajan, V. Mamakou, M. Mangino, A. Manichaikul, J. Marten, M. Mattheisen, L. Mavarani, A. F. McDaid, K. Meidtner, T. L. Melendez, J. M. Mercader, Y. Milaneschi, J. E. Miller, I. Y. Millwood, P. P. Mishra, R. E. Mitchell, L. T. Møllehave, A. Morgan, S. Mucha, M. Munz, M. Nakatochi, C. P. Nelson, M. Nethander, C. W. Nho, A. A. Nielsen, I. M. Nolte, S. S. Nongmaithem, R. Noordam, I. Ntalla, T. Nutile, A. Pandit, P. Christofidou, K. Pärna, M. Pauper, E. R. B. Petersen, L. V. Petersen, N. Pitkänen, O. Polašek, A. Poveda, M. H. Preuss, S. Pyarajan, L. M. Raffield, H. Rakugi, J. Ramirez, A. Rasheed, D. Raven, N. W. Rayner, C. Riveros, R. Rohde, D. Ruggiero, S. E. Ruotsalainen, K. A. Ryan, M. Sabater-Lleal, R. Saxena, M. Scholz, A. Sendamarai, B. Shen, J. Shi, J. H. Shin, C. Sidore, C. M. Sitlani, R. C. Slieker, R. A. J. Smit, A. V. Smith, J. A. Smith, L. J. Smyth, L. Southam, V. Steinthorsdottir, L. Sun, F. Takeuchi, D. S. P. Tallapragada, K. D. Taylor, B. O. Tayo, C. Tcheandjieu, N. Terzikhan, P. Tesolin, A. Teumer, E. Theusch, D. J. Thompson, G. Thorleifsson, P. R. H. J. Timmers, S. Trompet, C. Turman, S. Vaccargiu, S. W. van der Laan, P. J. van der Most, J. B. van Klinken, J. van Setten, S. S. Verma, N. Verweij, Y. Veturi, C. A. Wang, C. Wang, L. Wang, Z. Wang, H. R. Warren, W. Bin Wei, A. R. Wickremasinghe, M. Wielscher, K. L. Wiggins, B. S. Winsvold, A. Wong, Y. Wu, M. Wuttke, R. Xia, T. Xie, K. Yamamoto, J. Yang, J. Yao, H. Young, N. A. Yousri, L. Yu, L. Zeng, W. Zhang, X. Zhang, J.-H. Zhao, W. Zhao, W. Zhou, M. E. Zimmermann, M. Zoledziewska, L. S. Adair, H. H. H. Adams, C. A. Aguilar-Salinas, F. Al-Mulla, D. K. Arnett, F. W. Asselbergs, B. O. Åsvold, J. Attia, B. Banas, S. Bandinelli, D. A. Bennett, T. Bergler, D. Bharadwaj, G. Biino, H. Bisgaard, E. Boerwinkle, C. A. Böger, K. Bønnelykke, D. I. Boomsma, A. D. Børglum, J. B. Borja, C. Bouchard, D. W. Bowden, I. Brandslund, B. Brumpton, J. E. Buring, M. J. Caulfield, J. C. Chambers, G. R. Chandak, S. J. Chanock, N. Chaturvedi, Y.-D. I. Chen, Z. Chen, C.-Y. Cheng, I. E. Christophersen, M. Ciullo, J. W. Cole, F. S. Collins, R. S. Cooper, M. Cruz, F. Cucca, L. A. Cupples, M. J. Cutler, S. M. Damrauer, T. M. Dantoft, G. J. de Borst, L. C. P. G. M. de Groot, P. L. De Jager, D. P. V. de Kleijn, H. Janaka de Silva, G. V. Dedoussis, A. I. den Hollander, S. Du, D. F. Easton, P. J. M. Elders, A. H. Eliassen, P. T. Ellinor, S. Elmståhl, J. Erdmann, M. K. Evans, D. Fatkin, B. Feenstra, M. F. Feitosa, L. Ferrucci, I. Ford, M. Fornage, A. Franke, P. W. Franks, B. I. Freedman, P. Gasparini, C. Gieger, G. Girotto, M. E. Goddard, Y. M. Golightly, C. Gonzalez-Villalpando, P. Gordon-Larsen, H. Grallert, S. F. A. Grant, N. Grarup, L. Griffiths, V. Gudnason, C. Haiman, H. Hakonarson, T. Hansen, C. A. Hartman, A. T. Hattersley, C. Hayward, S. R. Heckbert, C.-K. Heng, C. Hengstenberg, A. W. Hewitt, H. Hishigaki, C. B. Hoyng, P. L. Huang, W. Huang, S. C. Hunt, K. Hveem, E. Hyppönen, W. G. Iacono, S. Ichihara, M. A. Ikram, C. R. Isasi, R. D. Jackson, M.-R. Jarvelin, Z.-B. Jin, K.-H. Jöckel, P. K. Joshi, P. Jousilahti, J. W. Jukema, M. Kähönen, Y. Kamatani, K. D. Kang, J. Kaprio, S. L. R. Kardia, F. Karpe, N. Kato, F. Kee, T. Kessler, A. V. Khera, C. C. Khor, L. A. L. M. Kiemeney, B.-J. Kim, E. K. Kim, H.-L. Kim, P. Kirchhof, M. Kivimaki, W.-P. Koh, H. A. Koistinen, G. D. Kolovou, J. S. Kooner, C. Kooperberg, A. Köttgen, P. Kovacs, A. Kraaijeveld, P. Kraft, R. M. Krauss, M. Kumari, Z. Kutalik, M. Laakso, L. A. Lange, C. Langenberg, L. J. Launer, L. Le Marchand, H. Lee, N. R. Lee, T. Lehtimäki, H. Li, L. Li, W. Lieb, X. Lin, L. Lind, A. Linneberg, C.-T. Liu, J. Liu, M. Loeffler, B. London, S. A. Lubitz, S. J. Lye, D. A. Mackey, R. Mägi, P. K. E. Magnusson, G. M. Marcus, P. M. Vidal, N. G. Martin, W. März, F. Matsuda, R. W. McGarrah, M. McGue, A. J. McKnight, S. E. Medland, D. Mellström, A. Metspalu, B. D. Mitchell, P. Mitchell, D. O. Mook-Kanamori, A. D. Morris, L. A. Mucci, P. B. Munroe, M. A. Nalls, S. Nazarian, A. E. Nelson, M. J. Neville, C. Newton-Cheh, C. S. Nielsen, M. M. Nöthen, C. Ohlsson, A. J. Oldehinkel, L. Orozco, K. Pahkala, P. Pajukanta, C. N. A. Palmer, E. J. Parra, C. Pattaro, O. Pedersen, C. E. Pennell, B. W. J. H. Penninx, L.

Perusse, A. Peters, P. A. Peyser, D. J. Porteous, D. Posthuma, C. Power, P. P. Pramstaller, M. A. Province, Q. Qi, J. Qu, D. J. Rader, O. T. Raitakari, S. Ralhan, L. S. Rallidis, D. C. Rao, S. Redline, D. F. Reilly, A. P. Reiner, S. Y. Rhee, P. M. Ridker, M. Rienstra, S. Ripatti, M. D. Ritchie, D. M. Roden, F. R. Rosendaal, J. I. Rotter, I. Rudan, F. Rutters, C. Sabanayagam, D. Saleheen, V. Salomaa, N. J. Samani, D. K. Sanghera, N. Sattar, B. Schmidt, H. Schmidt, R. Schmidt, M. B. Schulze, H. Schunkert, L. J. Scott, R. J. Scott, P. Sever, E. J. Shiroma, M. B. Shoemaker, X.-O. Shu, E. M. Simonsick, M. Sims, J. R. Singh, A. B. Singleton, M. F. Sinner, J. G. Smith, H. Snieder, T. D. Spector, M. J. Stampfer, K. J. Stark, D. P. Strachan, L. M. 't Hart, Y. Tabara, H. Tang, J.-C. Tardif, T. A. Thanaraj, N. J. Timpson, A. Tönjes, A. Tremblay, T. Tuomi, J. Tuomilehto, M.-T. Tusié-Luna, A. G. Uitterlinden, R. M. van Dam, P. van der Harst, N. Van der Velde, C. M. van Duijn, N. M. van Schoor, V. Vitart, U. Völker, P. Vollenweider, H. Völzke, N. H. Wacher-Rodarte, M. Walker, Y. X. Wang, N. J. Wareham, R. M. Watanabe, H. Watkins, D. R. Weir, T. M. Werge, E. Widen, L. R. Wilkens, G. Willemsen, W. C. Willett, J. F. Wilson, T.-Y. Wong, J.-T. Woo, A. F. Wright, J.-Y. Wu, H. Xu, C. S. Yajnik, M. Yokota, J.-M. Yuan, E. Zeggini, B. S. Zemel, W. Zheng, X. Zhu, J. M. Zmuda, A. B. Zonderman, J.-A. Zwart, 23andMe Research Team, VA Million Veteran Program, DiscovEHR (DiscovEHR and MyCode Community Health Initiative), eMERGE (Electronic Medical Records and Genomics Network), Lifelines Cohort Study, PRACTICAL Consortium, Understanding Society Scientific Group, D. I. Chasman, Y. S. Cho, I. M. Heid, M. I. McCarthy, M. C. Y. Ng, C. J. O'Donnell, F. Rivadeneira, U. Thorsteinsdottir, Y. V. Sun, E. S. Tai, M. Boehnke, P. Deloukas, A. E. Justice, C. M. Lindgren, R. J. F. Loos, K. L. Mohlke, K. E. North, K. Stefansson, R. G. Walters, T. W. Winkler, K. L. Young, P.-R. Loh, J. Yang, T. Esko, T. L. Assimes, A. Auton, G. R. Abecasis, C. J. Willer, A. E. Locke, S. I. Berndt, G. Lettre, T. M. Frayling, Y. Okada, A. R. Wood, P. M. Visscher, J. N. Hirschhorn, A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).

4. COVID-19 Host Genetics Initiative, A second update on mapping the human genetic architecture of COVID-19. *Nature* **621**, E7–E26 (2023).

5. K. G. Aragam, T. Jiang, A. Goel, S. Kanoni, B. N. Wolford, D. S. Atri, E. M. Weeks, M. Wang, G. Hindy, W. Zhou, C. Grace, C. Roselli, N. A. Marston, F. K. Kamanu, I. Surakka, L. M. Venegas, P. Sherliker, S. Koyama, K. Ishigaki, B. O. Åsvold, M. R. Brown, B. Brumpton, P. S. de Vries, O. Giannakopoulou, P. Giardoglou, D. F. Gudbjartsson, U. Güldener, S. M. I. Haider, A. Helgadottir, M. Ibrahim, A. Kastrati, T. Kessler, T. Kyriakou, T. Konopka, L. Li, L. Ma, T. Meitinger, S. Mucha, M. Munz, F. Murgia, J. B. Nielsen, M. M. Nöthen, S. Pang, T. Reinberger, G. Schnitzler, D. Smedley, G. Thorleifsson, M. von Scheidt, J. C. Ulirsch, Biobank Japan, EPIC-CVD, D. O. Arnar, N. P. Burtt, M. C. Costanzo, J. Flannick, K. Ito, D.-K. Jang, Y. Kamatani, A. V. Khera, I. Komuro, I. J. Kullo, L. A. Lotta, C. P. Nelson, R. Roberts, G. Thorgeirsson, U. Thorsteinsdottir, T. R. Webb, A. Baras, J. L. M. Björkegren, E. Boerwinkle, G. Dedoussis, H. Holm, K. Hveem, O. Melander, A. C. Morrison, M. Orho-Melander, L. S. Rallidis, A. Ruusalepp, M. S. Sabatine, K. Stefansson, P. Zalloua, P. T. Ellinor, M. Farrall, J. Danesh, C. T. Ruff, H. K. Finucane, J. C. Hopewell, R. Clarke, R. M. Gupta, J. Erdmann, N. J. Samani, H. Schunkert, H. Watkins, C. J. Willer, P. Deloukas, S. Kathiresan, A. S. Butterworth, CARDIoGRAMplusC4D Consortium, Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).

6. MVP-Finngen-UKBB meta-analysis. https://mvp-ukbb.finngen.fi/about.

7. U. Võsa, A. Claringbould, H.-J. Westra, M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Yazar, H. Brugge, R. Oelen, D. H. de Vries, M. G. P. van der Wijst, S. Kasela, N. Pervjakova, I. Alves, M.-J. Favé, M. Agbessi, M. W. Christiansen, R.

Jansen, I. Seppälä, L. Tong, A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sönmez Flitman, A. Brown, V. Kukushkina, A. Kalnapenkis, S. Rüeger, E. Porcu, J. Kronberg, J. Kettunen, B. Lee, F. Zhang, T. Qi, J. A. Hernandez, W. Arindrarto, F. Beutner, BIOS Consortium, i2QTL Consortium, J. Dmitrieva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, A. W. Hewitt, M. Kähönen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei, Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M. G. Nivard, B. W. J. H. Penninx, J. K. Pritchard, O. T. Raitakari, O. Rotzschke, E. P. Slagboom, C. D. A. Stehouwer, M. Stumvoll, P. Sullivan, P. A. C. 't Hoen, J. Thiery, A. Tönjes, J. van Dongen, M. van Iterson, J. H. Veldink, U. Völker, R. Warmerdam, C. Wijmenga, M. Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, E. Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. L. Pierce, T. Lehtimäki, D. I. Boomsma, B. M. Psaty, S. A. Gharib, P. Awadalla, L. Milani, W. H. Ouwehand, K. Downes, O. Stegle, A. Battle, P. M. Visscher, J. Yang, M. Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, 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).

8. B. B. Sun, J. Chiou, M. Traylor, C. Benner, Y.-H. Hsu, T. G. Richardson, P. Surendran, A. Mahajan, C. Robins, S. G. Vasquez-Grinnell, L. Hou, E. M. Kvikstad, O. S. Burren, J. Davitte, K. L. Ferber, C. E. Gillies, Å. K. Hedman, S. Hu, T. Lin, R. Mikkilineni, R. K. Pendergrass, C. Pickering, B. Prins, D. Baird, C.-Y. Chen, L. D. Ward, A. M. Deaton, S. Welsh, C. M. Willis, N. Lehner, M. Arnold, M. A. Wörheide, K. Suhre, G. Kastenmüller, A. Sethi, M. Cule, A. Raj, L. Burkitt-Gray, E. Melamud, M. H. Black, E. B. Fauman, J. M. M. Howson, H. M. Kang, M. I. McCarthy, P. Nioi, S. Petrovski, R. A. Scott, E. N. Smith, S. Szalma, D. M. Waterworth, L. J. Mitnaul, J. Szustakowski, B. W. Gibson, M. R. Miller, C. D. Whelan, Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*, 1–10 (2023).

9. S. E. Graham, S. L. Clarke, K.-H. H. Wu, S. Kanoni, G. J. M. Zajac, S. Ramdas, I. Surakka, I. Ntalla, S. Vedantam, T. W. Winkler, A. E. Locke, E. Marouli, M. Y. Hwang, S. Han, A. Narita, A. Choudhury, A. R. Bentley, K. Ekoru, A. Verma, B. Trivedi, H. C. Martin, K. A. Hunt, Q. Hui, D. Klarin, X. Zhu, G. Thorleifsson, A. Helgadottir, D. F. Gudbjartsson, H. Holm, I. Olafsson, M. Akiyama, S. Sakaue, C. Terao, M. Kanai, W. Zhou, B. M. Brumpton, H. Rasheed, S. E. Ruotsalainen, A. S. Havulinna, Y. Veturi, Q. Feng, E. A. Rosenthal, T. Lingren, J. A. Pacheco, S. A. Pendergrass, J. Haessler, F. Giulianini, Y. Bradford, J. E. Miller, A. Campbell, K. Lin, I. Y. Millwood, G. Hindy, A. Rasheed, J. D. Faul, W. Zhao, D. R. Weir, C. Turman, H. Huang, M. Graff, A. Mahajan, M. R. Brown, W. Zhang, K. Yu, E. M. Schmidt, A. Pandit, S. Gustafsson, X. Yin, J. Luan, J.-H. Zhao, F. Matsuda, H.-M. Jang, K. Yoon, C. Medina-Gomez, A. Pitsillides, J. J. Hottenga, G. Willemsen, A. R. Wood, Y. Ji, Z. Gao, S. Haworth, R. E. Mitchell, J. F. Chai, M. Aadahl, J. Yao, A. Manichaikul, H. R. Warren, J. Ramirez, J. Bork-Jensen, L. L. Kårhus, A. Goel, M. Sabater-Lleal, R. Noordam, C. Sidore, E. Fiorillo, A. F. McDaid, P. Marques-Vidal, M. Wielscher, S. Trompet, N. Sattar, L. T. Møllehave, B. H. Thuesen, M. Munz, L. Zeng, J. Huang, B. Yang, A. Poveda, A. Kurbasic, C. Lamina, L. Forer, M. Scholz, T. E. Galesloot, J. P. Bradfield, E. W. Daw, J. M. Zmuda, J. S. Mitchell, C. Fuchsberger, H. Christensen, J. A. Brody, M. F. Feitosa, M. K. Wojczynski, M. Preuss, M. Mangino, P. Christofidou, N. Verweij, J. W. Benjamins, J. Engmann, R. L. Kember, R. C. Slieker, K. S. Lo, N. R. Zilhao, P. Le, M. E. Kleber, G. E. Delgado, S. Huo, D. D. Ikeda, H. Iha, J. Yang, J. Liu, H. L. Leonard, J. Marten, B. Schmidt, M. Arendt, L. J. Smyth, M. Cañadas-Garre, C. Wang, M. Nakatochi, A. Wong, N. Hutri-Kähönen, X. Sim, R. Xia, A. Huerta-Chagoya, J. C. Fernandez-Lopez, V. Lyssenko, M. Ahmed, A. U. Jackson, M. R. Irvin, C. Oldmeadow, H.-N. Kim, S. Ryu, P. R. H. J. Timmers, L. Arbeeva, R. Dorajoo, L. A. Lange, X. Chai, G. Prasad, L. Lorés-Motta, M. Pauper, J. Long, X. Li, E. Theusch, F. Takeuchi, C. N. Spracklen, A. Loukola, S. Bollepalli, S. C. Warner, Y. X. Wang, W. B. Wei, T. Nutile, D. Ruggiero, Y. J. Sung, Y.-J. Hung, S. Chen,

F. Liu, J. Yang, K. A. Kentistou, M. Gorski, M. Brumat, K. Meidtner, L. F. Bielak, J. A. Smith, P. Hebbar, A.-E. Farmaki, E. Hofer, M. Lin, C. Xue, J. Zhang, M. P. Concas, S. Vaccargiu, P. J. van der Most, N. Pitkänen, B. E. Cade, J. Lee, S. W. van der Laan, K. N. Chitrala, S. Weiss, M. E. Zimmermann, J. Y. Lee, H. S. Choi, M. Nethander, S. Freitag-Wolf, L. Southam, N. W. Rayner, C. A. Wang, S.-Y. Lin, J.-S. Wang, C. Couture, L.-P. Lyytikäinen, K. Nikus, G. Cuellar-Partida, H. Vestergaard, B. Hildalgo, O. Giannakopoulou, Q. Cai, M. O. Obura, J. van Setten, X. Li, K. Schwander, N. Terzikhan, J. H. Shin, R. D. Jackson, A. P. Reiner, L. W. Martin, Z. Chen, L. Li, H. M. Highland, K. L. Young, T. Kawaguchi, J. Thiery, J. C. Bis, G. N. Nadkarni, L. J. Launer, H. Li, M. A. Nalls, O. T. Raitakari, S. Ichihara, S. H. Wild, C. P. Nelson, H. Campbell, S. Jäger, T. Nabika, F. Al-Mulla, H. Niinikoski, P. S. Braund, I. Kolcic, P. Kovacs, T. Giardoglou, T. Katsuya, K. F. Bhatti, D. de Kleijn, G. J. de Borst, E. K. Kim, H. H. H. Adams, M. A. Ikram, X. Zhu, F. W. Asselbergs, A. O. Kraaijeveld, J. W. J. Beulens, X.-O. Shu, L. S. Rallidis, O. Pedersen, T. Hansen, P. Mitchell, A. W. Hewitt, M. Kähönen, L. Pérusse, C. Bouchard, A. Tönjes, Y.-D. I. Chen, C. E. Pennell, T. A. Mori, W. Lieb, A. Franke, C. Ohlsson, D. Mellström, Y. S. Cho, H. Lee, J.-M. Yuan, W.-P. Koh, S. Y. Rhee, J.-T. Woo, I. M. Heid, K. J. Stark, H. Völzke, G. Homuth, M. K. Evans, A. B. Zonderman, O. Polasek, G. Pasterkamp, I. E. Hoefer, S. Redline, K. Pahkala, A. J. Oldehinkel, H. Snieder, G. Biino, R. Schmidt, H. Schmidt, Y. E. Chen, S. Bandinelli, G. Dedoussis, T. A. Thanaraj, S. L. R. Kardia, N. Kato, M. B. Schulze, G. Girotto, B. Jung, C. A. Böger, P. K. Joshi, D. A. Bennett, P. L. De Jager, X. Lu, V. Mamakou, M. Brown, M. J. Caulfield, P. B. Munroe, X. Guo, M. Ciullo, J. B. Jonas, N. J. Samani, J. Kaprio, P. Pajukanta, L. S. Adair, S. A. Bechayda, H. J. de Silva, A. R. Wickremasinghe, R. M. Krauss, J.-Y. Wu, W. Zheng, A. I. den Hollander, D. Bharadwaj, A. Correa, J. G. Wilson, L. Lind, C.-K. Heng, A. E. Nelson, Y. M. Golightly, J. F. Wilson, B. Penninx, H.-L. Kim, J. Attia, R. J. Scott, D. C. Rao, D. K. Arnett, M. Walker, H. A. Koistinen, G. R. Chandak, C. S. Yajnik, J. M. Mercader, T. Tusié-Luna, C. A. Aguilar-Salinas, C. G. Villalpando, L. Orozco, M. Fornage, E. S. Tai, R. M. van Dam, T. Lehtimäki, N. Chaturvedi, M. Yokota, J. Liu, D. F. Reilly, A. J. McKnight, F. Kee, K.-H. Jöckel, M. I. McCarthy, C. N. A. Palmer, V. Vitart, C. Hayward, E. Simonsick, C. M. van Duijn, F. Lu, J. Qu, H. Hishigaki, X. Lin, W. März, E. J. Parra, M. Cruz, V. Gudnason, J.-C. Tardif, G. Lettre, L. M. ’t Hart, P. J. M. Elders, S. M. Damrauer, M. Kumari, M. Kivimaki, P. van der Harst, T. D. Spector, R. J. F. Loos, M. A. Province, B. M. Psaty, I. Brandslund, P. P. Pramstaller, K. Christensen, S. Ripatti, E. Widén, H. Hakonarson, S. F. A. Grant, L. A. L. M. Kiemeney, J. de Graaf, M. Loeffler, F. Kronenberg, D. Gu, J. Erdmann, H. Schunkert, P. W. Franks, A. Linneberg, J. W. Jukema, A. V. Khera, M. Männikkö, M.-R. Jarvelin, Z. Kutalik, F. Cucca, D. O. Mook-Kanamori, K. W. van Dijk, H. Watkins, D. P. Strachan, N. Grarup, P. Sever, N. Poulter, J. I. Rotter, T. M. Dantoft, F. Karpe, M. J. Neville, N. J. Timpson, C.-Y. Cheng, T.-Y. Wong, C. C. Khor, C. Sabanayagam, A. Peters, C. Gieger, A. T. Hattersley, N. L. Pedersen, P. K. E. Magnusson, D. I. Boomsma, E. J. C. de Geus, L. A. Cupples, J. B. J. van Meurs, M. Ghanbari, P. Gordon-Larsen, W. Huang, Y. J. Kim, Y. Tabara, N. J. Wareham, C. Langenberg, E. Zeggini, J. Kuusisto, M. Laakso, E. Ingelsson, G. Abecasis, J. C. Chambers, J. S. Kooner, P. S. de Vries, A. C. Morrison, K. E. North, M. Daviglus, P. Kraft, N. G. Martin, J. B. Whitfield, S. Abbas, D. Saleheen, R. G. Walters, M. V. Holmes, C. Black, B. H. Smith, A. E. Justice, A. Baras, J. E. Buring, P. M. Ridker, D. I. Chasman, C. Kooperberg, W.-Q. Wei, G. P. Jarvik, B. Namjou, M. G. Hayes, M. D. Ritchie, P. Jousilahti, V. Salomaa, K. Hveem, B. O. Åsvold, M. Kubo, Y. Kamatani, Y. Okada, Y. Murakami, U. Thorsteinsdottir, K. Stefansson, Y.-L. Ho, J. A. Lynch, D. J. Rader, P. S. Tsao, K.-M. Chang, K. Cho, C. J. O’Donnell, J. M. Gaziano, P. Wilson, C. N. Rotimi, S. Hazelhurst, M. Ramsay, R. C. Trembath, D. A. van Heel, G. Tamiya, M. Yamamoto, B.-J. Kim, K. L. Mohlke, T. M. Frayling, J. N. Hirschhorn, S. Kathiresan, M. Boehnke, P. Natarajan, G. M. Peloso, C. D. Brown, A. P. Morris, T. L. Assimes, P. Deloukas, Y. V. Sun, C. J. Willer, The power of genetic diversity in genome-wide association studies of lipids. *Nature*, 1–11 (2021).

10. M. K. Karjalainen, S. Karthikeyan, C. Oliver-Williams, E. Sliz, E. Allara, W. T. Fung, P. Surendran, W. Zhang, P. Jousilahti, K. Kristiansson, V. Salomaa, M. Goodwin, D. A. Hughes, M. Boehnke, L. Fernandes Silva, X. Yin, A. Mahajan, M. J. Neville, N. R. van Zuydam, R. de Mutsert, R. Li-Gao, D. O. Mook-Kanamori, A. Demirkan, J. Liu, R. Noordam, S. Trompet, Z. Chen, C. Kartsonaki, L. Li, K. Lin, F. A. Hagenbeek, J. J. Hottenga, R. Pool, M. A. Ikram, J. van Meurs, T. Haller, Y. Milaneschi, M. Kähönen, P. P. Mishra, P. K. Joshi, E. Macdonald-Dunlop, M. Mangino, J. Zierer, I. E. Acar, C. B. Hoyng, Y. T. E. Lechanteur, L. Franke, A. Kurilshikov, A. Zhernakova, M. Beekman, E. B. van den Akker, I. Kolcic, O. Polasek, I. Rudan, C. Gieger, M. Waldenberger, F. W. Asselbergs, China Kadoorie Biobank Collaborative Group, Estonian Biobank Research Team, FinnGen, C. Hayward, J. Fu, A. I. den Hollander, C. Menni, T. D. Spector, J. F. Wilson, T. Lehtimäki, O. T. Raitakari, B. W. J. H. Penninx, T. Esko, R. G. Walters, J. W. Jukema, N. Sattar, M. Ghanbari, K. Willems van Dijk, F. Karpe, M. I. McCarthy, M. Laakso, M.-R. Järvelin, N. J. Timpson, M. Perola, J. S. Kooner, J. C. Chambers, C. van Duijn, P. E. Slagboom, D. I. Boomsma, J. Danesh, M. Ala-Korpela, A. S. Butterworth, J. Kettunen, Genome-wide characterization of circulating metabolic biomarkers. *Nature* **628**, 130–138 (2024).

11. T. G. Richardson, G. M. Leyden, Q. Wang, J. A. Bell, B. Elsworth, G. Davey Smith, M. V. Holmes, Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol.* **20**, e3001547 (2022).

12. C. J. Smith, N. Sinnott-Armstrong, A. Cichońska, H. Julkunen, E. B. Fauman, P. Würtz, J. K. Pritchard, Integrative analysis of metabolite GWAS illuminates the molecular basis of pleiotropy and genetic correlation. *Elife* **11** (2022).

13. D. van der Meer, Z. Rahman, A. Ottas, P. Parekh, G. Kutrolli, S. Stinson, M. Koromina, J. Rokicki, I. Sonderby, N. Parker, M. Tesfaye, G. Hindley, L. Rodevand, E. Koch, N. Steen, J. Berg, K. O'Connell, O. Smeland, O. Frei, A. Dale, S. Djurovic, K. Lehto, M. Alver, L. Milani, A. Shadrin, O. Andreassen, Pleiotropic and sex-specific genetic architecture of circulating metabolic markers, *medRxiv* (2024)p. 2024.07.30.24311254.

14. M. Zoodsma, C. Beuchel, S. Yasmeen, L. Kohleick, A. Nepal, M. Koprulu, F. Kronenberg, M. Mayr, A. Williamson, M. Pietzner, C. Langenberg, A genetic map of human metabolism across the allele frequency spectrum, *medRxiv* (2025)p. 2025.01.30.25321073.

15. I. Rahu, R. Tambets, E. B. Fauman, K. Alasoo, Mendelian randomisation with proxy exposures: challenges and opportunities, *medRxiv* (2024)p. 2024.10.21.24315891.

16. A. Nag, R. S. Dhindsa, L. Middleton, X. Jiang, D. Vitsios, E. Wigmore, E. L. Allman, A. Reznichenko, K. Carss, K. R. Smith, Q. Wang, B. Challis, D. S. Paul, A. R. Harper, S. Petrovski, Effects of protein-coding variants on blood metabolite measurements and clinical biomarkers in the UK Biobank. *Am. J. Hum. Genet.* **110**, 487–498 (2023).

17. S. Stender, H. Gellert-Kristensen, G. D. Smith, Reclaiming mendelian randomization from the deluge of papers and misleading findings. *Lipids Health Dis.* **23**, 286 (2024).

18. E. Sanderson, M. M. Glymour, M. V. Holmes, H. Kang, J. Morrison, M. R. Munafò, T. Palmer, C. M. Schooling, C. Wallace, Q. Zhao, G. Davey Smith, Mendelian randomization. *Nature Reviews Methods Primers* **2**, 1–21 (2022).

19. S. Burgess, B. Woolf, A. M. Mason, M. Ala-Korpela, D. Gill, Addressing the credibility crisis in Mendelian randomization. *BMC Med.* **22**, 374 (2024).

20. B. F. Voight, G. M. Peloso, M. Orho-Melander, R. Frikke-Schmidt, M. Barbalic, M. K.

Jensen, G. Hindy, H. Hólm, E. L. Ding, T. Johnson, H. Schunkert, N. J. Samani, R. Clarke, J. C. Hopewell, J. F. Thompson, M. Li, G. Thorleifsson, C. Newton-Cheh, K. Musunuru, J. P. Pirruccello, D. Saleheen, L. Chen, A. F. R. Stewart, A. Schillert, U. Thorsteinsdottir, G. Thorgeirsson, S. Anand, J. C. Engert, T. Morgan, J. Spertus, M. Stoll, K. Berger, N. Martinelli, D. Girelli, P. P. McKeown, C. C. Patterson, S. E. Epstein, J. Devaney, M.-S. Burnett, V. Mooser, S. Ripatti, I. Surakka, M. S. Nieminen, J. Sinisalo, M.-L. Lokki, M. Perola, A. Havulinna, U. de Faire, B. Gigante, E. Ingelsson, T. Zeller, P. Wild, P. I. W. de Bakker, O. H. Klungel, A.-H. Maitland-van der Zee, B. J. M. Peters, A. de Boer, D. E. Grobbee, P. W. Kamphuisen, V. H. M. Deneer, C. C. Elbers, N. C. Onland-Moret, M. H. Hofker, C. Wijmenga, W. M. M. Verschuren, J. M. A. Boer, Y. T. van der Schouw, A. Rasheed, P. Frossard, S. Demissie, C. Willer, R. Do, J. M. Ordovas, G. R. Abecasis, M. Boehnke, K. L. Mohlke, M. J. Daly, C. Guiducci, N. P. Burtt, A. Surti, E. Gonzalez, S. Purcell, S. Gabriel, J. Marrugat, J. Peden, J. Erdmann, P. Diemert, C. Willenborg, I. R. König, M. Fischer, C. Hengstenberg, A. Ziegler, I. Buysschaert, D. Lambrechts, F. Van de Werf, K. A. Fox, N. E. El Mokhtari, D. Rubin, J. Schrezenmeir, S. Schreiber, A. Schäfer, J. Danesh, S. Blankenberg, R. Roberts, R. McPherson, H. Watkins, A. S. Hall, K. Overvad, E. Rimm, E. Boerwinkle, A. Tybjaerg-Hansen, L. A. Cupples, M. P. Reilly, O. Melander, P. M. Mannucci, D. Ardissino, D. Siscovick, R. Elosua, K. Stefansson, C. J. O'Donnell, V. Salomaa, D. J. Rader, L. Peltonen, S. M. Schwartz, D. Altshuler, S. Kathiresan, Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380**, 572–580 (2012).

21. S. Burgess, G. Davey Smith, N. M. Davies, F. Dudbridge, D. Gill, M. M. Glymour, F. P. Hartwig, Z. Kutalik, M. V. Holmes, C. Minelli, J. V. Morrison, W. Pan, C. L. Relton, E. Theodoratou, Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res* **4**, 186 (2019).

22. V. W. Skrivankova, R. C. Richmond, B. A. R. Woolf, N. M. Davies, S. A. Swanson, T. J. VanderWeele, N. J. Timpson, J. P. T. Higgins, N. Dimou, C. Langenberg, E. W. Loder, R. M. Golub, M. Egger, G. Davey Smith, J. B. Richards, Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, n2233 (2021).

23. Z. E. Reed, R. E. Wootton, J. N. Khouja, T. G. Richardson, E. Sanderson, G. Davey Smith, M. R. Munafò, Exploring pleiotropy in Mendelian randomisation analyses: What are genetic variants associated with "cigarette smoking initiation" really capturing? *Genet. Epidemiol.* **49**, e22583 (2025).

24. E. Sanderson, D. Rosoff, T. Palmer, K. Tilling, G. D. Smith, G. Hemani, Bias from heritable confounding in Mendelian randomization studies, *medRxiv* (2024)p. 2024.09.05.24312293.

25. D. van der Meer, A. Shadrin, S. Stinson, E. Koch, J. Rokicki, Z. Rahman, J. Bergstedt, A. Ottas, I. Sonderby, L. Rodevand, J. Fuhrer, D. S. Quintana, A. Dale, K. O'Connell, S. Djurovic, K. Lehto, M. Alver, L. Milani, O. Andreassen, Divergent patterns of genetic overlap between severe mental disorders and metabolic markers, *medRxiv* (2024)p. 2024.11.04.24316693.

26. D. Gill, M.-J. Dib, H. T. Cronjé, V. Karhunen, B. Woolf, E. Gagnon, I. Daghlas, M. Nyberg, D. Drakeman, S. Burgess, Common pitfalls in drug target Mendelian randomization and how to avoid them. *BMC Medicine* **22**, 1–12 (2024).

27. N. Kerimov, J. D. Hayhurst, K. Peikova, J. R. Manning, P. Walter, L. Kolberg, M. Samoviča, M. P. Sakthivel, I. Kuzmin, S. J. Trevanion, T. Burdett, S. Jupp, H. Parkinson, I. Papatheodorou, A. D. Yates, D. R. Zerbino, K. Alasoo, A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**,

1290–1299 (2021).

28. N. Kerimov, R. Tambets, J. D. Hayhurst, I. Rahu, P. Kolberg, U. Raudvere, I. Kuzmin, A. Chowdhary, A. Vija, H. J. Teras, M. Kanai, J. Ulirsch, M. Ryten, J. Hardy, S. Guelfi, D. Trabzuni, S. Kim-Hellmuth, W. Rayner, H. Finucane, H. Peterson, A. Mosaku, H. Parkinson, K. Alasoo, eQTL Catalogue 2023: New datasets, X chromosome QTLs, and improved detection and visualisation of transcript-level QTLs. *PLoS Genet.* **19**, e1010932 (2023).

29. L. A. Lotta, R. A. Scott, S. J. Sharp, S. Burgess, J. ’an Luan, T. Tillin, A. F. Schmidt, F. Imamura, I. D. Stewart, J. R. B. Perry, L. Marney, A. Koulman, E. D. Karoly, N. G. Forouhi, R. J. O. Sjögren, E. Näslund, J. R. Zierath, A. Krook, D. B. Savage, J. L. Griffin, N. Chaturvedi, A. D. Hingorani, K.-T. Khaw, I. Barroso, M. I. McCarthy, S. O’Rahilly, N. J. Wareham, C. Langenberg, Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. *PLoS Med.* **13**, e1002179 (2016).

30. M. Doestzada, D. V. Zhernakova, I. C L van den Munckhof, D. Wang, A. Kurilshikov, L. Chen, V. W. Bloks, M. van Faassen, J. H. W. Rutten, L. A. B. Joosten, M. G. Netea, C. Wijmenga, N. P. Riksen, A. Zhernakova, F. Kuipers, J. Fu, Systematic analysis of relationships between plasma branched-chain amino acid concentrations and cardiometabolic parameters: an association and Mendelian randomization study. *BMC Med.* **20**, 485 (2022).

31. I. Ramzan, A. Ardavani, F. Vanweert, A. Mellett, P. J. Atherton, I. Idris, The association between circulating branched chain amino acids and the temporal risk of developing type 2 diabetes mellitus: A systematic review & meta-analysis. *Nutrients* **14**, 4411 (2022).

32. F. Vanweert, P. Schrauwen, E. Phielix, Role of branched-chain amino acid metabolism in the pathogenesis of obesity and type 2 diabetes-related metabolic disturbances BCAA metabolism in type 2 diabetes. *Nutr. Diabetes* **12**, 35 (2022).

33. K. J. Karczewski, R. Gupta, M. Kanai, W. Lu, K. Tsuo, Y. Wang, R. K. Walters, P. Turley, S. Callier, N. Baya, D. S. Palmer, J. I. Goldstein, G. Sarma, M. Solomonson, N. Cheng, S. Bryant, C. Churchhouse, C. M. Cusick, T. Poterba, J. Compitello, D. King, W. Zhou, C. Seed, H. K. Finucane, M. J. Daly, B. M. Neale, E. G. Atkinson, A. R. Martin, Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects, *medRxiv* (2024)p. 2024.03.13.24303864.

34. M. Mitt, M. Kals, K. Pärn, S. B. Gabriel, E. S. Lander, A. Palotie, S. Ripatti, A. P. Morris, A. Metspalu, T. Esko, R. Mägi, P. Palta, Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).

35. S. Shi, S. Rubinacci, S. Hu, L. Moutsianas, A. Stuckey, A. C. Need, P. F. Palamara, M. Caulfield, J. Marchini, S. Myers, A Genomics England haplotype reference panel and imputation of UK Biobank. *Nat. Genet.* **56**, 1800–1803 (2024).

36. D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S.-B. Lee, X. Tian, B. L. Browning, S. Das, A.-K. Emde, W. E. Clarke, D. P. Loesch, A. C. Shetty, T. W. Blackwell, A. V. Smith, Q. Wong, X. Liu, M. P. Conomos, D. M. Bobo, F. Aguet, C. Albert, A. Alonso, K. G. Ardlie, D. E. Arking, S. Aslibekyan, P. L. Auer, J. Barnard, R. G. Barr, L. Barwick, L. C. Becker, R. L. Beer, E. J. Benjamin, L. F. Bielak, J. Blangero, M. Boehnke, D. W. Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F. Casella, B.

Chalazan, D. I. Chasman, Y.-D. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A. Correa, J. E. Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K. Dutcher, P. T. Ellinor, L. S. Emery, C. Eng, D. Fatkin, T. Fingerlin, L. Forer, M. Fornage, N. Franceschini, C. Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E. Hall, J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson, R. Kaplan, S. L. R. Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A. Konkle, C. Kooperberg, A. Köttgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K.-H. Lin, C. Liu, R. J. F. Loos, L. Garman, R. Gerszten, S. A. Lubitz, K. L. Lunetta, A. C. Y. Mak, A. Manichaikul, A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs, D. A. Meyers, J. L. Mikulla, M. A. Minear, B. D. Mitchell, S. Mohanty, M. E. Montasser, C. Montgomery, A. C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E. North, J. R. O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, J. Pleiness, W. S. Post, B. M. Psaty, D. C. Rao, S. Redline, A. P. Reiner, D. Roden, J. I. Rotter, I. Ruczinski, C. Sarnowski, S. Schoenherr, D. A. Schwartz, J.-S. Seo, S. Seshadri, V. A. Sheehan, W. H. Sheu, M. B. Shoemaker, N. L. Smith, J. A. Smith, N. Sotoodehnia, A. M. Stilp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, R. P. Tracy, D. J. Van Den Berg, R. S. Vasan, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S. Weir, S. T. Weiss, L.-C. Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-Koch, K. C. Barnes, E. Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E. K. Silverman, P. Qasba, W. Gan, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, G. J. Papanicolaou, D. A. Nickerson, S. R. Browning, M. C. Zody, S. Zöllner, J. G. Wilson, L. A. Cupples, C. C. Laurie, C. E. Jaquish, R. D. Hernandez, T. D. O'Connor, G. R. Abecasis, Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

37. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

38. C. Wallace, A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).

39. B. A. Ference, J. G. Robinson, R. D. Brook, A. L. Catapano, M. J. Chapman, D. R. Neff, S. Voros, R. P. Giugliano, G. Davey Smith, S. Fazio, M. S. Sabatine, Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. *N. Engl. J. Med.* **375**, 2144–2153 (2016).

40. K. Alasoo, J. Rodrigues, S. Mukhopadhyay, A. J. Knights, A. L. Mann, K. Kundu, HIPSCI Consortium, C. Hale, G. Dougan, D. J. Gaffney, Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* **50**, 424–431 (2018).

41. R. Burkhardt, E. E. Kenny, J. K. Lowe, A. Birkeland, R. Josowitz, M. Noel, J. Salit, J. B. Maller, I. Pe'er, M. J. Daly, D. Altshuler, M. Stoffel, J. M. Friedman, J. L. Breslow, Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon 13. *Arterioscler. Thromb. Vasc. Biol.* **28**, 2078–2084 (2008).

42. S. Hwang, V. Karhunen, A. Patel, S. Lockhart, P. Carter, J. Whittaker, S. Burgess, Human genetics suggests differing causal pathways from HMGCR inhibition to coronary artery disease and type 2 diabetes, *medRxiv* (2025)p. 2025.02.10.25321990.

43. K. Musunuru, A. Strong, M. Frank-Kamenetsky, N. E. Lee, T. Ahfeldt, K. V. Sachs, X. Li, H. Li, N. Kuperwasser, V. M. Ruda, J. P. Pirruccello, B. Muchmore, L. Prokunina-Olsson, J. L. Hall, E. E. Schadt, C. R. Morales, S. Lund-Katz, M. C. Phillips, J. Wong, W.

Cantley, T. Racie, K. G. Ejebe, M. Orho-Melander, O. Melander, V. Koteliansky, K. Fitzgerald, R. M. Krauss, C. A. Cowan, S. Kathiresan, D. J. Rader, From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).

44. T. S. Roman, M. E. Cannon, S. Vadlamudi, M. L. Buchkovich, B. N. Wolford, R. P. Welch, M. A. Morken, G. J. Kwon, A. Varshney, R. Kursawe, Y. Wu, A. U. Jackson, National Institutes of Health Intramural Sequencing Center (NISC) Comparative Sequencing Program, M. R. Erdos, J. Kuusisto, M. Laakso, L. J. Scott, M. Boehnke, F. S. Collins, S. C. J. Parker, M. L. Stitzel, K. L. Mohlke, A type 2 diabetes-associated functional regulatory variant in a pancreatic islet enhancer at the ADCY5 locus. *Diabetes* **66**, 2521–2530 (2017).

45. L. Alonso, A. Piron, I. Morán, M. Guindo-Martínez, S. Bonàs-Guarch, G. Atla, I. Miguel-Escalada, R. Royo, M. Puiggròs, X. Garcia-Hurtado, M. Suleiman, L. Marselli, J. L. S. Esguerra, J.-V. Turatsinze, J. M. Torres, V. Nylander, J. Chen, L. Eliasson, M. Defrance, R. Amela, H. Mulder, A. L. Gloyn, L. Groop, P. Marchetti, D. L. Eizirik, J. Ferrer, J. M. Mercader, M. Cnop, D. Torrents, TIGER: The gene expression regulatory variation landscape of human pancreatic islets. *Cell Rep.* **37** (2021).

46. A. Ray, P. Alabarse, R. Malik, M. Sargurupremraj, J. Bernhagen, M. Dichgans, S.-E. Baumeister, M. K. Georgakis, Single-cell transcriptome-wide Mendelian randomization and colocalization analyses uncover cell-specific mechanisms in atherosclerotic cardiovascular disease, *medRxiv* (2024)p. 2024.12.19.24319316.

47. S. C. Ritchie, P. Würtz, A. P. Nath, G. Abraham, A. S. Havulinna, L. G. Fearnley, A.-P. Sarin, A. J. Kangas, P. Soininen, K. Aalto, I. Seppälä, E. Raitoharju, M. Salmi, M. Maksimow, S. Männistö, M. Kähönen, M. Juonala, S. Ripatti, T. Lehtimäki, S. Jalkanen, M. Perola, O. Raitakari, V. Salomaa, M. Ala-Korpela, J. Kettunen, M. Inouye, The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection. *Cell Syst* **1**, 293–301 (2015).

48. A. Ajoolabady, D. Pratico, L. Lin, C. S. Mantzoros, S. Bahijri, J. Tuomilehto, J. Ren, Inflammation in atherosclerosis: pathophysiology and mechanisms. *Cell Death Dis.* **15**, 817 (2024).

49. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

50. K. Jaganathan, S. Kyriazopoulou Panagiotopoulou, J. F. McRae, S. F. Darbandi, D. Knowles, Y. I. Li, J. A. Kosmicki, J. Arbelaez, W. Cui, G. B. Schwartz, E. D. Chow, E. Kanterakis, H. Gao, A. Kia, S. Batzoglou, S. J. Sanders, K. K.-H. Farh, Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535–548.e24 (2019).

51. T. Zeng, Y. I. Li, Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biol.* **23**, 103 (2022).

52. G. Mann, S. Mora, G. Madu, O. A. J. Adegoke, Branched-chain amino acids: Catabolism in skeletal muscle and implications for muscle and whole-body metabolism. *Front. Physiol.* **12**, 702826 (2021).

53. J. Cheng, G. Novati, J. Pan, C. Bycroft, A. Žemgulytė, T. Applebaum, A. Pritzel, L. H. Wong, M. Zielinski, T. Sargeant, R. G. Schneider, A. W. Senior, J. Jumper, D. Hassabis, P. Kohli, Ž. Avsec, Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).

54. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).

55. J. P. Spence, H. Mostafavi, M. Ota, N. Milind, T. Gjorgjieva, C. J. Smith, Y. B. Simons, G. Sella, J. K. Pritchard, Specificity, length, and luck: How genes are prioritized by rare and common variant association studies, *bioRxiv* (2024)p. 2024.12.12.628073.

56. L. A. Lotta, S. J. Sharp, S. Burgess, J. R. B. Perry, I. D. Stewart, S. M. Willems, J. 'an Luan, E. Ardanaz, L. Arriola, B. Balkau, H. Boeing, P. Deloukas, N. G. Forouhi, P. W. Franks, S. Grioni, R. Kaaks, T. J. Key, C. Navarro, P. M. Nilsson, K. Overvad, D. Palli, S. Panico, J.-R. Quirós, E. Riboli, O. Rolandsson, C. Sacerdote, E. C. Salamanca, N. Slimani, A. M. Spijkerman, A. Tjonneland, R. Tumino, D. L. van der A, Y. T. van der Schouw, M. I. McCarthy, I. Barroso, S. O'Rahilly, D. B. Savage, N. Sattar, C. Langenberg, R. A. Scott, N. J. Wareham, Association Between Low-Density Lipoprotein Cholesterol-Lowering Genetic Variants and Risk of Type 2 Diabetes: A Meta-analysis. *JAMA* **316**, 1383–1391 (2016).

57. S. Carugo, C. R. Sirtori, A. Corsini, L. Tokgozoglu, M. Ruscica, PCSK9 inhibition and risk of diabetes: Should we worry? *Curr. Atheroscler. Rep.* **24**, 995–1004 (2022).

58. J. D. Mosley, M. Shi, D. Agamasu, N. S. Vaitinadin, V. L. Murthy, R. V. Shah, M. Bagheri, J. F. Ferguson, Branched-chain amino acids and type 2 diabetes: a bidirectional Mendelian randomization analysis. *Obesity (Silver Spring)* **32**, 423–435 (2024).

59. K. J. Filipski, L. A. Martinez-Alsina, M. R. Reese, E. Evrard, L. M. Buzon, K. O. Cameron, Y. Zhang, K. J. Coffman, J. Bradow, B. L. Kormos, S. Liu, J. D. Knafels, P. V. Sahasrabudhe, J. Chen, A. S. Kalgutkar, A. J. Bessire, C. C. Orozco, A. Balesano, M. A. Cerny, E. Bollinger, A. R. Reyes, B. Laforest, A. Rosado, G. Williams, M. Marshall, K. Tam Neale, X. Chen, D. Hirenallur-Shanthappa, J. C. Stansfield, J. Groarke, R. Qiu, S. Karas, R. J. Roth Flach, W. P. Esler, Discovery of first branched-chain ketoacid dehydrogenase kinase (BDK) inhibitor clinical candidate PF-07328948. *J. Med. Chem.* **68**, 2466–2482 (2025).

60. V. Zuber, N. F. Grinberg, D. Gill, I. Manipur, E. A. W. Slob, A. Patel, C. Wallace, S. Burgess, Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am. J. Hum. Genet.* **109**, 767–782 (2022).

61. Y. Wu, J. Zeng, F. Zhang, Z. Zhu, T. Qi, Z. Zheng, L. R. Lloyd-Jones, R. E. Marioni, N. G. Martin, G. W. Montgomery, I. J. Deary, N. R. Wray, P. M. Visscher, A. F. McRae, J. Yang, Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).

62. E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. Güneş, P. Hall, J. Hayhurst, A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland, K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson, L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, L. W. Harris, The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).

63. M. I. Kurki, J. Karjalainen, P. Palta, T. P. Sipilä, K. Kristiansson, K. M. Donner, M. P. Reeve, H. Laivuori, M. Aavikko, M. A. Kaunisto, A. Loukola, E. Lahtela, H. Mattsson, P. Laiho, P. Della Briotta Parolo, A. A. Lehisto, M. Kanai, N. Mars, J. Rämö, T. Kiiskinen, H. O. Heyne, K. Veerapen, S. Rüeger, S. Lemmelä, W. Zhou, S. Ruotsalainen, K. Pärn, T. Hiekkalinna, S. Koskelainen, T. Paajanen, V. Llorens, J. Gracia-Tabuenca, H. Siirtola, K. Reis, A. G. Elnahas, B. Sun, C. N. Foley, K. Aalto-Setälä, K. Alasoo, M. Arvas, K. Auro,

S. Biswas, A. Bizaki-Vallaskangas, O. Carpen, C.-Y. Chen, O. A. Dada, Z. Ding, M. G. Ehm, K. Eklund, M. Färkkilä, H. Finucane, A. Ganna, A. Ghazal, R. R. Graham, E. M. Green, A. Hakanen, M. Hautalahti, Å. K. Hedman, M. Hiltunen, R. Hinttala, I. Hovatta, X. Hu, A. Huertas-Vazquez, L. Huilaja, J. Hunkapiller, H. Jacob, J.-N. Jensen, H. Joensuu, S. John, V. Julkunen, M. Jung, J. Junttila, K. Kaarniranta, M. Kähönen, R. Kajanne, L. Kallio, R. Kälviäinen, J. Kaprio, N. Kerimov, J. Kettunen, E. Kilpeläinen, T. Kilpi, K. Klinger, V.-M. Kosma, T. Kuopio, V. Kurra, T. Laisk, J. Laukkanen, N. Lawless, A. Liu, S. Longerich, R. Mägi, J. Mäkelä, A. Mäkitie, A. Malarstig, A. Mannermaa, J. Maranville, A. Matakidou, T. Meretoja, S. V. Mozaffari, M. E. K. Niemi, M. Niemi, T. Niiranen, C. J. O´Donnell, M. Obeidat, G. Okafo, H. M. Ollila, A. Palomäki, T. Palotie, J. Partanen, D. S. Paul, M. Pelkonen, R. K. Pendergrass, S. Petrovski, A. Pitkäranta, A. Platt, D. Pulford, E. Punkka, P. Pussinen, N. Raghavan, F. Rahimov, D. Rajpal, N. A. Renaud, B. Riley-Gillis, R. Rodosthenous, E. Saarentaus, A. Salminen, E. Salminen, V. Salomaa, J. Schleutker, R. Serpi, H.-Y. Shen, R. Siegel, K. Silander, S. Siltanen, S. Soini, H. Soininen, J. H. Sul, I. Tachmazidou, K. Tasanen, P. Tienari, S. Toppila-Salmi, T. Tukiainen, T. Tuomi, J. A. Turunen, J. C. Ulirsch, F. Vaura, P. Virolainen, J. Waring, D. Waterworth, R. Yang, M. Nelis, A. Reigo, A. Metspalu, L. Milani, T. Esko, C. Fox, A. S. Havulinna, M. Perola, S. Ripatti, A. Jalanko, T. Laitinen, T. P. Mäkelä, R. Plenge, M. McCarthy, H. Runz, M. J. Daly, A. Palotie, FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).

64. A. Verma, J. E. Huffman, A. Rodriguez, M. Conery, M. Liu, Y.-L. Ho, Y. Kim, D. A. Heise, L. Guare, V. A. Panickan, H. Garcon, F. Linares, L. Costa, I. Goethert, R. Tipton, J. Honerlaw, L. Davies, S. Whitbourne, J. Cohen, D. C. Posner, R. Sangar, M. Murray, X. Wang, D. R. Dochtermann, P. Devineni, Y. Shi, T. N. Nandi, T. L. Assimes, C. A. Brunette, R. J. Carroll, R. Clifford, S. Duvall, J. Gelernter, A. Hung, S. K. Iyengar, J. Joseph, R. Kember, H. Kranzler, C. M. Kripke, D. Levey, S.-W. Luoh, V. C. Merritt, C. Overstreet, J. D. Deak, S. F. A. Grant, R. Polimanti, P. Roussos, G. Shakt, Y. V. Sun, N. Tsao, S. Venkatesh, G. Voloudakis, A. Justice, E. Begoli, R. Ramoni, G. Tourassi, S. Pyarajan, P. Tsao, C. J. O'Donnell, S. Muralidhar, J. Moser, J. P. Casas, A. G. Bick, W. Zhou, T. Cai, B. F. Voight, K. Cho, J. M. Gaziano, R. K. Madduri, S. Damrauer, K. P. Liao, Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).

65. C. Ormond, N. M. Ryan, A. Corvin, E. A. Heron, Converting single nucleotide variants between genome builds: from cautionary tale to solution. *Brief. Bioinform.* **22** (2021).

66. Y. Mahendran, A. Jonsson, C. T. Have, K. H. Allin, D. R. Witte, M. E. Jørgensen, N. Grarup, O. Pedersen, T. O. Kilpeläinen, T. Hansen, Genetic evidence of a causal effect of insulin resistance on branched-chain amino acid levels. *Diabetologia* **60**, 873–878 (2017).

67. Q. Wang, M. V. Holmes, G. Davey Smith, M. Ala-Korpela, Genetic support for a causal role of insulin resistance on circulating branched-chain amino acids and inflammation. *Diabetes Care* **40**, 1779–1786 (2017).

68. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

69. G. Hawkes, H. I. W. Wright, R. N. Beaumont, K. Chundru, A. Hanson, L. Jackson, A. N. Murray, K. Patel, T. Frayling, C. Wright, A. R. Wood, M. N. Weedon, Whole-genome sequencing analysis of anthropometric traits in 672,976 individuals reveals convergence between rare and common genetic associations, *bioRxiv* (2025)p. 2025.02.24.639925.

70. M. Kanai, R. Elzur, W. Zhou, Global Biobank Meta-analysis Initiative, M. J. Daly, H. K. Finucane, Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genom* **2** (2022).

71. L. Milani, M. Alver, S. Laur, S. Reisberg, T. Haller, O. Aasmets, E. Abner, H. Alavere, A. Allik, T. Annilo, K. Fischer, R. Hofmeister, G. Hudjashov, M. Jõeloo, M. Kals, L. Karo-Astover, S. Kasela, A. Kolde, K. Krebs, K. L. Krigul, J. Kronberg, K. Kruusmaa, V. Kukuškina, K. Kõiv, K. Lehto, L. Leitsalu, S. Lind, L. B. Luitva, K. Läll, K. Lüll, K. Metsalu, M. Metspalu, R. Mõttus, M. Nelis, T. Nikopensius, M. Nurm, M. Nõukas, M. Oja, E. Org, M. Palover, P. Palta, V. Pankratov, K. Pantiukh, N. Pervjakova, N. Pujol-Gualdo, A. Reigo, E. Reimann, S. Smit, D. Rogozina, D. Särg, N. Taba, H.-A. Talvik, M. Teder-Laving, N. Tõnisson, M. Vaht, U. Vainik, U. Võsa, B. Yelmen, T. Esko, R. Kolde, R. Mägi, J. Vilo, T. Laisk, A. Metspalu, The Estonian Biobank's journey from biobanking to personalized medicine. *Nat. Commun.* **16**, 3270 (2025).

72. L. Leitsalu, T. Haller, T. Esko, M.-L. Tammesoo, H. Alavere, H. Snieder, M. Perola, P. C. Ng, R. Mägi, L. Milani, K. Fischer, A. Metspalu, Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).

73. H. Julkunen, A. Cichońska, M. Tiainen, H. Koskela, K. Nybo, V. Mäkelä, J. Nokso-Koivisto, K. Kristiansson, M. Perola, V. Salomaa, P. Jousilahti, A. Lundqvist, A. J. Kangas, P. Soininen, J. C. Barrett, P. Würtz, Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat. Commun.* **14**, 604 (2023).

74. B. L. Browning, X. Tian, Y. Zhou, S. R. Browning, Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).

75. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W.-M. Chen, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

76. O. Delaneau, J.-F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).

77. S. Rubinacci, O. Delaneau, J. Marchini, Genotype imputation using the Positional Burrows Wheeler Transform. *PLoS Genet.* **16**, e1009049 (2020).

78. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fuchsberger, Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

79. P.-R. Loh, P. Danecek, P. F. Palamara, C. Fuchsberger, Y. A Reshef, H. K Finucane, S. Schoenherr, L. Forer, S. McCarthy, G. R. Abecasis, R. Durbin, A. L Price, Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).

80. Nightingale Health Biobank Collaborative Group, Metabolomic and genomic prediction of common diseases in 700,217 participants in three national biobanks. *Nat. Commun.* **15**, 10092 (2024).

81. S. C. Ritchie, P. Surendran, S. Karthikeyan, S. A. Lambert, T. Bolton, L. Pennells, J. Danesh, E. Di Angelantonio, A. S. Butterworth, M. Inouye, Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. *Sci. Data* **10**, 64 (2023).

82. J. Mbatchou, L. Barnard, J. Backman, A. Marcketta, J. A. Kosmicki, A. Ziyatdinov, C. Benner, C. O'Dushlaine, M. Barber, B. Boutkov, L. Habegger, M. Ferreira, A. Baras, J. Reid, G. Abecasis, E. Maxwell, J. Marchini, Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).

83. B. Bulik-Sullivan, H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, P.-R. Loh, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, L. Duncan, J. R. B. Perry, N. Patterson, E. B. Robinson, M. J. Daly, A. L. Price, B. M. Neale, An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

84. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, e1004383 (2014).

85. G. Wang, A. Sarkar, P. Carbonetto, M. Stephens, A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020).

86. Y. I. Li, D. A. Knowles, J. Humphrey, A. N. Barbeira, S. P. Dickinson, H. K. Im, J. K. Pritchard, Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).

87. J. Ansel, E. Yang, H. He, N. Gimelshein, A. Jain, M. Voznesensky, B. Bao, P. Bell, D. Berard, E. Burovski, G. Chauhan, A. Chourdia, W. Constable, A. Desmaison, Z. DeVito, E. Ellison, W. Feng, J. Gong, M. Gschwind, B. Hirsh, S. Huang, K. Kalambarkar, L. Kirsch, M. Lazos, M. Lezcano, Y. Liang, J. Liang, Y. Lu, C. K. Luk, B. Maher, Y. Pan, C. Puhrsch, M. Reso, M. Saroufim, M. Y. Siraichi, H. Suk, S. Zhang, M. Suo, P. Tillet, X. Zhao, E. Wang, K. Zhou, R. Zou, X. Wang, A. Mathews, W. Wen, G. Chanan, P. Wu, S. Chintala, "PyTorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation" in *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2* (ACM, New York, NY, USA, 2024; http://dx.doi.org/10.1145/3620665.3640366).

88. O. O. Yavorska, S. Burgess, MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).

89. R. Tambets, A. Kolde, P. Kolberg, M. I. Love, K. Alasoo, Extensive co-regulation of neighboring genes complicates the use of eQTLs in target gene prioritization. *HGG Adv.* **5**, 100348 (2024).

90. A. Zhu, N. Matoba, E. P. Wilson, A. L. Tapia, Y. Li, J. G. Ibrahim, J. L. Stein, M. I. Love, MRLocus: Identifying causal genes mediating a trait through Bayesian estimation of allelic heterogeneity. *PLoS Genet.* **17**, e1009455 (2021).

91. J. Bowden, G. Davey Smith, S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

92. A. van der Graaf, R. Warmerdam, C. M. P. Auwerx, eQTLGen Consortium, U. Vosa, M. C. Borges, L. Franke, Z. Kutalik, MR-link-2: pleiotropy robust cis Mendelian randomization validated in four independent gold-standard datasets of causality, *medRxiv* (2024)p. 2024.01.22.24301400.

93. S. Burgess, V. Zuber, E. Valdes-Marquez, B. B. Sun, J. C. Hopewell, Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. *Genet. Epidemiol.* **41**, 714–725 (2017).

94. UK10K Consortium, K. Walter, J. L. Min, J. Huang, L. Crooks, Y. Memari, S. McCarthy, J. R. B. Perry, C. Xu, M. Futema, D. Lawson, V. Iotchkova, S. Schiffels, A. E. Hendricks, P. Danecek, R. Li, J. Floyd, L. V. Wain, I. Barroso, S. E. Humphries, M. E. Hurles, E. Zeggini, J. C. Barrett, V. Plagnol, J. B. Richards, C. M. T. Greenwood, N. J. Timpson, R. Durbin, N. Soranzo, The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).

95. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

96. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

97. M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Yeo, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A. Žídek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).

98. E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).