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

Ralf Tambets[1], Jaanika Kronberg[2], Erik Abner[2], Urmo Võsa[2], Ida Rahu[1], Nele Taba[2], Anastassia Kolde[2,3], Estonian Biobank Research Team[2], Krista Fischer[2,3], 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]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia
*These authors jointly supervised this work.
Correspondence: tonu.esko@ut.ee, kaur.alasoo@ut.ee, priit.palta@ut.ee

Abstract

Examining the upstream molecular consequences of genetic variation significantly enhances our understanding of the heritable determinants of complex traits and disease predisposition. Metabolites serve as key indicators of various biological processes and disease states, playing a crucial role in this systematic mapping, also providing opportunities for the discovery of new biomarkers for disease diagnosis and prognosis. Here, we present a genome-wide association study for 249 circulating metabolite traits quantified by nuclear magnetic resonance spectroscopy across various genetic ancestry groups from the Estonian Biobank and the UK Biobank. We generated mixed model associations in the Estonian Biobank and six major genetic ancestry groups of the UK Biobank and performed two separate meta-analyses across the predominantly European genetic ancestry samples (n = 599,249) and across all samples (n = 619,372). In total, we identified 89,489 locus-metabolite pairs and 8,917 independent lead variants, out of which 4,184 appear to be novel associated loci. Moreover, 12.4% of the independent lead variants had a minor allele frequency of less than 1%, highlighting the importance of including low-frequency and rare variants in metabolic biomarker studies. Our publicly available results provide a valuable resource for future GWAS interpretation and drug target prioritisation studies.

Introduction

Systematic mapping of the heritable determinants underlying complex traits and disease predisposition can be greatly improved by detailed understanding of upstream molecular consequences of genetic variation. Studying metabolite traits is crucial because 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 molecular human biology and the potential for novel therapeutic targets. This understanding can lead to the identification of new biomarkers for

disease diagnosis, prognosis, and treatment, as well as the development of personalised medical interventions.

Although genome-wide association studies (GWAS) for several traits and diseases now exceed the sample size of 1 million individuals (COVID-19 Host Genetics Initiative, 2023; Suzuki et al., 2024; Yengo et al., 2022; Zhou et al., 2022), studies of molecular traits such as gene expression (Võsa et al., 2021), plasma proteins (Sun et al., 2023) or circulating metabolites have lagged behind. Notable exceptions are five blood lipid traits where the largest meta-analysis now includes data from 1.65 million individuals (Graham et al., 2021). Therefore, recent large-scale metabolite GWAS continue to uncover novel associations and biological insights (Karjalainen et al., 2024; Richardson et al., 2022; Smith et al., 2022; van der Meer et al., 2024). For example, Karjalainen *et al* performed GWAS meta-analysis of 233 circulating metabolites from the Nightingale Health nuclear magnetic resonance (NMR) platform in up to 136,016 participants from 33 cohorts (Karjalainen et al., 2024), identifying 443 independent loci and revealing significant pleiotropy and polygenicity (Karjalainen et al., 2024). Additionally, using up to 115,082 samples from the phase 1 release of the UK Biobank Nightingale Health platform NMR data, two studies reported high levels of pleiotropy and genetic correlation between metabolites (Richardson et al., 2022; Smith et al., 2022). Similarly, the latest study using 207,836 unrelated White British UK Biobank participants from phase 2 release of the UK Biobank NMR data increased the number of discovered loci to 497 (van der Meer et al., 2024).

However, for more than half of the metabolites captured by NMR, the proportion of heritability explained by genome-wide significant variants remains below 50% in the largest GWAS to date (van der Meer et al., 2024), indicating that much larger sample sizes are needed to discover the remaining 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 low imputation accuracy of low-frequency variants. Therefore less attention has been paid to low-frequency and rare variation, which, while explaining less heritability over-all, could still provide important biological insights (Nag et al., 2023).

The large number of metabolites, their complexity and diversity create a challenge for their identification, compared to other omics measurements. Although the Nightingale Health NMR platform is highly reproducible and 39 of 249 inferred metabolites have now been clinically validated, there have been various changes over the years both in the number of metabolites quantified as well as their absolute quantification results (Bizzarri et al., 2023). This variability can pose additional challenges when meta-analysing metabolites across multiple cohorts. For example, although the latest NMR data release from the UK Biobank contains 249 metabolites, Karjalainen *et al* analysed only 233 metabolites from the same platform, 225 of which were shared (Karjalainen et al., 2024). Thus, to reduce unwanted variability and maximise statistical power, it is essential to ensure that metabolite quantification and normalisation is performed in a uniform manner across cohorts.

Here, we present a genome-wide association study for 249 circulating metabolites quantified by nuclear magnetic resonance spectroscopy 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 (Karjalainen et al., 2024; Richardson et al., 2022; Smith et al., 2022; van der Meer et al., 2024). These publicly available results will provide a valuable resource for future GWAS interpretation and target prioritisation studies.

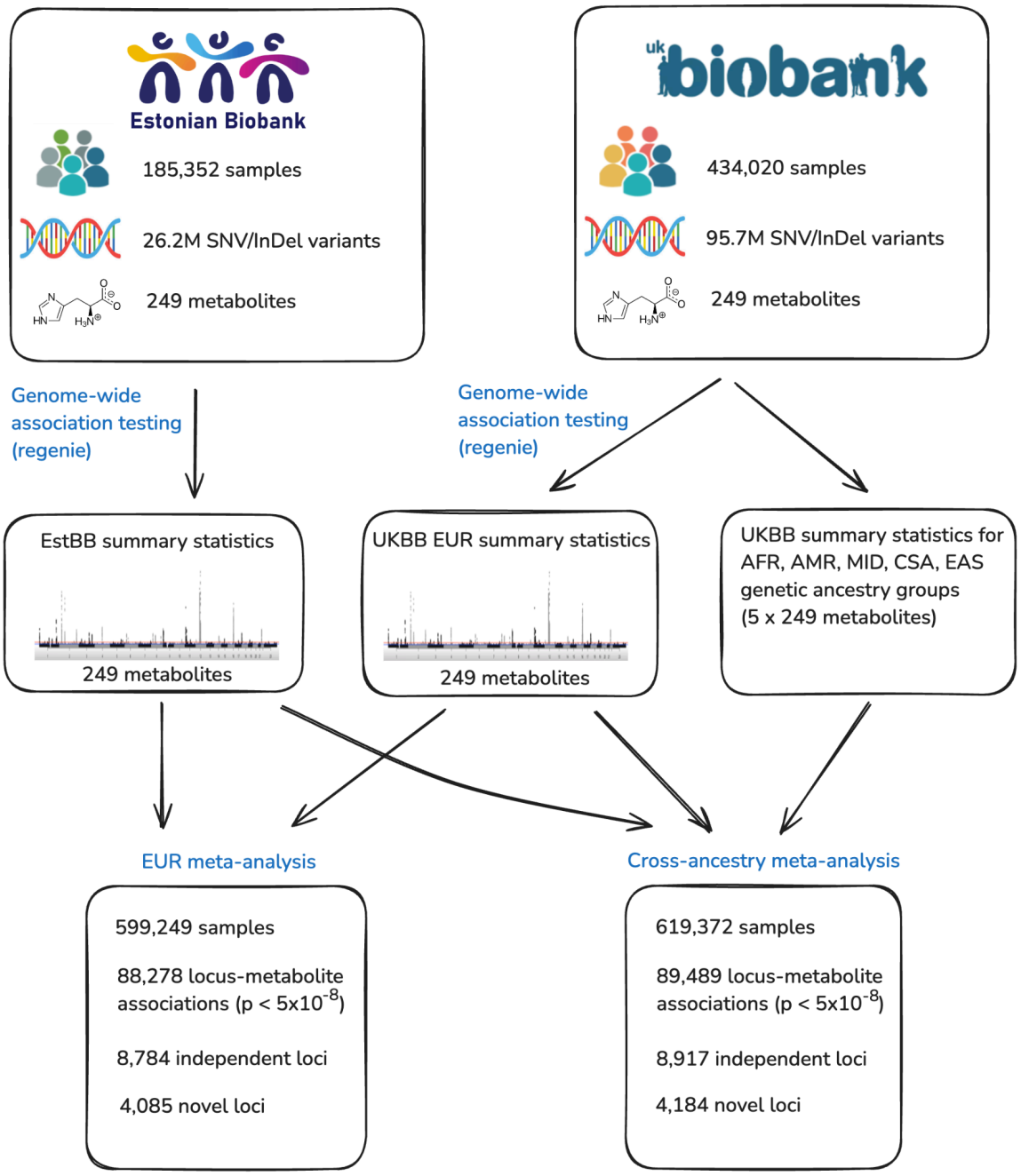

Figure 1. **Overview of the analysis plan.** First, a separate GWAS was performed for each metabolite in the Estonian Biobank and six ancestry groups of the UK Biobank: EUR (European), AFR (African), AMR (Admixed American), MID (Middle Eastern), EAS (East Asian), CSA (Central/South Asian). The definition of these broad genetic ancestry groups is discussed in (Karczewski et al., 2024). Second, inverse variance weighted meta-analysis was used to identify shared and ancestry-specific signals. InDel - insertion/deletion variant.

Results

Per-cohort analyses

We performed GWAS for 249 metabolites (Table S1) in the Estonian Biobank (EstBB) and six genetic ancestry groups from the UK Biobank (UKBB) (Figure 1). The UK Biobank genetic ancestry groups were defined previously by the Pan-UKBB project (Karczewski et al., 2024). Relying on population-specific genotype imputation panel for the EstBB (Mitt et al., 2017) and Genomics England imputation panel (Shi et al., 2024) for the UKBB allowed us to test 10-96 million variants for all ancestry-metabolite pairs (up to 9× more than (Karjalainen et al., 2024; van der Meer et al., 2024)). The number of trait-level genome-wide significant ($p < 5x10^{-8}$) locus-metabolite pairs ranged from 37 (UKBB_AMR) to 63,235 (UKBB_EUR) and the number of independent lead variants (r2 < 0.8) ranged from 24 to 6415, with most associations detected in the EstBB and UKBB_EUR subsets (Table 1). In comparison, applying the same filtering to Karjalainen *et al* summary statistics revealed 8,724 genome-wide significant locus-metabolite pairs corresponding to 669 independent loci.

Table 1. Number of significant locus-metabolite pairs ($p < 5x10^{-8}$) and unique lead variants (r2 > 0.8) detected in each genetic ancestry group and the two meta-analyses.

Genetic ancestry group		Locus-metabolite pairs	Unique lead variants	Sample size
EstBB		27,015	2,394	185,352
UKBB_EUR (European)		63,235	6,415	413,897
UKBB_AFR (African)		916	144	6,439
UKBB_AMR (Admixed American)		37	24	928
UKBB_MID (Middle Eastern)		92	44	1,500
UKBB_CSA (Central/South Asian)		1,070	112	8,652
UKBB_EAS (East Asian)		358	40	2,604
meta_EUR		88,278	8,784	599,249

meta_ALL		89,489	8,917	619,372

To assess the similarity of the associations detected in the two biobanks, we used LD score regression to calculate genetic correlations for each metabolite between the EstBB (n = 185,352) and the EUR subset of the UKBB (n = 413,897). We observed generally high genetic correlations for the matched metabolites between the two biobanks (median rg = 0.91, mean rg = 0.89, Bonferroni-corrected $p < 0.05$ for all comparisons; Table S2). Motivated by the high genetic correlation between the two biobanks, we proceeded with the meta-analyses.

Discovery of novel loci via meta-analysis

In the meta-analysis of EstBB and the EUR subset of the UKBB (meta_EUR, n = 599,249), we identified 88,278 locus-metabolite pairs corresponding to 8,784 independent lead variants ($r2 < 0.8$), representing approximately 10-fold increase as compared to the results reported by (Karjalainen et al., 2024). The heritability of individual metabolites ranged from 2.8% for Acetoacetate to 19.5% for HDL_size (median 10.2%) (Figure S1) and we observed a clear linear relationship between heritability and the number of loci discovered for each metabolite (Figure S2). In the following sections, we will focus on 56 selected metabolites representing amino acids, glycolysis related metabolites, ketone bodies, fluid balance, inflammation, and major lipid subclasses (Table S3). If not mentioned otherwise, we present the results from the meta-analysis of the EUR genetic ancestry groups from EstBB and UKBB (meta_EUR). Besides these results, we have publicly released the complete GWAS summary statistics for all 249 metabolites in all genetic ancestry groups as well as the two meta-analyses via the NHGRI-EBI GWAS Catalog (see Data availability).

We compared the associations detected in our meta_EUR meta-analysis to the largest independent metabolomics GWAS meta-analysis using the same NMR platform ((Karjalainen et al., 2024), n = 136,016). On average, 97% of the lead associations detected by Karjalainen *et al* also replicated in our meta_EUR analysis (Figure 2) with a highly concordant direction of effect (Figure S3). Identical analysis using results from our meta_ALL analysis is presented in Figure S4. 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 2). Altogether, we identified 4,085 novel loci not previously reported by Karjalainen *et al*, including 248 loci on chromosome X.

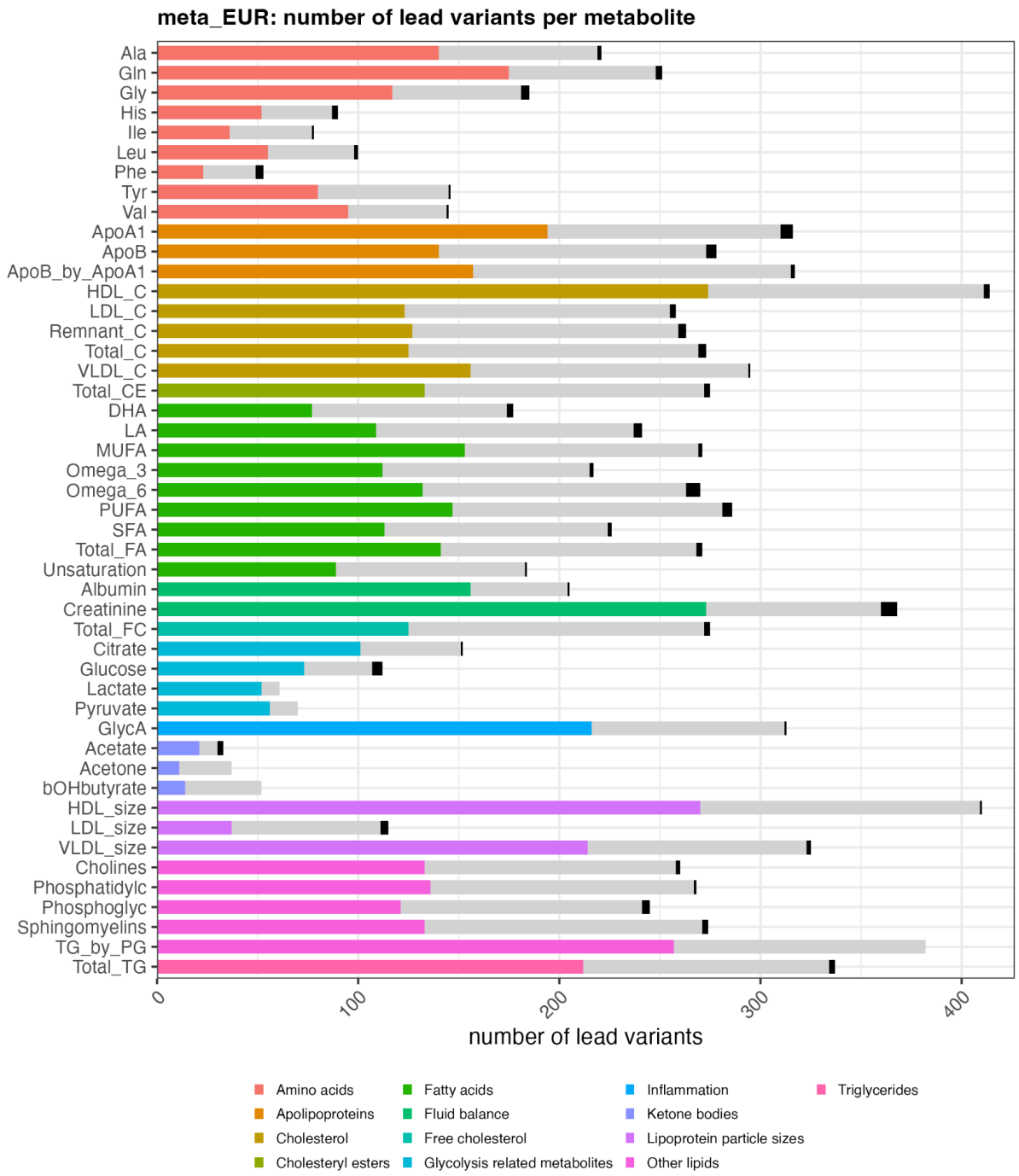

Figure 2. Number of genome-wide significant loci detected among selected metabolites shared with Karjalainen *et al*. Coloured bars represent novel loci. Grey bars represent loci detected by both our analysis and by Karjalainen *et al*. Black represents loci detected by Karjalainen *et al* and missed by us. Six metabolites included in our analysis (Clinical_LDL_C, non_HDL_C, Omega_6_by_Omega_3, PUFA_by_MUFA, Total_BCAA, Total_L, Total_P, Total_PL) were not profiled by Karjalainen *et al*. Identical analysis using results from our meta_ALL analysis is presented in Figure S4.

Associations with low-frequency variants

While the previous GWAS studies of NMR metabolites have focussed on common variation (MAF > 1%) (Karjalainen et al., 2024; van der Meer et al., 2024), we tested all variants with minor allele count (MAC) greater than 20. Thus, in our meta_EUR meta-analysis, 12.4% of the independent lead variants (n = 1,088 variants) had MAF < 1% (Figure 3). As expected, these low-frequency variants also had larger effect sizes than those at higher allele frequencies (Figure 3). This estimate is a lower bound as many loci with common lead variants are likely to contain secondary signals with lower allele frequencies.

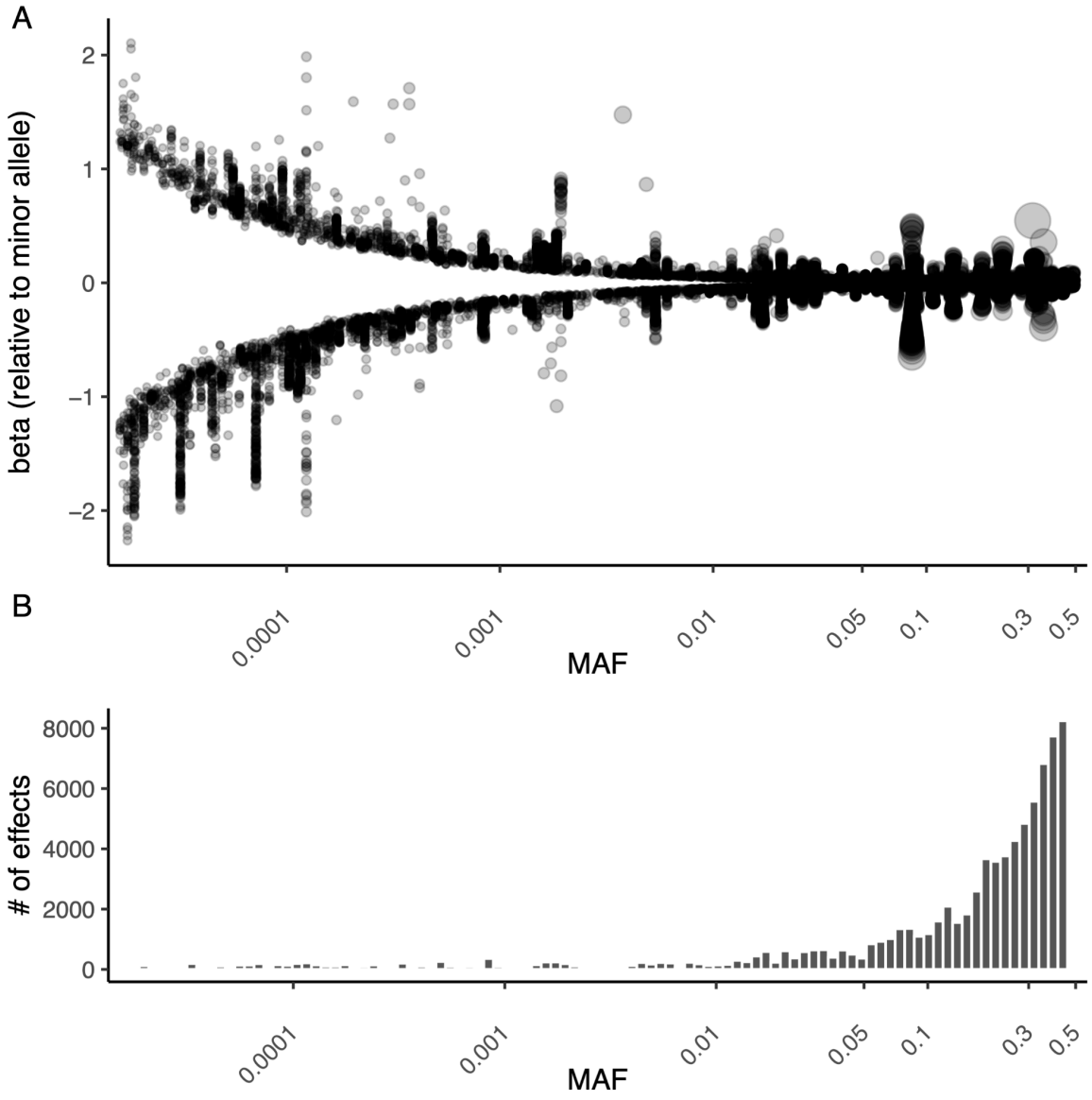

Figure 3. **Detected metabolite trait associations with low-frequency variants.** (**A**) Relationship between lead variant minor allele frequency (MAF) and effect size (beta). Each dot signifies the lead variant (+/- 1Mb window) from each locus-metabolite combination (meta_EUR). The size of each dot 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.

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 the meta-analysis increased the number of independent lead variants from 8,784 to 8,917 (Figure 1, 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 effects.

Genetic correlation and pleiotropy

To understand the shared genetic control of various classes of metabolites, we used LD score regression to estimate the pairwise genetic correlations between the selected set of 56 metabolites in our meta_EUR meta-analysis (Figure 4). Pairwise genetic correlations for all 249 metabolites are shown in Table S4. As expected, we observed high genetic correlations between 25 lipid traits. The three branched-chain amino acids (Leu, Val, Ile, total branched-chain amino acids), ketone bodies, and a group of lipid-related metabolites (LDL and HDL size, HDL cholesterol, ratio between polyunsaturated and monounsaturated fatty acids and unsaturation) also formed three distinct highly genetically correlated clusters (Figure 4). In contrast, other amino acids and glycolysis-related metabolites exhibited moderate genetic correlations with other metabolites. Finally, it is especially interesting that negative correlations with large clusters of other metabolites appear for metabolite ratios, such as PUFA_by_MUFA and omega_6_by_omega_3.

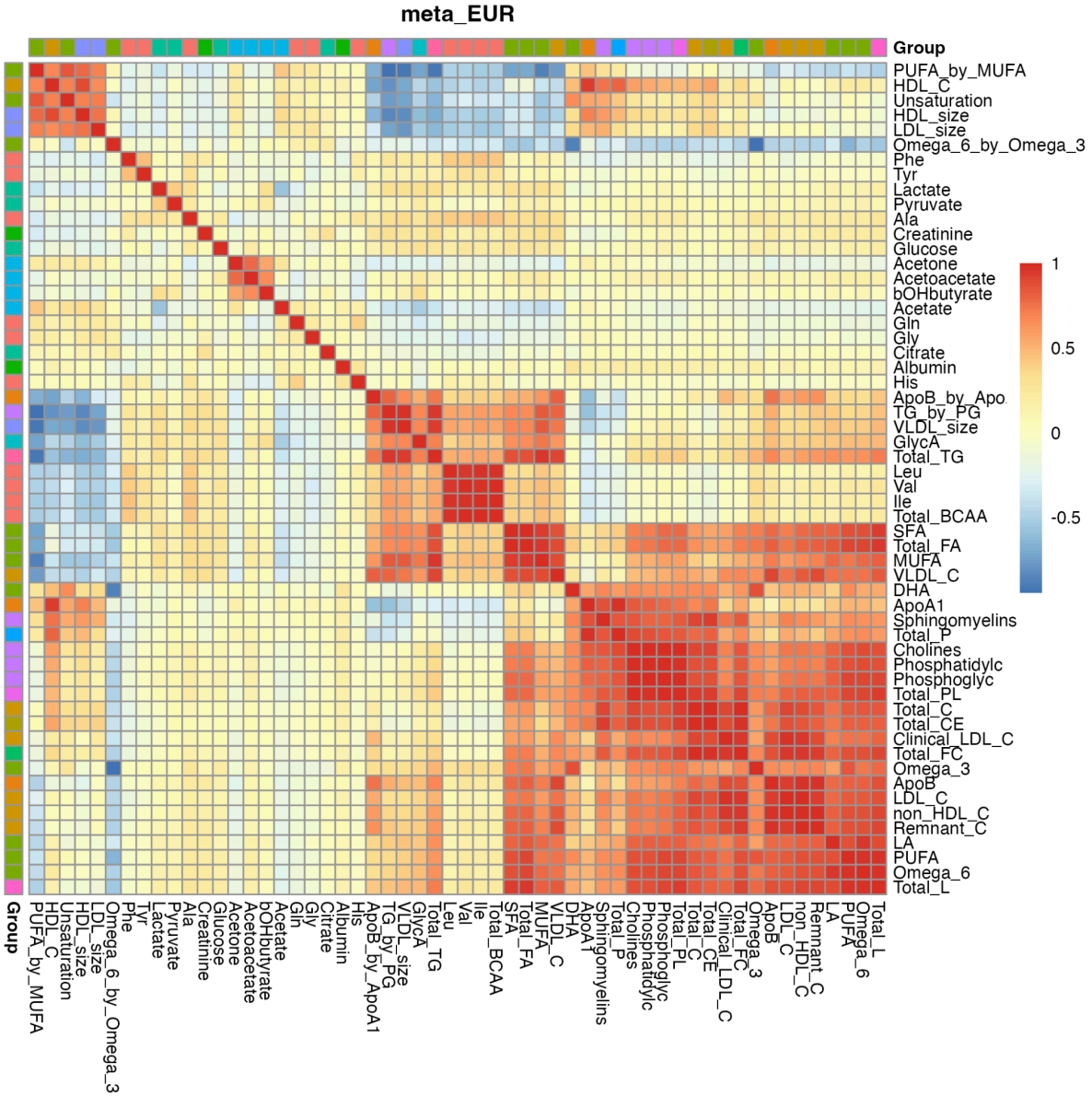

Figure 4. Genetic correlations for each of the 56 selected metabolites in the meta_EUR dataset. The highest correlation was observed between branched-chain amino acids, ketones, and lipid-related metabolites, respectively.

Additionally to the calculated genetic correlations, we identified clusters of lead variants that were shared ($r^2 > 0.8$) between different metabolite traits. Among the 249 metabolite traits, most lead variants were significantly associated ($p < 5x10^{-8}$) with multiple metabolites (mean = 10; median = 2). When focussing on the selected 56 metabolites, we detected 880 independent lead variants that were significantly associated ($p < 5x10^{-8}$) with five or more metabolites. Most prominently, a common missense variant (MAF = 40%) in the glucokinase regulatory protein (*GCKR*) (rs1260326, *GCKR*:p.Leu446Pro) was significantly associated ($p < 5x10^{-8}$) with 51 (out of 56) selected metabolites (Figure S5).

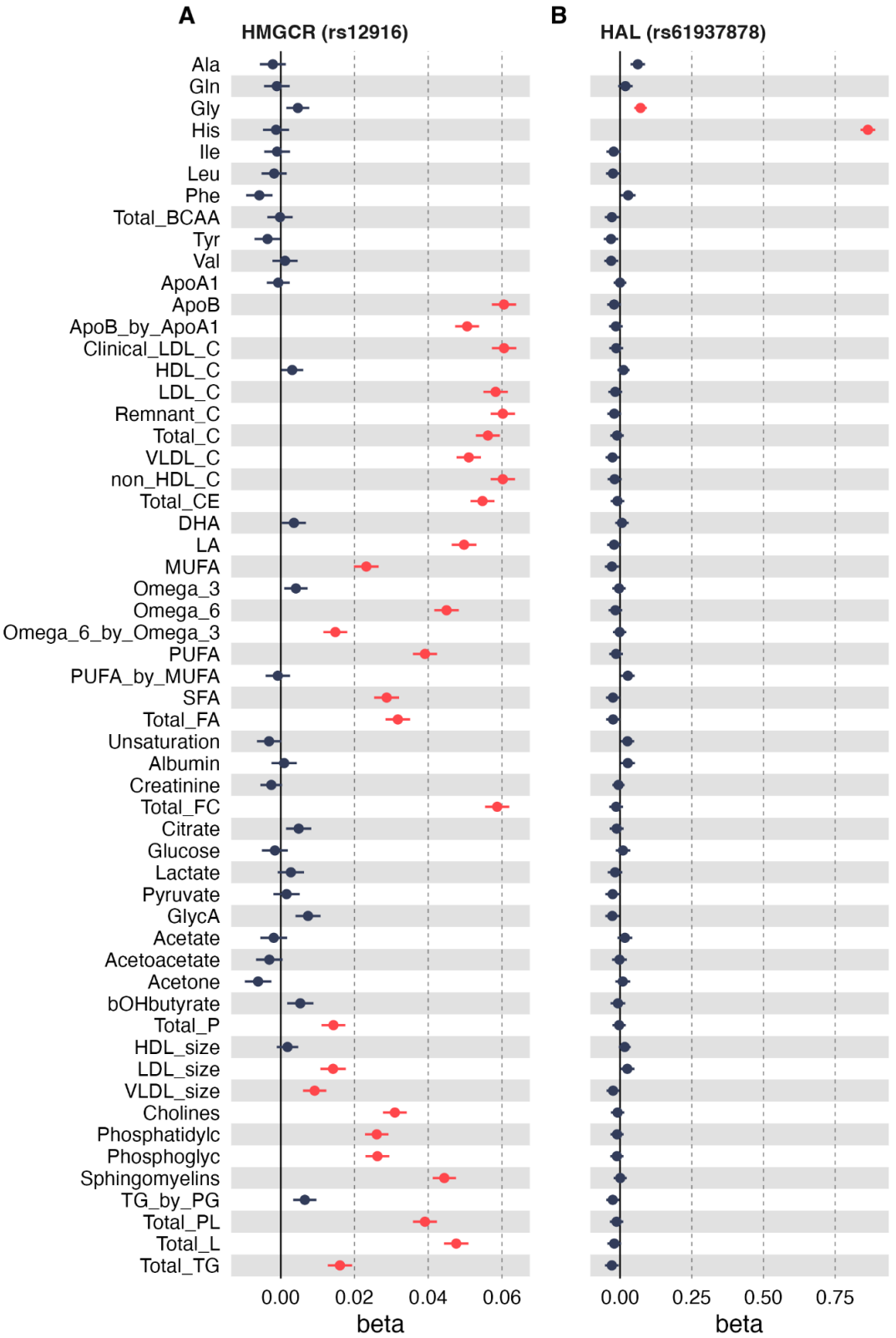

Figure 5. Examples of pleiotropy in the meta_EUR dataset. (**A**) Pleiotropic effects of the *HMGCR* locus lead variant rs12916 on many lipid-related metabolites. (**B**) Pleiotropy in the *HAL* locus (rs61937878) is limited to histidine and glycine amino acids. Red colour indicates statistically significant signals ($p < 5x10^{-8}$).

As another example of a highly pleiotropic locus, a non-coding variant (rs12916) in the HMG-CoA reductase (*HMGCR)* gene was associated with 27 (out of 56) selected metabolites and overall, with 180 (out of 249) tested metabolite traits (Figure 5A). Genetic variation in *HMGCR* has been robustly linked to cardiovascular disease (Ference et al., 2016) and *HMGCR* is also a known drug target for statin therapy.

Nevertheless, not all associations were equally pleiotropic. For example, a low-frequency (MAF = 0.4%) missense variant in the histidine ammonia lyase (*HAL)* gene (rs61937878) was significantly associated ($p < 5x10^{-8}$) with 2 out of 56 selected metabolites (histidine and glycine) and no additional significantly associated metabolites were detected among the full set of 249 tested (Figure 5B). HAL converts histidine to trans-urocanate (Hall, 1952), thus explaining the extremely strong association with histidine (beta = 0.864; $\log_{10}(p) = 951$). In contrast, the association with glycine was much weaker (beta = 0.071; $\log_{10}(p) = 9.60$) and could not be easily explained by a direct effect of the HAL enzymatic activity.

Discussion

A major advantage of our study is that we were able to use samples from two large biobanks assayed on the same metabolomics platform at the same time. This allowed us to ensure that all samples were processed and analysed in the same way, following identical procedures for sample processing, spectra generation, data acquisition in the same laboratory, thus reducing technical variability and increasing statistical power for discoveries. This low technical variability was exemplified by the high genetic correlation that we observed between the two biobanks (mean rg = 0.89). Furthermore, our meta-analyses involved 3-5x more samples and up to 9x more variants than two previous studies using the same NMR platform (Karjalainen et al., 2024; van der Meer et al., 2024). As a result, we were able to replicate 97% of previously known associations while simultaneously detecting more than 4000 novel associations (Figure 2). At the level of individual metabolites, the number of detected signals increased between 27-85%, Thus, our study provides the most comprehensive catalogue of genetic associations with these metabolites yet.

By utilising state-of-the-art population-specific genotype imputation panels (Mitt et al., 2017; Shi et al., 2024; Taliun et al., 2021), we were able to test proportionally more low-frequency variants than previous studies, leading to the discovery of numerous novel locus-metabolite associations. As a result, 12.4% of the independent lead variants detected in our analysis had a minor allele frequency (MAF) of less than 1% (Figure 3). This proportion is only likely to increase as we and others seek to identify independent low-frequency signals at established GWAS loci. Thus, while these low-frequency associations are unlikely to explain a large proportion of trait heritability, they can still be a valuable resource of genetic instruments for *cis*-Mendelian randomisation (*cis*-MR) drug target MR studies (Richardson et al., 2022). We expect future statistical fine mapping and rare variant analysis studies (Nag et al., 2023) to uncover many novel biological insights.

While large-scale biobanks 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 metabolite GWASs (Karjalainen et al., 2024; Richardson et al., 2022; Smith et al., 2022). Some of this pleiotropy is readily interpretable, such as co-regulation between various lipid traits (Richardson et al.,

2022) or opposing effects between substrates and products of enzymatic reactions (Smith et al., 2022). However, given our large sample size, we also detected more cryptic pleiotropic effects such as the histidine ammonia lyase (HAL) missense variant effect on glycine (Figure 5B). Such extensive overlaps exemplify that as cohorts grow larger, the detection of pleiotropic signals becomes more pronounced, complicating the disentanglement of direct and indirect genetic effects. This observation aligns with the omnigenic model, where all expressed genes in a cell contribute to complex traits through interconnected regulatory networks (Boyle et al., 2017; Sinnott-Armstrong et al., 2021; Smith et al., 2022).

Besides revealing molecular mechanisms of trait-associated genetic variation, the integration of genomic and metabolomic data from multiple biobanks has demonstrated the potential of metabolomics-based risk scores to estimate common disease risk more effectively than traditional polygenic scores (Buergel et al., 2022; Nightingale Health Biobank Collaborative Group et al., 2023). However, training these metabolomic risk scores requires longitudinal data from large biobanks, which might not be available for disease with low prevalence in the general population. Interestingly, a recent study demonstrated that it is possible to combine GWAS data for molecular biomarkers and disease outcomes to build predictive risk models in the absence of longitudinal data (Sens et al., 2024). These findings underscore the ongoing potential of large metabolite GWAS to improve our understanding of disease risk and progression, paving the way for personalised prevention and treatment strategies (Julkunen et al., 2023).

Our study also has several limitations. First, 97% of the samples included in our analysis were of predominantly European genetic ancestries. This demographic skew severely 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 1.3% (Table 1) when samples from other UKBB genetic ancestry groups (AFR, AMR, CSA, EAS, MID) were included in the analysis. Secondly, due to significant computational and methodological challenges, we did not perform statistical fine mapping of the identified loci. As a result, we are likely missing many of secondary signals at the genome-wide significant loci. We expect that cross-population fine mapping methods such as SuSiEx can help to resolve some of these issues in the future (Yuan et al., 2024). Lastly, we applied a global genome-wide significance threshold ($p < 5x10^{-8}$) tailored for common variants. To control for false-positive associations, this threshold may need adjustment to account for the large number of metabolites tested and the low allele frequency threshold (allele count > 20) utilised in our study.

Our comprehensive results, made publicly available, will serve as a valuable resource for the scientific community for future research, enabling more detailed analyses of genetic influences on circulating metabolite levels and paving the way for additional future studies towards improved understanding of genetic basis of metabolic traits and complex diseases, identify novel therapeutic targets, and develop personalised intervention strategies.

Methods

Cohorts

Estonian Biobank. The Estonian Biobank (EstBB) is a population-based biobank with 212,955 participants in the current data freeze (Milani et al., 2024). All biobank participants have signed a broad informed consent form and their blood sample collection was undertaken across the country between 2002 and 2021 (Leitsalu et al., 2015). The activities of EstBB are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of EstBB. 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 (Bycroft et al., 2018). 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 metabolomic biomarkers (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 (Julkunen et al., 2023). 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 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 (Mitt et al., 2017). 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 in build 37 and were lifted over to build 38 using Picard. Phasing was performed using the Beagle v5.4 software (Browning et al., 2021). Imputation was performed with Beagle v5.4 software (beagle.22Jul22.46e.jar) and default settings. Dataset was split into batches of 5,000. A population specific reference panel consisting of 2,695 WGS samples 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 (Manichaikul et al., 2010), 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 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) (Shi et al., 2024).

Briefly, the UK Biobank 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 (Delaneau et al., 2013) 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 (Rubinacci et al., 2020) (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. More details on the imp

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 (Das et al., 2016). 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 (Loh et al., 2016) and imputation was conducted with mimimac4 (Das et al., 2016). 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 (Nightingale Health Biobank Collaborative Group et al., 2023). 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 (Ritchie et al., 2023). 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

We conducted genome-wide association tests for each of the seven genetic ancestry groups separately using regenie v3.1.1 (Mbatchou et al., 2021), with sex, age, age squared and the top principal components 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.

Meta-analysis

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) (Bulik-Sullivan et al., 2015) 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 larger 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 $5x10^{-8}$. 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 r2 was at least 0.8. The variant with the lowest p-value was considered to be the only independent lead in each cluster.

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 (Sollis et al., 2023) (accessions GCST90449363 - GCST90451603, Table S5). GWAS lead variants are available from Zenodo (https://dx.doi.org/10.5281/zenodo.13937265). 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. 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. 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. The Estonian Biobank Research Team consists of Mari Nelis, Georgi Hudjasov, Reedik Mägi, Andres Metspalu, and Lili Milani.

Funding

K.A and I.R. were supported by a grant from 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. R.T., E.A, U.V., J.K. and T.E. were supported by the Estonian Research Council grant no PRG1291. 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. Project was supported by European Union's Horizon 2020 research and innovation programme under Grant Agreement No 101017802 (OPTOMICS).

Supplementary Figures

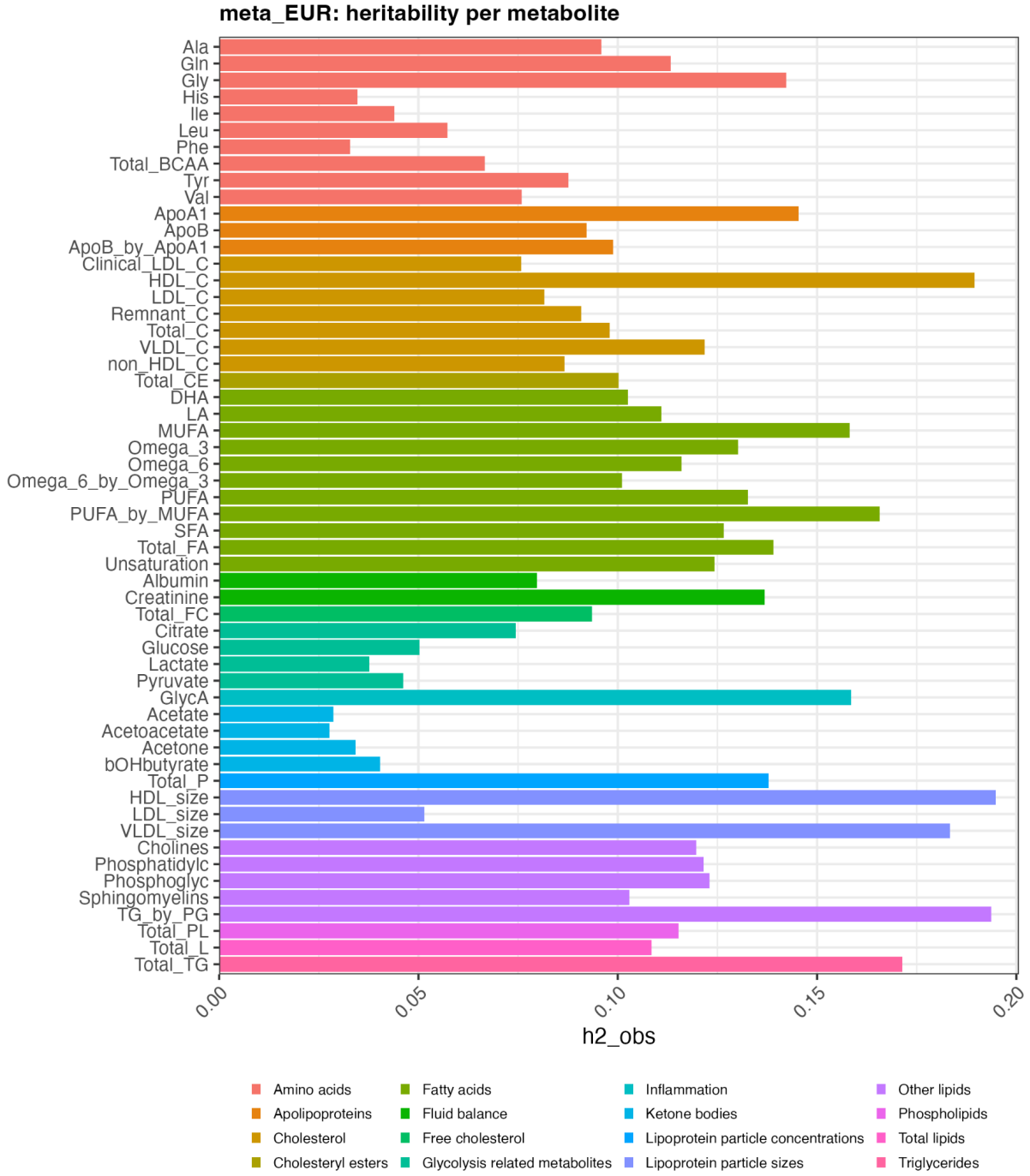

Figure S1. Heritability estimates from LD score regression for the 56 selected metabolites.

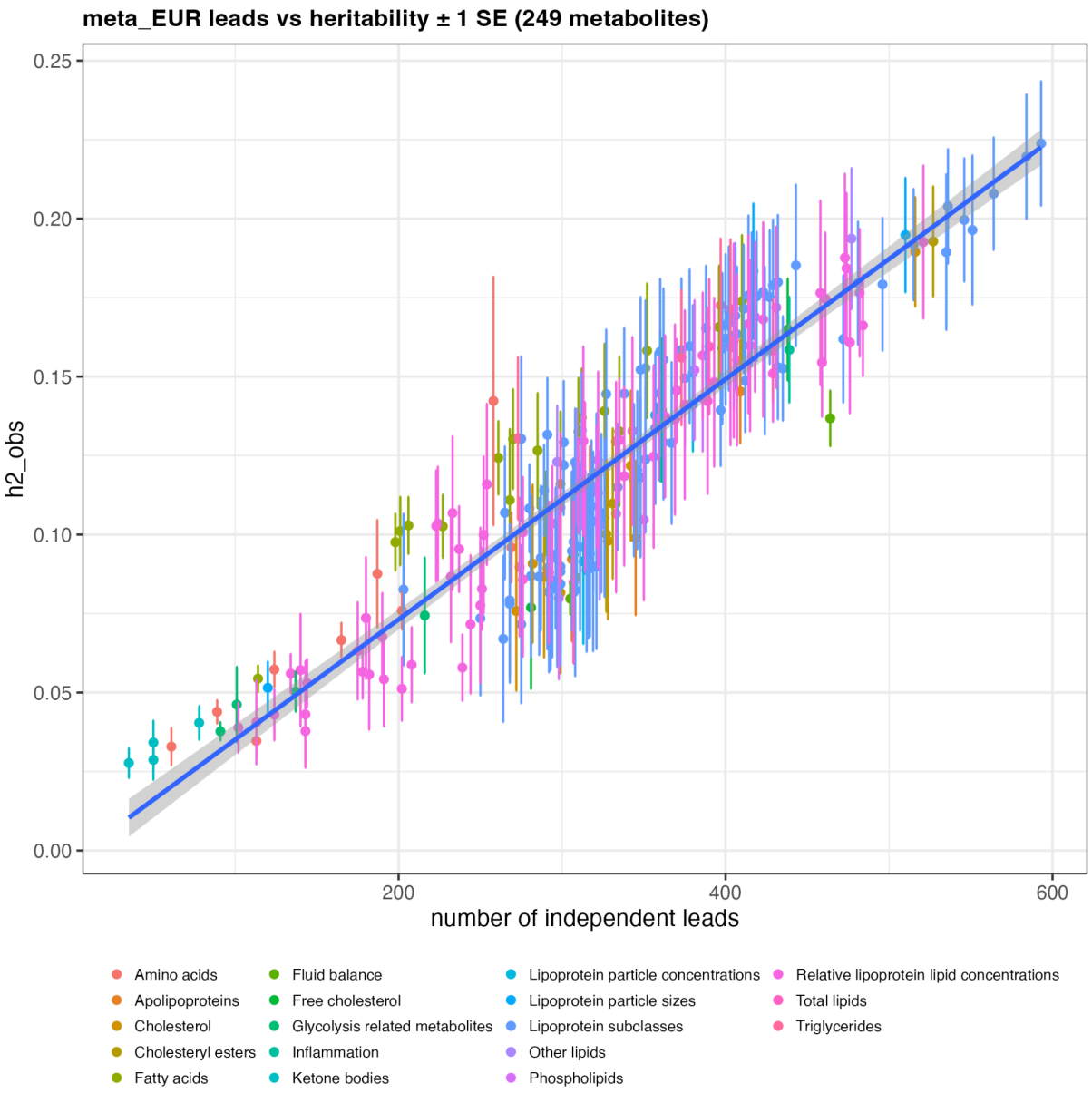

Figure S2. Relationship between heritability and number of genome-wide significant hits detected.

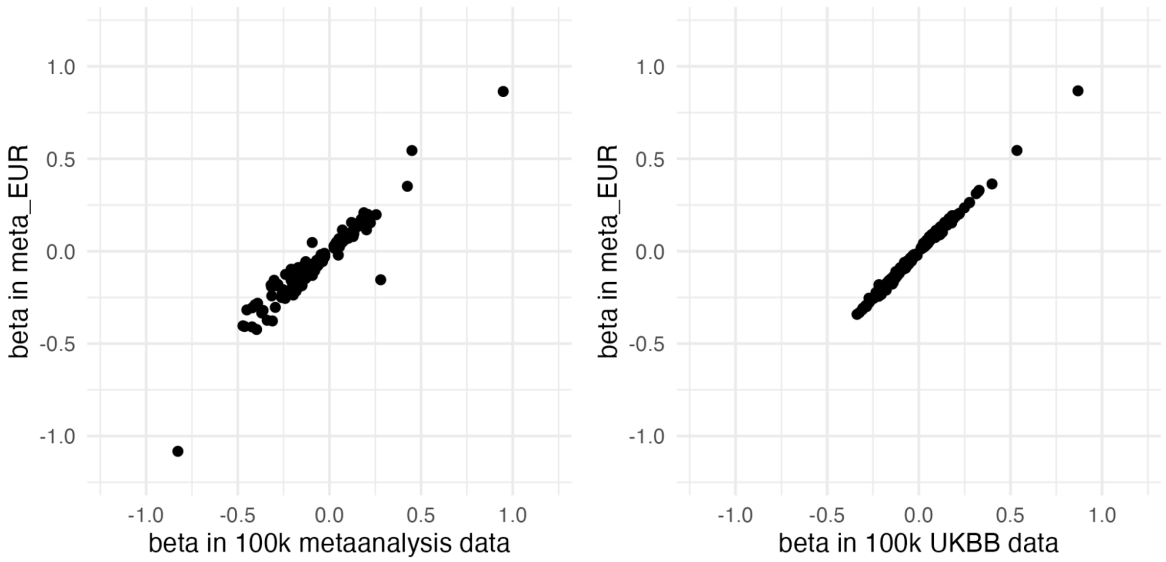

Figure S3. Comparison of shared lead variant betas in our European-ancestry meta-analysis (meta_EUR) and results presented by Karjalainen *et al*. Left panel, scatter plot of GWAS lead variant effect sizes from Karjalainen *et al* main analysis (n = 137k, 33 cohorts) and our meta_EUR. Right panel, GWAS 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. The 100k 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.

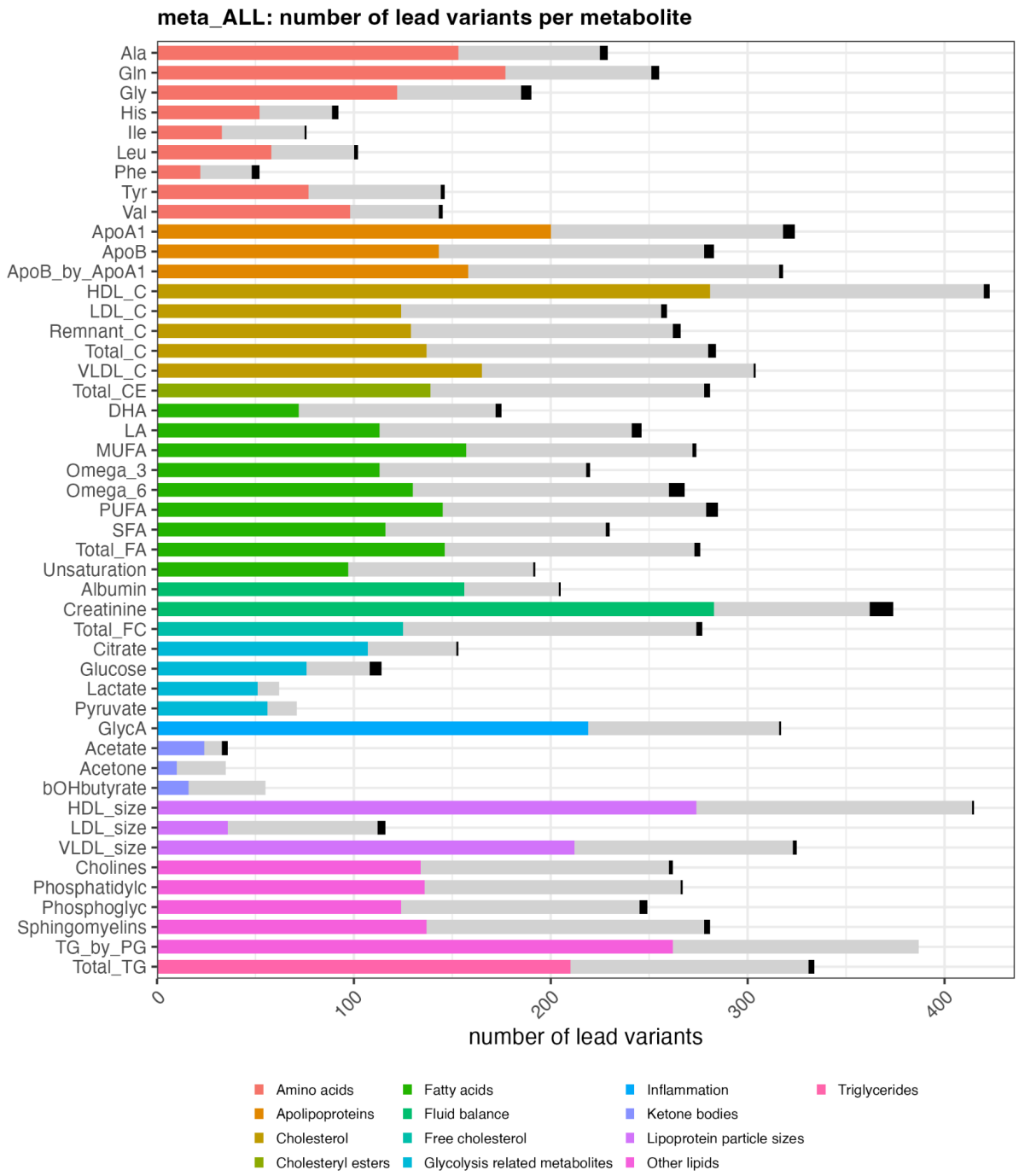

Figure S4. Number of genome-wide significant loci detected among selected metabolites shared with Karjalainen *et al* in the meta_ALL meta-analysis. Coloured bars represent novel loci. Grey bars represent loci detected by both our analysis and by Karjalainen *et al*. Black represents loci detected by Karjalainen *et al* and missed by us.

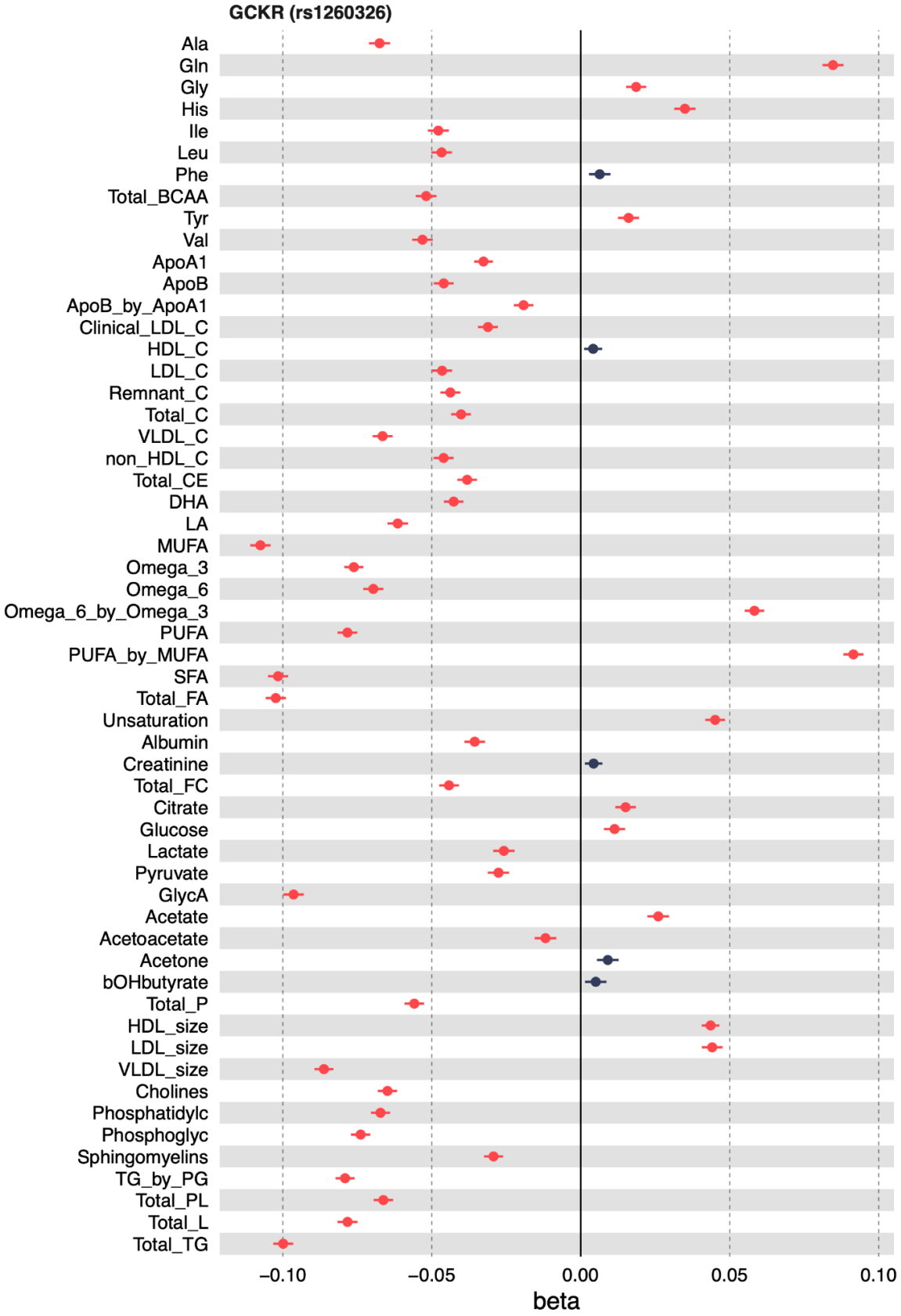

Figure S5. Pleiotropic effects of the *GCKR* missense variant rs1260326 on 56 selected metabolites.

References

Bizzarri D, Reinders MJT, Beekman M, Slagboom PE, van den Akker EB, On Behalf Of The BbmriNl. 2023. Technical report: A comprehensive comparison between different quantification versions of Nightingale health's 1H-NMR metabolomics platform. *Metabolites* **13**. doi:10.3390/metabo13121181

Boyle EA, Li YI, Pritchard JK. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**:1177–1186.

Browning BL, Tian X, Zhou Y, Browning SR. 2021. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet* **108**:1880–1890.

Buergel T, Steinfeldt J, Ruyoga G, Pietzner M, Bizzarri D, Vojinovic D, Upmeier zu Belzen J, Loock L, Kittner P, Christmann L, Hollmann N, Strangalies H, Braunger JM, Wild B, Chiesa ST, Spranger J, Klostermann F, van den Akker EB, Trompet S, Mooijaart SP, Sattar N, Jukema JW, Lavrijssen B, Kavousi M, Ghanbari M, Ikram MA, Slagboom E, Kivimaki M, Langenberg C, Deanfield J, Eils R, Landmesser U. 2022. Metabolomic profiles predict individual multidisease outcomes. *Nat Med* 1–12.

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L, Perry JRB, Patterson N, Robinson EB, Daly MJ, Price AL, Neale BM. 2015. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. doi:10.1038/ng.3406

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**:203–209.

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

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh P-R, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. 2016. Next-generation genotype imputation service and methods. *Nat Genet* **48**:1284–1287.

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

Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, Voros S, Giugliano RP, Davey Smith G, Fazio S, Sabatine MS. 2016. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. *N Engl J Med* **375**:2144–2153.

Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, Surakka I, Ntalla I, Vedantam S, Winkler TW, Locke AE, Marouli E, Hwang MY, Han S, Narita A, Choudhury A, Bentley AR, Ekoru K, Verma A, Trivedi B, Martin HC, Hunt KA, Hui Q, Klarin D, Zhu X, Thorleifsson G, Helgadottir A, Gudbjartsson DF, Holm H, Olafsson I, Akiyama M, Sakaue S, Terao C, Kanai M, Zhou W, Brumpton BM, Rasheed H, Ruotsalainen SE, Havulinna AS, Veturi Y, Feng Q, Rosenthal EA, Lingren T, Pacheco JA, Pendergrass SA, Haessler J, Giulianini F, Bradford Y, Miller JE, Campbell A, Lin K, Millwood IY, Hindy G, Rasheed A, Faul JD, Zhao W, Weir DR, Turman C, Huang H, Graff M, Mahajan A, Brown MR, Zhang W, Yu K, Schmidt EM, Pandit A, Gustafsson S, Yin X, Luan J, Zhao J-H, Matsuda F, Jang H-M, Yoon K, Medina-Gomez C, Pitsillides A, Hottenga JJ, Willemsen G, Wood AR, Ji Y, Gao Z, Haworth S, Mitchell RE, Chai JF, Aadahl M, Yao J, Manichaikul A, Warren HR, Ramirez J, Bork-Jensen J, Kårhus LL, Goel A, Sabater-Lleal M, Noordam R, Sidore C, Fiorillo E, McDaid AF, Marques-Vidal P, Wielscher M, Trompet S, Sattar N, Møllehave LT, Thuesen BH, Munz M, Zeng L, Huang J, Yang B, Poveda A, Kurbasic A, Lamina C, Forer L, Scholz M, Galesloot TE, Bradfield JP, Daw EW, Zmuda

JM, Mitchell JS, Fuchsberger C, Christensen H, Brody JA, Feitosa MF, Wojczynski MK, Preuss M, Mangino M, Christofidou P, Verweij N, Benjamins JW, Engmann J, Kember RL, Slieker RC, Lo KS, Zilhao NR, Le P, Kleber ME, Delgado GE, Huo S, Ikeda DD, Iha H, Yang J, Liu J, Leonard HL, Marten J, Schmidt B, Arendt M, Smyth LJ, Cañadas-Garre M, Wang C, Nakatochi M, Wong A, Hutri-Kähönen N, Sim X, Xia R, Huerta-Chagoya A, Fernandez-Lopez JC, Lyssenko V, Ahmed M, Jackson AU, Irvin MR, Oldmeadow C, Kim H-N, Ryu S, Timmers PRHJ, Arbeeva L, Dorajoo R, Lange LA, Chai X, Prasad G, Lorés-Motta L, Pauper M, Long J, Li X, Theusch E, Takeuchi F, Spracklen CN, Loukola A, Bollepalli S, Warner SC, Wang YX, Wei WB, Nutile T, Ruggiero D, Sung YJ, Hung Y-J, Chen S, Liu F, Yang J, Kentistou KA, Gorski M, Brumat M, Meidtner K, Bielak LF, Smith JA, Hebbar P, Farmaki A-E, Hofer E, Lin M, Xue C, Zhang J, Concas MP, Vaccargiu S, van der Most PJ, Pitkänen N, Cade BE, Lee J, van der Laan SW, Chitrala KN, Weiss S, Zimmermann ME, Lee JY, Choi HS, Nethander M, Freitag-Wolf S, Southam L, Rayner NW, Wang CA, Lin S-Y, Wang J-S, Couture C, Lyytikäinen L-P, Nikus K, Cuellar-Partida G, Vestergaard H, Hildalgo B, Giannakopoulou O, Cai Q, Obura MO, van Setten J, Li X, Schwander K, Terzikhan N, Shin JH, Jackson RD, Reiner AP, Martin LW, Chen Z, Li L, Highland HM, Young KL, Kawaguchi T, Thiery J, Bis JC, Nadkarni GN, Launer LJ, Li H, Nalls MA, Raitakari OT, Ichihara S, Wild SH, Nelson CP, Campbell H, Jäger S, Nabika T, Al-Mulla F, Niinikoski H, Braund PS, Kolcic I, Kovacs P, Giardoglou T, Katsuya T, Bhatti KF, de Kleijn D, de Borst GJ, Kim EK, Adams HHH, Ikram MA, Zhu X, Asselbergs FW, Kraaijeveld AO, Beulens JWJ, Shu X-O, Rallidis LS, Pedersen O, Hansen T, Mitchell P, Hewitt AW, Kähönen M, Pérusse L, Bouchard C, Tönjes A, Chen Y-DI, Pennell CE, Mori TA, Lieb W, Franke A, Ohlsson C, Mellström D, Cho YS, Lee H, Yuan J-M, Koh W-P, Rhee SY, Woo J-T, Heid IM, Stark KJ, Völzke H, Homuth G, Evans MK, Zonderman AB, Polasek O, Pasterkamp G, Hoefer IE, Redline S, Pahkala K, Oldehinkel AJ, Snieder H, Biino G, Schmidt R, Schmidt H, Chen YE, Bandinelli S, Dedoussis G, Thanaraj TA, Kardia SLR, Kato N, Schulze MB, Girotto G, Jung B, Böger CA, Joshi PK, Bennett DA, De Jager PL, Lu X, Mamakou V, Brown M, Caulfield MJ, Munroe PB, Guo X, Ciullo M, Jonas JB, Samani NJ, Kaprio J, Pajukanta P, Adair LS, Bechayda SA, de Silva HJ, Wickremasinghe AR, Krauss RM, Wu J-Y, Zheng W, den Hollander AI, Bharadwaj D, Correa A, Wilson JG, Lind L, Heng C-K, Nelson AE, Golightly YM, Wilson JF, Penninx B, Kim H-L, Attia J, Scott RJ, Rao DC, Arnett DK, Walker M, Koistinen HA, Chandak GR, Yajnik CS, Mercader JM, Tusié-Luna T, Aguilar-Salinas CA, Villalpando CG, Orozco L, Fornage M, Tai ES, van Dam RM, Lehtimäki T, Chaturvedi N, Yokota M, Liu J, Reilly DF, McKnight AJ, Kee F, Jöckel K-H, McCarthy MI, Palmer CNA, Vitart V, Hayward C, Simonsick E, van Duijn CM, Lu F, Qu J, Hishigaki H, Lin X, März W, Parra EJ, Cruz M, Gudnason V, Tardif J-C, Lettre G, 't Hart LM, Elders PJM, Damrauer SM, Kumari M, Kivimaki M, van der Harst P, Spector TD, Loos RJF, Province MA, Psaty BM, Brandslund I, Pramstaller PP, Christensen K, Ripatti S, Widén E, Hakonarson H, Grant SFA, Kiemeney LALM, de Graaf J, Loeffler M, Kronenberg F, Gu D, Erdmann J, Schunkert H, Franks PW, Linneberg A, Jukema JW, Khera AV, Männikkö M, Jarvelin M-R, Kutalik Z, Cucca F, Mook-Kanamori DO, van Dijk KW, Watkins H, Strachan DP, Grarup N, Sever P, Poulter N, Rotter JI, Dantoft TM, Karpe F, Neville MJ, Timpson NJ, Cheng C-Y, Wong T-Y, Khor CC, Sabanayagam C, Peters A, Gieger C, Hattersley AT, Pedersen NL, Magnusson PKE, Boomsma DI, de Geus EJC, Cupples LA, van Meurs JBJ, Ghanbari M, Gordon-Larsen P, Huang W, Kim YJ, Tabara Y, Wareham NJ, Langenberg C, Zeggini E, Kuusisto J, Laakso M, Ingelsson E, Abecasis G, Chambers JC, Kooner JS, de Vries PS, Morrison AC, North KE, Daviglus M, Kraft P, Martin NG, Whitfield JB, Abbas S, Saleheen D, Walters RG, Holmes MV, Black C, Smith BH, Justice AE, Baras A, Buring JE, Ridker PM, Chasman DI, Kooperberg C, Wei W-Q, Jarvik GP, Namjou B, Hayes MG, Ritchie MD, Jousilahti P, Salomaa V, Hveem K, Åsvold BO, Kubo M, Kamatani Y, Okada Y, Murakami Y, Thorsteinsdottir U, Stefansson K, Ho Y-L, Lynch JA, Rader DJ, Tsao PS, Chang K-M, Cho K, O'Donnell CJ, Gaziano JM, Wilson P, Rotimi CN, Hazelhurst S, Ramsay M, Trembath RC, van Heel DA, Tamiya G, Yamamoto M, Kim B-J, Mohlke KL, Frayling TM,

Hirschhorn JN, Kathiresan S, Boehnke M, Natarajan P, Peloso GM, Brown CD, Morris AP, Assimes TL, Deloukas P, Sun YV, Willer CJ. 2021. The power of genetic diversity in genome-wide association studies of lipids. *Nature* 1–11.
Hall DA. 1952. Histidine alpha-deaminase and the production of urocanic acid in the mammal. *Biochem J* **51**:499–504.
Julkunen H, Cichońska A, Tiainen M, Koskela H, Nybo K, Mäkelä V, Nokso-Koivisto J, Kristiansson K, Perola M, Salomaa V, Jousilahti P, Lundqvist A, Kangas AJ, Soininen P, Barrett JC, Würtz P. 2023. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun* **14**:604.
Karczewski KJ, Gupta R, Kanai M, Lu W, Tsuo K, Wang Y, Walters RK, Turley P, Callier S, Baya N, Palmer DS, Goldstein JI, Sarma G, Solomonson M, Cheng N, Bryant S, Churchhouse C, Cusick CM, Poterba T, Compitello J, King D, Zhou W, Seed C, Finucane HK, Daly MJ, Neale BM, Atkinson EG, Martin AR. 2024. Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. *medRxiv*. doi:10.1101/2024.03.13.24303864
Karjalainen MK, Karthikeyan S, Oliver-Williams C, Sliz E, Allara E, Fung WT, Surendran P, Zhang W, Jousilahti P, Kristiansson K, Salomaa V, Goodwin M, Hughes DA, Boehnke M, Fernandes Silva L, Yin X, Mahajan A, Neville MJ, van Zuydam NR, de Mutsert R, Li-Gao R, Mook-Kanamori DO, Demirkan A, Liu J, Noordam R, Trompet S, Chen Z, Kartsonaki C, Li L, Lin K, Hagenbeek FA, Hottenga JJ, Pool R, Ikram MA, van Meurs J, Haller T, Milaneschi Y, Kähönen M, Mishra PP, Joshi PK, Macdonald-Dunlop E, Mangino M, Zierer J, Acar IE, Hoyng CB, Lechanteur YTE, Franke L, Kurilshikov A, Zhernakova A, Beekman M, van den Akker EB, Kolcic I, Polasek O, Rudan I, Gieger C, Waldenberger M, Asselbergs FW, China Kadoorie Biobank Collaborative Group, Estonian Biobank Research Team, FinnGen, Hayward C, Fu J, den Hollander AI, Menni C, Spector TD, Wilson JF, Lehtimäki T, Raitakari OT, Penninx BWJH, Esko T, Walters RG, Jukema JW, Sattar N, Ghanbari M, Willems van Dijk K, Karpe F, McCarthy MI, Laakso M, Järvelin M-R, Timpson NJ, Perola M, Kooner JS, Chambers JC, van Duijn C, Slagboom PE, Boomsma DI, Danesh J, Ala-Korpela M, Butterworth AS, Kettunen J. 2024. Genome-wide characterization of circulating metabolic biomarkers. *Nature* **628**:130–138.
Leitsalu L, Haller T, Esko T, Tammesoo M-L, Alavere H, Snieder H, Perola M, Ng PC, Mägi R, Milani L, Fischer K, Metspalu A. 2015. Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* **44**:1137–1147.
Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, Schoenherr S, Forer L, McCarthy S, Abecasis GR, Durbin R, L Price A. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**:1443–1448.
Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**:2867–2873.
Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, Benner C, O'Dushlaine C, Barber M, Boutkov B, Habegger L, Ferreira M, Baras A, Reid J, Abecasis G, Maxwell E, Marchini J. 2021. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**:1097–1103.
Milani L, Alver M, Laur S, Reisberg S, Haller T, Aasmets O, Abner E, Alavere H, Allik A, Annilo T, Fischer K, Hudjashov G, Jõeloo M, Kals M, Karo-Astover L, Kasela S, Kolde A, Krebs K, Krigul KL, Kronberg J, Kruusmaa K, Kukuškina V, Kõiv K, Lehto K, Leitsalu L, Lind S, Luitva LB, Läll K, Lüll K, Metsalu K, Metspalu M, Mõttus R, Nelis M, Nikopensius T, Nurm M, Nõukas M, Oja M, Org E, Palover M, Palta P, Pankratov V, Pantiukh K, Pervjakova N, Pujol-Gualdo N, Reigo A, Reimann E, Smit S, Sokurova D, Taba N, Talvik H-A, Teder-Laving M, Tõnisson N, Vaht M, Vainik U, Võsa U, Esko T, Kolde R, Mägi R, Vilo J, Laisk T, Metspalu A. 2024. From Biobanking to Personalized Medicine: the journey of the Estonian Biobank. *medRxiv*. doi:10.1101/2024.09.22.24313964
Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, Ripatti S, Morris AP, Metspalu A, Esko T, Mägi R, Palta P. 2017. 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.
Nag A, Dhindsa RS, Middleton L, Jiang X, Vitsios D, Wigmore E, Allman EL, Reznichenko A, Carss K, Smith KR, Wang Q, Challis B, Paul DS, Harper AR, Petrovski S. 2023. Effects of protein-coding variants on blood metabolite measurements and clinical biomarkers in the UK Biobank. *Am J Hum Genet* **110**:487–498.
Nightingale Health Biobank Collaborative Group, Barrett JC, Esko T, Fischer K, Jostins-Dean L, Jousilahti P, Julkunen H, Jaaskelainen T, Kerimov N, Kerminen S, Kolde A, Koskela H, Kronberg J, Lundgren SN, Lundqvist A, Makela V, Nybo K, Perola M, Salomaa V, Schut K, Soikkeli M, Soininen P, Tiainen M, Tillmann` T, Wurtz P. 2023. Metabolomic and genomic prediction of common diseases in 477,706 participants in three national biobanks. *medRxiv*. doi:10.1101/2023.06.09.23291213
Richardson TG, Leyden GM, Wang Q, Bell JA, Elsworth B, Davey Smith G, Holmes MV. 2022. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol* **20**:e3001547.
Ritchie SC, Surendran P, Karthikeyan S, Lambert SA, Bolton T, Pennells L, Danesh J, Di Angelantonio E, Butterworth AS, Inouye M. 2023. Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. *Sci Data* **10**:64.
Rubinacci S, Delaneau O, Marchini J. 2020. Genotype imputation using the Positional Burrows Wheeler Transform. *PLoS Genet* **16**:e1009049.
Sens D, Shilova L, Gräf L, Grebenshchikova M, Eskofier BM, Casale FP. 2024. Genetics-driven risk predictions leveraging the Mendelian randomization framework. *Genome Res*. doi:10.1101/gr.279252.124
Shi S, Rubinacci S, Hu S, Moutsianas L, Stuckey A, Need AC, Palamara PF, Caulfield M, Marchini J, Myers S. 2024. A Genomics England haplotype reference panel and imputation of UK Biobank. *Nat Genet* **56**:1800–1803.
Sinnott-Armstrong N, Naqvi S, Rivas M, Pritchard JK. 2021. GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background. *Elife* **10**. doi:10.7554/eLife.58615
Smith CJ, Sinnott-Armstrong N, Cichońska A, Julkunen H, Fauman EB, Würtz P, Pritchard JK. 2022. Integrative analysis of metabolite GWAS illuminates the molecular basis of pleiotropy and genetic correlation. *Elife* **11**. doi:10.7554/eLife.79348
Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Güneş O, Hall P, Hayhurst J, Ibrahim A, Ji Y, John S, Lewis E, MacArthur JAL, McMahon A, Osumi-Sutherland D, Panoutsopoulou K, Pendlington Z, Ramachandran S, Stefancsik R, Stewart J, Whetzel P, Wilson R, Hindorff L, Cunningham F, Lambert SA, Inouye M, Parkinson H, Harris LW. 2023. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* **51**:D977–D985.
Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, Hou L, Kvikstad EM, Burren OS, Davitte J, Ferber KL, Gillies CE, Hedman ÅK, Hu S, Lin T, Mikkilineni R, Pendergrass RK, Pickering C, Prins B, Baird D, Chen C-Y, Ward LD, Deaton AM, Welsh S, Willis CM, Lehner N, Arnold M, Wörheide MA, Suhre K, Kastenmüller G, Sethi A, Cule M, Raj A, Burkitt-Gray L, Melamud E, Black MH, Fauman EB, Howson JMM, Kang HM, McCarthy MI, Nioi P, Petrovski S, Scott RA, Smith EN, Szalma S, Waterworth DM, Mitnaul LJ, Szustakowski J, Gibson BW, Miller MR, Whelan CD. 2023. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* 1–10.
Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, Mandla R, Huerta-Chagoya A, Melloni GEM, Kanoni S, Rayner NW, Bocher O, Arruda AL, Sonehara K, Namba S, Lee SSK, Preuss MH, Petty LE, Schroeder P, Vanderwerff B, Kals M, Bragg F, Lin K, Guo X, Zhang W, Yao J, Kim YJ, Graff M, Takeuchi F, Nano J, Lamri A, Nakatochi M, Moon S, Scott RA, Cook JP, Lee J-J, Pan I, Taliun D, Parra EJ, Chai J-F, Bielak LF, Tabara Y, Hai Y, Thorleifsson G, Grarup N, Sofer T, Wuttke M, Sarnowski C, Gieger C, Nousome D, Trompet S, Kwak S-H, Long J, Sun M, Tong L,

Chen W-M, Nongmaithem SS, Noordam R, Lim VJY, Tam CHT, Joo YY, Chen C-H, Raffield LM, Prins BP, Nicolas A, Yanek LR, Chen G, Brody JA, Kabagambe E, An P, Xiang AH, Choi HS, Cade BE, Tan J, Broadaway KA, Williamson A, Kamali Z, Cui J, Thangam M, Adair LS, Adeyemo A, Aguilar-Salinas CA, Ahluwalia TS, Anand SS, Bertoni A, Bork-Jensen J, Brandslund I, Buchanan TA, Burant CF, Butterworth AS, Canouil M, Chan JCN, Chang L-C, Chee M-L, Chen J, Chen S-H, Chen Y-T, Chen Z, Chuang L-M, Cushman M, Danesh J, Das SK, de Silva HJ, Dedoussis G, Dimitrov L, Doumatey AP, Du S, Duan Q, Eckardt K-U, Emery LS, Evans DS, Evans MK, Fischer K, Floyd JS, Ford I, Franco OH, Frayling TM, Freedman BI, Genter P, Gerstein HC, Giedraitis V, González-Villalpando C, González-Villalpando ME, Gordon-Larsen P, Gross M, Guare LA, Hackinger S, Hakaste L, Han S, Hattersley AT, Herder C, Horikoshi M, Howard A-G, Hsueh W, Huang M, Huang W, Hung Y-J, Hwang MY, Hwu C-M, Ichihara S, Ikram MA, Ingelsson M, Islam MT, Isono M, Jang H-M, Jasmine F, Jiang G, Jonas JB, Jørgensen T, Kamanu FK, Kandeel FR, Kasturiratne A, Katsuya T, Kaur V, Kawaguchi T, Keaton JM, Kho AN, Khor C-C, Kibriya MG, Kim D-H, Kronenberg F, Kuusisto J, Läll K, Lange LA, Lee KM, Lee M-S, Lee NR, Leong A, Li L, Li Y, Li-Gao R, Ligthart S, Lindgren CM, Linneberg A, Liu C-T, Liu J, Locke AE, Louie T, Luan J ’an, Luk AO, Luo X, Lv J, Lynch JA, Lyssenko V, Maeda S, Mamakou V, Mansuri SR, Matsuda K, Meitinger T, Melander O, Metspalu A, Mo H, Morris AD, Moura FA, Nadler JL, Nalls MA, Nayak U, Ntalla I, Okada Y, Orozco L, Patel SR, Patil S, Pei P, Pereira MA, Peters A, Pirie FJ, Polikowsky HG, Porneala B, Prasad G, Rasmussen-Torvik LJ, Reiner AP, Roden M, Rohde R, Roll K, Sabanayagam C, Sandow K, Sankareswaran A, Sattar N, Schönherr S, Shahriar M, Shen B, Shi J, Shin DM, Shojima N, Smith JA, So WY, Stančáková A, Steinthorsdottir V, Stilp AM, Strauch K, Taylor KD, Thorand B, Thorsteinsdottir U, Tomlinson B, Tran TC, Tsai F-J, Tuomilehto J, Tusie-Luna T, Udler MS, Valladares-Salgado A, van Dam RM, van Klinken JB, Varma R, Wacher-Rodarte N, Wheeler E, Wickremasinghe AR, van Dijk KW, Witte DR, Yajnik CS, Yamamoto K, Yamamoto K, Yoon K, Yu C, Yuan J-M, Yusuf S, Zawistowski M, Zhang L, Zheng W, VA Million Veteran Program, Raffel LJ, Igase M, Ipp E, Redline S, Cho YS, Lind L, Province MA, Fornage M, Hanis CL, Ingelsson E, Zonderman AB, Psaty BM, Wang Y-X, Rotimi CN, Becker DM, Matsuda F, Liu Y, Yokota M, Kardia SLR, Peyser PA, Pankow JS, Engert JC, Bonnefond A, Froguel P, Wilson JG, Sheu WHH, Wu J-Y, Hayes MG, Ma RCW, Wong T-Y, Mook-Kanamori DO, Tuomi T, Chandak GR, Collins FS, Bharadwaj D, Paré G, Sale MM, Ahsan H, Motala AA, Shu X-O, Park K-S, Jukema JW, Cruz M, Chen Y-DI, Rich SS, McKean-Cowdin R, Grallert H, Cheng C-Y, Ghanbari M, Tai E-S, Dupuis J, Kato N, Laakso M, Köttgen A, Koh W-P, Bowden DW, Palmer CNA, Kooner JS, Kooperberg C, Liu S, North KE, Saleheen D, Hansen T, Pedersen O, Wareham NJ, Lee J, Kim B-J, Millwood IY, Walters RG, Stefansson K, Ahlqvist E, Goodarzi MO, Mohlke KL, Langenberg C, Haiman CA, Loos RJF, Florez JC, Rader DJ, Ritchie MD, Zöllner S, Mägi R, Marston NA, Ruff CT, van Heel DA, Finer S, Denny JC, Yamauchi T, Kadowaki T, Chambers JC, Ng MCY, Sim X, Below JE, Tsao PS, Chang K-M, McCarthy MI, Meigs JB, Mahajan A, Spracklen CN, Mercader JM, Boehnke M, Rotter JI, Vujkovic M, Voight BF, Morris AP, Zeggini E. 2024. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature*. doi:10.1038/s41586-024-07019-6

Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, Pitsillides AN, LeFaive J, Lee S-B, Tian X, Browning BL, Das S, Emde A-K, Clarke WE, Loesch DP, Shetty AC, Blackwell TW, Smith AV, Wong Q, Liu X, Conomos MP, Bobo DM, Aguet F, Albert C, Alonso A, Ardlie KG, Arking DE, Aslibekyan S, Auer PL, Barnard J, Barr RG, Barwick L, Becker LC, Beer RL, Benjamin EJ, Bielak LF, Blangero J, Boehnke M, Bowden DW, Brody JA, Burchard EG, Cade BE, Casella JF, Chalazan B, Chasman DI, Chen Y-DI, Cho MH, Choi SH, Chung MK, Clish CB, Correa A, Curran JE, Custer B, Darbar D, Daya M, de Andrade M, DeMeo DL, Dutcher SK, Ellinor PT, Emery LS, Eng C, Fatkin D, Fingerlin T, Forer L, Fornage M, Franceschini N, Fuchsberger C, Fullerton SM, Germer S, Gladwin MT, Gottlieb DJ, Guo X, Hall ME, He J, Heard-Costa NL, Heckbert SR, Irvin MR, Johnsen JM, Johnson AD,

Kaplan R, Kardia SLR, Kelly T, Kelly S, Kenny EE, Kiel DP, Klemmer R, Konkle BA, Kooperberg C, Köttgen A, Lange LA, Lasky-Su J, Levy D, Lin X, Lin K-H, Liu C, Loos RJF, Garman L, Gerszten R, Lubitz SA, Lunetta KL, Mak ACY, Manichaikul A, Manning AK, Mathias RA, McManus DD, McGarvey ST, Meigs JB, Meyers DA, Mikulla JL, Minear MA, Mitchell BD, Mohanty S, Montasser ME, Montgomery C, Morrison AC, Murabito JM, Natale A, Natarajan P, Nelson SC, North KE, O'Connell JR, Palmer ND, Pankratz N, Peloso GM, Peyser PA, Pleiness J, Post WS, Psaty BM, Rao DC, Redline S, Reiner AP, Roden D, Rotter JI, Ruczinski I, Sarnowski C, Schoenherr S, Schwartz DA, Seo J-S, Seshadri S, Sheehan VA, Sheu WH, Shoemaker MB, Smith NL, Smith JA, Sotoodehnia N, Stilp AM, Tang W, Taylor KD, Telen M, Thornton TA, Tracy RP, Van Den Berg DJ, Vasan RS, Viaud-Martinez KA, Vrieze S, Weeks DE, Weir BS, Weiss ST, Weng L-C, Willer CJ, Zhang Y, Zhao X, Arnett DK, Ashley-Koch AE, Barnes KC, Boerwinkle E, Gabriel S, Gibbs R, Rice KM, Rich SS, Silverman EK, Qasba P, Gan W, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Papanicolaou GJ, Nickerson DA, Browning SR, Zody MC, Zöllner S, Wilson JG, Cupples LA, Laurie CC, Jaquish CE, Hernandez RD, O'Connor TD, Abecasis GR. 2021. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**:290–299.

van der Meer D, Rahman Z, Ottas A, Parekh P, Kutrolli G, Stinson S, Koromina M, Rokicki J, Sonderby I, Parker N, Tesfaye M, Hindley G, Rodevand L, Koch E, Steen N, Berg J, O'Connell K, Smeland O, Frei O, Dale A, Djurovic S, Lehto K, Alver M, Milani L, Shadrin A, Andreassen O. 2024. Pleiotropic and sex-specific genetic architecture of circulating metabolic markers. *medRxiv*. doi:10.1101/2024.07.30.24311254

Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, Brugge H, Oelen R, de Vries DH, van der Wijst MGP, Kasela S, Pervjakova N, Alves I, Favé M-J, Agbessi M, Christiansen MW, Jansen R, Seppälä I, Tong L, Teumer A, Schramm K, Hemani G, Verlouw J, Yaghootkar H, Sönmez Flitman R, Brown A, Kukushkina V, Kalnapenkis A, Rüeger S, Porcu E, Kronberg J, Kettunen J, Lee B, Zhang F, Qi T, Hernandez JA, Arindrarto W, Beutner F, BIOS Consortium, i2QTL Consortium, Dmitrieva J, Elansary M, Fairfax BP, Georges M, Heijmans BT, Hewitt AW, Kähönen M, Kim Y, Knight JC, Kovacs P, Krohn K, Li S, Loeffler M, Marigorta UM, Mei H, Momozawa Y, Müller-Nurasyid M, Nauck M, Nivard MG, Penninx BWJH, Pritchard JK, Raitakari OT, Rotzschke O, Slagboom EP, Stehouwer CDA, Stumvoll M, Sullivan P, 't Hoen PAC, Thiery J, Tönjes A, van Dongen J, van Iterson M, Veldink JH, Völker U, Warmerdam R, Wijmenga C, Swertz M, Andiappan A, Montgomery GW, Ripatti S, Perola M, Kutalik Z, Dermitzakis E, Bergmann S, Frayling T, van Meurs J, Prokisch H, Ahsan H, Pierce BL, Lehtimäki T, Boomsma DI, Psaty BM, Gharib SA, Awadalla P, Milani L, Ouwehand WH, Downes K, Stegle O, Battle A, Visscher PM, Yang J, Scholz M, Powell J, Gibson G, Esko T, Franke L. 2021. 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.

Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, Graff M, Eliasen AU, Jiang Y, Raghavan S, Miao J, Arias JD, Graham SE, Mukamel RE, Spracklen CN, Yin X, Chen S-H, Ferreira T, Highland HH, Ji Y, Karaderi T, Lin K, Lüll K, Malden DE, Medina-Gomez C, Machado M, Moore A, Rüeger S, Sim X, Vrieze S, Ahluwalia TS, Akiyama M, Allison MA, Alvarez M, Andersen MK, Ani A, Appadurai V, Arbeeva L, Bhaskar S, Bielak LF, Bollepalli S, Bonnycastle LL, Bork-Jensen J, Bradfield JP, Bradford Y, Braund PS, Brody JA, Burgdorf KS, Cade BE, Cai H, Cai Q, Campbell A, Cañadas-Garre M, Catamo E, Chai J-F, Chai X, Chang L-C, Chang Y-C, Chen C-H, Chesi A, Choi SH, Chung R-H, Cocca M, Concas MP, Couture C, Cuellar-Partida G, Danning R, Daw EW, Degenhard F, Delgado GE, Delitala A, Demirkan A, Deng X, Devineni P, Dietl A, Dimitriou M, Dimitrov L, Dorajoo R, Ekici AB, Engmann JE, Fairhurst-Hunter Z, Farmaki A-E, Faul JD, Fernandez-Lopez J-C, Forer L, Francescatto M, Freitag-Wolf S, Fuchsberger C, Galesloot TE, Gao Y, Gao Z, Geller F, Giannakopoulou O, Giulianini F, Gjesing AP, Goel A, Gordon SD, Gorski M, Grove J, Guo X, Gustafsson S, Haessler J, Hansen TF, Havulinna AS, Haworth SJ, He J,

Heard-Costa N, Hebbar P, Hindy G, Ho Y-LA, Hofer E, Holliday E, Horn K, Hornsby WE, Hottenga J-J, Huang H, Huang J, Huerta-Chagoya A, Huffman JE, Hung Y-J, Huo S, Hwang MY, Iha H, Ikeda DD, Isono M, Jackson AU, Jäger S, Jansen IE, Johansson I, Jonas JB, Jonsson A, Jørgensen T, Kalafati I-P, Kanai M, Kanoni S, Kårhus LL, Kasturiratne A, Katsuya T, Kawaguchi T, Kember RL, Kentistou KA, Kim H-N, Kim YJ, Kleber ME, Knol MJ, Kurbasic A, Lauzon M, Le P, Lea R, Lee J-Y, Leonard HL, Li SA, Li X, Li X, Liang J, Lin H, Lin S-Y, Liu J, Liu X, Lo KS, Long J, Lores-Motta L, Luan J 'an, Lyssenko V, Lyytikäinen L-P, Mahajan A, Mamakou V, Mangino M, Manichaikul A, Marten J, Mattheisen M, Mavarani L, McDaid AF, Meidtner K, Melendez TL, Mercader JM, Milaneschi Y, Miller JE, Millwood IY, Mishra PP, Mitchell RE, Møllehave LT, Morgan A, Mucha S, Munz M, Nakatochi M, Nelson CP, Nethander M, Nho CW, Nielsen AA, Nolte IM, Nongmaithem SS, Noordam R, Ntalla I, Nutile T, Pandit A, Christofidou P, Pärna K, Pauper M, Petersen ERB, Petersen LV, Pitkänen N, Polašek O, Poveda A, Preuss MH, Pyarajan S, Raffield LM, Rakugi H, Ramirez J, Rasheed A, Raven D, Rayner NW, Riveros C, Rohde R, Ruggiero D, Ruotsalainen SE, Ryan KA, Sabater-Lleal M, Saxena R, Scholz M, Sendamarai A, Shen B, Shi J, Shin JH, Sidore C, Sitlani CM, Slieker RC, Smit RAJ, Smith AV, Smith JA, Smyth LJ, Southam L, Steinthorsdottir V, Sun L, Takeuchi F, Tallapragada DSP, Taylor KD, Tayo BO, Tcheandjieu C, Terzikhan N, Tesolin P, Teumer A, Theusch E, Thompson DJ, Thorleifsson G, Timmers PRHJ, Trompet S, Turman C, Vaccargiu S, van der Laan SW, van der Most PJ, van Klinken JB, van Setten J, Verma SS, Verweij N, Veturi Y, Wang CA, Wang C, Wang L, Wang Z, Warren HR, Bin Wei W, Wickremasinghe AR, Wielscher M, Wiggins KL, Winsvold BS, Wong A, Wu Y, Wuttke M, Xia R, Xie T, Yamamoto K, Yang J, Yao J, Young H, Yousri NA, Yu L, Zeng L, Zhang W, Zhang X, Zhao J-H, Zhao W, Zhou W, Zimmermann ME, Zoledziewska M, Adair LS, Adams HHH, Aguilar-Salinas CA, Al-Mulla F, Arnett DK, Asselbergs FW, Åsvold BO, Attia J, Banas B, Bandinelli S, Bennett DA, Bergler T, Bharadwaj D, Biino G, Bisgaard H, Boerwinkle E, Böger CA, Bønnelykke K, Boomsma DI, Børglum AD, Borja JB, Bouchard C, Bowden DW, Brandslund I, Brumpton B, Buring JE, Caulfield MJ, Chambers JC, Chandak GR, Chanock SJ, Chaturvedi N, Chen Y-DI, Chen Z, Cheng C-Y, Christophersen IE, Ciullo M, Cole JW, Collins FS, Cooper RS, Cruz M, Cucca F, Cupples LA, Cutler MJ, Damrauer SM, Dantoft TM, de Borst GJ, de Groot LCPGM, De Jager PL, de Kleijn DPV, Janaka de Silva H, Dedoussis GV, den Hollander AI, Du S, Easton DF, Elders PJM, Eliassen AH, Ellinor PT, Elmståhl S, Erdmann J, Evans MK, Fatkin D, Feenstra B, Feitosa MF, Ferrucci L, Ford I, Fornage M, Franke A, Franks PW, Freedman BI, Gasparini P, Gieger C, Girotto G, Goddard ME, Golightly YM, Gonzalez-Villalpando C, Gordon-Larsen P, Grallert H, Grant SFA, Grarup N, Griffiths L, Gudnason V, Haiman C, Hakonarson H, Hansen T, Hartman CA, Hattersley AT, Hayward C, Heckbert SR, Heng C-K, Hengstenberg C, Hewitt AW, Hishigaki H, Hoyng CB, Huang PL, Huang W, Hunt SC, Hveem K, Hyppönen E, Iacono WG, Ichihara S, Ikram MA, Isasi CR, Jackson RD, Jarvelin M-R, Jin Z-B, Jöckel K-H, Joshi PK, Jousilahti P, Jukema JW, Kähönen M, Kamatani Y, Kang KD, Kaprio J, Kardia SLR, Karpe F, Kato N, Kee F, Kessler T, Khera AV, Khor CC, Kiemeney LALM, Kim B-J, Kim EK, Kim H-L, Kirchhof P, Kivimaki M, Koh W-P, Koistinen HA, Kolovou GD, Kooner JS, Kooperberg C, Köttgen A, Kovacs P, Kraaijeveld A, Kraft P, Krauss RM, Kumari M, Kutalik Z, Laakso M, Lange LA, Langenberg C, Launer LJ, Le Marchand L, Lee H, Lee NR, Lehtimäki T, Li H, Li L, Lieb W, Lin X, Lind L, Linneberg A, Liu C-T, Liu J, Loeffler M, London B, Lubitz SA, Lye SJ, Mackey DA, Mägi R, Magnusson PKE, Marcus GM, Vidal PM, Martin NG, März W, Matsuda F, McGarrah RW, McGue M, McKnight AJ, Medland SE, Mellström D, Metspalu A, Mitchell BD, Mitchell P, Mook-Kanamori DO, Morris AD, Mucci LA, Munroe PB, Nalls MA, Nazarian S, Nelson AE, Neville MJ, Newton-Cheh C, Nielsen CS, Nöthen MM, Ohlsson C, Oldehinkel AJ, Orozco L, Pahkala K, Pajukanta P, Palmer CNA, Parra EJ, Pattaro C, Pedersen O, Pennell CE, Penninx BWJH, Perusse L, Peters A, Peyser PA, Porteous DJ, Posthuma D, Power C, Pramstaller PP, Province MA, Qi Q, Qu J, Rader DJ, Raitakari OT, Ralhan S, Rallidis LS, Rao DC, Redline S, Reilly DF, Reiner AP, Rhee

SY, Ridker PM, Rienstra M, Ripatti S, Ritchie MD, Roden DM, Rosendaal FR, Rotter JI, Rudan I, Rutters F, Sabanayagam C, Saleheen D, Salomaa V, Samani NJ, Sanghera DK, Sattar N, Schmidt B, Schmidt H, Schmidt R, Schulze MB, Schunkert H, Scott LJ, Scott RJ, Sever P, Shiroma EJ, Shoemaker MB, Shu X-O, Simonsick EM, Sims M, Singh JR, Singleton AB, Sinner MF, Smith JG, Snieder H, Spector TD, Stampfer MJ, Stark KJ, Strachan DP, 't Hart LM, Tabara Y, Tang H, Tardif J-C, Thanaraj TA, Timpson NJ, Tönjes A, Tremblay A, Tuomi T, Tuomilehto J, Tusié-Luna M-T, Uitterlinden AG, van Dam RM, van der Harst P, Van der Velde N, van Duijn CM, van Schoor NM, Vitart V, Völker U, Vollenweider P, Völzke H, Wacher-Rodarte NH, Walker M, Wang YX, Wareham NJ, Watanabe RM, Watkins H, Weir DR, Werge TM, Widen E, Wilkens LR, Willemsen G, Willett WC, Wilson JF, Wong T-Y, Woo J-T, Wright AF, Wu J-Y, Xu H, Yajnik CS, Yokota M, Yuan J-M, Zeggini E, Zemel BS, Zheng W, Zhu X, Zmuda JM, Zonderman AB, Zwart J-A, 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, Chasman DI, Cho YS, Heid IM, McCarthy MI, Ng MCY, O'Donnell CJ, Rivadeneira F, Thorsteinsdottir U, Sun YV, Tai ES, Boehnke M, Deloukas P, Justice AE, Lindgren CM, Loos RJF, Mohlke KL, North KE, Stefansson K, Walters RG, Winkler TW, Young KL, Loh P-R, Yang J, Esko T, Assimes TL, Auton A, Abecasis GR, Willer CJ, Locke AE, Berndt SI, Lettre G, Frayling TM, Okada Y, Wood AR, Visscher PM, Hirschhorn JN. 2022. A saturated map of common genetic variants associated with human height. *Nature* **610**:704–712.

Yuan K, Longchamps RJ, Pardiñas AF, Yu M, Chen T-T, Lin S-C, Chen Y, Lam M, Liu R, Xia Y, Guo Z, Shi W, Shen C, Schizophrenia Workgroup of Psychiatric Genomics Consortium, Daly MJ, Neale BM, Feng Y-CA, Lin Y-F, Chen C-Y, O'Donovan MC, Ge T, Huang H. 2024. Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nat Genet* **56**:1841–1850.

Zhou W, Kanai M, Wu K-HH, Rasheed H, Tsuo K, Hirbo JB, Wang Y, Bhattacharya A, Zhao H, Namba S, Surakka I, Wolford BN, Lo Faro V, Lopera-Maya EA, Läll K, Favé M-J, Partanen JJ, Chapman SB, Karjalainen J, Kurki M, Maasha M, Brumpton BM, Chavan S, Chen T-T, Daya M, Ding Y, Feng Y-CA, Guare LA, Gignoux CR, Graham SE, Hornsby WE, Ingold N, Ismail SI, Johnson R, Laisk T, Lin K, Lv J, Millwood IY, Moreno-Grau S, Nam K, Palta P, Pandit A, Preuss MH, Saad C, Setia-Verma S, Thorsteinsdottir U, Uzunovic J, Verma A, Zawistowski M, Zhong X, Afifi N, Al-Dabhani KM, Al Thani A, Bradford Y, Campbell A, Crooks K, de Bock GH, Damrauer SM, Douville NJ, Finer S, Fritsche LG, Fthenou E, Gonzalez-Arroyo G, Griffiths CJ, Guo Y, Hunt KA, Ioannidis A, Jansonius NM, Konuma T, Lee MTM, Lopez-Pineda A, Matsuda Y, Marioni RE, Moatamed B, Nava-Aguilar MA, Numakura K, Patil S, Rafaels N, Richmond A, Rojas-Muñoz A, Shortt JA, Straub P, Tao R, Vanderwerff B, Vernekar M, Veturi Y, Barnes KC, Boezen M, Chen Z, Chen C-Y, Cho J, Smith GD, Finucane HK, Franke L, Gamazon ER, Ganna A, Gaunt TR, Ge T, Huang H, Huffman J, Katsanis N, Koskela JT, Lajonchere C, Law MH, Li L, Lindgren CM, Loos RJF, MacGregor S, Matsuda K, Olsen CM, Porteous DJ, Shavit JA, Snieder H, Takano T, Trembath RC, Vonk JM, Whiteman DC, Wicks SJ, Wijmenga C, Wright J, Zheng J, Zhou X, Awadalla P, Boehnke M, Bustamante CD, Cox NJ, Fatumo S, Geschwind DH, Hayward C, Hveem K, Kenny EE, Lee S, Lin Y-F, Mbarek H, Mägi R, Martin HC, Medland SE, Okada Y, Palotie AV, Pasaniuc B, Rader DJ, Ritchie MD, Sanna S, Smoller JW, Stefansson K, van Heel DA, Walters RG, Zöllner S, 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, Martin AR, Willer CJ, Daly MJ, Neale BM. 2022. Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genom* **2**:100192.