Article

FNIP1 variants are associated with favourable metabolism in 1 million humans

https://doi.org/10.1038/s41586-026-10864-2

Received: 22 October 2025

Accepted: 30 June 2026

Published online: 5 August 2026

Open access

Check for updates

George Hindy[1,26], Rene C. Adam[2,26], Olukayode Sosina[1], Dwaine Pryce[2], David Blair[1], Joseph Herman[1], Joseph Lee[2], Peter Dornbos[1], Ernst Mayerhofer[1], Arthur Gilly[1], Charleen Hunt[2], Benjamin Geraghty[1], Karl Landheer[1], Liron Ganel[1], Antoine Baldassari[1], Chuanyi Zhang[1], Ivory Mintah[2], Daphne Sun[2], Angel Coppola[2], Kyle Brown[2], Tram Nguyen[2], Xin Xia[2], Regeneron Genetics Center*, GHS-RGC DiscovEHR Collaboration*, Penn Medicine Biobank*, Colorado Center for Personalized Medicine*, Mayo Clinic-RGC Project Generation*, Mexico City Prospective Study*, Malmö Diet and Cancer Study*, Indiana University School of Medicine (Indiana-CLDB)*, UCLA-RGC ATLAS Collaboration*, Mount Sinai Million Health Discoveries Program*, UTSW-Dallas Biobank*, South Asia Biobank Study*, Daniel J. Rader[3], Olle Melander[4,5], Christopher D. Still[6], Jaime Berumen[7], Pablo Kuri-Morales[8], Jesus Alegre-Díaz[7], Jason M. Torres[9], Jonathan R. Emberson[9], Rory Collins[9], Roberto Tapia-Conyer[7], Suganthi Balasubramanian[1], Marcus B. Jones[1], Michelle G. LeBlanc[1], Andrew J. Murphy[2], Christos A. Kyratsous[2], John D. Overton[1], Jeffrey G. Reid[1], Goncalo R. Abecasis[1], Jonathan Marchini[1], Cristen Willer[1], George D. Yancopoulos[2], Mark W. Sleeman[2], Jonas Bovijn[1], Adam Locke[1], Aris Baras[1], Niek Verweij[1], Viktoria Gusarova[2,27] ✉ & Luca A. Lotta[1,27] ✉

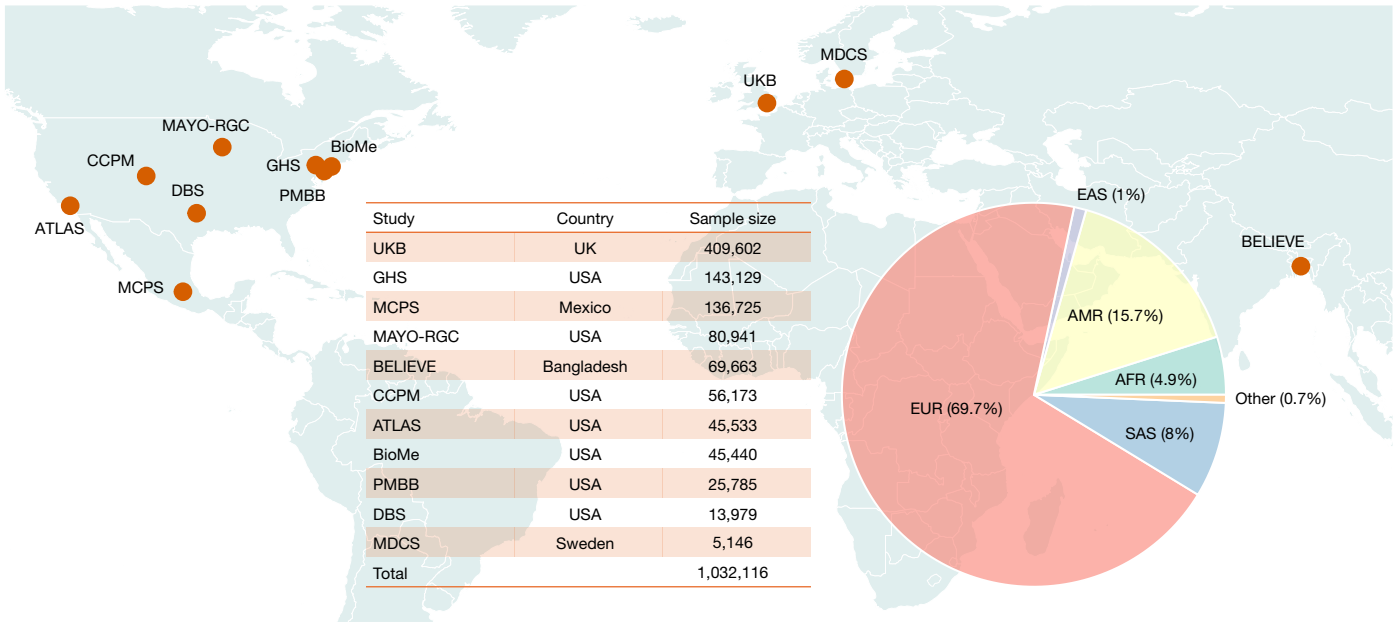
Altered energy metabolism is a shared driver across cardiometabolic diseases—the leading cause of death globally[1]. Energy metabolism varies between individuals and is partly heritable[2–9]. Here, to investigate the genetic basis of energy metabolism, we perform an exome-sequencing analysis of 1,032,116 people from America, Europe and Asia, and estimate associations between rare protein-coding variants and the ratio of triglyceride to high-density-lipoprotein cholesterol (TG:HDL)—an energy-state biomarker that we associate with diverse cardiometabolic risk factors and diseases. We identify 59 independent genes ($P < 1.04 \times 10^{-7}$) that are enriched for liver- and adipose-expressed master regulators of energy balance, storage and metabolism; 23 (39%) of these genes encode approved or clinical-stage drug targets. Ultra-rare protein-truncating variants in *FNIP1* (allele frequency, 0.01%), which encodes a suppressor of energy expenditure and mitochondrial metabolism, are associated with a lower TG:HDL ratio, lower liver fat, lower glycaemia, favourable fat distribution and around 60% lower odds of cardiometabolic disease. *FNIP1* knockdown in primary human hepatocytes induces lipid breakdown and lysosomal gene expression, while combined hepatic knockdown of *Fnip1* with its paralogue *Fnip2* or knockdown of its interactor *Flcn* protect against weight gain, reduce liver fat and enhance insulin sensitivity in mice fed a high-fat diet. Our study implicates the FNIP1 pathway in human energy metabolism and highlights its inhibition as a potential therapeutic strategy in cardiometabolic disease.

Altered energy and lipid metabolism contribute to cardiometabolic diseases such as coronary artery disease (CAD), obesity, type 2 diabetes and metabolic-dysfunction-associated steatotic liver disease (MASLD), which collectively represent the leading cause of death worldwide and account for a substantial and growing burden of morbidity[1,10]. The TG:HDL ratio has been proposed as a biomarker of the metabolic state of the body and its levels are associated with insulin resistance, type 2 diabetes and CAD[7,11–14].

Large exome-sequencing association studies of rare coding variants have led to the identification of therapeutic targets[8,9,15–18]. Here we use data from over 1 million people with exome sequencing, genome-wide common variant imputation and detailed health phenotypes from 11 cohorts across the world (Fig. 1 and Supplementary Table 1) to validate the TG:HDL ratio as a biomarker of the energy state of the body and risk of various cardiometabolic diseases.

We also perform a genetic discovery analysis of the TG:HDL ratio, implicating 59 genes of which coding variants are independently associated with this metabolic biomarker, including known and new therapeutic targets. With a combination of human genetics, omics data, and in vitro and in vivo validation in experimental models, we implicate the

A list of affiliations appears at the end of the paper. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: viktoria.gusarova@regeneron.com; luca.lotta@regeneron.com



Study	Country	Sample size
UKB	UK	409,602
GHS	USA	143,129
MCPS	Mexico	136,725
MAYO-RGC	USA	80,941
BELIEVE	Bangladesh	69,663
CCPM	USA	56,173
ATLAS	USA	45,533
BioMe	USA	45,440
PMBB	USA	25,785
DBS	USA	13,979
MDCS	Sweden	5,146
Total		1,032,116

Fig. 1 | Overview of the cohorts contributing to the exome-wide association analysis of the TG:HDL ratio. In total, 1,032,116 individuals were included in the exome-wide association analysis of rare variants with a TG:HDL ratio. The contributing cohorts were as follows: UCLA ATLAS Precision Health Biobank (ATLAS); BangladEsh Longitudinal Investigation of Emerging Vascular and nonvascular Events (BELIEVE); Mount Sinai BioMe BioBank (BioMe); Colorado Center for Personalized Medicine Biobank (CCPM); Dallas Biobank Study at UT Southwestern Medical Center (DBS); Geisinger Health System MyCode (GHS); Mayo Clinic Project Generation (MAYO-RGC); Mexico City Prospective Study (MCPS); Malmö Diet and Cancer Study (MDCS); University of Pennsylvania Penn Medicine BioBank (PMBB); and UK Biobank (UKB). AFR, African ancestry; AMR, admixed American ancestry; EAS, East Asian ancestry; EUR, European ancestry; SAS, South Asian ancestry. World maps were generated in R (v.4.5.2) using the maps package (v.3.4.3).

folliculin interacting protein 1 (*FNIP1*) and folliculin (*FLCN*) pathway in energy metabolism and the aetiology of cardiometabolic disease in humans as an illustrative example of how massive rare-coding-variant discovery analyses of this circulating metabolic biomarker can lead to translational and therapeutically relevant insights.

Epidemiology of the TG:HDL ratio

In an in-depth epidemiological analysis of refined anthropometric and metabolic traits, a higher TG:HDL ratio was associated with higher total and visceral adiposity across body imaging modalities (Fig. 2a, Supplementary Fig. 2 and Supplementary Table 2). This was accompanied by greater ectopic fat deposition in the liver and skeletal muscle, higher levels of liver damage biomarkers (for example, alanine aminotransferase (ALT)) and higher levels of liver-biopsy-defined steatosis, inflammation and fibrosis (Fig. 2b,c). There were statistically significant associations with higher fasting insulin, higher glycated haemoglobin, higher blood pressure and higher circulating C-reactive protein (CRP; Fig. 2d). Finally, a higher baseline TG:HDL ratio was predictive of an increased risk of incident type 2 diabetes, myocardial infarction, MASLD and liver cirrhosis (Fig. 2e). These associations were consistently observed across African, admixed American, East Asian, South Asian and European ancestries (Supplementary Figs. 3–7).

Thus, the TG:HDL ratio is a simple and powerful biomarker of energy metabolism and cardiometabolic disease risk, which can be used for genetic and biologic discovery.

Genomics of the TG:HDL ratio

We next performed a genetic discovery analysis of the TG:HDL ratio in a diverse set of 1,032,116 individuals from 11 cohorts and three continents (Fig. 1 and Supplementary Table 1).

In multiancestry genome-wide association and fine-mapping analyses of common variants (minor allele frequency ≥ 1%), we identified 1,617 independent signals at 801 physical loci associated with the TG:HDL ratio (Methods, Supplementary Note 1, Supplementary Tables 3 and 4 and Supplementary Fig. 8). In ancestry-specific fine-mapping analyses, we identified 1,406 independent signals in European, 209 in admixed American, 63 in South Asian, 44 in African and 9 in East Asian analyses (Supplementary Note 1, Supplementary Table 5 and Supplementary Figs. 9–13). Common variant signals showed enrichment for genes expressed in adipose and liver tissues (Supplementary Fig. 14) and we found strong genetic associations with anthropometric and metabolic phenotypes (Supplementary Fig. 15).

In an exome-wide gene-level association analysis of rare coding variants (minor allele frequency < 1%), 60 genes were associated with the TG:HDL ratio at the exome-wide level of statistical significance ($P < 1.04 \times 10^{-7}$), in an analysis adjusted for the 1,617 common variant signals identified in the same individuals (Fig. 3a,b, Extended Data Table 1 and Supplementary Note 2). Of the 60 genes, 59 showed an independent association when adjusting for other gene-burden signals (Supplementary Tables 6 and 7 and Supplementary Note 2), whereas the association of the *AC126283.2* gene disappeared in the conditional analysis, reflecting the sequence overlap with *PLA2G12A* (Supplementary Tables 6 and 8 and Supplementary Note 2).

Of the 59 independent genes, 15 (25%) had been identified in the previous largest rare variant analysis of TG:HDL ratio in around 170,000 people[19], while the remaining 44 (75%) were novel rare-coding-variant associations with the TG:HDL ratio (Supplementary Tables 7 and 9 and Supplementary Note 2).

The associations of the 59 genes with the TG:HDL ratio were robust and consistent in a series of sensitivity and subgroup analyses using different common variant adjustment, ancestry adjustment, phenotype definition approaches, as well as sex-, ancestry- and fasting-state subgroups analyses (Supplementary Tables 10–16, Supplementary Figs. 16–18 and Supplementary Note 2). Gene-burden associations were directionally consistent across ancestry groups, and the higher frequency in non-European ancestry groups of rare variants in some of

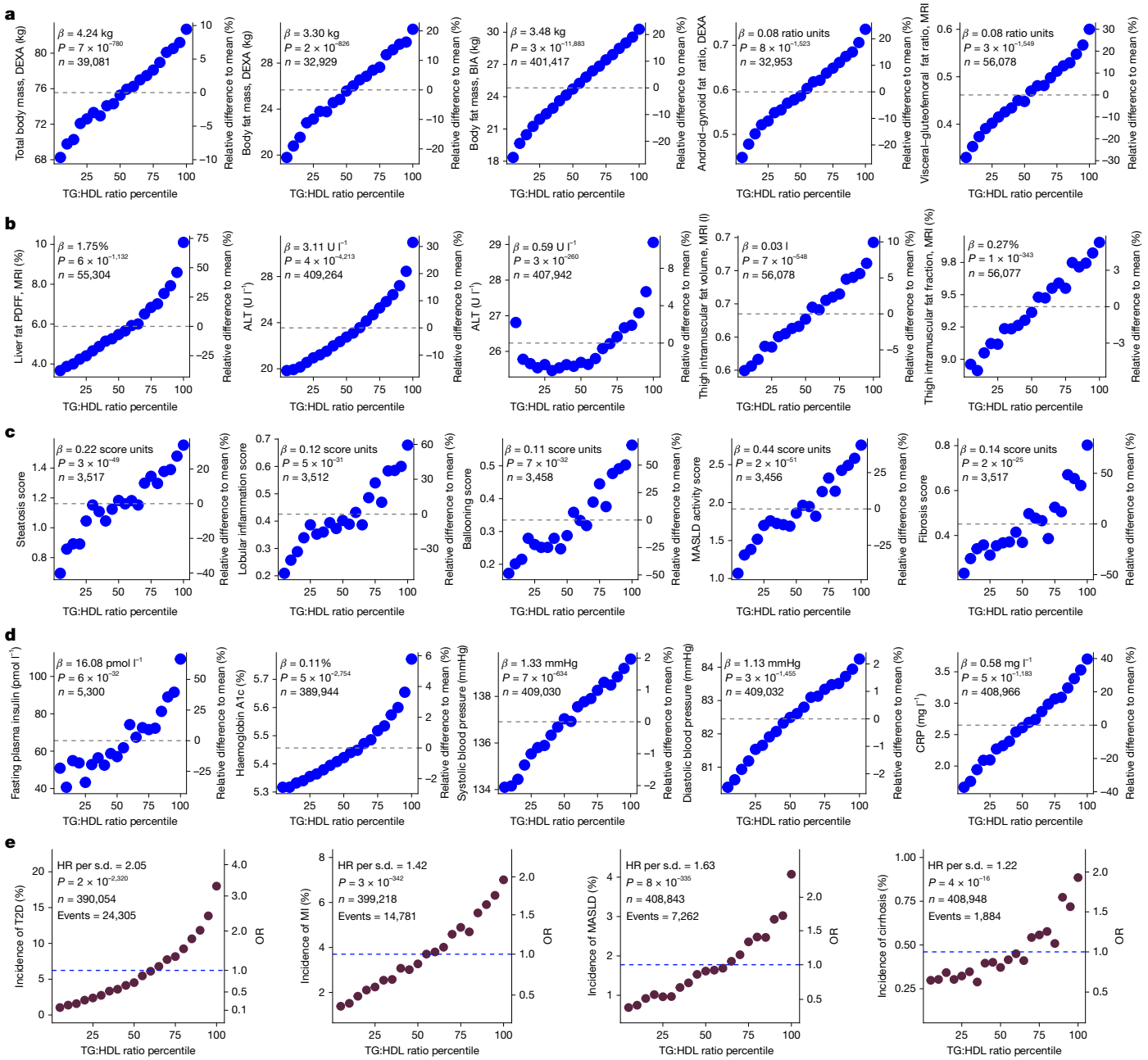


Fig. 2 | Association of the TG:HDL ratio with cardiometabolic phenotypes and disease risk. Epidemiological associations of the TG:HDL ratio with body mass and fat distribution as measured using dual-energy X-ray absorptiometry (DEXA), bioimpedance (BIA) and magnetic resonance imaging (MRI) (**a**); ectopic fat deposition in the liver estimated by MRI-proton density fat fraction (MRI-PDFF), liver injury measured by circulating aminotransferases and intramuscular fat estimated by MRI (**b**); liver steatosis, inflammation, ballooning, MASLD activity score and fibrosis assessed at liver biopsy (**c**); fasting plasma insulin as a measure of insulin resistance, haemoglobin A1c as a measure of glycaemic control, systolic and diastolic blood pressure, and CRP as a measure of inflammation (**d**); and risk of incident type 2 diabetes (T2D), myocardial infarction (MI), MASLD and liver cirrhosis (**e**). Associations of the TG:HDL ratio with phenotypes and disease outcomes were estimated in the UKB, except for fasting plasma insulin, which was estimated in the MDCS cohort. Associations with liver biopsy measures (**c**) were estimated in the GHS bariatric surgery cohort. In **a**–**d**, associations between the TG:HDL ratio (per 1 s.d. increase) and each outcome were estimated using linear regression adjusted for age and sex; effect estimates are shown with two-sided *P* values. In **e**, the associations between the TG:HDL ratio (per 1 s.d. increase) and incident events were estimated using Cox proportional hazards models adjusted for age and sex; hazard ratios (HR) are shown with two-sided *P* values. No adjustment was made for multiple comparisons.

the genes drove the identification of several associations (for example, *LIPE*, *PCSK9* and *GALNT2*; Supplementary Table 11, Supplementary Fig. 19 and Supplementary Note 2). In sex-stratified analyses, we noted an interesting sex interaction for *PDE3B* ($I^2 = 93\%$; $P_{interaction} = 1.8 \times 10^{-4}$; β_{women}, −0.40 versus β_{men}, −0.28 s.d. units per allele; Supplementary Table 13, Supplementary Fig. 17 and Supplementary Note 2)—a fat-distribution-associated gene[9].

In a replication analysis in 114,942 participants from the All of Us Research program, association estimates for the 59 genes showed strong correlations with those of our discovery analysis (Pearson's $r = 0.97$, $P = 2.1 \times 10^{-38}$; Supplementary Table 14, Supplementary Fig. 18 and Supplementary Note 2).

Using proteomics and immunoassay data, we studied the impact on circulating protein levels for predicted loss-of-function (pLOF) variants

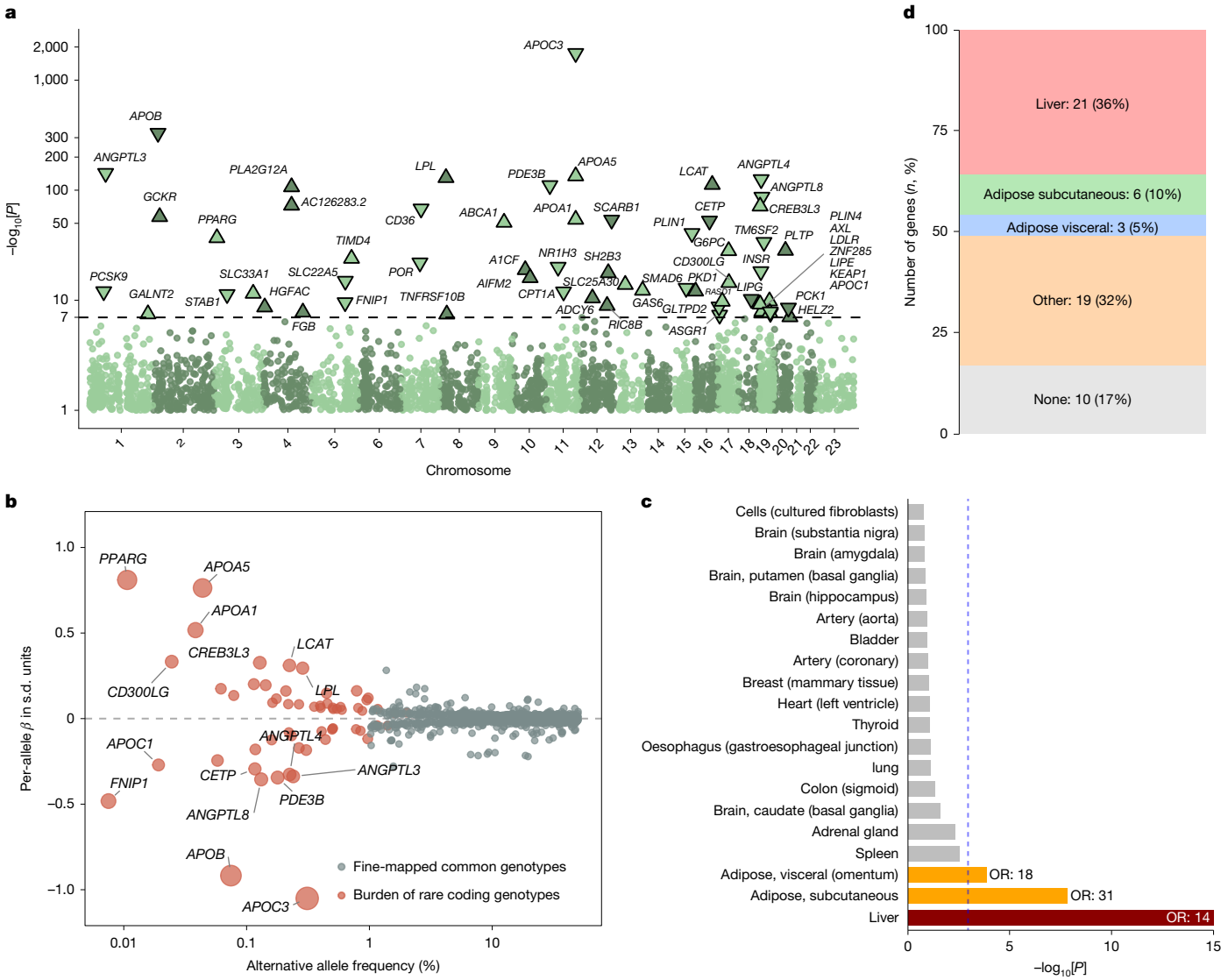


Fig. 3 | Associations with TG:HDL ratio in the exome-wide analysis and tissue enrichment for the identified genes. a, Gene-based exome-wide associations with the TG:HDL ratio. Triangles indicate association with a higher (upward) or lower (downward) TG:HDL ratio. Gene-level *P* values were determined using the Cauchy combination test (ACAT), combining *P* values from 24 gene-burden additive models per gene estimated by two-sided Wald tests. **b**, The relationship between the AAF and per-allele β for 1,617 fine-mapped common variants (grey) and 60 gene-burden genotypes (coral) associated with the TG:HDL ratio in our study. Symbol size is proportional to effect size. Per-allele β estimates are from the gene-burden additive model with the lowest two-sided Wald test *P* value per gene. **c**, Tissue expression enrichment among the 59 TG:HDL-ratio-associated genes as compared to the rest of the genes in the genome. The OR represents the likelihood of observing a TG:HDL ratio analysis $P \leq 1.04 \times 10^{-7}$ among genes exhibiting expression enrichment in a given tissue, relative to genes lacking such enrichment in that tissue. For each gene, the tissue with expression enrichment was defined as the tissue with the largest upward deviation in expression from the median across tissues and with an expression of at least 6 s.d. above the median expression across tissues. The significance threshold ($P = 0.0011$), designated by the dashed blue line, was derived after Bonferroni correction for 43 tissues. The plot shows the top 20 tissues. The enrichment ORs and *P* values were obtained using a one-sided penalized likelihood ratio test from the Firth-corrected logistic model. **d**, The number and proportion of genes among the 59 TG:HDL-ratio-associated genes enriched for specific tissues in GTEx.

in 13 genes with available measurements and found strong associations with lower levels of the encoded proteins for all genes, consistent with a loss-of-function effect (Supplementary Fig. 20). We also intersected protein quantitative trait loci (pQTL) with common variant signals at the gene-burden loci and identified 25 colocalizing signals between pQTLs and TG:HDL fine-mapped peaks, 14 of which converged on the exome-wide significant genes from the rare coding variant analysis (Supplementary Table 17).

Among the 59 genes, there was a considerable over-representation of genes with enriched expression in liver and adipose tissues (odds ratio (OR) for enrichment, 13.6; $P_{enrichment} = 9 \times 10^{-16}$ for liver; OR of 31.2 and 18.1, $P = 1 \times 10^{-8}$ and $P = 1 \times 10^{-4}$ for subcutaneous and visceral adipose tissue, respectively; Fig. 3c,d), further emphasizing the enrichment for the same energy metabolism tissues also observed in the common variant analysis (Supplementary Fig. 14).

The 59 associated genes included: (1) 11 (19%) genes for which coding variants are known to cause monogenic diseases of lipid metabolism (Supplementary Table 7); (2) 31 (53%) genes encoding known drug targets, including 23 for drugs in clinical development in humans or approved by health authorities (Supplementary Table 7); (3) 7 genes in the lipoprotein lipase (LPL) pathway, which regulates tissue energy availability (Extended Data Fig. 1 and Supplementary Table 7); and (4) 5 genes encoding apo-lipoproteins regulating the transport of lipids in the circulation, including a novel association for rare coding variants in the *APOC1* gene with a lower TG:HDL ratio (Extended Data Table 1 and Supplementary Tables 6 and 7).

We performed pathway enrichment analyses using DEPICT[20], which revealed 761 enriched pathways (at a Bonferroni-corrected $P < 5.9 \times 10^{-6}$; Supplementary Table 18) dominated by energy metabolism pathways, including lipid and glucose homeostasis, oxygen and energy consumption, lipoprotein metabolism, insulin resistance, liver fat accumulation, lipolysis regulation and fat storage (Supplementary Table 18 and Supplementary Note 2). For these pathways, an average of 16 (interquartile range, 11–19) of our 59 genes contributed to the pathway enrichment association (Supplementary Table 18 and Supplementary Note 2), reflecting a high concentration of true biological signal. Pathway analysis for gene subsets defined by the relative contribution of triglycerides versus HDL-C to the associations with the TG:HDL ratio revealed enrichment for pathways relating to the respective lipoprotein particles (Supplementary Tables 19–21 and Supplementary Note 2).

More than half of the 59 TG:HDL-ratio-associated genes showed gene-burden associations with glycaemic, liver or atherosclerosis related phenotypes (Supplementary Fig. 21), consistent with their central role in energy balance, storage and metabolism (Extended Data Fig. 1). There were also extensive associations with refined lipoprotein particle composition phenotypes measured using nuclear magnetic resonance in 469,068 individuals (Extended Data Fig. 2).

The FNIP1 pathway in energy metabolism

Ultra-rare protein-truncating variants in *FNIP1* were observed in around 1 out of every 7,000 people who we sequenced (combined alternative allele frequency (AAF) = 0.01%) and were associated with around 0.5 s.d. lower TG:HDL ratio ($P = 1.1 \times 10^{-10}$), representing a novel rare variant association and one of the largest effect associations identified in our analysis (Fig. 3, Extended Data Table 1 and Supplementary Tables 6 and 7). Associations with favourable phenotypes and disease protection for rare pLOF variants have highlighted a number of therapeutic targets[21–26], leading to new approved medicines[27–30], so we examined this association in more depth.

A total of 86 distinct ultra-rare pLOF variants in *FNIP1* contributed to the association (Supplementary Table 22), including a validated[31] loss-of-function variant implicated in immunodeficiency-93 and hypertrophic cardiomyopathy (Online Mendelian Inheritance in Man ID (MIM) 619705), an autosomal recessive disease caused by complete loss of FNIP1 function[31–33] (Supplementary Fig. 22a). Associations with TG:HDL ratio were robust when restricting to variants predicted to cause nonsense-mediated decay and for both N- and C-terminal variants (Supplementary Fig. 22b), consistent with a loss-of-function effect.

In over 1 million people with available exome-sequencing and clinical data, *FNIP1* pLOF variants were associated with lower atherogenic blood lipids, lower body-mass index, a lower visceral–gluteofemoral fat volume ratio, lower body fat percentage, higher body lean percentage, lower glycated haemoglobin, lower liver fat and lower liver enzyme levels, highlighting a favourable cardiometabolic phenotype across a broad range of traits (Fig. 4a and Extended Data Fig. 3).

In 227,636 cases and 265,114 controls, *FNIP1* pLOF variants were associated with around 60% lower odds of cardiometabolic disease outcomes (that is, CAD, type 2 diabetes, MASLD and liver cirrhosis combined; OR = 0.39; 95% confidence interval (CI) = 0.22–0.69; $P = 0.0011$; Fig. 4a), in the heterozygous state.

Heterozygous *FNIP1* variants did not show a statistically significant association with a composite electronic health record phenotype for the clinical features of immunodeficiency due to biallelic loss-of-function of FNIP1 (refs. 31,33,34), consistent with the recessive inheritance of this Mendelian syndrome (Supplementary Note 3, Supplementary Table 23 and Supplementary Figs. 23 and 24).

FNIP1 encodes folliculin-interacting protein 1, which interacts with folliculin (encoded by *FLCN*) to inhibit mitochondrial biogenesis, oxidative phosphorylation and energy expenditure downstream of AMPK[35].

In our human genetic analysis, *FLCN* heterozygous pLOF plus deleterious missense variants were associated with a lower TG:HDL ratio (per-allele $\beta = -0.13$ ratio units; 95% CI = −0.23 to −0.02; $P = 0.017$) and with an overall pattern of association with cardiometabolic traits that was similar in directionality, but weaker in magnitude compared with *FNIP1* variants (Supplementary Fig. 25). Heterozygous variants of *FLCN* cause autosomal dominant Birt–Hogg–Dubé syndrome (MIM 607273), an adult-onset disorder characterized by pulmonary cysts, pneumothorax and tumour predisposition, particularly renal[36]. Compared with non-carriers, heterozygous carriers of pLOF variants in *FLCN* had an 18-fold higher risk of pneumothorax and a 9-fold higher risk of kidney cancer and were at increased risk for a composite phenotype of Birt–Hogg–Dubé symptoms (Supplementary Figs. 23 and 26a, Supplementary Table 24 and Supplementary Note 3).

We did not observe associations with the TG:HDL ratio for pLOF variants in *FNIP2* (Supplementary Fig. 27), which encodes another member of the folliculin pathway with functions that have been hypothesized to be similar to those of FNIP1.

Mechanistically, the FNIP1–FLCN complex inhibits TFEB and TFE3, which are transcription factors that drive mitochondrial and lysosomal gene expression[37–40]. Accordingly, FNIP1–FLCN pathway inhibition promotes energy use and expenditure[35] (Extended Data Fig. 4a).

In this context, our human genetics findings suggest that heterozygous loss-of-function of *FNIP1* and, to a lesser degree, *FLCN* may induce mitochondrial biogenesis and energy expenditure in metabolically active tissues resulting in a favourable phenotype. Consistent with this hypothesis, previous literature has proposed a role for *Flcn* in mouse models of MASLD and liver inflammation, in which liver-specific knockdown has shown benefit[41,42], whereas *Fnip1* has not been studied in such models.

FNIP1 and *FLCN* are expressed in human and mouse hepatocytes (Supplementary Fig. 28), suggesting that at least part of the metabolic benefit observed in human variant carriers may be due to liver energy metabolism. This is important as there is precedent for the successful development and regulatory approval of *N*-acetylgalactosamine (GalNAc)-conjugated short interfering RNA (siRNA) therapeutics that effectively, durably and specifically silence a variety of genes selectively in hepatocytes in humans[29,30,43–45]. Thus, selective silencing of *FNIP1* in hepatocytes with siRNA may be a tractable approach for exploiting the favourable metabolic effects of variants in this pathway, while avoiding potential side effects of systemic loss-of-function.

To validate this hypothesis, we used siRNA to knockdown *FNIP1* in primary human hepatocytes. *FNIP1* silencing (>90% mRNA reduction) led to induction of lysosomal and lipid catabolism genes (Fig. 4b, Extended Data Fig. 5 and Supplementary Fig. 1), consistent with the reported biology of this pathway[35,46] and with the hypothesis that knockdown of *FNIP1* in human hepatocytes affects metabolic gene expression.

We next turned to in vivo mouse models to evaluate the impact on whole body energy homeostasis of FNIP–FLCN pathway inhibition in the liver. We used guide RNAs (gRNAs) delivered by adeno-associated virus serotype 8 (AAV8-gRNAs) to ablate *Flcn*, *Fnip1*, *Fnip2* or *Fnip1* together with *Fnip2* in the liver of *Cas9*-expressing mice challenged with a high-fat, high-fructose diet (HFHFD). DNA editing, mRNA and protein analysis confirmed the targeting approach (Extended Data Fig. 4b,c,e), which led to robust induction of mitochondrial and lysosomal genes in liver (Extended Data Fig. 4d and Extended Data Fig. 6a–c). Mice with hepatic *Flcn* deletion were substantially protected against weight and fat gain and had a higher proportion of lean mass (Fig. 4c,d and Extended Data Fig. 7a–e). Individually, *Fnip1* or *Fnip2* inhibition had no impact on weight gain, but combined *Fnip1* + *Fnip2* inhibition protected against diet-induced obesity similarly to *Flcn* inhibition (Fig. 4c,d and Extended Data Fig. 7a–e). The lack of phenotype in mouse with selective *Fnip1* liver inhibition probably reflects interspecies differences between mice and humans, given that (1) in our human genetics analysis, none of the 155 carriers of pLOF variants in *FNIP1* had any *FNIP2* pLOF variants, as

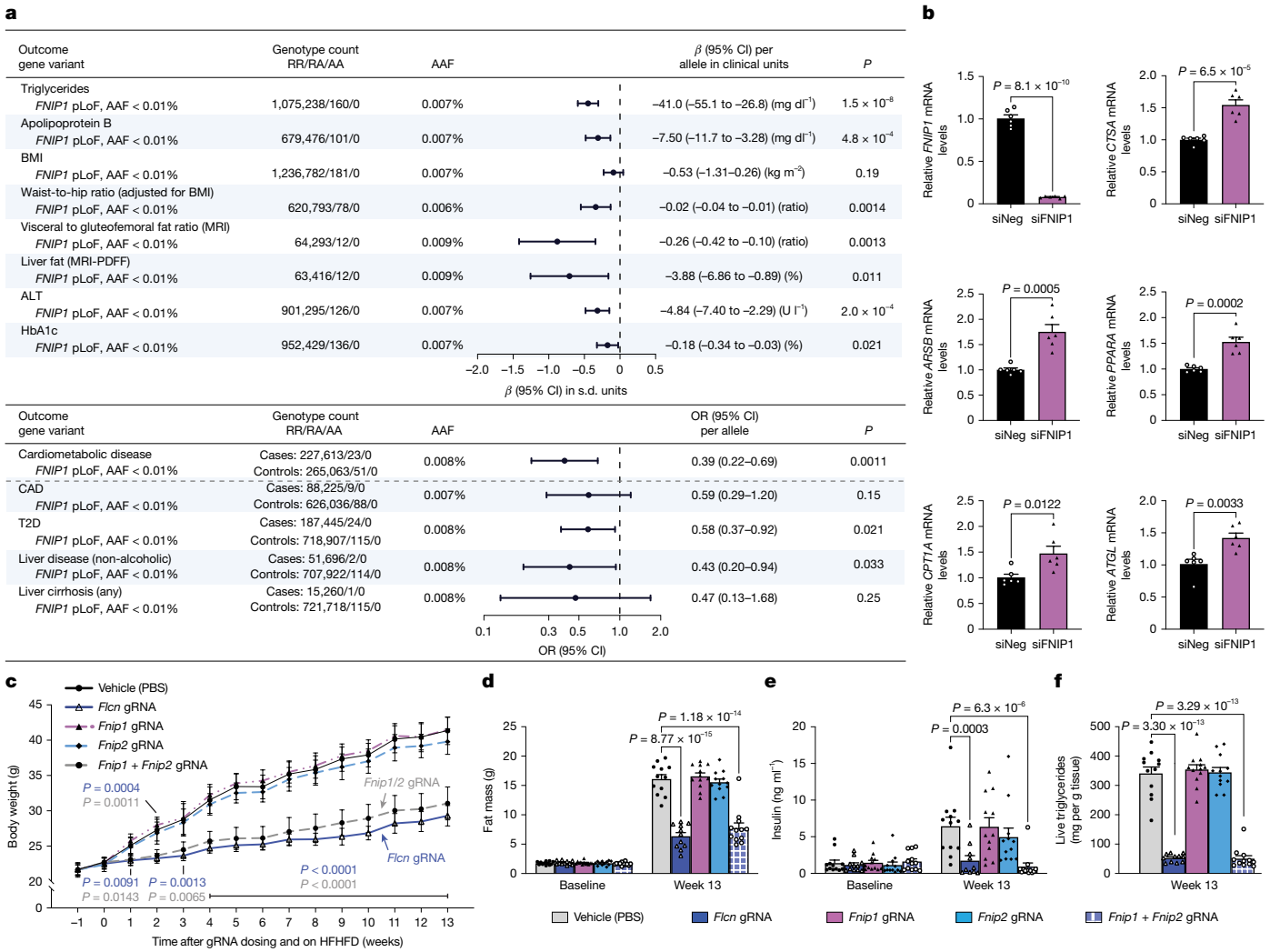
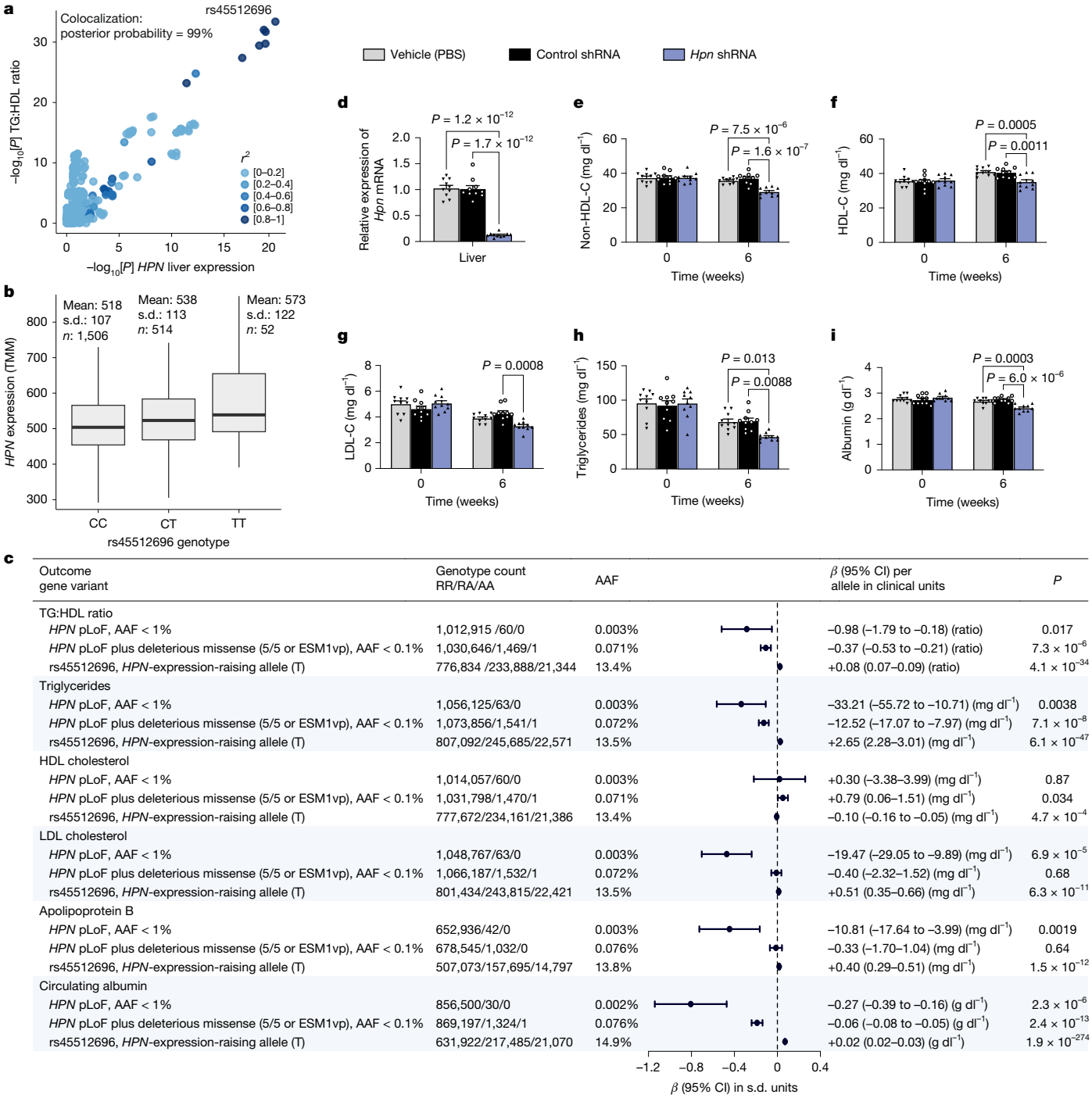


Fig. 4 | Associations with cardiometabolic traits for *FNIP1* pLOF variants in humans, in vitro knockdown of *FNIP1* in primary human hepatocytes and in vivo liver knockdown of *Fnip1*, *Fnip2* and *Flcn* in mice. a, Association of rare pLOF variants (AAF < 0.01%) in *FNIP1* with quantitative cardiometabolic traits ($\beta \pm 95\%$ CI) (top) and disease outcomes (OR ± 95% CI) (bottom). Association estimates are from inverse-variance weighted fixed-effect meta-analysis and *P* values were calculated using two-sided Wald tests. Cases of composite cardiometabolic disease had a diagnosis of CAD, type 2 diabetes or liver disease (non-alcoholic liver disease (MASLD) or cirrhosis); controls were controls for each component disease. **b**, mRNA levels of *FNIP1* and lysosomal/lipid catabolic genes in primary human hepatocytes after siRNA knockdown of *FNIP1* (siFNIP1) or non-targeting control (siNeg). $n = 6$ biological replicates per group. Data are mean ± s.e.m. *P* values were calculated using unpaired two-tailed *t*-tests. **c**–**f**, Cas9-Ready mice (male, 10 weeks of age) received AAV8-gRNA or vehicle (PBS) and were maintained on an HFHFD. Animal experiments (**c**–**f**) involved $n = 12$ (vehicle), 11 (*Flcn* gRNA), 12 (*Fnip1* gRNA), 12 (*Fnip2* gRNA) and 12 (*Fnip1* + *Fnip2* gRNAs) mice (biologically independent subjects). **c**, Mouse body weights. Data are mean ± 95% CI. **d**, Mouse fat mass was determined by echoMRI at the baseline and 13 weeks after administration. Data are mean ± s.e.m. **e**, The non-fasted plasma insulin was measured at the baseline and 13 weeks after administration. Data are mean ± s.e.m. In **c**–**e**, *P* values were calculated using two-way analysis of variance (ANOVA) with Dunnett's correction relative to the vehicle control. **f**, Hepatic triglyceride content after 13 weeks on the diet. Data are mean ± s.e.m. Statistical analysis was performed using one-way ANOVA with Dunnett's correction relative to the vehicle control. Exact *P* values are shown in the panels (**b**–**f**) and provided as source data (**c**, weeks 4–13). RR, homozygous reference; RA, heterozygous; AA, homozygous alternative.

expected from their genotype frequency (Supplementary Table 25), which means that the phenotype is solely due to *FNIP1* variants; (2) *FNIP2* variants did not show robust associations with metabolic traits; and (3) human hepatocyte experiments showed the induction of lipid catabolism and lysosomal genes with the specific knockdown of *FNIP1*.

Mice with liver *Flcn* or *Fnip1* + *Fnip2* inhibition exhibited improved insulin tolerance, lower circulating insulin, reduced liver triglyceride content and lower ALT levels on an HFHFD (Fig. 4e,f and Extended Data Fig. 7f,g). These mice also displayed metabolically beneficial gene expression profiles in adipose tissue, with higher transcript levels of adipose master regulator *Pparg* and its target genes, as well as higher transcript levels of lipolysis and oxidative phosphorylation genes[47] (Extended Data Fig. 8a–d). The adaptations in adipose tissue appeared to be secondary to changes in hepatic energy metabolism, given that our AAV8-gRNA targeting approach did not perturb the *Flcn*, *Fnip1* and *Fnip2* genes in non-hepatic tissues, including adipose (Extended Data Fig. 8a and Supplementary Fig. 29). Notably, hepatic *Inhbe* transcript levels were reduced by around 52–55% after *Flcn* or *Fnip1* + *Fnip2* inhibition (Extended Data Fig. 8e). Activin E, encoded by *Inhbe*, is a hepatokine that inhibits adipose lipolysis and regulates adipose depots through the ALK7 receptor[9,47], and similar levels of *Inhbe* reduction have previously been shown to elicit weight loss, fat loss and increased lipid oxidation in obese mice[48]. Thus, the favourable adipose gene expression signatures observed with hepatic FNIP–FLCN pathway inhibition may be partly mediated by a reduction in hepatic activin E expression.



c

Outcome gene variant	Genotype count RR/RA/AA	AAF	β (95% CI) per allele in clinical units	P
TG:HDL ratio				
HPN pLoF, AAF < 1%	1,012,915 /60/0	0.003%	−0.98 (−1.79 to −0.18) (ratio)	0.017
HPN pLoF plus deleterious missense (5/5 or ESM1vp), AAF < 0.1%	1,030,646/1,469/1	0.071%	−0.37 (−0.53 to −0.21) (ratio)	7.3×10^{-6}
rs45512696, *HPN*-expression-raising allele (T)	776,834 /233,888/21,344	13.4%	+0.08 (0.07–0.09) (ratio)	4.1×10^{-34}
Triglycerides				
HPN pLoF, AAF < 1%	1,056,125/63/0	0.003%	−33.21 (−55.72 to −10.71) (mg dl^{-1})	0.0038
HPN pLoF plus deleterious missense (5/5 or ESM1vp), AAF < 0.1%	1,073,856/1,541/1	0.072%	−12.52 (−17.07 to −7.97) (mg dl^{-1})	7.1×10^{-8}
rs45512696, *HPN*-expression-raising allele (T)	807,092/245,685/22,571	13.5%	+2.65 (2.28–3.01) (mg dl^{-1})	6.1×10^{-47}
HDL cholesterol				
HPN pLoF, AAF < 1%	1,014,057/60/0	0.003%	+0.30 (−3.38–3.99) (mg dl^{-1})	0.87
HPN pLoF plus deleterious missense (5/5 or ESM1vp), AAF < 0.1%	1,031,798/1,470/1	0.071%	+0.79 (0.06–1.51) (mg dl^{-1})	0.034
rs45512696, *HPN*-expression-raising allele (T)	777,672/234,161/21,386	13.4%	−0.10 (−0.16 to −0.05) (mg dl^{-1})	4.7×10^{-4}
LDL cholesterol				
HPN pLoF, AAF < 1%	1,048,767/63/0	0.003%	−19.47 (−29.05 to −9.89) (mg dl^{-1})	6.9×10^{-5}
HPN pLoF plus deleterious missense (5/5 or ESM1vp), AAF < 0.1%	1,066,187/1,532/1	0.072%	−0.40 (−2.32–1.52) (mg dl^{-1})	0.68
rs45512696, *HPN*-expression-raising allele (T)	801,434/243,815/22,421	13.5%	+0.51 (0.35–0.66) (mg dl^{-1})	6.3×10^{-11}
Apolipoprotein B				
HPN pLoF, AAF < 1%	652,936/42/0	0.003%	−10.81 (−17.64 to −3.99) (mg dl^{-1})	0.0019
HPN pLoF plus deleterious missense (5/5 or ESM1vp), AAF < 0.1%	678,545/1,032/0	0.076%	−0.33 (−1.70–1.04) (mg dl^{-1})	0.64
rs45512696, *HPN*-expression-raising allele (T)	507,073/157,695/14,797	13.8%	+0.40 (0.29–0.51) (mg dl^{-1})	1.5×10^{-12}
Circulating albumin				
HPN pLoF, AAF < 1%	856,500/30/0	0.002%	−0.27 (−0.39 to −0.16) (g dl^{-1})	2.3×10^{-6}
HPN pLoF plus deleterious missense (5/5 or ESM1vp), AAF < 0.1%	869,197/1,324/1	0.076%	−0.06 (−0.08 to −0.05) (g dl^{-1})	2.4×10^{-13}
rs45512696, *HPN*-expression-raising allele (T)	631,922/217,485/21,070	14.9%	+0.02 (0.02–0.03) (g dl^{-1})	1.9×10^{-274}

β (95% CI) in s.d. units

Fig. 5 | Genetic associations of *HPN* variants with cardiometabolic traits in humans and in vivo experimental results for *Hpn* liver knockdown in mice. **a**, Correlation of $-\log_{10}[P]$ for liver *HPN* gene expression and the TG:HDL ratio at the *HPN* locus. Linkage disequilibrium (r^2) with index rs45512696 variant is shown. **b**, *HPN* gene expression levels in liver tissue by rs45512696 genotype. The box plots show the median (centre line), 25th and 75th percentiles (box bounds) and the whiskers extend to the most extreme datapoints within 1.5× the interquartile range from the box bounds. TMM, trimmed mean of *M*-values. **c**, The association of rare pLOF variants (AAF < 1%), pLOF plus missense variants predicted to be deleterious in all five algorithms or by ESM1vp (AAF < 0.1%), or the common rs45512696 variant with blood lipid and albumin levels (β ± 95% CI). Association estimates are from inverse-variance weighted fixed-effect meta-analysis and *P* values were calculated using two-sided Wald tests. **d**, Liver mRNA levels of *Hpn* in AAV8-control shRNA, AAV8-Hpn shRNA or vehicle-treated (PBS) WT mice fed a standard diet. Data are mean ± s.e.m. $n = 10$ mice per group. *P* values were calculated using one-way ANOVA with Tukey correction post-test relative to the vehicle control or control shRNA. **e**–**h**, Non-fasted plasma lipids of WT mice before (week 0) and 6 weeks after AAV-shRNA or vehicle administration on standard diet. *Hpn* knockdown lowers plasma non HDL-C (**e**), HDL-C (**f**), LDL-C (**g**) and triglycerides (**h**). **i**, Non-fasted plasma albumin levels of WT mice before (week 0) and 6 weeks after AAV-shRNA or vehicle administration. For all plasma measurements (**e**–**i**), data are mean ± s.e.m. $n = 10$ mice per group representing biologically independent data points. *P* values were calculated using two-way ANOVA with Tukey correction post-test relative to the vehicle control or control shRNA. For **d**–**i**, exact *P* values are shown. ESM1vp, evolutionary scale modelling.

To validate that the energy metabolism changes observed in our in vivo experiments were due to specific gene targeting in hepatocytes as opposed to other liver cells, we used GalNAc-conjugated *Flcn* siRNAs and demonstrated that selective hepatocyte *Flcn* mRNA and protein knockdown was associated with a phenotype analogous to that observed in the AAV8-gRNA studies (Extended Data Fig. 9).

We further performed distinct series of AAV8-gRNAs experiments investigating liver *Flcn* or *Fnip1* + *Fnip2* inhibition at thermoneutrality or in a chow diet model (Supplementary Figs. 30 and 31). In experiments at thermoneutrality, we observed similar protective phenotypes in the treated mice as those observed in the ambient temperature conditions (Supplementary Fig. 30), indicating that enhanced energy metabolism, rather than cold stress, confers metabolic benefits on *Flcn* or *Fnip1* + *Fnip2* hepatic inhibition. In chow diet experiments, we observed that *Flcn* or *Fnip1* + *Fnip2* inhibition protected against fat mass gain and lowered liver triglycerides, although, as expected, the phenotype was less pronounced than that observed in the HFHFD challenge model (Supplementary Fig. 31).

Finally, to evaluate the long-term consequences of hepatic silencing, we performed a 30-week *Flcn* AAV-gRNA experiment in the setting of HFHFD, which demonstrated reduced liver lipids and markedly lower ALT levels, indicative of favourable metabolism and liver protection (Extended Data Fig. 10a–e). This was further supported by gene expression profiling demonstrating that long-term *Flcn* inhibition prevented the induction of fibrosis and liver damage/repair genes that was observed in control animals on an HFHFD (Extended Data Fig. 10f).

Liver hepsin in blood lipid metabolism

Given the enrichment for liver-expressed genes in our discovery analysis, we used liver RNA-sequencing (RNA-seq) and expression QTL (eQTL) data from a bariatric surgery cohort of 1,946 people for deeper mining of causative genes.

Of 1,617 independent common variant associations with TG:HDL ratio in our analysis, 151 (9%) at 150 genes co-localized with a sentinel liver eQTL (posterior probability > 0.9; Supplementary Table 26). Of the 150 genes implicated by the co-localized eQTLs, 9 (15%) were among the 59 exome-wide significant genes (a 21-fold enrichment compared with what is expected by chance; $P_{enrichment} = 1.4 \times 10^{-9}$), while an additional 9 genes met a more liberal Bonferroni correction for 150 liver eQTL-implicated genes (gene-burden *P* value greater than 1.04×10^{-7}, but smaller than 3.3×10^{-4}) (Supplementary Table 26).

Among the 18 genes with both gene-burden and common variant liver eQTL evidence, 15 showed directional concordance, whereby the liver-expression-lowering allele was associated with the TG:HDL ratio in the same direction as the rare deleterious coding variants (observed percentage, 83%; expected percentage under null hypothesis, 50%; $P_{binomial} = 0.0075$) (Supplementary Table 27), indicative of true signal enrichment.

Among these genes was hepsin (*HPN*), which encodes a hepatocyte-specific transmembrane serine protease that is known to curb energy expenditure and prevent adipocyte browning in mice, in part due to regulation of HGF/MET signalling[49,50] (Supplementary Figs. 32 and 33a,b). An eQTL for higher liver expression of hepsin co-localized with an association peak for higher TG:HDL ratio (rs45512696) (Fig. 5a,b, Supplementary Fig. 34 and Supplementary Table 26). Conversely, ultra-rare pLOF variants in *HPN* showed large-effect associations with lower levels of TG:HDL ratio, lower triglycerides, lower LDL-C and lower circulating albumin (Fig. 5c and Supplementary Fig. 35).

On the basis of these orthogonal lines of human genetics evidence, we hypothesized that reducing liver expression of hepsin might lead to lower atherogenic lipid levels and tested this hypothesis in mice. AAV8-delivered short-hairpin RNAs (AAV8-shRNAs) targeting *Hpn* achieved 87% mRNA knockdown in the liver and were associated with lower plasma cholesterol and triglycerides, but had no impact on body weight (Fig. 5d–h and Supplementary Fig. 33c,d). Furthermore, consistent with human genetics associations, *Hpn* knockdown led to reduced circulating albumin and total protein levels (Fig. 5i and Supplementary Fig. 33e).

Discussion

Our large epidemiological and genomic study of the TG:HDL ratio provides substantial evidence advancing understanding of the genetic basis of lipid and energy metabolism in humans.

Important insights from this analysis include (1) the validation of the TG:HDL ratio as a biomarker of body energy state and risk of cardiometabolic disease in the general population and across ancestries; (2) the identification of several dozen associations with the TG:HDL ratio for rare coding variants, greatly expanding the catalogue of genes and molecular players in lipid and energy metabolism in humans; (3) the description of ultra-rare coding variants in the FNIP1 pathway, showing large-effect protective associations with a broad range of cardiometabolic traits and diseases (atherogenic lipids, glycaemia, liver fat and damage, and cardiometabolic disease outcomes); and (4) the demonstration of convergence of rare coding and common variant associations, showing a central role of liver genes in TG:HDL-ratio biology, including the identification of rare variants in hepsin conferring a favourable lipidemic profile.

Our genetic findings demonstrate the importance of the FNIP1 pathway in human energy metabolism and in the risk of cardiometabolic diseases in the general population. Previous work[35,41,46] and the experimental data presented here suggest that the protective effects of *FNIP1* variants may be due to enhanced lipid breakdown, mitochondrial fatty acid oxidation and energy expenditure in metabolically active tissues. It is therefore possible that therapeutic modulation of FNIP1 might be beneficial in a range of cardiometabolic disease settings through enhanced fat breakdown and energy expenditure.

Given the precedent for the development of hepatocyte targeted siRNA therapeutics in humans[29,30,43–45], we specifically explored the possibility that liver inhibition of this pathway may induce a favourable metabolic profile. In human hepatocytes, silencing of *FNIP1* induced lipid breakdown and lysosomal genes and, across various mouse models, hepatic silencing of *Flcn* and of *Fnip1* + *Fnip2* induced a similar expression profile in liver and adipose tissues, with beneficial impacts on weight gain, circulating lipids, liver fat, insulin sensitivity and glycaemic control.

Conceptually, selective inhibition in hepatocytes may also help to avoid potential adverse phenotypes of systemic loss-of-function of this pathway, such as those observed in humans with heterozygous *FLCN* variants, homozygous *FNIP1* variants or, to a milder degree, heterozygous *FNIP1* variants. Liver safety is another important consideration and previous experiments in mouse models of metabolic liver disease have shown benefits for liver-specific inhibition of *Flcn*[41,42]. However, a recent study has shown susceptibility to liver injury and carcinogenesis for knockout of *Flcn* in hepatocytes and cholangiocytes in mice[51]. In our experiments, liver knockdown was associated with a protective liver phenotype, particularly in the HFHFD setting. In humans, liver cancer has not been reported as a clinical feature of homozygous *FNIP1* deficiency[31–34,52,53] or Birt–Hogg–Dubé syndrome[36] and, in our large dataset, *FNIP1* variants were associated with lower ALT, lower liver fat and protection from metabolic liver disease consistent with a favourable liver phenotype. Neither *FNIP1* nor *FLCN* heterozygous variants showed statistically significant associations with hepatic cancers. Besides the liver, previous literature on mouse models has suggested a role for the FNIP1–folliculin pathway on skeletal muscle and adipose tissue metabolism[54–59]. These findings are broadly consistent with results from our human genetics study and it is therefore possible that muscle- or adipose-specific inhibition of this pathway may also be beneficial to human metabolism.

Taken together, our data demonstrate the importance of the FNIP1 pathway in human energy metabolism. Future work across

experimental settings will ultimately clarify the full phenotypic landscape, as well as the potential for efficacy and safety of modulating this pathway in humans.

In conclusion, our human genetic analysis of the energy-state biomarker TG:HDL ratio in over 1 million people provides an illustration of the power of mega-scale exome sequencing and rare coding variant analysis to help map biologically important pathways in human disease, including the identification of potential therapeutic targets for common cardiometabolic diseases that collectively still represent the most common cause of death worldwide.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-10864-2.

1. Naghavi, M. et al. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **403**, 2100–2132 (2024).
2. Bouchard, C. et al. Genetic effect in resting and exercise metabolic rates. *Metabolism* **38**, 364–370 (1989).
3. Bouchard, C. et al. The response to long-term overfeeding in identical twins. *N. Engl. J. Med.* **322**, 1477–1482 (1990).
4. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
5. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
6. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
7. Lotta, L. A. et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).
8. Akbari, P. et al. Sequencing of 640,000 exomes identifies *GPR75* variants associated with protection from obesity. *Science* https://doi.org/10.1126/science.abf8683 (2021).
9. Akbari, P. et al. Multiancestry exome sequencing reveals *INHBE* mutations associated with favorable fat distribution and protection from diabetes. *Nat. Commun.* **13**, 4844 (2022).
10. Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V. & Roth, G. A. The global burden of cardiovascular diseases and risk: a compass for future health. *J. Am. Coll. Cardiol.* **80**, 2361–2371 (2022).
11. Oliveri, A. et al. Comprehensive genetic study of the insulin resistance marker TG:HDL-C in the UK Biobank. *Nat. Genet.* **56**, 212–221 (2024).
12. DeForest, N. et al. Genome-wide discovery and integrative genomic characterization of insulin resistance loci using serum triglycerides to HDL-cholesterol ratio as a proxy. *Nat. Commun.* **15**, 8068 (2024).
13. Salazar, M. R. et al. Comparison of the abilities of the plasma triglyceride/high-density lipoprotein cholesterol ratio and the metabolic syndrome to identify insulin resistance. *Diab. Vasc. Dis. Res.* **10**, 346–352 (2013).
14. Kosmas, C. E. et al. The triglyceride/high-density lipoprotein cholesterol (TG/HDL-C) ratio as a risk marker for metabolic syndrome and cardiovascular disease. *Diagnostics* https://doi.org/10.3390/diagnostics13050929 (2023).
15. Abul-Husn, N. S. et al. A protein-truncating *HSD17B13* variant and protection from chronic liver disease. *N. Engl. J. Med.* **378**, 1096–1106 (2018).
16. Verweij, N. et al. Germline mutations in *CIDEB* and protection against liver disease. *N. Engl. J. Med.* **387**, 332–344 (2022).
17. Nag, A. et al. Human genetics uncovers *MAP3K15* as an obesity-independent therapeutic target for diabetes. *Sci. Adv.* **8**, eadd5430 (2022).
18. Deaton, A. M. et al. Rare loss of function variants in the hepatokine gene *INHBE* protect from abdominal obesity. *Nat. Commun.* **13**, 4319 (2022).
19. Hindy, G. et al. Rare coding variants in 35 genes associate with circulating lipid levels—a multi-ancestry analysis of 170,000 exomes. *Am. J. Hum. Genet.* **109**, 81–96 (2022).
20. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
21. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr & Hobbs, H. H. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
22. Musunuru, K. et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **363**, 2220–2227 (2010).
23. Dewey, F. E. et al. Genetic and pharmacologic inactivation of ANGPTL3 and cardiovascular disease. *N. Engl. J. Med.* **377**, 211–221 (2017).
24. Pollin, T. I. et al. A null mutation in human *APOC3* confers a favorable plasma lipid profile and apparent cardioprotection. *Science* **322**, 1702–1705 (2008).
25. Crosby, J. et al. Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. *N. Engl. J. Med.* **371**, 22–31 (2014).
26. Jorgensen, A. B., Frikke-Schmidt, R., Nordestgaard, B. G. & Tybjaerg-Hansen, A. Loss-of-function mutations in *APOC3* and risk of ischemic vascular disease. *N. Engl. J. Med.* **371**, 32–41 (2014).
27. Schwartz, G. G. et al. Alirocumab and cardiovascular outcomes after acute coronary syndrome. *N. Engl. J. Med.* **379**, 2097–2107 (2018).
28. Raal, F. J. et al. Evinacumab for homozygous familial hypercholesterolemia. *N. Engl. J. Med.* **383**, 711–720 (2020).
29. Watts, G. F. et al. Plozasiran for managing persistent chylomicronemia and pancreatitis risk. *N. Engl. J. Med.* **392**, 127–137 (2025).
30. Ray, K. K. et al. Two phase 3 trials of inclisiran in patients with elevated LDL cholesterol. *N. Engl. J. Med.* **382**, 1507–1519 (2020).
31. Saettini, F. et al. Absent B cells, agammaglobulinemia, and hypertrophic cardiomyopathy in folliculin-interacting protein 1 deficiency. *Blood* **137**, 493–499 (2021).
32. Moreno-Corona, N., Valagussa, A., Thouenon, R., Fischer, A. & Kracker, S. A case report of folliculin-interacting protein 1 deficiency. *J. Clin. Immunol.* **43**, 1751–1753 (2023).
33. Niehues, T. et al. Mutations of the gene *FNIP1* associated with a syndromic autosomal recessive immunodeficiency with cardiomyopathy and pre-excitation syndrome. *Eur. J. Immunol.* **50**, 1078–1080 (2020).
34. Roncareggi, S., Iritani, B. M. & Saettini, F. FNIP1 deficiency: pathophysiology and clinical manifestations of a rare syndromic primary immunodeficiency. *Curr. Issues Mol. Biol.* https://doi.org/10.3390/cimb47040290 (2025).
35. Malik, N. et al. Induction of lysosomal and mitochondrial biogenesis by AMPK phosphorylation of FNIP1. *Science* **380**, eabj5559 (2023).
36. Sattler, E. C. & Steinlein, O. K. in *GeneReviews* (eds Adam, M. P. et al.) https://www.ncbi.nlm.nih.gov/books/NBK1522/ (Univ. Washington, 1993).
37. Tsun, Z. Y. et al. The folliculin tumor suppressor is a GAP for the RagC/D GTPases that signal amino acid levels to mTORC1. *Mol. Cell* **52**, 495–505 (2013).
38. Lawrence, R. E. et al. Structural mechanism of a Rag GTPase activation checkpoint by the lysosomal folliculin complex. *Science* **366**, 971–977 (2019).
39. Shen, K. et al. Cryo-EM structure of the human FLCN-FNIP2-Rag-Ragulator complex. *Cell* **179**, 1319–1329 (2019).
40. Napolitano, G. et al. A substrate-specific mTORC1 pathway underlies Birt–Hogg–Dubé syndrome. *Nature* **585**, 597–602 (2020).
41. Gosis, B. S. et al. Inhibition of nonalcoholic fatty liver disease in mice by selective inhibition of mTORC1. *Science* **376**, eabf8271 (2022).
42. Paquette, M. et al. Loss of hepatic *Flcn* protects against fibrosis and inflammation by activating autophagy pathways. *Sci. Rep.* **11**, 21268 (2021).
43. Garrelfs, S. F. et al. Lumasiran, an RNAi therapeutic for primary hyperoxaluria type 1. *N. Engl. J. Med.* **384**, 1216–1226 (2021).
44. Balwani, M. et al. Phase 3 trial of RNAi therapeutic givosiran for acute intermittent porphyria. *N. Engl. J. Med.* **382**, 2289–2301 (2020).
45. Adams, D. et al. Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *N. Engl. J. Med.* **379**, 11–21 (2018).
46. Li, J. et al. Myeloid Folliculin balances mTOR activation to maintain innate immunity homeostasis. *JCI Insight* **5**, https://doi.org/10.1172/jci.insight.126939 (2019).
47. Adam, R. C. et al. Activin E-ACVR1C cross talk controls energy storage via suppression of adipose lipolysis in mice. *Proc. Natl Acad. Sci. USA* **120**, e2309967120 (2023).
48. Sugiyama, M. et al. Inhibin betaE (INHBE) is a possible insulin resistance-associated hepatokine identified by comprehensive gene expression analysis in human liver biopsy samples. *PLoS ONE* **13**, e0194798 (2018).
49. Li, S. et al. Hepsin enhances liver metabolism and inhibits adipocyte browning in mice. *Proc. Natl Acad. Sci. USA* **117**, 12359–12367 (2020).
50. Kirchhofer, D. et al. Hepsin activates pro-hepatocyte growth factor and is inhibited by hepatocyte growth factor activator inhibitor-1B (HAI-1B) and HAI-2. *FEBS Lett.* **579**, 1945–1950 (2005).
51. Custode, B. M. et al. Folliculin depletion results in liver cell damage and cholangiocarcinoma through MiT/TFE activation. *Cell Death Differ.* **32**, 1460–1472 (2025).
52. Ulas, S. et al. Clinical and immunologic features of a patient with homozygous *FNIP1* variant. *J Pediatr. Hematol. Oncol.* **46**, e472–e475 (2024).
53. Spivak, I. et al. A novel mutation in *FNIP1* associated with a syndromic immunodeficiency and cardiomyopathy. *Immunogenetics* **77**, 2 (2024).
54. Reyes, N. L. et al. Fnip1 regulates skeletal muscle fiber type specification, fatigue resistance, and susceptibility to muscular dystrophy. *Proc. Natl Acad. Sci. USA* **112**, 424–429 (2015).
55. Xiao, L. et al. AMPK phosphorylation of FNIP1 (S220) controls mitochondrial function and muscle fuel utilization during exercise. *Sci. Adv.* **10**, eadj2752 (2024).
56. Hasumi, H. et al. Regulation of mitochondrial oxidative metabolism by tumor suppressor FLCN. *J. Natl Cancer Inst.* **104**, 1750–1764 (2012).
57. Yan, M. et al. Chronic AMPK activation via loss of FLCN induces functional beige adipose tissue through PGC-1α/ERRα. *Genes Dev.* **30**, 1034–1046 (2016).
58. Wada, S. et al. The tumor suppressor FLCN mediates an alternate mTOR pathway to regulate browning of adipose tissue. *Genes Dev.* **30**, 2551–2564 (2016).
59. Yin, Y. et al. FNIP1 regulates adipocyte browning and systemic glucose homeostasis in mice by shaping intracellular calcium dynamics. *J. Exp. Med.* https://doi.org/10.1084/jem.20212491 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

[1]Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY, USA. [2]Regeneron Pharmaceuticals, Tarrytown, NY, USA. [3]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [4]Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden. [5]Department of Internal Medicine, Skåne University Hospital, Malmö, Sweden. [6]Geisinger Obesity Institute, Geisinger Health System, Danville, PA, USA. [7]Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico. [8]Proyecto OriGen, Tecnológico de Monterrey, Monterrey, Mexico. [9]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [26]These authors contributed equally: George Hindy, Rene C. Adam. [27]These authors jointly supervised this work: Viktoria Gusarova, Luca A. Lotta.

Regeneron Genetics Center

George Hindy[1,26], Olukayode Sosina[1], David Blair[1], Joseph Herman[1], Peter Dornbos[1], Ernst Mayerhofer[1], Arthur Gilly[1], Benjamin Geraghty[1], Karl Landheer[1], Liron Ganel[1], Antoine Baldassari[1], Chuanyi Zhang[1], Andrew Deubler[1], Adolfo Ferrando[1], Giovanni Coppola[1], Alan Shuldiner[1], Katherine Siminovitch[1], Jason Portnoy[1], Lyndon Mitnaul[1], Alison Fenney[1], Manuel Allen Revez Ferreira[1], Maya Ghoussaini[1], Mona Nafde[1], William Salerno[1], Lourdes Crane[1], Christina Beechert[1], Erin Fuller[1], Laura M. Cremona[1], Eugene Kalyuskin[1], Hang Du[1], Caitlin Forsythe[1], Zhenhua Gu[1], Kristy Guevara[1], Michael Lattari[1], Alexander Lopez[1], Kia Manoochehri[1], Prathyusha Challa[1], Manasi Pradhan[1], Raymond Reynoso[1], Ricardo Schiavo[1], Maria Sotiropoulos Padilla[1], Chenggu Wang[1], Sarah E. Wolf[1], Manan Goyal[1], George Mitra[1], Sanjay Sreeram[1], Rouel Lanche[1], Vrushali Mahajan[1], Sai Lakshmi Vasireddy[1], Gisu Eom[1], Krishna Pawan Punuru[1], Sujit Gokhale[1], Benjamin Sultan[1], Pooja Mule[1], Mudasar Sarwar[1], Muhammad Aqeel[1], Xiaodong Bai[1], Lance Zhang[1], Sean O'Keeffe[1], Razvan Panea[1], Evan Edelstein[1], Ayesha Rasool[1], Evan K. Maxwell[1], Boris Boutkov[1], Alexander Gorovits[1], Ju Guan[1], Lukas Habegger[1], Alicia Hawes[1], Olga Krasheninina[1], Samantha Zarate[1], Adam J. Mansfield[1], Joshua Backman[1], Kathy Burch[1], Adrian Campos[1], Sheila Gaynor[1], Arkopravo Ghosh[1], Salvador Romero Martinez[1], Christopher Gillies[1], Lauren Gurski[1], Eric Jorgenson[1], Tyler Joseph[1], Michael Kessler[1], Jack Kosmicki[1], Priyanka Nakka[1], Olivier Delaneau[1], Anthony Marcketta[1], Joelle Mbatchou[1], Arden Moscati[1], Anita Pandit[1], Jonathan Ross[1], Carlo Sidore[1], Eli Stahl[1], Timothy Thornton[1], Sailaja Vedantam[1], Rujin Wang[1], Kuan-Han Wu[1], Bin Ye[1], Blair Zhang[1], Andrey Ziyatdinov[1], Yuxin Zou[1], Jingning Zhang[1], Kyoko Watanabe[1], Mira Tang[1], Frank Wendt[1], Suying Bao[1], Kathie Sun[1], Sean Yu[1], Aaron Zhang[1], David Corrigan[1], Dhruv Shidhaye[1], Chen Wang[1], Keyrun Adhikari[1], Alexander Lachmann[1], Mark Weiner[1], Brian Hobbs[1], Jon Silver[1], William Palmer[1], Rita Guerreiro[1], Amit Joshi[1], Sarah Graham[1], Erola Pairo Castineira[1], Tanima De[1], Luanluan Sun[1], Mary Haas[1], Juan Rodriguez-Flores[1], Moeen Riaz[1], Manav Kapoor[1], Gannie Tzoneva[1], Momodou W. Jallow[1], Anna Alkelai[1], Ariane Ayer[1], Veera Rajagopal[1], Sahar Gelfman[1], Vijay Kumar[1], Jacqueline Otto[1], Jose Bras[1], Silvia Alvarez[1], Jessie Brown[1], Hossein Khiabanian[1], Joana Revez[1], Kimberly Skead[1], Valentina Zavala[1], Jae Soon Sul[1], Lei Chen[1], Sam Choi[1], Amy Damask[1], Nan Lin[1], Charles Paulding[1], Sameer Malhotra[1], Nadia Rana[1], Jennifer Rico-Varela[1], Jaimee Hernandez[1], Larizbeth Romero[1], Ashley Paynter[1], Randi Schwartz[1], Jody Hankins[1], Anna Han[1], Samuel Hart[1], Ryan Smith[1], Ann Perez-Beals[1], Gina Solari[1], Johannie Rivera-Picart[1], Michelle Pagan[1], Sunilbe Siceron[1], Suganthi Balasubramanian[1], Marcus B. Jones[1], Michelle G. LeBlanc[1], John D. Overton[1], Jeffrey G. Reid[1], Goncalo R. Abecasis[1], Jonathan Marchini[1], Cristen Willer[1], Jonas Bovijn[1], Adam Locke[1], Aris Baras[1], Niek Verweij[1] & Luca A. Lotta[1,27]

GHS-RGC DiscovEHR Collaboration
Adam Buchanan[10], David J. Carey[10], Christa L. Martin[10], Michelle Meyer[10], Kyle Retterer[10] & David Rolston[10]

[10]Geisinger Health System, Danville, PA, USA.

Penn Medicine Biobank
Daniel J. Rader[3], Marylyn D. Ritchie[11], JoEllen Weaver[11], Nawar Naseer[11], Giorgio Sirugo[11], Afiya Poindexter[11], Yi-An Ko[11], Kyle P. Nerz[11], Jenna Dever[11], Aidan Harvey[11], Sydney Linn[11], Meghan Livingstone[11], Fred Vadivieso[11], Stephanie DerOhannessian[11], Teo Tran[11], Julia Stephanowski[11], Salma Santos[11], Ned Haubein[11], Joseph Dunn[11], Anurag Verma[11], Colleen Morse Kripke[11], Marjorie Risman[11], Renae Judy[11], Colin Wollack[11], Shefali S. Verma[11], Scott Damrauer[11], Yuki Bradford[11], Scott Dudek[11] & Theodore Drivas[11]

[11]University of Pennsylvania, Philadelphia, PA, USA.

Colorado Center for Personalized Medicine
Heather D. Anderson[12], Christina L. Aquilante[12], Kelsey Arbogast[12], Ian M. Brooks[12], Elizabeth E. Burke[12], Emily M. Casteel[12], Joanne B. Cole[12], Curtis R. Coughlin II[12], Jacob Crawford[12], Kristy Crooks[12], Erin Culver[12], Matthew J. Fisher[12], Teresa C. Frye[12], Hunter George[12], Chris R. Gignoux[12], Elizabeth K. Gilliland[12], Casey S. Greene[12], Emily Hearst[12], Audrey E. Hendricks[12], Randi K. Johnson[12], Shelby Jones[12], Dave Kao[12], Gabrielle A. Knortz[12], Danielle Koffenberger[12], Santhanagopalan Krishnamoorthy[12], Lisa Ku[12], Elizabeth L. Kudron[12], Rashawnda Lacy[12], Ethan M. Lange[12], Joe A. Lesny[12], Meng Lin[12], James L. Martin[12], Adrian M. Stewart[12], Alaa Radwan[12], Anna Tanaka[12], Carolina Sanchez-Wild[12], Carolyn T. Swartz[12], Elise L. Shalowitz[12], Emily B. Todd[12], Hoda Sherif[12], Jack Pattee[12], Jonathan A. Shortt[12], Katy E. Trinkley[12], Laura K. Wiley[12], Neda Rasouli[12], Nick Rafaels[12], Nicole L. McDaniel[12], Nikita Pozdeyev[12], Sridharan Raghavan[12] & Vendant Vohra[12]

[12]University of Colorado Anschutz Medical Campus, Aurora, CO, USA.

Mayo Clinic-RGC Project Generation
James R. Cerhan[13], Fergus J. Couch[13], Janet E. Olson[13], Nicholas B. Larson[13], Zachary S. Fredericksen[13], Mine Cicek[13], Joanna M. Biernacka[13], Victor M. Karpyak[13], Prashanthi Vemuri[13], Vijay K. Ramanan[13], Mark A. Frye[13], Sarah A. McLaughlin[13], Kathryn J. Ruddy[13], Suzette J. Bielinski[13], John C. Lieske[13], W. Michael Hooten[13], Lisa A. Boardman[13], Richard B. Kennedy[13], Andrew D. Badley[13], Sean C. Dowdy[13], Jamie N. Bakkum-Gamez[13], Gretchen E. Glaser[13], Ping Yang[13], Celine M. Vachon[13], Angela Dispenzieri[13], Samuel O. Antwi[13], Ann L. Oberg[13], Kari G. Rabe[13], Scott H. Kaufmann[13], Ellen L. Goode[13], William A. Cliby[13], J. Eric Ahlskog[13], James H. Bower[13], Peter C. Harris[13], Naveen L. Pereira[13], Daniel J. Ma[13], Robert W. Mutter[13], Jeanette E. Eckel-Passow[13], Lisa K. Colborn[13], Andrew J. Danielsen[13], Jonathan J. Harrington[13] & Jennifer M. Kushwaha[13]

[13]Mayo Clinic, Rochester, MN, USA.

Mexico City Prospective Study
Jaime Berumen[7], Pablo Kuri-Morales[8], Jesus Alegre-Díaz[7], Jason M. Torres[9], Jonathan R. Emberson[9], Rory Collins[9] & Roberto Tapia-Conyer[7]

Malmö Diet and Cancer Study
Olle Melander[4,5]

Indiana University School of Medicine (Indiana-CLDB)
Naga P. Chalasani[14], Tae-Hwi L. Schwantes-An[14] & Andrew J. Saykin[14]

[14]Indiana University School of Medicine, Indianapolis, IN, USA.

UCLA-RGC ATLAS Collaboration
Alex Bui[15], Antonia Petruse[15], Arash Naeim[15], Chris Denny[15], Clara Lajonchere[15], Clara Magyar[15], Dan Geschwind[15], Maryam Ariannejad[15], Paul Boutros[15], Paul Spellman[15], Sarah Dry[15], Stan Nelson[15], Albert Duntugan[15], Bogdan Pasaniuc[15], Paul Tung[15], Tim Chang[15], Yael Berkovich[15], Ankur Jain[15], Danielle Martinez[15], Ghouse Mohammed[15], Lora Illiev[15], Michael Broudy[15], Roni Haas[15], Shaiful Alam[15], Taka Yamaguchi[15] & Yash Patel[15]

[15]University of California, Los Angeles, Los Angeles, CA, USA.

Mount Sinai Million Health Discoveries Program
Alexander W. Charney[16]

[16]Icahn School of Medicine at Mount Sinai, New York, NY, USA.

UTSW-Dallas Biobank
Julia Kozlitina[17] & Jonathan Cohen[17]

[17]University of Texas Southwestern Medical Center, Dallas, TX, USA.

South Asia Biobank Study
Marie Loh[18], Prasad Katulanda[19], Vinitaa Jha[20], Anuradhani Kasturiratne[21], Khadija Irfan Khawaja[22], Malay Kanti Mridha[23], Ranjit Mohan Anjana[24] & John C. Chambers[18,25]

[18]Nanyang Technological University, Singapore, Singapore. [19]University of Colombo, Colombo, Sri Lanka. [20]Devki Devi Foundation, New Delhi, India. [21]University of Kelaniya, Ragama, Sri Lanka. [22]Services Institute of Medical Sciences, Lahore, Pakistan. [23]BRAC University, Dhaka, Bangladesh. [24]Madras Diabetes Research Foundation, Chennai, India. [25]Imperial College London, London, UK.

Methods

Study population

The genetic discovery analysis of TG:HDL ratio encompassed multiple cohorts with a cumulative sample size of 1,032,116 participants. Study participants were recruited from various population-based and hospital-based health systems across three continents (Fig. 1 and Supplementary Table 1). These cohorts included: the UK Biobank population-based study (UKB, $n = 409{,}602$)[60]; the Geisinger Health System MyCode cohort (GHS, $n = 143{,}129$)[61]; the Mexico City Prospective Study (MCPS, $n = 136{,}725$), a study of residents in two districts of Mexico City (Coyoacán and Iztapalapa)[62] with a genetically admixed population with components of Indigenous American, European and African ancestry as described in detail previously[63]; the Mayo Clinic Project Generation health system-based cohort (MAYO-RGC, $n = 80{,}941$)[64]; the population-based BangladEsh Longitudinal Investigation of Emerging Vascular and nonvascular Events (BELIEVE, $n = 69{,}663$)[65]; the Colorado Center for Personalized Medicine biobank, a health system-based cohort (CCPM, $n = 56{,}173$)[66]; the UCLA ATLAS Precision Health Biobank, a health system-based cohort (ATLAS, $n = 45{,}533$)[67]; the Mount Sinai BioMe BioBank, a health system-based cohort (BioMe, $n = 45{,}440$)[68]; the University of Pennsylvania Penn Medicine BioBank, a health system-based cohort (PMBB, $n = 25{,}785$)[69]; the population-based Dallas Biobank Study at UT Southwestern Medical Center (DBS, $n = 13{,}979$)[70]; and the population-based Malmö Diet and Cancer Study (MDCS, $n = 5{,}146$)[71]. Analyses of other cardiometabolic phenotypes included participants from the aforementioned cohorts and also 4,755 participants from the Indiana University School of Medicine (Indiana-CLDB) study and 53,325 participants from the South Asia Biobank[72]. All studies received approval from the relevant ethics committees as listed below, and the participants gave their informed consent to take part in the research. UKB cohort: ethical approval for the UKB was obtained from the North West Centre for Research Ethics Committee (11/NW/0382). The work described here was approved by the UKB under application number 26041. GHS cohort: The MyCode Community Health Initiative was approved by the Geisinger institutional review board (IRB; study 2006-0258). MCPS cohort: the study was approved by scientific and ethics committees within the Mexican National Council of Science and Technology (0595 P-M), the Mexican Ministry of Health and the Central Oxford Research Ethics Committee (C99.260). MAYO-RGC cohort: the study was approved by the Mayo Clinic IRB under protocol number 09-007763. BELIEVE cohort: BELIEVE has received approvals from the relevant institutional review boards of the Bangladesh Medical Research Council, the National Heart Foundation Hospital and Research Institute, icddr,b and Bangabandhu Sheikh Mujib Medical University (BMRC/NREC/2013- 2016/390; BMRC/NREC/2016-2019/243; BSMMU/2019/1184; BSMMU/2019/1185; PR-18051; HBREC.2019.09). CCPM cohort: biospecimens and associated data used in this study were obtained from the biobank at the Colorado Center for Personalized Medicine (CCPM) at the University of Colorado Anschutz Medical Campus (CU AMC). All samples and data were collected under IRB approved protocol (15-0461). ATLAS cohort: ATLAS is an approved study by the University of California Los Angeles (UCLA) IRB (UCLA IRB 17-001013). BioMe/Mount Sinai Million Health Discoveries Program: The Icahn School of Medicine at Mount Sinai's IRB, Program for the Protection of Human Subjects (PPHS), approved the BioMe Biobank and Mount Sinai Million Health Discoveries Program (PPHS IRB 11-01139 and 21-01743). PMBB cohort: the PMBB is approved by the University of Pennsylvania IRB under protocol number 813913. DBS cohort: The University of Texas Southwestern (UTSW) Biobank was reviewed and approved by the UTSW IRB under the study STU 022011-116: "Genetic Variation in a Multi-Ethnic Population". MDCS cohort: IRB approval for MDCS was obtained from the Ethics committee of Lund University (LU 51-90) and the Regional Board of Ethics in Lund (Dnr 2016/479). Indiana-CLDB: the study has been reviewed by Indiana University IRB and approved under the protocol number 1105005445. South Asia Biobank: the study received ethical approval from the institutional Research Ethics Committee (18IC4698).

Clinical phenotypes

Clinical biomarkers. Measurements of blood lipids, glycaemic markers, transaminases, CRP and blood pressure were obtained either through direct clinical assessments or electronic health record (EHR) extraction, depending on the cohort type. For participants in population-based cohort studies, these measurements were conducted during standardized visits to designated assessment centres. By contrast, for participants from health-system-based cohorts, data were extracted retrospectively from EHR datasets and participant-specific median values were used. Blood lipid levels were primarily assessed using routine clinical chemistry methods across most cohorts. In the MCPS and BELIEVE studies, lipid measurements were derived from nuclear magnetic resonance metabolomics profiling, performed using the Nightingale Health platform[73]. Blood lipids were corrected for medication use by dividing by a class-specific correction factor obtained from previously reported clinical trials[19,74–78]. The TG:HDL ratio was calculated using triglyceride and HDL cholesterol values obtained from the same blood draw at baseline in epidemiological cohort studies (5 cohorts and 635,115 participants). In EHR-based studies (6 cohorts and 397,001 participants), where multiple measurements were available, median values across clinical encounters were calculated for TG and HDL separately and then the ratio of the median values was calculated. The use of median values across encounters is a standard approach in EHR-based genetic research to define a patient's typical lipid profile[79–81]. Importantly, this avoids the exclusion of participants with valid but non-concurrent lipid measurements, thereby preserving sample size and discovery power. A sensitivity analysis was conducted restricting the TG:HDL ratio to values derived from TG and HDL measured concurrently from the same blood draw. A stratified analysis by fasting status was also conducted in the UKB, comparing TG:HDL ratios derived from fasting samples (defined as at least 6 h since the last meal) with those from non-fasting samples (defined as at most 2 h since the last meal). Blood pressure levels were corrected by adding 15 mmHg to systolic and 10 mmHg to diastolic blood pressure in those individuals receiving anti-hypertensive treatment as previously reported[82].

Liver biopsy. The bariatric surgery cohort at Geisinger Health System (GHS) comprises 3,599 participants of European ancestry who underwent bariatric surgery procedures[61]. Intraoperative wedge liver biopsies were systematically obtained 10 cm left of the falciform ligament before any manipulation of the liver or stomach. Biopsies were sectioned, with the principal portion allocated for clinical histopathology—fixed in 10% neutral buffered formalin and stained with haematoxylin and eosin for general assessment and Masson's trichrome for fibrosis evaluation. The remaining tissue was preserved in a research biobank, either stabilized in the RNAlater tissue collection system (Thermo Fisher Scientific) or snap-frozen in liquid nitrogen. Histopathological evaluation was conducted by an experienced pathologist and independently reviewed by a second pathologist. Scoring followed the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) system[83], with steatosis graded as 0 (<5% parenchymal involvement), 1 (5 to <34%), 2 (34 to <67%) or 3 (>67%); lobular inflammation as 0 (none), 1 (mild, <2 foci per ×200 field), 2 (moderate, 2–4 foci per ×200 field) or 3 (severe, >4 foci per ×200 field); hepatocyte ballooning as 0 (none), 1 (few ballooned cells), or 2 (many/prominent ballooned cells); and fibrosis staged from 0 (none) to 4 (cirrhosis) according to established criteria. The MASLD activity score was generated by summing of the steatosis, lobular inflammation and ballooning scores.

Imaging phenotypes

MRI. MRI images were obtained from the UKB imaging cohort. For quantification of liver fat, we used data from a two-dimensional abdominal

MRI scan that included the liver; a subset of individuals was scanned using a Dixon gradient echo protocol, while those imaged from 2016 onward were scanned using the IDEAL (iterative decomposition of water and fat with echo asymmetry and least-squares estimation) protocol. The resulting data have an in-plane pixel size of 2.5 × 2.5 mm and a slice thickness of 6 mm.

Whole-body MRI scans, spanning from the neck to the knees, were obtained through a two-point Dixon[84] spoiled gradient-echo (FLASH) T1-weighted protocol with echo times of 2.39 ms (out of phase) and 4.77 ms (in phase), a repetition time of 6.69 ms and an excitation flip angle of 10 degrees[85]. Image acquisition was split across six distinct anatomical stages, each representing a separate 3D volume, with an in-plane spatial resolution of 2.23 mm × 2.23 mm and an out-of-plane resolution ranging from 3.00 mm to 4.50 mm according to the specific stage.

All images were captured using Siemens MAGNETOM clinical MRI scanners (Siemens Healthineers).

MRI-derived phenotypes. To assess total body adipose tissue volumes, we used an automatic segmentation model developed to delineate major adipose depots from Dixon MRI data into visceral and subcutaneous compartments[86]; visceral fat was further subdivided into mediastinal and abdominal components, and subcutaneous fat into the upper-body, abdominal and gluteofemoral components. The volume of fat within these regions was computed by summing the fat fractions for voxels within the segmented regions. The visceral-to-gluteofemoral fat ratio—a marker associated with poor metabolic outcomes[9]—was computed as the abdominal visceral fat volume divided by the subcutaneous gluteofemoral fat volume; the abdominal-to-gluteofemoral subcutaneous fat ratio was similarly computed using the respective total volumes. All data processing and phenotype extraction was conducted using the UKB Research Analysis Platform.

The liver fat percentage, measured as proton density liver fat fraction (PDFF), represents the proportion of fat content in the liver. A deterministic image processing algorithm was used to segment the liver on MRI images, using a multithresholding approach to exclude vessels. The methodology was validated using a publicly available phantom dataset containing vials with varying fat concentrations. PDFF within the segmented region was determined as the ratio of fat signal to the total fat and water signals[16,87]. To account for any systematic differences between PDFF estimates from gradient echo and IDEAL, the mean values were shifted to be equal for the two modalities, and an indicator variable encoding the imaging modality was included as an additional covariate for genetic analyses to account for any residual differences.

BIA. BIA for body composition was performed at the baseline using the Tanita BC 418ma Body Fat Analyzer. Fat and fat-free mass estimates derived from BIA were obtained from UKB fields 23100 and 23101, respectively; these are defined such that fat-free mass is equal to total body weight minus fat mass. The body fat percentage was calculated using the BIA-derived fat mass and baseline weight (UKB field 23098).

DEXA. DEXA measures were obtained using the GE-Lunar iDXA instrument. DEXA-derived measures of lean mass and android/gynoid fat mass were obtained from the UKB (UKB fields 23280, 23245 and 23262 respectively).

Disease definitions

Type 2 diabetes. Individuals diagnosed with type 2 diabetes were ascertained according to established methodologies[9]. Case identification was based on the presence of (1) at least one inpatient or two outpatient EHR entries with type 2 diabetes diagnosis codes (ICD-10: E11, O24.1; or ICD-9 equivalents), or documentation as a cause of death; (2) glycaemic biomarker measurements (HbA1c, random or fasting glucose) consistent with the diabetic range[88] or algorithmically defined type 2 diabetes status derived from self-reported medical history and medication use[89]. Participants were excluded from cases if they met criteria for type 1 diabetes, indicated by ICD-10 codes E10, O24.0 or algorithmic classification based on self-reported and medication data[89]. Control participants were those not fulfilling type 2 diabetes criteria, with further exclusions applied for any EHR evidence of diabetes (any type), family history of diabetes or glycaemic biomarkers within the prediabetic range.

Liver disease. Cases of liver disease, encompassing MASLD and liver cirrhosis, were identified according to previously reported criteria[16]. Case inclusion required fulfilment of at least one of the following: (1) documentation of liver disease in EHRs from at least one inpatient encounter, two or more outpatient encounters or as a recorded cause of death; (2) self-reported diagnosis at the baseline; or (3) documented history of surgical or medical interventions related to liver disease. Individuals not meeting these criteria were classified as controls. Exclusion from the control group was applied if any of the following were present: (1) diagnosis of a liver disease not qualifying as a case; (2) outpatient encounter for the liver disease of interest; (3) elevated ALT levels (>25 IU l^{-1} in women, >33 IU l^{-1} in men); or (4) diagnosis of ascites attributed to liver pathology.

CAD. Cases of CAD were identified according to previously described criteria[9]. Inclusion as a case required fulfilment of at least one of the following: (1) documentation of CAD and/or myocardial infarction in EHRs from at least one inpatient encounter, two or more outpatient visits or as a recorded cause of death; (2) self-reported diagnosis of CAD or myocardial infarction at the baseline; or (3) documented history of surgical or medical interventions for CAD, including coronary artery bypass grafting or percutaneous coronary intervention. Individuals not meeting case criteria were classified as controls. Control samples were further excluded if a family history of CAD was identified, on the basis of EHR or self-reported information at the study baseline.

Epidemiological analysis methods

The associations between the TG:HDL ratio and various phenotypes—including body weight, fat distribution, glycaemic indices, hepatic steatosis, liver injury markers and histopathological features—were estimated using correlation or linear regression models. TG:HDL ratio levels were stratified in percentiles. Cox proportional hazards models were used to estimate the time-to-event relationship between the TG:HDL ratio and the incidence of type 2 diabetes, myocardial infarction, MASLD and liver cirrhosis in at-risk individuals, that is, without the outcome at or before baseline.

DNA preparation and exome sequencing

As previously outlined[90], genomic DNA (gDNA) libraries were constructed by enzymatically fragmenting high-molecular-mass DNA to achieve an average fragment length of 200 bp. To facilitate multiplexed exome capture and sequencing, unique 10 bp asymmetric barcodes were incorporated into the DNA fragments of individual samples during library amplification. Equimolar quantities of these barcoded libraries were combined and subjected to exome capture using either a modified xGen Exome Research Panel probe library (Integrated DNA Technology) or a modified version of the human comprehensive exome panel (Twist Bioscience). After PCR amplification and quantification, the enriched libraries were multiplexed and sequenced on Illumina platforms, generating 75 bp paired-end reads. Sequencing was performed on the Illumina HiSeq 2500 and NovaSeq 6000 instruments, using S2 or S4 flow cells.

Read mapping and variant calling

Sequencing reads in FASTQ format were generated from Illumina image data using bcl2fastq (v.2.19.0) (Illumina). In accordance with the original

quality functional equivalent (OQFE) protocol[91], reads were aligned to the GRCh38 reference genome using BWA MEM (v.0.7.15)[92] in an alt-aware configuration. Duplicate reads were identified and flagged, and additional per-read annotations were incorporated. Variant detection was performed using a Parabricks-accelerated implementation of DeepVariant (v.0.10), using a custom model to call single-nucleotide variants and short insertions and deletions, resulting in per-sample genome variant call format (VCF) files[93]. These individual VCF files were subsequently merged using GLnexus (v.1.4.3)[94] to produce a joint-genotyped, multisample cohort-level VCF (pVCF). For downstream analyses, PLINK (v.1.9)[95] was used to convert the pVCF file into PLINK-compatible files.

Exome sequencing quality control

For each cohort, we performed exome sequencing quality control with the following parameters: sample missingness <10%, variant missingness <10%, ±10 bp buffer window around target regions, a low heterozygosity HWE $P > 1 \times 10^{-100}$ and an excess heterozygosity HWE $P > 1 \times 10^{-30}$.

Variant annotation

Variant annotation was performed using Variant Effect Predictor (VEP, v.100.4)[96], using Ensembl release 100 human protein-coding transcript models[97]. For all analyses, a single functional consequence was assigned to each variant on the basis of its annotation on the canonical transcript. The canonical transcript for each gene was determined using a combination of MANE[98], APPRIS[99] and Ensembl canonical tags, following previously described procedures[98].

Nonsense-mediated decay (NMD) predictions were performed using the NMD plugin in VEP[96]. A variant was predicted to escape NMD if it satisfies any of the four canonical rules: (1) the variant is located in the last exon of the transcript; (2) the variant is located within 50 bases upstream of the penultimate (second to last) exon; (3) the variant is located within the first 100 coding bases of the transcript; and (4) the transcript is intronless, containing only a single exon. A variant is predicted to escape NMD if it satisfies any of these four rules[100,101].

Genetic association analyses

Genetic association analyses were conducted in each cohort using the linear whole-genome regression approach implemented in REGENIE (v.3.4 or older)[102]. Two primary analyses were performed across autosomes (chromosomes 1–22) and chromosome X in the discovery cohort: a genome-wide association study (GWAS) of common TOPMed-imputed variants[103] (AAF ≥ 1%) and a gene-based association analysis of rare coding variants. In the first step of REGENIE (trait prediction based on genetic data), directly genotyped variants were included if they had an AAF ≥ 1%, missingness < 10%, Hardy–Weinberg equilibrium $P > 10^{-15}$ and passed linkage disequilibrium (LD) pruning (window size, 1,000 variants; sliding window, 100 variants; r^2 threshold, 0.1). The association model in the second step of REGENIE incorporated the following covariates: age, age^2, sex, age × sex, age^2 × sex, the first 10 principal components (PCs) based on common variants (derived from a set of LD-pruned array variants: 1,000-variant windows, 50-variant step size, r^2 threshold 0.1), the first 20 PCs based on rare variants (the same pruning parameters) and cohort-specific sequencing batch covariates. The second step of REGENIE also included as a covariate a genome-wide leave-one-chromosome-out polygenic score generated in step 1, which adjusts for relatedness and residual population stratification[102]. For each variant or gene burden being tested in step 2, the polygenic score used leaves out the chromosome of the variant or the gene burden being tested (https://rgcgithub.github.io/regenie/)[102]. The main analysis was a pooled multiancestry analysis (that is, not stratified by ancestry). We also separately performed sensitivity analyses stratified by ancestry (that is, European, admixed American, South Asian, African and East Asian ancestries). For the non-pseudoautosomal regions of chromosome X, a dosage compensation model was used: homozygous reference males were coded as 0, hemizygous males as 2 and heterozygous males set as missing.

To ensure independence between rare and common variants signals in the same chromosome, exome-wide gene-burden analyses were additionally adjusted for common variant signals (minor allele frequency ≥ 1%) identified by fine-mapping in the chromosome of the gene-burden being tested, a procedure that is orthogonal to REGENIE's leave-one-chromosome-out polygenic score approach and ensures control for local linkage disequilibrium between common and rare variants. We also performed a sensitivity analysis in which the adjustment was performed using common variants identified by ancestry-specific fine-mapping. For the 59 independent gene-burden signals identified in the main analysis, we also performed ancestry-, fasting-state- and sex-stratified gene-burden analyses. To assess whether the inclusion of age and sex covariates introduced bias through differential contributions to TG versus HDL cholesterol estimates, for the 59 independent gene-burden signals, we conducted quality control analyses by removing each covariate group one at a time (that is, age, age^2, sex, age × sex, and age^2 × sex) from the genetic association models, finding correlations between effect estimates across the 59 genes $r > 0.99$ for each leave-one-covariate-out analysis for TG, HDL and TG:HDL ratio, consistent with minimal co-variate impact and consistent influences across TG and HDL.

Common variant analysis. A GWAS of common variants (minor allele frequency ≥ 1%) was conducted for the TG:HDL ratio. Association analyses were performed within each cohort by fitting linear regression models in REGENIE. Subsequently, cohort-specific results were combined using fixed-effect inverse variance-weighted meta-analysis. After meta-analysis, SuSiE (v.0.12.35)[104] was used to identify the most probable causal variants for each independent signal surpassing the genome-wide significance threshold of $P < 5 \times 10^{-8}$. To define fine-mapping regions, we first identified variants with $P < 1 \times 10^{-6}$ and separated them into regions at least 100 kb apart. Only regions containing at least one genome-wide significant variant ($P < 5 \times 10^{-8}$) were retained. Each region was then extended to the nearest recombination hotspot, with extensions ranging from 25 kb to 250 kb. To ensure adequate coverage for fine-mapping, regions smaller than 100 kb were symmetrically expanded to meet the minimum length requirement. Finally, a 100 kb buffer was applied around each region and regions with overlapping buffers were merged into a single region to form the final loci for causal variant identification. For each locus, an exact LD matrix was computed on the basis of the individuals included in the association analysis using the pooled estimate of the covariance across cohorts. This was calculated using all cohorts in the meta-analysis and the exact set of individuals used for analysis in each cohort. Specifically, the ith cohort's covariance matrix was weighted by its degrees of freedom ($n_i - 1$), values across all cohorts were then summed and the sum was divided by the total degrees of freedom ($N - k$) where N is the overall sample size and k is the number of cohorts in the meta-analysis. The generated pooled covariance matrix was then used for fine-mapping using SuSiE (v.0.12.35)[104]. SuSiE generates credible sets of common variants at each locus, representing a high likelihood of causality for the observed signal. Within these credible sets, variants are assigned posterior inclusion probabilities (PIP), summing to unity, with higher PIPs indicating a greater probability of causal association. For the main analysis, we set the L parameter of maximum theoretical independent signals per region to 30. For signal within each region, we identified the minimum set of variants capturing 95% of the cumulative PIP. This set represents the 95% credible set, and the variant with the highest PIP at each locus was designated as the sentinel variant. Sentinel variants with frequency ≥1% were adjusted for in the gene-based association analysis as described above.

Rare variant analysis. Variants predicted to induce frameshifts, premature stop codons or disrupt canonical splice donor or acceptor sites were classified as pLOF. Missense variants were further annotated using

six computational pathogenicity prediction tools: ESM1vp[105], SIFT[106], PolyPhen2+HumDiv[107], PolyPhen2+HumVar[107], MutationTaster[108] and LRT[109]. Variant inclusion for gene-burden testing was determined by AAF, pLOF status (including stop gained, frameshift, splice donor and splice acceptor variants) and, for missense variants, the number of algorithms predicting a deleterious effect. Burden tests were conducted across all combinations of four AAF thresholds (maximum AAF across ancestries)—singleton, 0.0001, 0.001 and 0.01—and six variant class groupings: (1) pLOF; (2) pLOF plus deleterious missense in 1 out of 5 algorithms (non-ESM1vp); (3) pLOF plus deleterious missense in 5 out of 5 algorithms (non-ESM1vp); (4) pLOF plus deleterious missense according to ESM1vp; (5) pLOF plus deleterious missense in 5 out of 5 algorithms and according to ESM1vp; (6) pLOF plus any missense. For each gene, only the canonical transcript was considered for variant annotation and gene-burden analysis. For each gene, this approach yielded a total of 24 gene-burden exposures, the *P* values of which were subsequently aggregated using ACAT to derive the overall BURDEN-ACAT *P* value. BURDEN-ACAT[110] uses the ACAT[111] method to integrate results from 24 distinct burden tests into a single aggregated *P* value. $P < 1.04 \times 10^{-7}$ (Bonferroni corrected for 20,000 genes and 24 different gene-burden exposures per gene) was used as the exome-wide significance threshold in the exome-wide gene-burden discovery analysis, consistent with previous literature[8,16]. A within-chromosome conditional analysis was also performed adjusting each gene-burden signal for other gene-burden signals on the same chromosome.

Tissue-enrichment analysis

Expression tissue enrichment for identified genes was calculated using gene expression values from the V8 data freeze from GTEx, as previously described[8]. After RNA-seq quality-control steps described in GTEx, the median expression level across individuals was calculated for each tissue and gene. For each tissue, z_{tissue} scores were calculated by standardizing TMM-normalized gene-expression values for each gene using the median and median absolute deviation of all gene expression values within that tissue. Within a gene, z_{gene} values were generated using the same standardization approach, across tissues, on z_{tissue} values. This second layer of standardization ensures that comparisons for a gene across tissues are easier to make and interpret. For each gene, expression enhanced tissues are defined as those where z_{gene} scores are at least 6 s.d. greater than the median z_{gene} scores observed for that gene across tissues, while expression enriched tissues are defined as those that meet the enhanced definition and have the largest positive deviation. To quantify per tissue enrichment of gene-burden associations from the TG:HDL ratio exome-wide analyses, we first aggregated all of the *P* values across all burden tests using the Cauchy combination test[111] per gene. The OR is based on assessing the relationship between the outcome indicator and the tissue specific indicator variable of the observed expression enriched status per gene. The outcome indicator variable is 1 if the per gene aggregated $P \leq 1.04 \times 10^{-7}$ and 0 otherwise. The tissue specific indicator variable is set to 1 for a given gene and tissue if that tissue is identified as an 'expression enriched tissue' for that gene (as defined above); otherwise, it is 0. The enrichment OR is then calculated using the Firth-corrected approach[112] to account for cells with zero counts after excluding all primary sexual and reproductive tissues. Each tissue enrichment OR can therefore be interpreted as the odds of seeing a TG:HDL ratio analyses $P \leq 1.04 \times 10^{-7}$ comparing genes with expression-enriched status in the tissue to those that are not expression enriched in the tissue. For the common-variant tissue enrichment analysis based on autosomal protein-coding genes, we defined the outcome for each tissue as 1/0 if the gene was identified as expression enriched in the tissue. The exposure variable was set to 1 if the gene was identified in the variant to gene prioritization approach and 0 otherwise (see the 'Variant to gene prioritization' section below). To minimize the potential confounding, gene length was included as a covariate. The enrichment OR was then estimated using the Firth-corrected logistic model[112] to account for cells with zero counts after excluding all primary sexual and reproductive tissues.

Liver RNA-seq

We conducted liver RNA-seq analysis of samples obtained from 1,946 participants enrolled in the Geisinger Health System (GHS) cohort, all of whom underwent perioperative wedge liver biopsies during bariatric surgery.

RNA preparation and sequencing. Total RNA was extracted and processed using the NEBNext Poly(A) mRNA Magnetic Isolation Module and the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs), followed by amplification with Kapa HiFi polymerase (Roche) and custom barcoded primers (IDT). Sequencing was performed on the Illumina NovaSeq 6000 platform using S2 flow cells, generating paired-end 75 bp reads. The mean sequencing depth was 72 million reads per sample (median, 68 million), with 93% of samples yielding at least 50 million reads and 99% exceeding 45 million reads, indicating high coverage.

RNA data quality control. RNA-seq data processing was performed according to protocols similar to the GTEx v8 analysis pipeline (https://gtexportal.org/home/documentationPage#staticTextAnalysisMethods). Reads were aligned to the GRCh38/hg38 human reference genome using STAR (v.2.5.3a)[113]. Optical duplicate reads were marked using Picard v.2.9.0 with OPTICAL_DUPLICATE_PIXEL_DISTANCE set to 15,000. Gene quantification utilized GENCODE Release 32 annotation, collapsed to a single transcript model per gene. Gene-level expression was quantified with RNA-SeQC (v.1.1.9)[114], applying the following read filters: (1) uniquely mapped reads; (2) properly paired reads; (3) alignment distance ≤ 6; and (4) reads fully contained within exon boundaries. For tissue-specific quality control, gene expression values (read counts) were normalized across samples using the TMM method[115,116]. Genes were retained for downstream analyses if they met expression thresholds of ≥0.1 TPM in ≥20% of samples and ≥6 unnormalized reads in ≥20% of samples.

eQTL analysis. eQTL analyses followed the GTEx workflow. Covariates included age, sex, the first four PCs derived from common genetic variants and the first 100 PCs from gene expression data to account for potential batch effects. Gene expression levels were transformed using a rank-inverse normal distribution prior to analysis.

eQTL to TG:HDL ratio colocalization. We performed an integrative analysis of TG:HDL ratio fine-mapped sentinel variants and liver eQTLs in the GHS bariatric surgery cohort. By cross-linking 1,617 fine-mapped sentinel variants associated with the TG:HDL ratio to liver eQTL sentinel variants, and including proxies in strong LD ($r^2 > 0.8$), we systematically evaluated the potential for shared causal variants. Bayesian statistical co-localization was conducted within 500 kb flanking regions around each matched variant, estimating the posterior probability of a shared causal signal between liver eQTLs and the TG:HDL ratio using the coloc software (v.5.2.3) implemented in R[117].

Proteomics and pQTL analysis

For UKB, we used summary-level data from the UKB Plasma Proteomics Project (PPP)[118]. For GHS, serum samples were obtained for 9,941 individuals and proteomic NPX values were generated by Olink based on the Olink Explore 3072 assay and using the same protocol as used for the UKB PPP[118]. Analytes were measured across eight different disease specific panels (cardiometabolic I/II, inflammation I/II, neurology I/II and oncology I/II). The samples were processed according to the manufacturer's protocol with custom automation at the Regeneron Genetics Center. The samples were sequenced on 10B flow cells on the Illumina NovaSeq X platform. The following quality-control steps

were applied: PC outliers were removed, putative sample swaps were excluded and data were normalized using NPX Intensity Normalization, which subtracts plate medians per assay. For pQTL analyses, protein levels were rank-inverse normal transformed, then adjusted for age (defined as age at serum collection), sex, age^2, age × sex and age^2 × sex. The final pQTL results comprised inverse-variance-weighted meta-analysed data from the UKB PPP project and pQTL data generated from individuals in the GHS cohort.

Variant to gene prioritization

In the TG:HDL-ratio common-variant GWAS, putative causal genes were prioritized with the following criteria: (1) colocalization with GTEx eQTL fine-mapped peaks in any tissue (defined as $r^2 \geq 0.8$); (2) colocalization with pQTL fine-mapped peaks ($r^2 \geq 0.8$); or (3) physical proximity to the nearest gene.

Pathway-enrichment analysis

We used DEPICT (v.1)[20] gene-level pathway *z*-scores as an input. The 59 target genes (mapped to Ensembl IDs) were intersected with the DEPICT gene set, and pathway enrichment was computed through a Stouffer score: $z_p = \sum_i z_{i,p}/\sqrt{n}$, with one-sided *P* values ($P = \Phi(-z_p)$) for enrichment. Multiple-testing control by Bonferroni correction was performed for the total number of pathways tested, yielding $P < 5.9 \times 10^{-6}$. For reporting, we summarized mean pathway *z* scores and listed the top contributing genes per pathway (highest positive *z* scores among the 59 genes).

Mouse models and procedures

C57BL/6NTac mice were from Taconic. Cas9-Ready (v.2.5, 2673) mice, expressing *Cas9* transgene under the CAG promoter, were previously reported[119]. Male mice were housed under a 12 h–12 h light–dark cycle at 22 ± 1 °C or 30 °C (for experiments at thermoneutral conditions), under humidity-controlled conditions (30–70% humidity) in static cages (≤5 mice per cage) with free access to food and water and fed either control chow diet (PicoLab Rodent Diet 20, LabDiet 5053) or HFHFD (Research Diets, D09100310). Mouse experiments were performed on age-matched and strain-matched pairs (littermates). Mice were between 6 and 11 weeks of age at the beginning of the studies. Animal sample size (*n*) was chosen on the basis of previous experience with similar experiments, experimental feasibility, availability of samples and the number necessary to obtain definitive, significant results. No statistical methods were used to predetermine sample sizes. In all studies, mice were allocated into experimental groups on the basis of matching body weight, body composition and lipid levels. The goal was to balance weight, body composition and lipid levels at the baseline (similar values between groups), to allow evaluation of treatment efficacy after intervention. Investigators were blinded to group allocation during serum chemistry analysis, liver lipid analysis, protein analysis and DNA editing analysis. Blinding was not possible during the in-live part of the mouse studies due to the need to treat mice with different reagents according to experimental designs. All of the animals were monitored for changes in body weight on a weekly basis. No adverse reactions or signs of discomfort were observed during the course of the experiment. Body mass composition was assessed in awake mice using echoMRI. Plasma was collected at various timepoints after AAV administration. The studies involved male mice only, which historically have shown more robust metabolic disease phenotypes compared with female mice[120–122]. We acknowledge that this experimental design choice may limit the generalizability of mouse model findings to female mice. All animal procedures were conducted in compliance with protocols approved by the Regeneron Pharmaceuticals Institutional Animal Care and Use Committee.

AAV administration. Non-fasted baseline serum chemistry was established on a chow diet, and mice were sorted into treatment groups on the basis of body weight and lipid levels. AAVs encoding shRNAs were of serotype AAV8 designed for liver-targeted mRNA knockdown in mice. AAV-shRNAs were diluted in saline and administered to mice by intravenous injection at a dose of 2.5×10^{11} viral genomes per mouse. *Hpn* targeting AAV (AAV8-GFP-U6-m-Hpn-shRNA, Vector Biolabs, shAAV-261634) encoded for a 1:1 mixture of 2 shRNAs: shRNA 1, 5′-CCGGGTGGATCTTCAAGGCCATAAACTCGAGTTTATGGCCTTGAAGATCCACTTTTT-3′; shRNA 2, 5′-CCGGCAACGGCACATCGGGCTTCTTCTCGAGAAGAAGCCCGATGTGCCGTTGTTTTT-3′.

Scramble (non-targeting) shRNA (AAV8-GFP-U6-scrmb-shRNA, Vector Biolabs) was used as control AAV, while saline alone was used as a vehicle control.

For studies using gRNAs, AAV8-multi-gRNAs targeting *Flcn*, *Fnip1*, *Fnip2* or *Fnip1* + *Fnip2* (combined AAVs) were intravenously administered to recipient mice at 2.5×10^{10} viral genomes per mouse. Multi-gRNA constructs included five distinct gRNAs targeting the same gene encoded by the same vector, with each gRNA driven by its own U6 promoter. Co-delivery of multiple, distinct gRNAs was previously shown to result in efficient gene perturbations and has been used to dissect molecular pathways in vivo[123].

gRNA sequences (name, gRNA sequence, protospacer-adjacent motif): mm_Fnip1_g15, GTAAGTGCTTGTGGATGCAG, GGG; mm_Fnip1_g19, AAGAGGCACTCCTGATCAGG, CGG; mm_Fnip1_g24, GCCAGGAAGAGAACTGAATG, AGG; mm_Fnip1_g25, CTGAATGAGGACAGAGACAG, CGG; mm_Fnip1_g28, AAAGTACCTGAACTCAGTCA, GGG; mm_Fnip2_g2, GCTACAATGGGTAGCTTCTG, TGG; mm_Fnip2_g3, ATTAACCAAGATCCTCAGGC, TGG; mm_Fnip2_g4, GAAGAGCTTGGAGACGGGAA, GGG; mm_Fnip2_g5, TTGTCTGACTTCGGAGCCAG, CGG; mm_Fnip2_g6, TATGTAGTGTATCTTCAGGG, TGG; mm_Flcn_g21, TTACACCAGAGGGTGCTGAA, GGG; mm_Flcn_g25, ATGGTCTGTGGACAACACGG, CGG; mm_Flcn_g26, AGCATGGTCTGTGGACAACA, CGG; mm_Flcn_g28, GTGTCTCACACACTTACCTG, AGG; mm_Flcn_g30, TATGAGTTTGTGGTGACCAG, TGG; mm_Ttr_g26, TTACAGCCACGTCTACAGCA, GGG.

siRNA administration in vivo. Non-fasted baseline serum chemistry was established on chow diet, and mice were sorted into treatment groups on the basis of body weight and body composition. GalNAc-conjugated siRNAs targeting *Flcn*, or a non-targeting control sequence, were administered through subcutaneous injections at 10 mg per kg every 10 days. GalNAc-conjugated *Flcn* or control siRNAs were provided by Alnylam Pharmaceuticals.

Amplicon library preparation. gDNA was extracted from liver samples. Target-specific oligos were designed (21–27 bp) to generate a maximum amplicon size of 350 bp with a primer melting temperature (T_m) of 60–65 °C. Barcode adapter sequences were added to the target specific oligo and the full sequence was ordered from Integrated DNA Technologies (IDT). PCR was completed on each gDNA sample. In brief, in each reaction, 4 ng of gDNA was combined with IDT oligos, Q5 polymerase (M0491, New England Biolabs), 10 µM dNTPs, buffer and water according to the manufacturer's specifications. The amplification products were then diluted 1:100 and used for the PCR barcoding reaction to create the final sequencing library. Each barcoding reaction contained a single amplified target with a forward and reverse primer containing a unique barcode and index. Each plate of PCRs was pooled in equal volumes and then purified in a single tube using AMPure XP reagent (A63881, Beckmann-Coulter) according to the manufacturer's instructions. The final library concentration was measured using the Qubit fluorometer (Q32866, Invitrogen). Then, 4 nmol of the prepared library was loaded onto the Illumina MiSeq system according to the manufacturer's instructions using the 2×300 read kit (MS-102-3003, Illumina).

Sequence mapping and characterization. Barcoded samples were demultiplexed to individual reads (FASTQ format). Forward and reverse reads of each FASTQ file were then merged using PEAR (v.0.9.8)[124]. Merged reads were mapped to the *Mus musculus* genome version 10 (mm10)

using Bowtie2 (v.2.5.4)[125]. Each sample was sequenced with a minimum of 20,000 merged reads across the expected guide cleavage location. Finally, characterization of barcoded samples was performed using a custom perl script (available at https://rgc-community.regeneron.com/ on the page dedicated to this manuscript). In brief, all insertions, deletions or base changes within a window of 20 bases upstream and downstream of the expected cut site were considered to be CRISPR-induced modifications. The number of reads containing insertions, deletions or base changes was compared to the number of reads with wild-type sequence to determine the percentage of editing per group.

Liver and lipid analysis. Blood was collected in EDTA tubes and plasma was obtained by centrifugation at 10,000 rpm for 10 min at 4 °C. Circulating total cholesterol, LDL-C, HDL-C, NEFA, albumin, total protein and ALT levels were measured in plasma using the ADVIA Chemistry XPT blood chemistry analyzer (Bayer). Non-HDL-C levels were calculated by subtracting HDL-C from total cholesterol values. To determine liver lipid levels, snap-frozen liver samples were weighed and homogenized in chloroform:methanol (2:1) solution, followed by addition of saline and centrifugation to achieve phase separation. The organic phase (bottom layer) was transferred into a new tube and evaporated with nitrogen gas. The dried lipids were then solubilized with chloroform:Triton X-100 (3:1) solution. Triglyceride and cholesterol content was measured enzymatically (Infinity, Thermo Fisher Scientific) according to the manufacturer's instructions and normalized to wet tissue weight, as previously described[47].

Glycaemic control. To assess glycaemic control, an insulin tolerance test was performed. Mice were fasted for 4 h, followed by insulin administration through intraperitoneal injection (HumulinR, Lilly, 0.75 U per kg body weight). The tip of the tail of each mouse was scratched to draw blood. Blood samples were collected at 0, 30, 60, 90 and 120 min, and glucose was measured using the Accu-check blood glucose monitoring system (Roche). For insulin measurements, blood was collected in capillary tubes containing protease inhibitors. Blood was centrifuged at 10,000 rpm for 10 min to separate the plasma and the mouse insulin ELISA kit (Mercodia, 10-1247-01) was used to determine insulin levels.

Quantitative PCR with reverse transcription. Tissue samples were collected in RNAlater (Thermo Fisher Scientific) and frozen. Total RNA was extracted using TRIzol reagent and Direct-zol RNA miniprep kits according to the manufacturer's instructions (Thermo Fisher Scientific, Zymo Research). gDNA was removed using the DNase I provided in the Direct-zol RNA kit (Zymo Research). mRNA (up to 2 µg) was reverse transcribed into cDNA using SuperScript VILOTM Master Mix (Thermo Fisher Scientific). cDNA was amplified in duplicate reactions containing 5 µl of 2× TaqMan Gene Expression Master Mix (Thermo Fisher Scientific), 0.5 µl TaqMan assay probes (20×, Thermo Fisher Scientific), 3.5 µl nuclease-free H_2O and 1 µl cDNA using the QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) and data were analysed using the $2^{-\Delta\Delta C_t}$ method.

RNA-seq and read mapping. Total RNA was purified from liver ($n = 5$ mice fed a chow diet for 30 weeks, $n = 7$ mice fed an HFHFD for 30 weeks (vehicle control group), $n = 8$ mice treated with *Flcn* AAV8-gRNA on an HFHFD for 30 weeks). Tissue samples were transferred from RNAlater (Thermo Fisher Scientific, AM7021) to 1–3 ml Trizol reagent (Thermo Fisher Scientific, 15596026). The samples were homogenized on the custom Omni homogenizer (Omni, 51-000-1) at 20,000 rpm for 180 s. The lysates were phase-separated with chloroform and the aqueous phase was purified on the KingFisher Flex (Thermo Fisher Scientific, 5400630) system with the MagMAX-96 for Microarrays Total RNA Isolation Kit (Thermo Fisher Scientific, AM1839) with an additional DNase (Qiagen, 79254) step added between the first and second washes. RNA was quantified on the Lunatic (Unchained Labs, 700-2000) system using a standard Lunatic plate (Unchained Labs, 701-2019), and the integrity was read on the 5300 Fragment Analyzer (Agilent, M5311AA) using an RNA Kit (Agilent, DNF-471-1000) according to the manufacturer's protocol. mRNA-seq libraries were generated using the KAPA mRNA HyperPrep Kit (Roche Sequencing). Starting material was 500 ng RNA and fragmentation was done at 85 °C for 6 min. cDNA was ligated with 1.5 µM xGen Dual Index UMI Adapters (Integrated DNA Technologies) and amplified using 12 PCR cycles. Sequencing of the resulting libraries was done on NovaSeq 6000 (Illumina) using a 51 cycle, single-end sequencing recipe. Raw sequence data (BCL files) were converted to FASTQ format by Illumina BCL Convert (v.4.3.6). Reads were decoded on the basis of their barcodes, and read quality was evaluated with FastQC (v.0.12.1) (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the mouse genome (GRCm38.REGN) using ArrayStudio software (v.12.5) (OmicSoft) allowing two mismatches. Reads mapped to the exons of a gene were summed at the gene level. Differential gene expression analysis was performed using the DESeq2 (v.1.34.0) package[126]. Benjamini–Hochberg multiple testing correction was applied to the obtained *P* values. Significant genes were determined using cut-offs of FDR < 0.05 and fold change > 2.

scRNA-seq. Five human single-cell RNA-seq (scRNA-seq) datasets were downloaded from the Gene Expression Omnibus (GEO): GSE136103 (ref. 127), GSE168933 (ref. 128), GSE115469 (ref. 129), GSE158723 (ref. 130) and GSE185477 (ref. 131). Only samples from human liver where scRNA-seq was performed were included from these datasets. For GSE136103 (ref. 127), counts tables were used as provided by the authors. For the other four datasets, raw data were processed with 10x Genomics Cell Ranger (v.7). Doublets were detected and removed using Scrublet (v.0.2.3)[132]. Cells with <200 detected genes and/or mitochondrial counts >10% were removed. Counts from all datasets were merged into a single AnnData object and downstream processing was performed using scanpy (v.1.9.1)[133]. Genes not expressed in any cell in the merged object were filtered out. Counts were normalized to 10,000 counts per cell then log-transformed. The top 5,000 highly variable genes were identified, and principal component analysis was performed using the top 100 PCs. Batch effects across datasets were corrected using the python implementation of Harmony, Harmonypy (v.0.0.10)[134]. A shared nearest-neighbour graph was constructed on the corrected embedding using 20 neighbours and 100 PCs and a uniform manifold approximation and projection (UMAP)[135] was computed. Leiden clustering was performed at a resolution of 0.1, and clusters were annotated to major cell types using canonical markers. Myeloid and lymphoid compartments were extracted and further subclustered to resolve additional cell types. Mouse liver scRNA data were obtained from a previous study (GSE156052)[136]. Author-provided UMAP coordinates and cell type annotations were used without modification.

Protein detection using liquid chromatography–mass spectrometry. Protein lysates were generated from livers using SDS buffer containing protease inhibitors. After reduction and alkylation, proteins were enzymatically digested into peptides using trypsin/Lys-C. Synthetic stable isotope labelled peptide ($^{13}C_6$,$^{15}N_4$-arginine or $^{13}C_6$,$^{15}N_2$-lysine, AQUA QuantProHeavy, Thermo Fisher Scientific) standards for FLCN and FNIP2 were added for validation. Peptides were separated on a C18 column using a 70 min gradient and analysed on the Thermo Scientific Exploris480 Mass Spectrometer using parallel reaction monitoring. Extracted ion chromatograms for target peptides (FLCN, LLEGAPT EDTLVQMEK; FNIP2, GPSSEPVPNR) were generated using Thermo Scientific Freestyle software (v.1.8.65.0) and the area under the curve was used for quantification. Data were normalized to histone H4C1 protein (DNIQGITKPAIR) to account for input differences.

Knockdown in primary human hepatocytes. Cryopreserved primary human hepatocytes were obtained from Thermo Fisher Scientific (donor

HU8450, male, HMCPTS). Cells were authenticated by the manufacturer by evaluating phase I enzyme activities, transporter activity and bile canaliculi formation. Cells were not tested for mycoplasma contamination. Cells were thawed and cryopreservation medium was exchanged for plating medium (William's E medium (A1217601, Thermo Fisher Scientific) supplemented with primary hepatocyte thawing/plating supplements (CM3000, Thermo Fisher Scientific)). Cells were seeded into collagen-precoated 12-well tissue culture plates (356500, Corning) at a density of 600,000 cells per well. After 6 h, when cells had settled and formed a monolayer, the plating medium was exchanged for maintenance medium (William's E medium supplemented with primary hepatocyte maintenance supplements, serum-free (CM4000, Thermo Fisher Scientific)). Hepatocytes were transfected with 100 nM siRNAs (Silencer Select siRNA: FNIP1 s41304, 4392421; siNeg Silencer siRNA 1, AM4635, Thermo Fisher Scientific) using lipofectamine RNAiMax reagent (13778150, Invitrogen) diluted in OptiMEM medium (51985034, Gibco), according to the manufacturer's protocol. Cells were collected after 96 h and collected for RNA extraction using the RNeasy Mini Kit (74104, Qiagen), followed by gene expression analysis and immunoblotting.

Immunoblot analysis. Primary human hepatocyte cell pellets were resuspended in 1× RIPA buffer (20-188, EMD Millipore) supplemented with protease inhibitors (PIA32955, Thermo Fisher Scientific). The lysates were centrifuged at 10,000*g* for 10 min at 4 °C. The supernatant was collected and protein concentrations were determined using the DC assay (5000111, Bio-Rad). Identical amounts (8 µg) of protein were size-fractionated on 4–20% gradient SDS–PAGE gels (5671093, Bio-Rad) under reducing conditions and transferred to PVDF membranes (1704157, Bio-Rad). The membranes were blocked with 5% milk in Tris-buffered saline with 0.1% Tween-20. After blocking, the membranes were probed with primary antibodies (anti-human FNIP1, Cell Signaling, 36892, 1:1,000; anti-HSP90, Cell Signaling, 4877, 1:1,000) and the signal was detected using an enhanced chemiluminescent detection system (WBULS0500, EMD Millipore) and imaged on the Amersham Imager 600 (GE Healthcare Life Sciences). Full scans of the immunoblots are provided in Supplementary Fig. 1.

Statistical analysis and reproducibility. In cell and animal experiments, statistical and graphical data analyses were performed using Microsoft Excel and Prism 10 (GraphPad). Data are expressed as mean ± s.e.m. or mean ± 95% CIs. Mean values were compared using one-way or two-way ANOVA as implemented in GraphPad Prism 10 (GraphPad). $P < 0.05$ was considered to be significant. All experiments with AAV8-gRNAs (*Flcn*, *Fnip1* and *Fnip2*) in mice fed high-fat diets for up to 13 weeks, studies in primary human hepatocytes (*Fnip1*) and studies with AAV-shRNAs (*Hpn*) were independently performed twice and produced similar results. Studies with long-term (30 weeks) inhibition of *Flcn*, *Fnip1* and *Fnip2*, as well as studies of mice fed chow, at thermoneutrality or using GalNAc-conjugated siRNAs were performed once, but yielded results that were similar to the AAV8-gRNA studies in mice fed high-fat diets.

Reporting summary

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

Data availability

Summary association statistics for the TG:HDL genome-wide discovery analysis are available online (https://rgc-community.regeneron.com/). All other association results are reported in the Article and its Supplementary Information. Data from the MCPS are available to bona fide researchers. For more details, the study's data and sample sharing policy is available (in English or Spanish; https://www.ctsu.ox.ac.uk/research/mcps). Study data can be examined through the study's data showcase (https://datashare.ndph.ox.ac.uk/mexico/) and MCPS ancestry-specific allele frequencies are available online (https://rgc-mcps.regeneron.com). Data from the BELIEVE study are available on reasonable request to the data access committee. Further information on the application process can be found on the BELIEVE study website (https://www.believestudy-bangladesh.org/accessing-believe-data/). UKB genotypic and phenotypic data are available to approved investigators through the UK Biobank study (www.ukbiobank.ac.uk/). Additional information about how to access the data are available on the UK Biobank website (https://www.ukbiobank.ac.uk/use-our-data/). GHS data are available to qualified academic non-commercial researchers for reproducibility of data, results and conclusions of the Article through the portal (https://regeneron.envisionpharma.com/vt_regeneron/) under a data access agreement. For the UCLA-ATLAS (ATLAS) cohort, due to privacy regulations, individual-level data cannot be published. Comprehensive processed data and summary statistics are available in the Supplementary Tables or through the UCLA-ATLAS browser (https://atlas-phewas.mednet.ucla.edu). For PMBB, BioMe, DBS, MAYO-RGC and CCPM, academic, non-commercial researchers interested in reproducing the results reported in this Article may request access to data through a data agreement by reaching out to the associated biobank at the following links: PMBB (https://pmbb.med.upenn.edu/contact.php); BioMe (https://mountsinaimillion.org/en/contact); DBS, J. Cohen (jonathan.cohen@utsouthwestern.edu); MAYO-RGC, Center for Individualized Medicine, Mayo Clinic Research (https://www.mayo.edu/research/centers-programs/mayo-clinic-biobank/for-researchers); and CCPM (https://medschool.cuanschutz.edu/cobiobank/contact-us). MDCS data may be available to qualified academic non-commercial researchers to reproduce results reported in this Article through the portal (https://www.malmo-kohorter.lu.se/malmo-cohorts), following the principles outlined in this policy (https://www.malmo-kohorter.lu.se/sites/malmo-kohorter.lu.se/files/mdcs_mpp_mos_request_form_vermar20.doc). Indiana-CLDB data can be made available for those requesting through Indiana Biobank's data request process. Applications are reviewed by three members of the Indiana Biobank's Scientific Review Committee and, if approved, made available on a secure-enclave for analysis. South Asia Biobank data may be accessible to researchers on request to the data access committee. Contact forms and emails are provided on the Global Health Research Unit (GHRU) on Diabetes and Cardiovascular Disease in South Asia website (https://www.ghru-southasia.org). RNA-seq data have been deposited at Gene Expression Omnibus (GEO) under accession number GSE330407. Source data are provided with this paper.

Code availability

The custom perl script used for the characterization of barcoded samples in the mouse experiments is available for download (https://rgc-community.regeneron.com/).

60. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
61. Carey, D. J. et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet. Med.* **18**, 906–913 (2016).
62. Tapia-Conyer, R. et al. Cohort profile: the Mexico City Prospective Study. *Int. J. Epidemiol.* **35**, 243–249 (2006).
63. Ziyatdinov, A. et al. Genotyping, sequencing and analysis of 140,000 adults from Mexico City. *Nature* **622**, 784–793 (2023).
64. Olson, J. E. et al. Characteristics and utilisation of the Mayo Clinic Biobank, a clinic-based prospective collection in the USA: cohort profile. *BMJ Open* **9**, e032707 (2019).
65. Chowdhury, R. et al. Cohort profile: the BangladEsh Longitudinal Investigation of Emerging Vascular and nonvascular Events (BELIEVE) cohort study. *BMJ Open* **15**, e088338 (2025).
66. Wiley, L. K. et al. Building a vertically integrated genomic learning health system: the biobank at the Colorado Center for Personalized Medicine. *Am. J. Hum. Genet.* **111**, 11–23 (2024).
67. Lajonchere, C. et al. An integrated, scalable, electronic video consent process to power precision health research: large, population-based, cohort implementation and scalability study. *J. Med. Internet Res.* **23**, e31121 (2021).
68. Abul-Husn, N. S. & Kenny, E. E. Personalized medicine and the power of electronic health records. *Cell* **177**, 58–69 (2019).

69. Park, J. et al. A genome-first approach to aggregating rare genetic variants in *LMNA* for association with electronic health record phenotypes. *Genet. Med.* **22**, 102–111 (2020).
70. Kozlitina, J. et al. Exome-wide association study identifies a *TM6SF2* variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **46**, 352–356 (2014).
71. Berglund, G., Elmstahl, S., Janzon, L. & Larsson, S. A. The Malmo Diet and Cancer Study. Design and feasibility. *J. Intern. Med.* **233**, 45–51 (1993).
72. Song, P. et al. Data resource profile: understanding the patterns and determinants of health in South Asians—the South Asia Biobank. *Int. J. Epidemiol.* **50**, 717–718 (2021).
73. Wurtz, P. et al. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am. J. Epidemiol.* **186**, 1084–1096 (2017).
74. Baigent, C. et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet* **376**, 1670–1681 (2010).
75. Cannon, C. P. et al. Ezetimibe added to statin therapy after acute coronary syndromes. *N. Engl. J. Med.* **372**, 2387–2397 (2015).
76. Sudhop, T. et al. Inhibition of intestinal cholesterol absorption by ezetimibe in humans. *Circulation* **106**, 1943–1948 (2002).
77. Lloyd-Jones, D. M. et al. 2017 Focused Update of the 2016 ACC Expert Consensus Decision Pathway on the Role of Non-Statin Therapies for LDL-Cholesterol Lowering in the Management of Atherosclerotic Cardiovascular Disease Risk: a report of the American College of Cardiology Task Force on Expert Consensus Decision Pathways. *J. Am. Coll. Cardiol.* **70**, 1785–1822 (2017).
78. Birjmohun, R. S., Hutten, B. A., Kastelein, J. J. & Stroes, E. S. Efficacy and safety of high-density lipoprotein cholesterol-increasing compounds: a meta-analysis of randomized controlled trials. *J. Am. Coll. Cardiol.* **45**, 185–197 (2005).
79. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
80. Dewey, F. E. et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* https://doi.org/10.1126/science.aaf6814 (2016).
81. Dennis, J. K. et al. Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. *Genome Med.* **13**, 6 (2021).
82. Tobin, M. D., Sheehan, N. A., Scurrah, K. J. & Burton, P. R. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat. Med.* **24**, 2911–2935 (2005).
83. Kleiner, D. E. et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
84. Dixon, W. T. Simple proton spectroscopic imaging. *Radiology* **153**, 189–194 (1984).
85. Littlejohns, T. J. et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat. Commun.* **11**, 2624 (2020).
86. Herman, J. L. et al. Humans with function-disrupting variants in the myostatin gene (*MSTN*) have increased skeletal muscle mass and strength, and less adiposity. *Nat. Commun.* https://doi.org/10.1038/s41467-026-70422-2 (2026).
87. O'Dushlaine, C. et al. Genome-wide association study of liver fat, iron, and extracellular fluid fraction in the UK Biobank. Preprint at *medRxiv* https://doi.org/10.1101/2021.10.25.21265127 (2021).
88. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2021. *Diabetes Care* **44**, S15–S33 (2021).
89. Eastwood, S. V. et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. *PLoS ONE* **11**, e0162388 (2016).
90. Sun, K. Y. et al. A deep catalogue of protein-coding variation in 983,578 individuals. *Nature* **631**, 583–592 (2024).
91. Krasheninina, O. et al. Open-source mapping and variant calling for large-scale NGS data from original base-quality scores. Preprint at *bioRxiv* https://doi.org/10.1101/2020.12.15.356360 (2020).
92. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
93. Yun, T. et al. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. *Bioinformatics* **36**, 5582–5589 (2021).
94. Lin, M. F. et al. GLnexus: joint variant calling for large cohort sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/343970 (2018).
95. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
96. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
97. Cunningham, F. et al. Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).
98. Morales, J. et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* **604**, 310–315 (2022).
99. Rodriguez, J. M. et al. APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* **41**, D110–D117 (2013).
100. Coban-Akdemir, Z. et al. Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *Am. J. Hum. Genet.* **103**, 171–187 (2018).
101. Lindeboom, R. G., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* **48**, 1112–1118 (2016).
102. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
103. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
104. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. B* **82**, 1273–1300 (2020).
105. Gillies, C. E. et al. Variant classification using proteomics-informed large language models increases power of rare variant association studies and enhances target discovery. *Genet. Epidemiol.* **49**, e70023 (2025).
106. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
107. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* https://doi.org/10.1002/0471142905.hg0720s76 (2013).
108. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
109. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
110. Ziyatdinov, A. et al. Joint testing of rare variant burden scores using non-negative least squares. *Am J. Hum. Genet.* **111**, 2139–2149 (2024).
111. Liu, Y. & Xie, J. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.* **115**, 393–402 (2020).
112. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38 (1993).
113. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
114. DeLuca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
115. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
116. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
117. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
118. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
119. Smagris, E. et al. Divergent role of mitochondrial amidoxime reducing component 1 (MARC1) in human and mouse. *PLoS Genet.* **20**, e1011179 (2024).
120. Pettersson, U. S., Walden, T. B., Carlsson, P. O., Jansson, L. & Phillipson, M. Female mice are protected against high-fat diet induced metabolic syndrome and increase the regulatory T cell population in adipose tissue. *PLoS ONE* **7**, e46057 (2012).
121. de Souza, G. O., Wasinski, F. & Donato, J. Jr. Characterization of the metabolic differences between male and female C57BL/6 mice. *Life Sci.* **301**, 120636 (2022).
122. Huang, K. P. et al. Sex differences in response to short-term high fat diet in mice. *Physiol. Behav.* **221**, 112894 (2020).
123. Thurkauf, M. et al. Fast, multiplexable and efficient somatic gene deletions in adult mouse skeletal muscle fibers using AAV-CRISPR/Cas9. *Nat. Commun.* **14**, 6116 (2023).
124. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
125. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
126. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
127. Ramachandran, P. et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512–518 (2019).
128. Buonomo, E. L. et al. Liver stromal cells restrict macrophage maturation and stromal IL-6 limits the differentiation of cirrhosis-linked macrophages. *J. Hepatol.* **76**, 1127–1137 (2022).
129. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
130. Payen, V. L. et al. Single-cell RNA sequencing of human liver reveals hepatic stellate cell heterogeneity. *JHEP Rep.* **3**, 100278 (2021).
131. Andrews, T. S. et al. Single-cell, single-nucleus, and spatial RNA sequencing of the human liver identifies cholangiocyte and mesenchymal heterogeneity. *Hepatol. Commun.* **6**, 821–840 (2022).
132. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
133. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
134. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
135. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426 (2020).
136. Su, Q. et al. Single-cell RNA transcriptome landscape of hepatocytes and non-parenchymal cells in healthy and NAFLD mouse liver. *iScience* **24**, 103233 (2021).

Acknowledgements We thank all of the participants and personnel of all of the studies included in the Article; all of the patient participants who engaged in the MyCode Community Health Initiative and the members of the MyCode Research Team; the members of the Geisinger-Regeneron DiscovEHR Collaboration who have been critical in the generation of the data used for this study; the participants and members of the BELIEVE Study Consortium (a full list of consortium members is provided in the Supplementary Information); the participants of the Mount Sinai BioMe Biobank and the Mount Sinai Million Health Discoveries Program for making this research possible; the participants who have made CCPM a reality; the participants who took part in the South Asia Biobank study and study team members for their time and energy spent on data collection; and the staff at the National Institutes of Health's All of Us Research Program for making data available that were included in the replication analysis of this Article. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the UK Department of Health and Social Care. We acknowledge Alnylam Pharmaceuticals for providing GalNAc-conjugated *Flcn* and control siRNAs.

Author contributions Conceptualization of the project, study design, interpretation of results and drafting of the manuscript and its revisions: G.H., R.C.A., V.G. and L.A.L. Project coordination: G.H., R.C.A., V.G., L.A.L., M.B.J. and M.G.L. Human genetics analysis: G.H., O.S., D.B., J.H., P.D., E.M., A.G., B.G., K.L., L.G., A. Baldassari, C.Z., S.B., J. Bovijn, A.L., N.V. and L.A.L. In vitro and in vivo experiments: R.C.A., D.P., J.L., C.H., I.M., D.S., A.C., K.B., T.N., X.X. and V.G. Sample collection, data generation, results interpretation and provision of critical information for the manuscript finalization: D.J.R., O.M., C.D.S., J. Berumen, P.K.-M., J.A.-D., J.M.T., J.R.E., R.C.

and R.T.-C. Human genetics sequencing and analytical infrastructure: J.D.O., J.G.R., G.R.A., J.M., C.W., A. Baras and L.A.L. Experimental programs infrastructure: A.J.M., C.A.K., G.D.Y. and M.W.S. All of the authors reviewed the manuscript for important intellectual content and approved the final version for submission. V.G. and L.A.L. jointly supervised this work. The contributions of the consortium members were as follows. The members of the Regeneron Genetics Center contributed to exome sequencing, sequence-variant calling, sample handling, genomic data generation, refined phenotype generation, data formatting, processing and quality control, development of analytical infrastructure, review, quality control and visualization of the data and results contributing the manuscript, research program management, coordination and administrative support, review of the manuscript and provision of necessary information for the finalization of the manuscript all of which made this work possible. Members of the GHS-RGC DiscovEHR Collaboration, Penn Medicine Biobank, Colorado Center for Personalized Medicine, Mayo Clinic-RGC Project Generation, Mexico City Prospective Study, Malmö Diet and Cancer Study, Indiana University School of Medicine (Indiana-CLDB), UCLA-RGC ATLAS Collaboration, Mount Sinai Million Health Discoveries Program, UTSW-Dallas Biobank and South Asia Biobank Study performed cohort design, participant consenting and recruitment, phenotype data collection and quality control, sample collection, cohort data generation, quality control and interpretation, review of the manuscript and provision of critical information necessary for the finalization of the manuscript.

Funding This research received funding from Regeneron Pharmaceuticals. The Mexico City Prospective Study (MCPS) has received funding from the Mexican Health Ministry, the National Council of Science and Technology for Mexico, Wellcome, Cancer Research UK, British Heart Foundation, Kidney Research UK and the UK Medical Research Council. Genotyping and exome sequencing in MCPS was funded through an academic partnership between the National Autonomous University of Mexico, the University of Oxford, Regeneron and AstraZeneca. The BangladEsh Longitudinal Investigation of Emerging Vascular and nonvascular Events (BELIEVE) study has been supported by an award to the CAPABLE consortium from the UK Medical Research Council/UK Research and Innovation Global Challenge Research Fund 'Grow' call (MR/P02811X/1). The coordinating centre has also been supported by underpinning support from the British Heart Foundation (BHF Chair Award CH/12/2/29428; BHF programme grant RG/18/13/33946) and the National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre (BRC-121520014; NIHR203312). This work was also supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. This research was also funded by a University of Cambridge Global Challenge Research Fund grant (RG88576). The PMBB study was supported by the Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878. CCPM is supported through a partnership between UCHealth, the University of Colorado (CU) Anschutz Medical Campus and the CU School of Medicine. MAYO-RGC Project Generation was supported in part by Mayo Clinic's Center for Individualized Medicine. The Indiana-CLDB study was made possible, in part, with support from the Indiana Biobank and the Indiana Clinical and Translational Sciences Institute funded, in part by, grant number UM1TR004402 and the Lilly Endowment. This UCLA-ATLAS study was supported by NIH National Center for Advancing Translational Science (NCATS) UCLA CTSI grant number UL1TR001881. The UCLA IPH acknowledges funding through the DGSOM and UCLAHealth and thanks the patients who participated in the ATLAS Community Health Initiative. The BioMe Biobank and the Mount Sinai Million Health Discoveries Program are funded by the Charles Bronfman Institute for Personalized Medicine at the Icahn School of Medicine and genetic data were generated in a partnership with the Regeneron Genetics Center. The DBS was funded by a generous donor. South Asia Biobank was funded by the National Institute for Health Research (NIHR) (16/136/68 and 132960) using UK aid from the UK Government to support global health research.

Competing interests Regeneron and Regeneron Genetics Center affiliated authors receive salaries from and own options and/or stock of the company. G.D.Y. is the chief scientific officer and member of the board of directors at Regeneron Pharmaceuticals. A.J.M. is an executive officer of Regeneron Pharmaceuticals. G.H., R.C.A., M.W.S., V.G. and L.A.L. are listed as inventors on pending patent applications relating to *FNIP1* and *FLCN* genetics (US-2025-0129366, WO/2025/049558, PCT/US2025/016730). J.R.E. reports grants to the University of Oxford from AstraZeneca and Regeneron Pharmaceuticals. The other authors declare no competing interests.

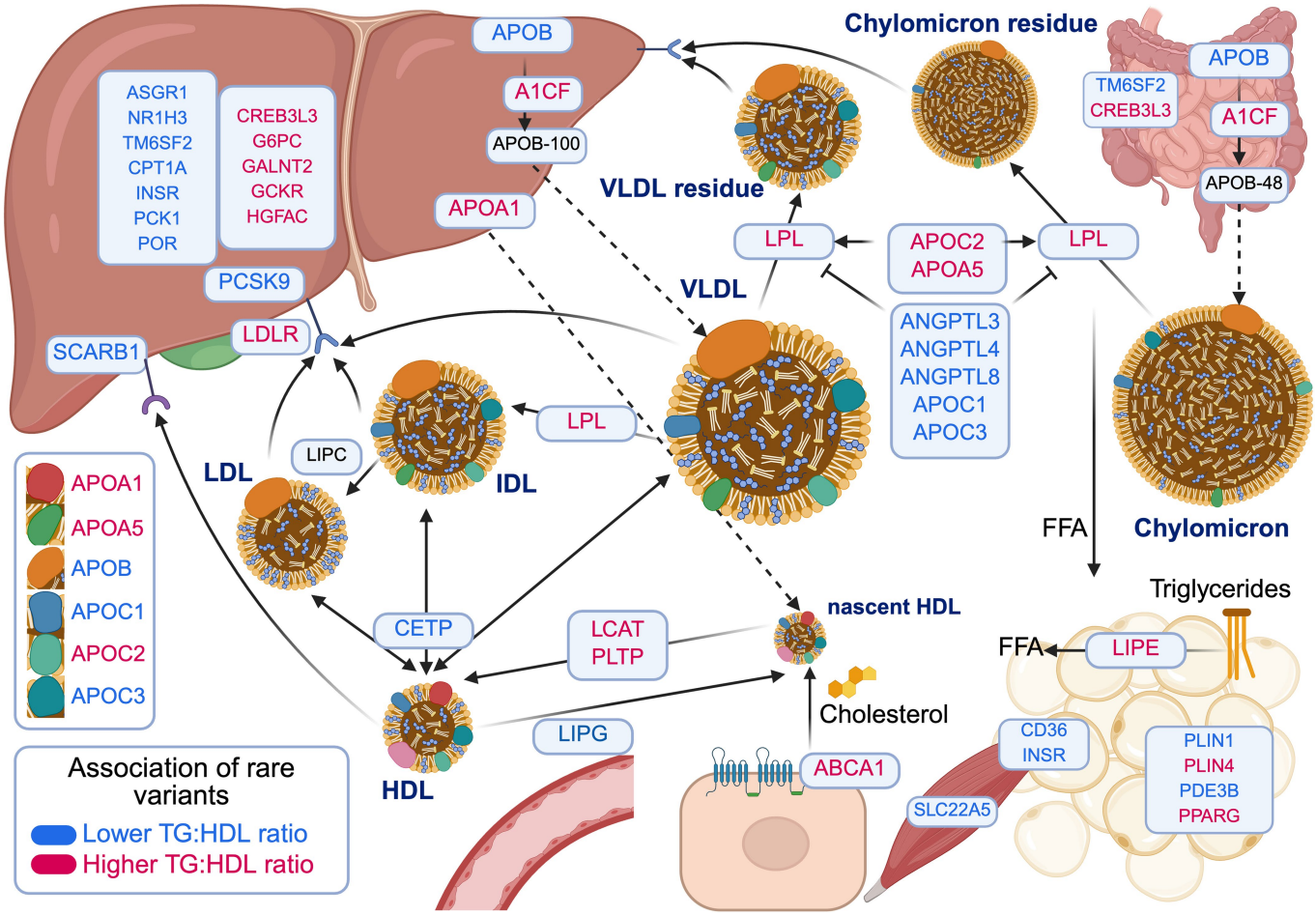
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10864-2.

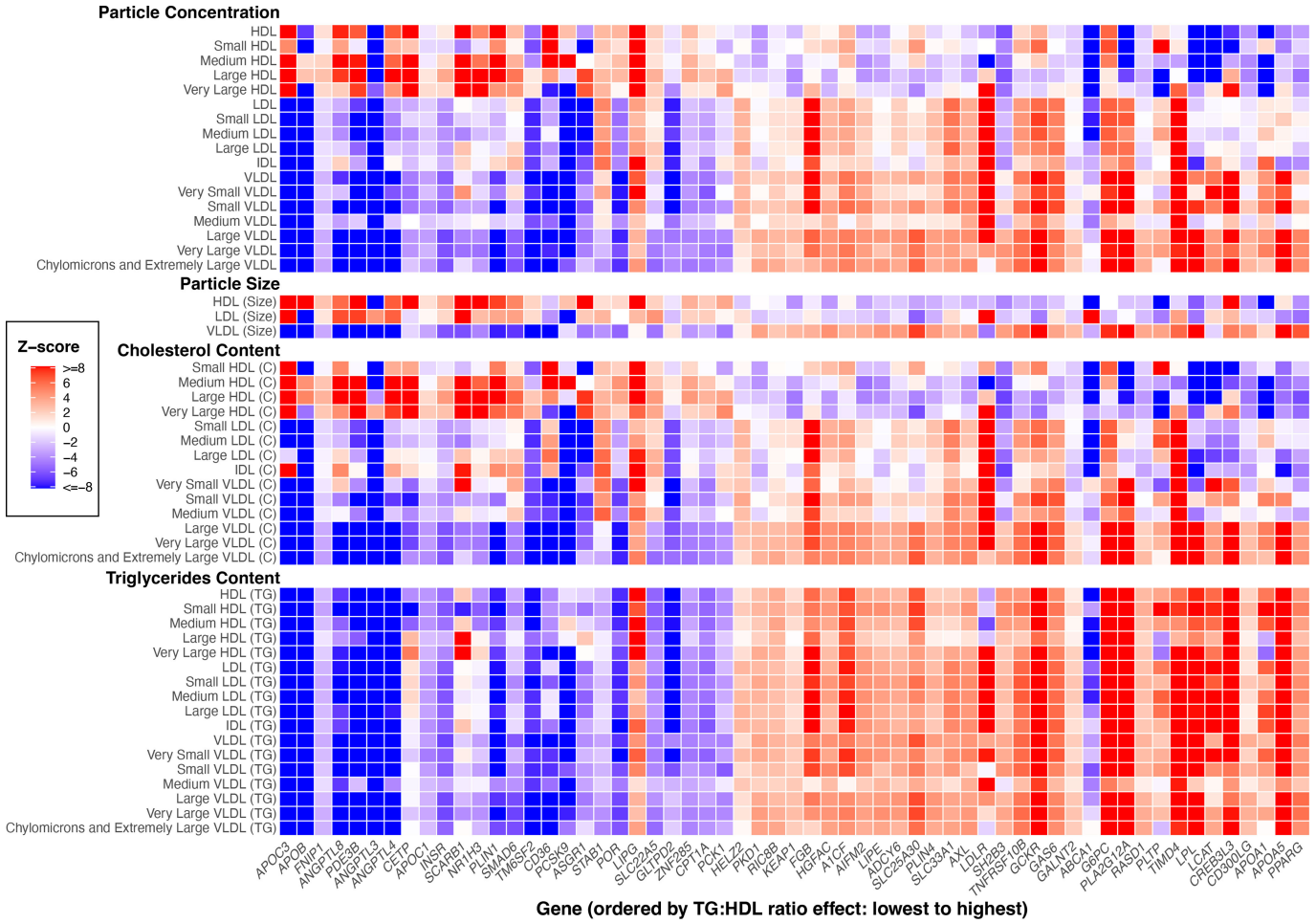
Correspondence and requests for materials should be addressed to Viktoria Gusarova or Luca A. Lotta.

Peer review information *Nature* thanks David Carling, who co-reviewed with Eneko Pascual-Navarro; Satoshi Koyama; and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.



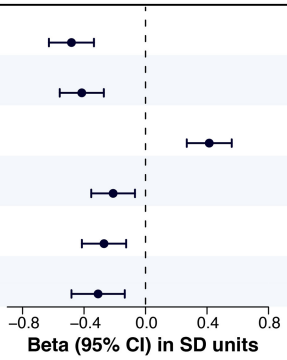
Extended Data Fig. 1 | Interplay of genes identified in the exome-wide analysis of TG:HDL ratio and overview of their role in energy metabolism. HDL; high-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; TG:HDL, triglycerides to HDL cholesterol ratio; VLDL, very low-density lipoprotein. Created in BioRender. Hindy, G. https://biorender.com/znkf5px (2026).



Extended Data Fig. 2 | Associations of the 59 TG:HDL ratio genes with detailed lipid and lipoprotein particle phenotypes. Direction of effect is shown on a spectrum where positive associations are shown in red and negative associations are shown in blue. Darker colours indicate a larger absolute z-score values or higher significance for *P* values. Association estimates are from inverse-variance weighted fixed-effect meta-analysis and *P* values are from the gene-burden test with the lowest two-sided Wald test *P* value per gene. HDL; high-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; TG:HDL, triglycerides to HDL cholesterol ratio; VLDL, very low-density lipoprotein.

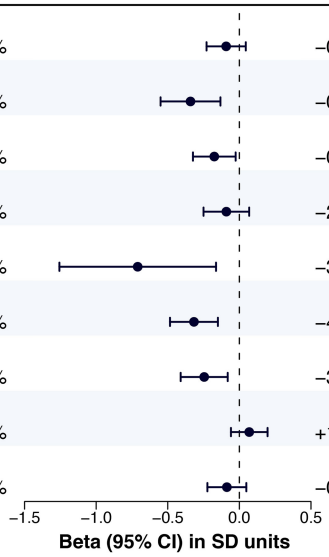
a

Outcome Gene Variant	Genotype Count RR / RA / AA	AAF	Beta (95% CI) per Allele in Clinical Units	*P*
Triglyceride/HDL Ratio				
FNIP1 pLoF, AAF<0.01%	1,031,961 / 155 / 0	0.008%	−1.66 (−2.17, −1.16) ratio	1.1×10^{-10}
Triglycerides				
FNIP1 pLoF, AAF<0.01%	1,075,238 / 160 / 0	0.007%	−41.0 (−55.1, −26.8) mg/dL	1.5×10^{-8}
HDL cholesterol				
FNIP1 pLoF, AAF<0.01%	1,033,114 / 155 / 0	0.008%	+6.38 (4.13, 8.63) mg/dL	2.8×10^{-8}
LDL cholesterol				
FNIP1 pLoF, AAF<0.01%	1,067,562 / 158 / 0	0.007%	−8.70 (−14.6, −2.83) mg/dL	0.0037
Non−HDL cholesterol				
FNIP1 pLoF, AAF<0.01%	1,032,502 / 155 / 0	0.008%	−11.9 (−18.2, −5.53) mg/dL	2.4×10^{-4}
Apolipoprotein B				
FNIP1 pLoF, AAF<0.01%	679,476 / 101 / 0	0.007%	−7.50 (−11.71, −3.28) mg/dL	4.8×10^{-4}



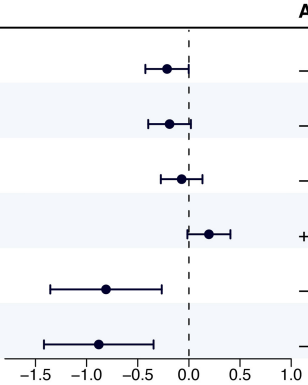
b

Outcome Gene Variant	Genotype Count RR / RA / AA	AAF	Beta (95% CI) per Allele in Clinical Units	*P*
Body Mass Index				
FNIP1 pLoF, AAF<0.01%	1,236,782 / 181 / 0	0.007%	−0.53 (−1.31, 0.26) kg/m2	0.19
Waist to Hip Ratio (Adjusted for BMI)				
FNIP1 pLoF, AAF<0.01%	620,793 / 78 / 0	0.006%	−0.02 (−0.03, −0.01) ratio	0.0014
HbA1c				
FNIP1 pLoF, AAF<0.01%	952,429 / 136 / 0	0.007%	−0.18 (−0.34, −0.03) %	0.021
Random glucose				
FNIP1 pLoF, AAF<0.01%	969,570 / 134 / 0	0.007%	−2.42 (−6.66, 1.83) mg/dL	0.26
Liver Fat (MRI−PDFF)				
FNIP1 pLoF, AAF<0.01%	63,416 / 12 / 0	0.009%	−3.88 (−6.86, −0.89) %	0.011
Alanine Aminotransferase				
FNIP1 pLoF, AAF<0.01%	901,295 / 126 / 0	0.007%	−4.84 (−7.40, −2.29) U/L	2.0×10^{-4}
Aspartate Aminotransferase				
FNIP1 pLoF, AAF<0.01%	919,423 / 130 / 0	0.007%	−3.20 (−5.34, −1.06) U/L	0.0034
Systolic Blood Pressure				
FNIP1 pLoF, AAF<0.01%	1,252,057 / 183 / 0	0.007%	+1.24 (−1.06, 3.54) mmHg	0.29
Diastolic Blood Pressure				
FNIP1 pLoF, AAF<0.01%	1,252,038 / 183 / 0	0.007%	−0.87 (−2.22, 0.49) mmHg	0.21



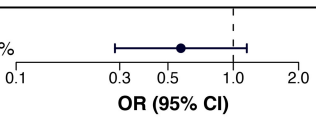
c

Outcome Gene Variant	Genotype Count RR / RA / AA	AAF	Beta (95% CI) per Allele in Clinical Units	*P*
(BIA) Body fat mass				
FNIP1 pLoF, AAF<0.01%	598,281 / 75 / 0	0.006%	−1.99 (−3.97, −0.02) kg	0.048
(BIA) Body fat %				
FNIP1 pLoF, AAF<0.01%	598,848 / 75 / 0	0.006%	−1.67 (−3.52, 0.17) %	0.076
(BIA) Body lean mass				
FNIP1 pLoF, AAF<0.01%	582,140 / 75 / 0	0.006%	−0.78 (−3.04, 1.47) kg	0.49
(BIA) Body lean %				
FNIP1 pLoF, AAF<0.01%	582,125 / 75 / 0	0.006%	+1.73 (−0.13, 3.59) %	0.068
(MRI) Abdominal to Gluteofemoral Fat Ratio				
FNIP1 pLoF, AAF<0.01%	64,293 / 12 / 0	0.009%	−0.16 (−0.26, −0.05) ratio	0.0036
(MRI) Visceral to Gluteofemoral Fat Ratio				
FNIP1 pLoF, AAF<0.01%	64,293 / 12 / 0	0.009%	−0.26 (−0.42, −0.10) ratio	0.0013



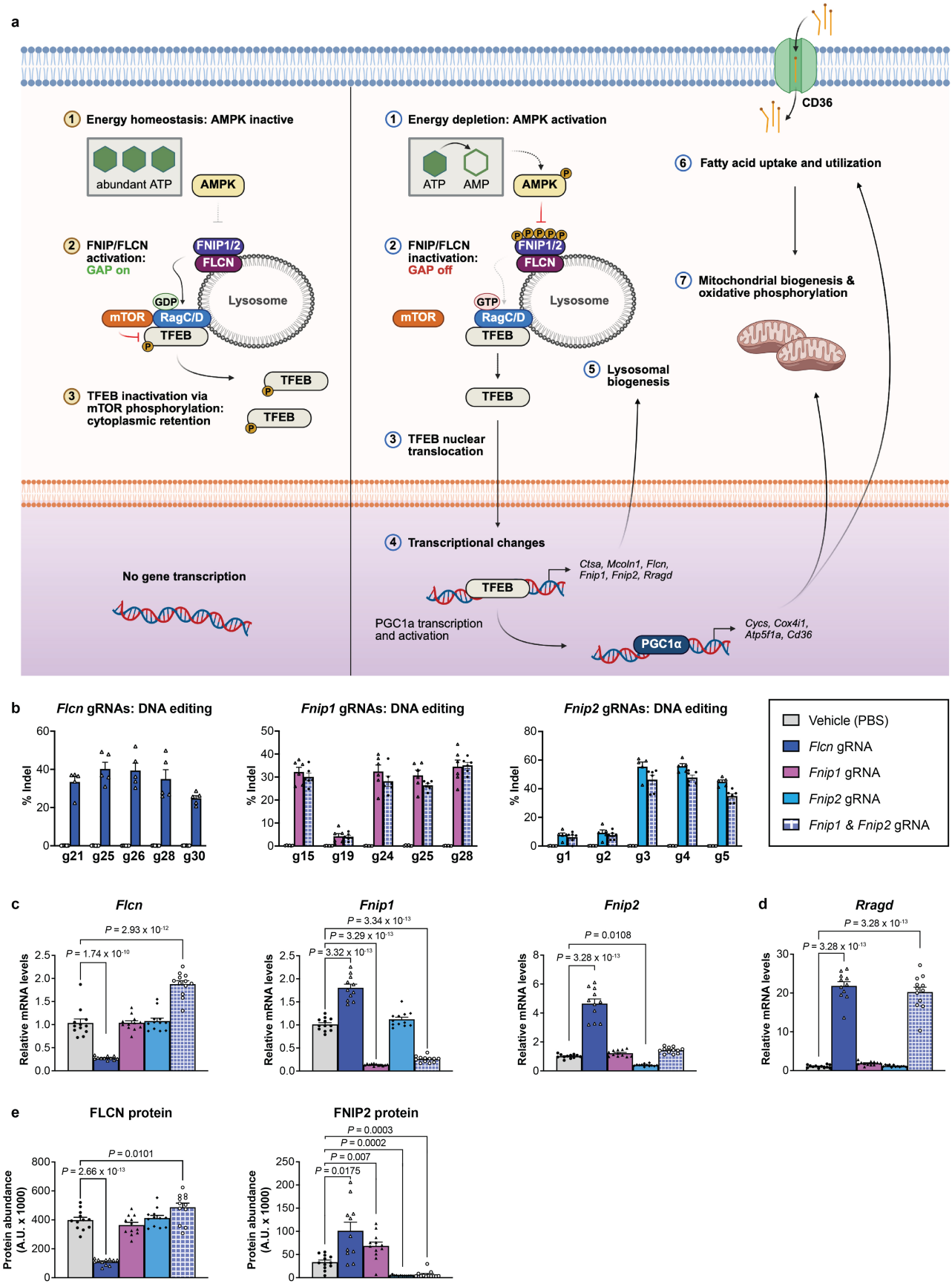
d

Outcome Gene Variant	Genotype Count RR / RA / AA	AAF	Odds Ratio (95% CI) per Allele	*P*
Myocardial Infarction	Cases: 60,973 / 7 / 0			
FNIP1 pLoF, AAF<0.01%	Controls: 670,765 / 95 / 0	0.007%	0.57 (0.28, 1.15)	0.12



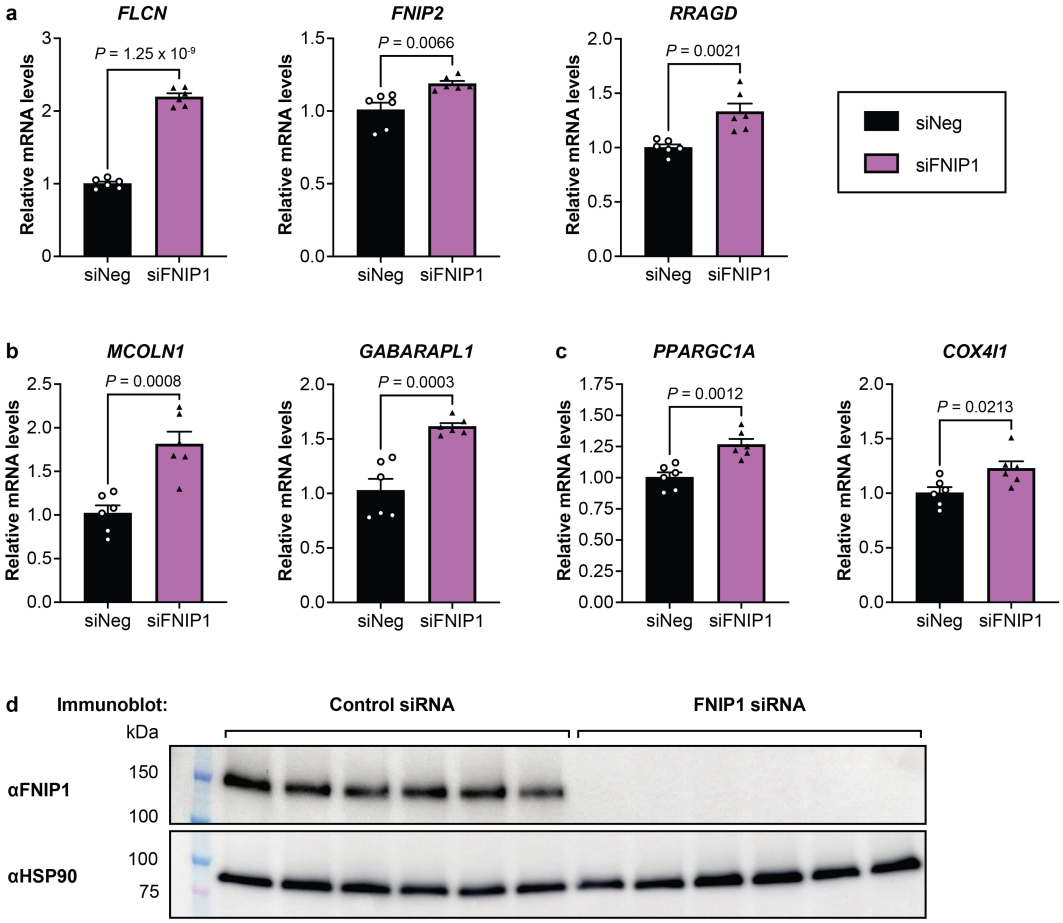
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Associations with cardiometabolic traits for ultra-rare pLOF variants in *FNIP1*. Association of rare predicted loss-of-function (pLOF) variants (alternative allele frequency [AAF] <0.01%) in *FNIP1* with cardiometabolic traits with blood lipids (**a**), anthropometrics, glycaemia, liver steatosis and injury, blood pressure (**b**), measures of body fat distribution (**c**) and myocardial infarction (**d**). Data are presented as Beta ± 95% CI (**a**-**c**) or OR ± 95% CI (**d**). Association estimates are from inverse-variance weighted fixed-effect meta-analysis and *P* values are from two-sided Wald tests. AAF, alternative allele frequency; BIA, bioimpedance; CI, confidence interval; ESM1vp, Evolutionary Scale Modelling; OR, odds ratio; pLOF, predicted loss-of-function; MRI, magnetic resonance imaging; PDFF, proton density fat fraction; RR, homozygous for the reference allele; RA, heterozygous; AA, homozygous for the alternative allele.

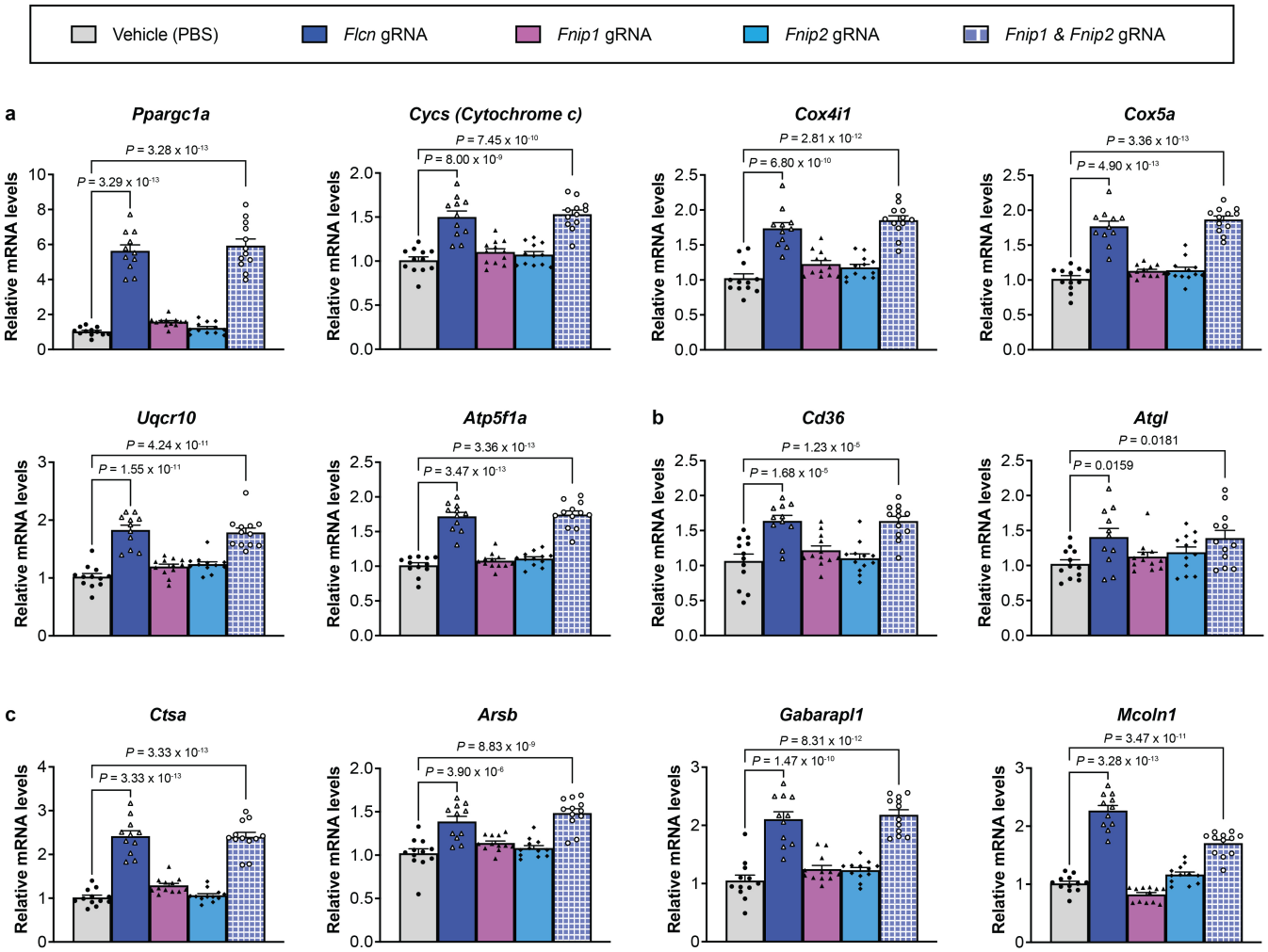


Extended Data Fig. 4 | See next page for caption.

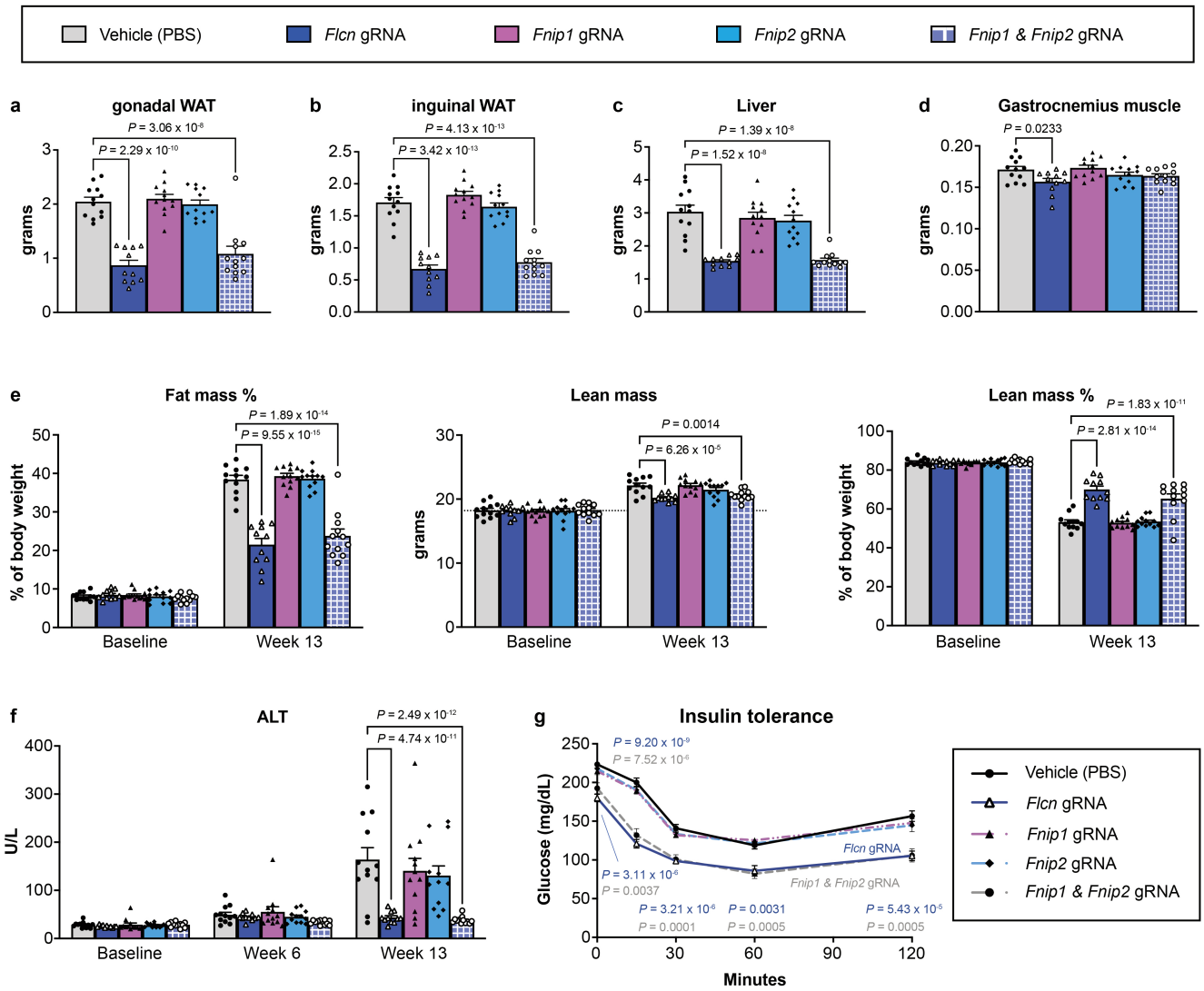
Extended Data Fig. 4 | The FNIP–FLCN pathway regulates mitochondrial and lysosomal function. a, Schematic overview of the FNIP–FLCN (FNIP/FLCN) signalling pathway. (left), During homeostasis, the active FNIP–FLCN complex exhibits GTPase-activating protein (GAP) function towards RagC/D, promoting GTP → GDP hydrolysis and leading to TFEB recruitment to the lysosomal membrane. mTOR subsequently phosphorylates TFEB, resulting in its cytoplasmic retention and transcriptional inactivation. (right), Upon energy depletion, energy sensor AMP-activated kinase (AMPK) phosphorylates FNIP1, thus inhibiting the GAP activity and allowing for release of TFEB from the lysosomal surface. TFEB then translocates to the nucleus to activate its target genes, including lysosomal genes and PGC1-alpha to promote mitochondrial gene expression in order to restore energy balance. Notably, TFEB also induces expression of *Fnip1*, *Fnip2*, *Flcn* and *Rrag* genes as part of a negative feedback loop. In liver, *Rragd* (encoding for RagD) is particularly sensitive to TFEB transcriptional regulation and can serve as readout for TFEB activity. Created in BioRender. Adam, R. https://biorender.com/0ceojsp (2026). **b**, Sanger sequencing of the gRNA cleavage sites revealed characteristic small insertions and/or deletions ('indels') in livers of *Cas9-Ready* mice. The number of reads containing *Flcn*, *Fnip1* or *Fnip2* indels for each gRNA was compared to the number of reads with wild-type sequence and expressed as percentage of editing (Indel %) of genomic sequences in liver. While individual gRNAs often lead to >30% editing in liver, the combination of 5 gRNAs targeting each gene (co-packaged, see Methods and Thürkauf et al., 2023)[123] results in efficient gene perturbations in liver cells targeted by the AAV8-gRNAs. Mean ± s.e.m. are shown. Data are from n = 6 (vehicle control), n = 5 (*Flcn* gRNA), n = 6 (*Fnip1* gRNA), n = 5 (*Fnip2* gRNA) and n = 6 (*Fnip1* & *Fnip2* gRNAs) biologically independent mice. **c**, mRNA levels of *Flcn*, *Fnip1* and *Fnip2* genes in liver following AAV8-gRNA administration. Since TFEB induces the *Fnip*/*Flcn* genes as part of a negative feedback loop, perturbation of *Flcn* or *Fnip & Fnip2* and subsequent TFEB activation is expected to increase expression of *Flcn*, *Fnip1* and *Fnip2* (e.g. see induction of *Flcn* mRNA in *Fnip1 & Fnip2* gRNA group). **d**, Hepatic mRNA levels of TFEB target gene *Rragd* (RagD). The robust transcriptional induction upon *Flcn* or dual *Fnip1 & Fnip2* inhibition confirms efficient FNIP–FLCN pathway perturbation and TFEB induction in liver. For all mRNA expression profiles, mean ± s.e.m. are shown. *P* values are from one-way ANOVA with Dunnett's correction posttest relative to vehicle control. **e**, Hepatic protein levels of FLCN and FNIP2, as measured via liquid chromatography-mass spectrometry (LC-MS) assay. Mean ± s.e.m. are shown. *P* values are from one-way ANOVA with Dunnett's correction posttest relative to vehicle control. In panels **c**-**e**, data are from n = 12 (vehicle control), n = 11 (*Flcn* gRNA), n = 12 (*Fnip1* gRNA), n = 12 (*Fnip2* gRNA) and n = 12 (*Fnip1* & *Fnip2* gRNAs) biologically independent mice. In panels **c**-**e**, exact *P* values are shown. See also Source Data.



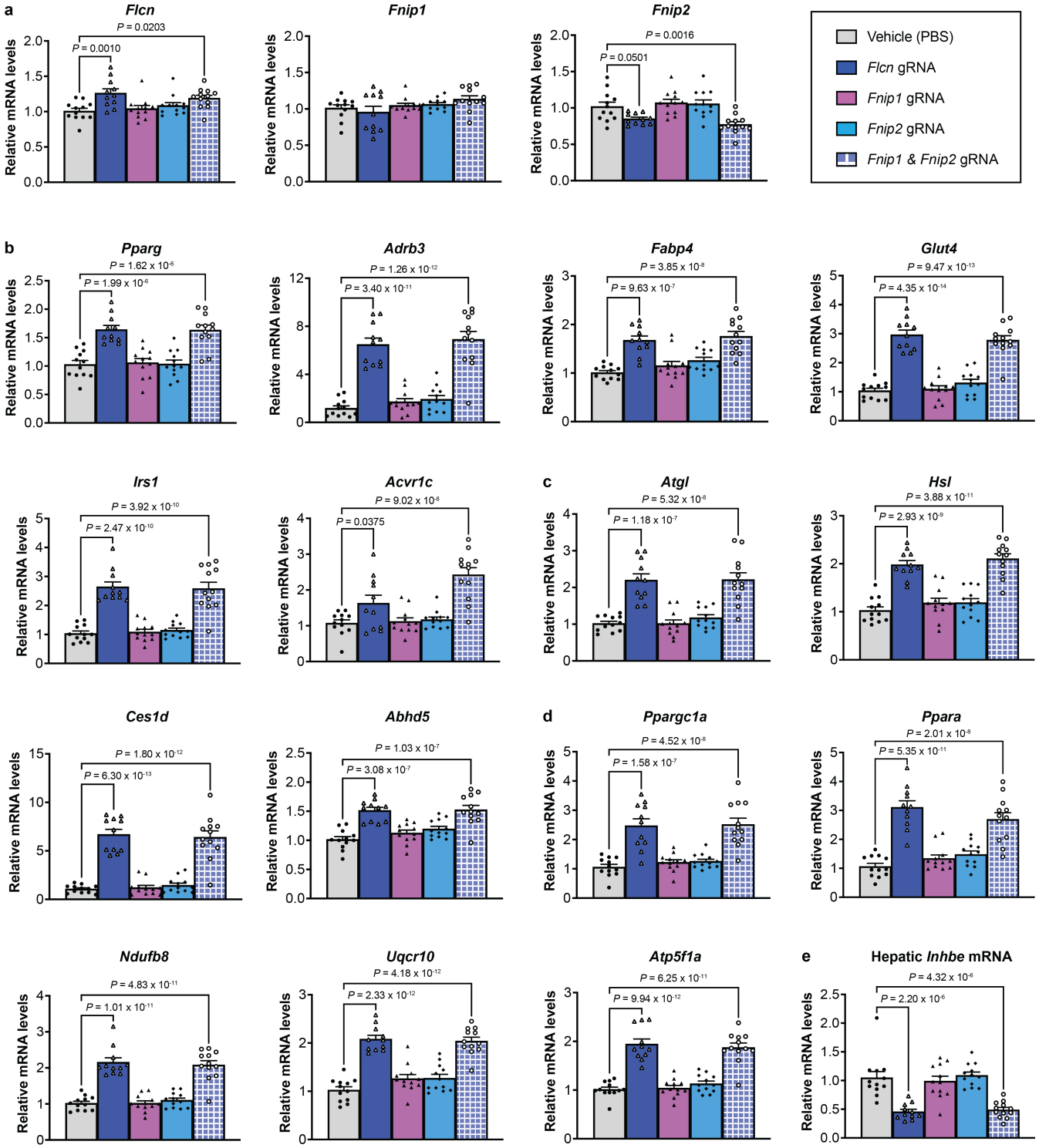
Extended Data Fig. 5 | *FNIP1* siRNA knockdown promotes TFEB pathway activation in primary human hepatocytes. mRNA levels of *FLCN*, *FNIP2* and *RRAGD* (**a**), lysosomal (**b**) and lipid catabolic (**c**) genes in primary human hepatocytes 96 h after *FNIP1* siRNA (siFNIP1) or non-targeting control siRNA (siNeg) transfection in primary human hepatocytes (n = 6 biological replicates per group). Mean ± s.e.m. are shown. *P* values are from unpaired, two-tailed t-test. Exact *P* values are shown in the corresponding panels. **d**, Immunoblot analysis of human primary hepatocyte cell lysates probed with anti-FNIP1 or anti-HSP90 (loading control) antibodies. See also Source Data and gel source data in Supplementary Fig. 1. The knockdown experiments in primary human hepatocytes were independently performed twice and produced similar results.



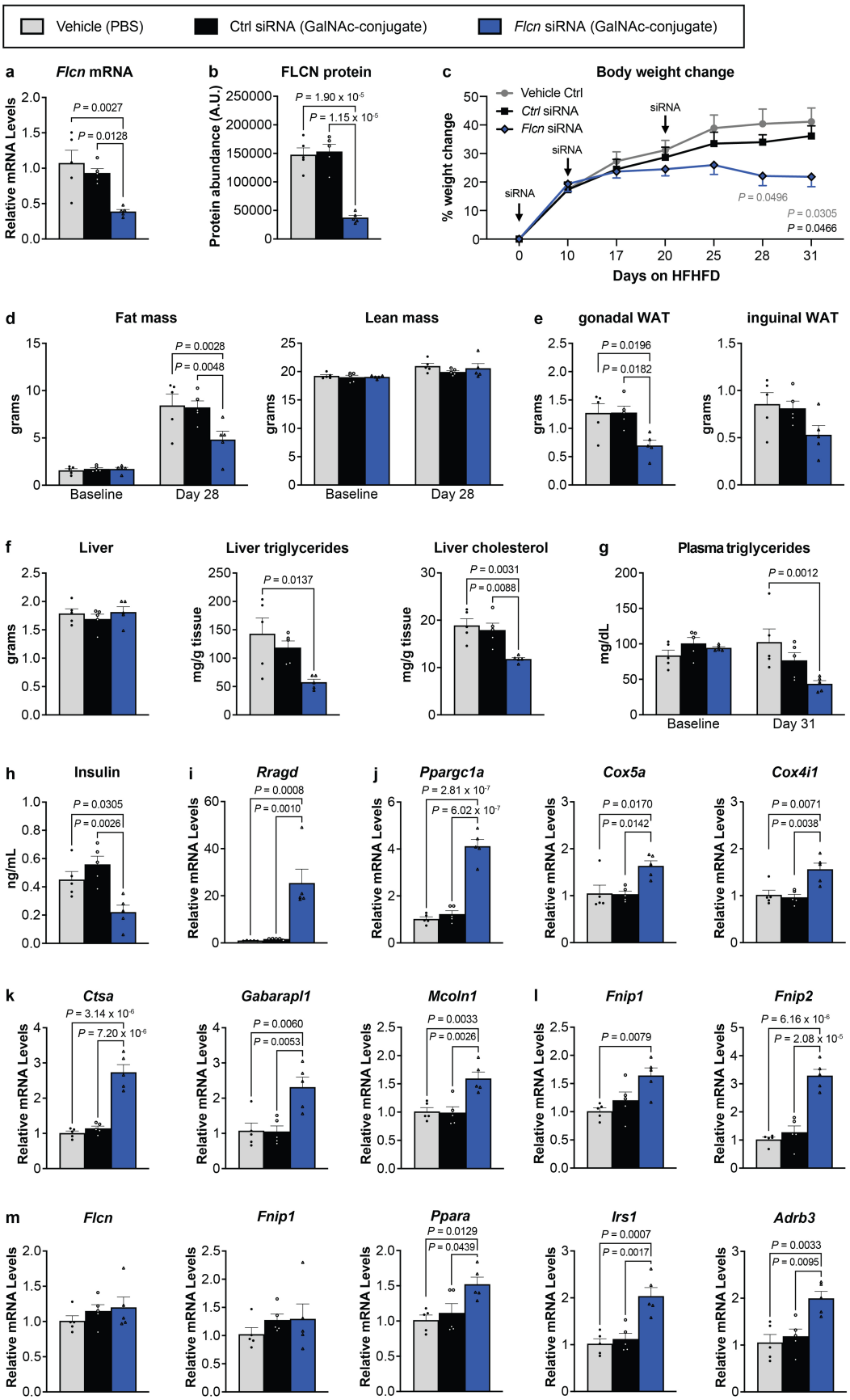
Extended Data Fig. 6 | *Flcn* or *Fnip1 & Fnip2* inhibition in liver of *Cas9-Ready* mice increases expression of mitochondrial, lysosomal and lipid metabolic genes. a, Hepatic mRNA levels of genes involved in mitochondrial biogenesis and oxidative phosphorylation. **b**, mRNA levels of fatty acid transporter *Cd36* and triglyceride lipase *Atgl* in liver. **c**, Expression of genes related to lysosomal function in liver. For all mRNA levels, mean ± s.e.m. are shown. Data are from n = 12 (vehicle control), n = 11 (*Flcn* gRNA), n = 12 (*Fnip1* gRNA), n = 12 (*Fnip2* gRNA) and n = 12 (*Fnip1* & *Fnip2* gRNAs) biologically independent mice. *P* values are from one-way ANOVA with Dunnett's correction posttest relative to vehicle control. Exact *P* values are shown in the corresponding panels. See also Source Data.



Extended Data Fig. 7 | *Flcn* or *Fnip1&Fnip2* inhibition in liver protects against diet-induced obesity, liver damage and insulin resistance. a-d, Tissue weights of *Cas9-Ready* mice 13 weeks following AAV-gRNA administration and while on high fat, high fructose diet throughout the study. Weights of gonadal white adipose tissue (**a**), inguinal white adipose tissue (**b**), liver (**c**) and gastrocnemius muscle (**d**). Mean ± s.e.m. are shown. *P* values are from one-way ANOVA with Dunnett's correction posttest relative to vehicle control. **e**, Body composition of male *Cas9-Ready* mice was determined by echoMRI before (week 0) and 13 weeks post-AAV-gRNA administration and while on high fat, high fructose diet. Mean ± s.e.m. are shown. *P* values are from two-way ANOVA with Dunnett's correction posttest relative to vehicle control. **f**, Plasma alanine aminotransferase (ALT) of *Cas9-Ready* mice were determined at baseline, 6 and 13 weeks following AAV-gRNA administration and while on high fat, high fructose diet. Mean ± s.e.m. are shown. *P* values are from two-way ANOVA with Dunnett's correction posttest relative to vehicle control. **g**, Insulin tolerance test in *Cas9-Ready* mice performed at 8 weeks after AAV-gRNAs or vehicle administration and while on high fat, high fructose diet. Mean ± s.e.m. are shown. *P* values are from two-way ANOVA with Dunnett's correction posttest relative to vehicle control (colour of *P* values indicates relevant group). For all measurements, data are from n = 12 (vehicle control), n = 11 (*Flcn* gRNA), n = 12 (*Fnip1* gRNA), n = 12 (*Fnip2* gRNA) and n = 12 (*Fnip1* & *Fnip2* gRNAs) biologically independent mice, and exact *P* values are shown in the corresponding panels. See also Source Data.

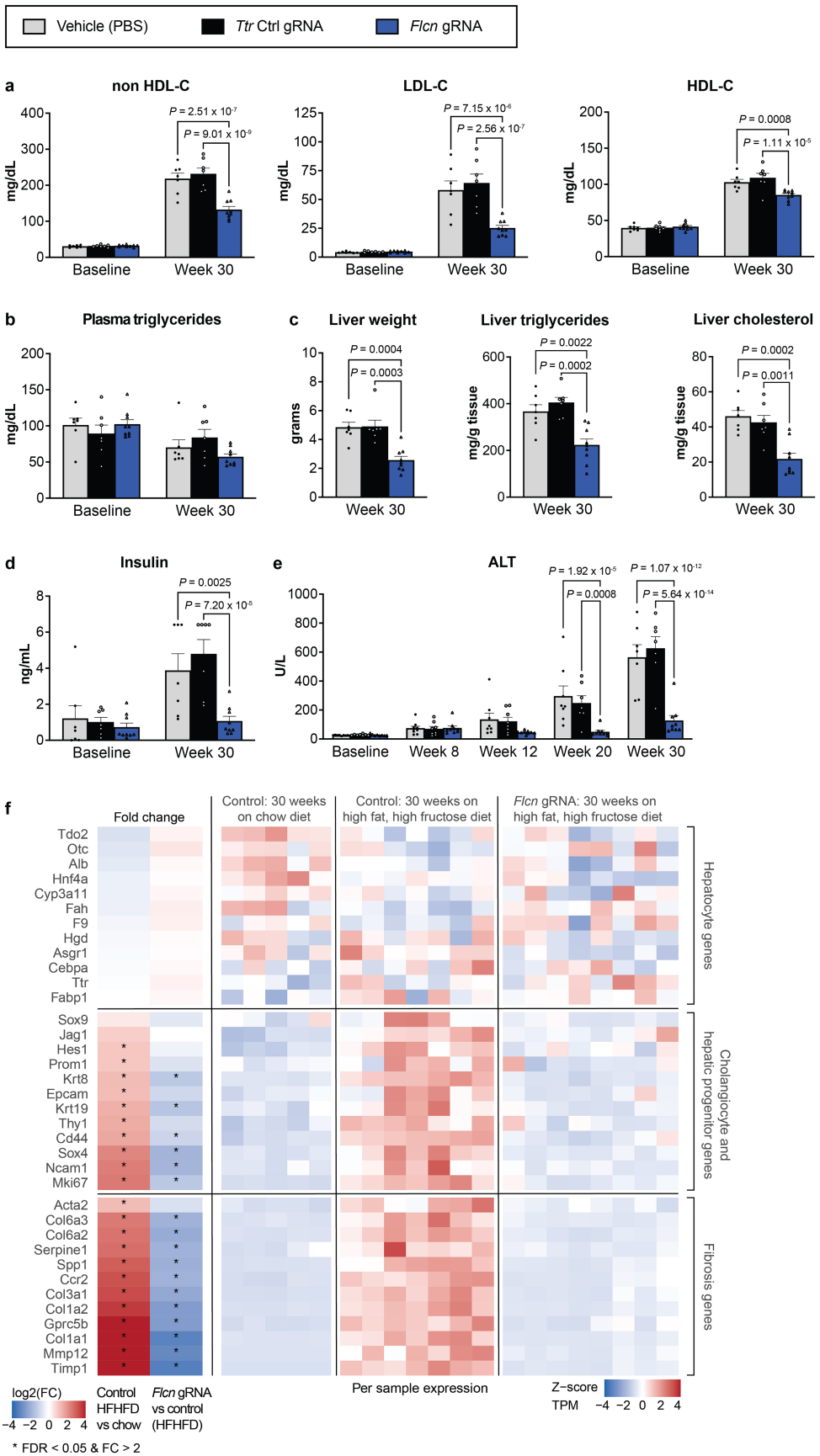


Extended Data Fig. 8 | Hepatic *Flcn* or *Fnip1* & *Fnip2* inhibition alters metabolic gene expression in white adipose tissue. mRNA levels of *Flcn*, *Fnip1* and *Fnip2* (**a**), *Pparg* and its target genes (**b**), lipolysis genes (**c**) and lipid catabolic genes (**d**) in gonadal white adipose tissue. **e**, Hepatic mRNA levels of *Inhbe*, a gene involved in liver-adipose cross-talk. For all mRNA levels, mean ± s.e.m. are shown. *P* values are from one-way ANOVA with Dunnett's correction posttest relative to vehicle control. Data are from n = 12 (vehicle control), n = 11 (*Flcn* gRNA), n = 12 (*Fnip1* gRNA), n = 12 (*Fnip2* gRNA) and n = 12 (*Fnip1* & *Fnip2* gRNAs) biologically independent mice, and exact *P* values are shown in the corresponding panels. See also Source Data.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Hepatocyte inhibition of *Flcn* using a GalNAc-siRNA phenocopies the effects of *Flcn* inhibition using AAV8-gRNA. Male WT mice were dosed with hepatocyte-targeting GalNAc-conjugated *Flcn* or control siRNA and challenged with high fat, high fructose diet for 31 days. **a-b**, Hepatic *Flcn* mRNA (**a**) and FLCN protein (**b**, determined by LC-MS) levels on day 31. Mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from one-way ANOVA with Tukey's correction posttest relative to vehicle control. **c**, Body weight % change of 11–15 week old male siRNA or vehicle (PBS)-treated mice. Mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from two-way ANOVA with Tukey's correction posttest relative to vehicle control. **d**, Fat and lean mass was determined by echoMRI at baseline and 28 days post-siRNA administration. Mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from two-way ANOVA with Tukey's correction posttest relative to vehicle control. **e**, Adipose tissue weights at day 31. Mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from one-way ANOVA with Tukey's correction posttest relative to vehicle control. **f**, Liver weights and liver lipid content at day 31. Mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from one-way ANOVA with Tukey's correction posttest relative to vehicle control. **g**, Plasma triglyceride levels were determined at baseline and 31 days post-siRNA administration. Mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from two-way ANOVA with Tukey's correction posttest relative to vehicle control. **h**, Plasma insulin levels at day 31. Mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from one-way ANOVA with Tukey's correction posttest relative to vehicle control. **i-l**, Hepatic mRNA levels of TFEB target gene *Rragd* (**i**), genes related to oxidative phosphorylation (**j**), lysosomal genes (**k**) and *Fnip1* and *Fnip2* (**l**). **m**, mRNA levels of *Flcn*, *Fnip1* and metabolic genes in gonadal white adipose tissue. For all mRNA levels, mean ± s.e.m. are shown (n = 5 mice per group). *P* values are from one-way ANOVA with Tukey's correction posttest relative to vehicle control. Exact *P* values are shown in the corresponding panels. See also Source Data.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Long-term inhibition of hepatic *FLCN* protects against diet-induced hyperlipidemia, insulin resistance and liver damage. Non-fasted plasma non HDL-C, LDL-C, HDL-C (**a**) and triglycerides (**b**) of *Cas9-ready* mice at baseline and 30 weeks after AAV8-gRNA or vehicle administration and while on high fat, high fructose diet (HFHFD) throughout the study. **c**, Liver weights and hepatic triglyceride and cholesterol content were measured after 30 weeks on HFHFD. **d**, Non-fasted plasma insulin of *Cas9-ready* mice at baseline and 30 weeks after AAV8-gRNA or vehicle administration and while on HFHFD. **e**, Plasma ALT levels were determined at baseline, 8, 12, 20 and 30 weeks after AAV8-gRNA or vehicle administration and while on HFHFD. In panels **a**-**e**, mean ± s.e.m are shown from n = 7 (vehicle), n = 7 (*Ttr* Ctrl gRNA) and n = 9 (*Flcn* gRNA) biologically independent mice. *P* values are from one-way (liver phenotypes) or two-way ANOVA (plasma measurements) with Tukey correction posttest relative to vehicle or control gRNA. In panels **a**-**e**, exact *P* values are shown. **f**, Heatmap of bulk RNA-seq expression of representative hepatocyte, cholangiocyte/hepatic progenitor and fibrosis genes in livers of 30 week chow-fed mice, 30 week HFHFD-fed mice, and mice treated with *Flcn* gRNA and on 30 week HFHFD. Data are from n = 5 (chow), n = 7 (Control HFD) and n = 8 (*Flcn* gRNA on HFD) biologically independent mice. (left), gene fold changes between control HFHFD vs chow groups, and *Flcn* gRNA vs control HFHFD groups. * false discovery rate <0.05 and fold change >2. (right), per sample expression of each gene. Expression TPM values were z-score scaled across each gene (rows). FC, fold change; TPM, transcripts per million. See also Source Data.

Extended Data Table 1 | Gene-burden associations with TG:HDL ratio in the exome-wide discovery analysis

Gene burden genotype, allele frequency cut-off	Cauchy-corrected discovery gene *P*	Best gene burden genotype				
		Beta (95% CI) in ratio units	Beta (95% CI) in SD units	*P*	Genotype Count (RR I RA I AA)	AAF
APOC3 pLOF plus deleterious missense (1/5), AAF<1%	2.4×10^{-1752}	−3.62 (−3.70, −3.54)	−1.05 (−1.07, −1.03)	1.2×10^{-1753}	1,025,695 / 6,412 / 9	0.31%
APOB pLOF plus deleterious missense (5/5), AAF<0.01%	1.6×10^{-330}	−3.16 (−3.32, −3.00)	−0.92 (−0.96, −0.87)	1.1×10^{-330}	1,030,577 / 1,539 / 0	0.08%
ANGPTL3 pLOF plus deleterious missense (ESM1vp), AAF<1%	3.0×10^{-143}	−1.16 (−1.25, −1.07)	−0.34 (−0.36, −0.31)	1.2×10^{-144}	1,027,173 / 4,941 / 2	0.24%
APOA5 pLOF plus deleterious missense (ESM1vp), AAF<0.1%	3.2×10^{-135}	2.63 (2.42, 2.84)	0.76 (0.70, 0.82)	2.7×10^{-136}	1,031,215 / 898 / 3	0.04%
LPL pLOF plus deleterious missense (1/5), AAF<0.1%	1.9×10^{-130}	1.02 (0.93, 1.10)	0.29 (0.27, 0.32)	8.0×10^{-132}	1,026,216 / 5,893 / 7	0.29%
ANGPTL4 pLOF plus deleterious missense (ESM1vp), AAF<0.1%	7.0×10^{-126}	−1.13 (−1.22, −1.03)	−0.33 (−0.35, −0.30)	2.9×10^{-127}	1,027,494 / 4,618 / 4	0.22%
LCAT pLOF plus deleterious missense (1/5), AAF<1%	4.2×10^{-113}	1.07 (0.98, 1.16)	0.31 (0.28, 0.34)	2.0×10^{-114}	1,027,512 / 4,597 / 7	0.22%
PDE3B pLOF plus deleterious missense (ESM1vp), AAF<1%	4.2×10^{-111}	−1.19 (−1.29, −1.08)	−0.34 (−0.37, −0.31)	2.4×10^{-112}	1,028,424 / 3,691 / 1	0.18%
PLA2G12A pLOF plus deleterious missense (1/5), AAF<1%	1.0×10^{-107}	0.56 (0.51, 0.61)	0.16 (0.15, 0.18)	4.3×10^{-109}	1,015,892 / 16,170 / 54	0.79%
ANGPTL8 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	2.3×10^{-87}	−1.22 (−1.34, −1.10)	−0.36 (−0.39, −0.32)	1.7×10^{-88}	1,029,405 / 2,708 / 3	0.13%
AC126283.2 pLOF plus deleterious missense (1/5), AAF<1%	1.7×10^{-73}	0.42 (0.37, 0.46)	0.12 (0.11, 0.13)	7.1×10^{-75}	1,011,985 / 20,071 / 60	0.98%
CREB3L3 pLOF plus deleterious missense (ESM1vp), AAF<0.1%	3.3×10^{-72}	1.13 (1.00, 1.25)	0.33 (0.29, 0.36)	2.9×10^{-73}	1,029,471 / 2,644 / 1	0.13%
CD36 pLOF plus deleterious missense (5/5), AAF<1%	2.0×10^{-68}	−0.41 (−0.45, −0.36)	−0.12 (−0.13, −0.10)	1.0×10^{-69}	1,012,166 / 19,897 / 53	0.97%
GCKR pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	4.9×10^{-58}	0.38 (0.33, 0.42)	0.11 (0.10, 0.12)	4.1×10^{-59}	1,012,490 / 19,591 / 35	0.95%
APOA1 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<0.1%	6.1×10^{-55}	1.78 (1.56, 2.00)	0.52 (0.45, 0.58)	1.0×10^{-55}	1,031,324 / 792 / 0	0.04%
SCARB1 pLOF plus deleterious missense (ESM1vp), AAF<1%	2.3×10^{-54}	−0.63 (−0.71, −0.55)	−0.18 (−0.21, −0.16)	1.4×10^{-55}	1,025,819 / 6,282 / 15	0.31%
CETP pLOF plus deleterious missense (ESM1vp), AAF<0.1%	2.5×10^{-53}	−1.01 (−1.14, −0.89)	−0.29 (−0.33, −0.26)	1.1×10^{-54}	1,029,706 / 2,408 / 2	0.12%
ABCA1 pLOF plus deleterious missense (5/5), AAF<0.1%	1.1×10^{-51}	0.51 (0.45, 0.58)	0.15 (0.13, 0.17)	4.6×10^{-53}	1,022,875 / 9,229 / 12	0.45%
PLIN1 pLOF plus deleterious missense (ESM1vp), AAF<1%	3.0×10^{-41}	−0.59 (−0.68, −0.51)	−0.17 (−0.20, −0.15)	1.4×10^{-42}	1,026,616 / 5,491 / 9	0.27%
PPARG pLOF plus deleterious missense (5/5 or ESM1vp), AAF<0.01%	1.4×10^{-37}	2.79 (2.37, 3.22)	0.81 (0.69, 0.93)	1.5×10^{-38}	1,031,896 / 220 / 0	0.01%
TM6SF2 pLOF plus deleterious missense (1/5), AAF<0.1%	2.3×10^{-34}	−0.42 (−0.49, −0.35)	−0.12 (−0.14, −0.10)	9.8×10^{-36}	1,023,070 / 9,037 / 9	0.44%
PLTP pLOF plus deleterious missense (ESM1vp), AAF<0.1%	4.9×10^{-29}	0.67 (0.56, 0.79)	0.20 (0.16, 0.23)	2.0×10^{-30}	1,029,166 / 2,948 / 2	0.14%
G6PC pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	7.4×10^{-29}	0.56 (0.46, 0.65)	0.16 (0.13, 0.19)	3.1×10^{-30}	1,027,803 / 4,312 / 1	0.21%
TIMD4 pLOF plus deleterious missense (ESM1vp), AAF<1%	1.1×10^{-24}	0.69 (0.56, 0.82)	0.20 (0.16, 0.24)	5.8×10^{-26}	1,029,754 / 2,359 / 3	0.11%
POR pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	1.2×10^{-22}	−0.25 (−0.29, −0.20)	−0.07 (−0.09, −0.06)	1.0×10^{-23}	1,014,685 / 17,409 / 22	0.85%
NR1H3 pLOF plus deleterious missense (ESM1vp), AAF<0.1%	9.7×10^{-21}	−0.62 (−0.75, −0.49)	−0.18 (−0.22, −0.14)	8.1×10^{-22}	1,029,686 / 2,427 / 3	0.12%
A1CF pLOF plus deleterious missense (1/5), AAF<1%	1.3×10^{-19}	0.20 (0.15, 0.24)	0.06 (0.04, 0.07)	5.4×10^{-21}	1,008,049 / 23,988 / 79	1.20%
INSR pLOF plus deleterious missense (ESM1vp), AAF<0.1%	5.7×10^{-19}	−0.84 (−1.02, −0.66)	−0.24 (−0.30, −0.19)	5.3×10^{-20}	1,030,919 / 1,197 / 0	0.06%
SH2B3 pLOF plus deleterious missense (5/5), AAF<1%	1.8×10^{-18}	0.30 (0.24, 0.37)	0.09 (0.07, 0.11)	7.6×10^{-20}	1,022,731 / 9,374 / 11	0.46%
AIFM2 pLOF plus deleterious missense (5/5), AAF<1%	1.0×10^{-16}	0.21 (0.17, 0.26)	0.06 (0.05, 0.08)	4.3×10^{-18}	1,015,436 / 16,630 / 50	0.81%
SLC22A5 pLOF plus any deleterious missense, AAF<0.1%	1.2×10^{-15}	−0.21 (−0.26, −0.16)	−0.06 (−0.08, −0.05)	5.1×10^{-17}	1,016,066 / 16,030 / 20	0.78%
CD300LG pLOF plus deleterious missense (5/5), AAF<0.01%	3.7×10^{-15}	1.15 (0.87, 1.42)	0.33 (0.25, 0.41)	5.7×10^{-16}	1,031,610 / 505 / 1	0.03%
SLC25A30 pLOF plus deleterious missense (ESM1vp), AAF<1%	1.3×10^{-14}	0.23 (0.18, 0.29)	0.07 (0.05, 0.08)	1.1×10^{-15}	1,020,131 / 11,953 / 32	0.58%
SMAD6 pLOF plus deleterious missense (5/5), AAF<0.01%	1.4×10^{-13}	−0.43 (−0.54, −0.32)	−0.12 (−0.16, −0.09)	1.2×10^{-14}	1,028,813 / 3,303 / 0	0.16%
GAS6 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<0.1%	5.2×10^{-13}	0.40 (0.30, 0.50)	0.12 (0.09, 0.15)	5.3×10^{-14}	1,028,512 / 3,599 / 5	0.17%
PKD1 pLOF plus deleterious missense (5/5), AAF<1%	8.4×10^{-13}	0.12 (0.09, 0.15)	0.03 (0.03, 0.04)	5.1×10^{-14}	990,597 / 41,407 / 112	2.00%
PCSK9 pLOF plus deleterious missense (5/5), AAF<1%	1.0×10^{-12}	−0.35 (−0.44, −0.26)	−0.10 (−0.13, −0.07)	5.4×10^{-14}	1,027,317 / 4,792 / 7	0.23%
CPT1A pLOF plus deleterious missense (1/5), AAF<1%	1.1×10^{-12}	−0.14 (−0.18, −0.11)	−0.04 (−0.05, −0.03)	6.2×10^{-14}	1,003,238 / 28,790 / 88	1.40%
SLC33A1 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	3.0×10^{-12}	0.26 (0.19, 0.33)	0.07 (0.05, 0.09)	2.5×10^{-13}	1,023,930 / 8,167 / 19	0.40%
STAB1 pLOF plus deleterious missense (ESM1vp), AAF<1%	5.0×10^{-12}	−0.26 (−0.32, −0.19)	−0.07 (−0.09, −0.05)	2.4×10^{-13}	1,023,697 / 8,411 / 8	0.41%
ADCY6 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	3.5×10^{-11}	0.22 (0.16, 0.28)	0.06 (0.05, 0.08)	2.1×10^{-12}	1,021,741 / 10,367 / 8	0.50%
LIPG pLOF plus deleterious missense (1/5), AAF<1%	6.2×10^{-11}	−0.22 (−0.28, −0.16)	−0.06 (−0.08, −0.05)	2.6×10^{-12}	1,021,876 / 10,225 / 15	0.50%
AXL pLOF plus deleterious missense (5/5 or ESM1vp), AAF<0.1%	1.5×10^{-10}	0.29 (0.20, 0.37)	0.08 (0.06, 0.11)	2.3×10^{-11}	1,026,608 / 5,505 / 3	0.27%
RASD1 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<0.01%	1.7×10^{-10}	0.60 (0.43, 0.78)	0.17 (0.12, 0.23)	1.7×10^{-11}	1,030,840 / 1,276 / 0	0.06%
FNIP1 pLOF, AAF<0.01%	2.9×10^{-10}	−1.66 (−2.17, −1.16)	−0.48 (−0.63, −0.34)	1.1×10^{-10}	1,031,961 / 155 / 0	0.01%
PLIN4 pLOF plus deleterious missense (5/5), AAF<1%	6.2×10^{-10}	0.24 (0.17, 0.31)	0.07 (0.05, 0.09)	7.8×10^{-11}	1,024,923 / 7,049 / 144	0.36%
RIC8B pLOF plus any deleterious missense, AAF<1%	1.1×10^{-09}	0.16 (0.11, 0.20)	0.04 (0.03, 0.06)	9.4×10^{-11}	1,014,135 / 17,908 / 73	0.87%
GLTPD2 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	1.9×10^{-09}	−0.20 (−0.26, −0.14)	−0.06 (−0.08, −0.04)	1.4×10^{-10}	1,021,793 / 10,294 / 29	0.50%
HGFAC pLOF plus deleterious missense (ESM1vp), AAF<1%	2.2×10^{-09}	0.20 (0.14, 0.25)	0.06 (0.04, 0.07)	1.1×10^{-10}	1,020,941 / 11,121 / 54	0.54%
PCK1 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<1%	3.1×10^{-09}	−0.13 (−0.17, −0.09)	−0.04 (−0.05, −0.03)	2.6×10^{-10}	1,008,273 / 23,819 / 24	1.20%
LDLR pLOF plus deleterious missense (5/5), AAF<0.1%	1.2×10^{-08}	0.29 (0.20, 0.38)	0.08 (0.06, 0.11)	9.0×10^{-10}	1,027,611 / 4,499 / 6	0.22%
ZNF285 pLOF plus deleterious missense (1/5), AAF<1%	1.4×10^{-08}	−0.19 (−0.26, −0.13)	−0.06 (−0.07, −0.04)	5.9×10^{-10}	1,021,685 / 10,420 / 11	0.51%
FGB pLOF plus deleterious missense (5/5), AAF<1%	1.7×10^{-08}	0.19 (0.13, 0.25)	0.06 (0.04, 0.07)	1.1×10^{-09}	1,021,663 / 10,433 / 20	0.51%
LIPE pLOF plus deleterious missense (5/5), AAF<1%	1.9×10^{-08}	0.22 (0.15, 0.28)	0.06 (0.04, 0.08)	1.1×10^{-09}	1,023,844 / 8,256 / 16	0.40%
KEAP1 pLOF plus any deleterious missense, AAF<1%	2.3×10^{-08}	0.17 (0.12, 0.23)	0.05 (0.03, 0.07)	1.5×10^{-09}	1,019,823 / 12,273 / 20	0.60%
APOC1 pLOF plus deleterious missense (5/5 or ESM1vp), AAF<0.1%	2.3×10^{-08}	−0.93 (−1.25, −0.62)	−0.27 (−0.36, −0.18)	6.1×10^{-09}	1,026,577 / 392 / 1	0.02%
GALNT2 pLOF plus deleterious missense (5/5), AAF<0.1%	3.2×10^{-08}	0.47 (0.31, 0.62)	0.14 (0.09, 0.18)	4.2×10^{-09}	1,030,495 / 1,619 / 2	0.08%
TNFRSF10B pLOF plus deleterious missense (5/5 or ESM1vp), AAF<0.1%	3.4×10^{-08}	0.32 (0.21, 0.43)	0.09 (0.06, 0.12)	6.0×10^{-09}	1,028,764 / 3,350 / 2	0.16%
ASGR1 pLOF plus deleterious missense (1/5), AAF<1%	4.5×10^{-08}	−0.28 (−0.37, −0.19)	−0.08 (−0.11, −0.05)	4.0×10^{-09}	1,027,561 / 4,551 / 4	0.22%
HELZ2 pLOF plus deleterious missense (ESM1vp), AAF<1%	8.9×10^{-08}	0.10 (0.07, 0.14)	0.03 (0.02, 0.04)	4.9×10^{-09}	999,023 / 33,038 / 55	1.60%

AAF, alternative allele frequency; CI, confidence interval; ESM1vp, Evolutionary Scale Modelling; pLOF, predicted loss-of-function; RR, homozygous for the reference allele; RA, heterozygous; AA, homozygous for the alternative allele.

nature portfolio

Corresponding author(s): Luca A. Lotta

Last updated by author(s): 06/11/2026

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No software or code was used for data collection.
Data analysis	bcl2fastq v2.19.0 (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) BWA MEM v0.7.15 (https://github.com/lh3/bwa) DeepVariant v0.10 (https://github.com/google/deepvariant) GLnexus v1.4.3 (https://github.com/dnanexus-rnd/GLnexus) PLINK 1.9 (https://www.cog-genomics.org/plink2/) Variant Effect Predictor (VEP) v100.4 (https://github.com/Ensembl/ensembl-vep/) REGENIE v3.4 or older (https://github.com/rgcgithub/regenie/) SuSiE v0.12.35 (https://github.com/stephenslab/susieR) STAR v2.5.3a (https://github.com/alexdobin/STAR) Picard v2.9.0 (https://github.com/broadinstitute/picard) RNA-SeQC v1.1.9 (https://github.com/getzlab/rnaseqc) COLOC v5.2.3 (https://github.com/chr1swallace/coloc) DEPICT v1 (https://github.com/perslab/depict) DESeq2 v1.34.0 (https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html) PEAR v0.9.8 (https://cme.h-its.org/exelixis/web/software/pear/doc.html) Bowtie2 v2.5.4 (https://github.com/benlangmead/bowtie2) Prism v10 (https://www.graphpad.com) Illumina BCL Convert v4.3.6 (https://www.10xgenomics.com) FastQC v0.12.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc)

10X Genomics Cell Ranger v7 (https://www.10xgenomics.com/support/software/cell-ranger/latest?gad_source=1)
ArrayStudio v12.5 (https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-omicsoft-suite/)
Scrublet v0.2.3 (https://github.com/swolock/scrublet)
Scanpy v1.9.1 (https://scanpy.scverse.org/en/stable/)
Harmonypy v0.0.10 (https://github.com/slowkow/harmonypy/releases/tag/v0.0.10)
Freestyle software v1.8.65.0 (https://www.thermofisher.com/)
The custom perl script used for the characterization of barcoded samples in the mouse experiments is available for download at: https://rgc-community.regeneron.com/.

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

Data

Policy information about availability of data

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

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

Summary association statistics for the TG-HDL genome-wide discovery analysis are available at https://rgc-community.regeneron.com/. All other association results are reported in the manuscript and its associated Supplementary Materials. Data from the MCPS are available to bona fide researchers. For more details, the study's Data and Sample Sharing policy is available (in English or Spanish) at https://www.ctsu.ox.ac.uk/research/mcps. Study data can be examined through the study's Data Showcase at https://datashare.ndph.ox.ac.uk/mexico/ and MCPS ancestry-specific allele frequencies are available at https://rgc-mcps.regeneron.com. Data from the BELIEVE study are available on reasonable request to the Data Access Committee. Further information on the application process can be found on the BELIEVE study website https://www.believestudy-bangladesh.org/accessing-believe-data/. UKB genotypic and phenotypic data are available to approved investigators via the UK Biobank study (www.ukbiobank.ac.uk/). Additional information about how to access the data are available at https://www.ukbiobank.ac.uk/use-our-data/. GHS data are available to qualified academic noncommercial researchers for reproducibility of data, results and conclusions of the manuscript through the portal at https://regeneron.envisionpharma.com/vt_regeneron/ under a data access agreement. For UCLA-ATLAS (ATLAS) cohort, due to privacy regulations, individual-level data cannot be published. Comprehensive processed data and summary statistics are available in the supplementary tables or through the UCLA-ATLAS browser (atlas-phewas.mednet.ucla.edu). For PMBB, BioMe, DBS, MAYO-RGC and CCPM, academic, non-commercial researchers interested in reproducing the results reported in this manuscript may request access to data via a data access agreement by reaching out to the associated biobank at the following links: PMBB, https://pmbb.med.upenn.edu/contact.php; BioMe, https://mountsinaimillion.org/en/contact; DBS, Contact: JONATHAN.COHEN@UTSOUTHWESTERN.EDU; MAYO-RGC, https://www.mayo.edu/research/centers-programs/mayo-clinic-biobank/for-researchers; CCPM, https://medschool.cuanschutz.edu/cobiobank/contact-us. MDCS data may be available to qualified academic non-commercial researchers to reproduce results reported in this manuscript through the portal at https://www.malmo-kohorter.lu.se/malmo-cohorts, following the principles outlined in this policy https://www.malmo-kohorter.lu.se/sites/malmo-kohorter.lu.se/files/mdcs_mpp_mos_request_form_vermar20.doc. Indiana-CLDB data can be made available for those requesting via Indiana Biobank's data request process. Application is reviewed by three members of the Indiana Biobank's Scientific Review Committee and if approved, made available on a secure-enclave for analysis. South Asia Biobank data may be accessible to researchers upon request to the data access committee. Contact forms and emails are provided on the Global Health Research Unit (GHRU) on Diabetes and Cardiovascular Disease in South Asia website (www.ghru-southasia.org).

Raw values for all experimental data are provided in the Source Data file. RNA-sequencing data have been deposited at Gene Expression Omnibus (GEO) with accession number GSE330407.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Biologic sex phenotype was derived from genetic data and included as covariate as described in the methods section.
Reporting on race, ethnicity, or other socially relevant groupings	Ancestry estimation was derived from genetic data and reported in descriptive statistics of Supplementary Table S1.
Population characteristics	Population characteristics were provided in Supplementary Table S1.
Recruitment	Our study included participants from both population-based and health-system based cohorts and are described in the methods section with relevant references for additional details on recruitment.
Ethics oversight	All studies received approval from the relevant ethics committees, and participants gave their informed consent to take part in the research. More details are provided in the references for respective cohorts in the Methods section.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	-Human studies: Sample size was determined by the maximal set of participants in each cohort with genotypic and phenotypic data after quality control. -Animal Studies: Sample size was chosen based on previous experience with similar experiments, experimental feasibility, availability of samples and N necessary to obtain definitive, significant results. No statistical methods were used to predetermine sample sizes. The sample size information of experiments has been provided in each figure legend.
Data exclusions	-Human Studies: Exclusions within each cohort were done based on missing genotypic and phenotypic data. Also filtering based on established quality control protocols was done. -Animal Studies: No data were excluded from the analysis.
Replication	-Human Studies: No replication analyses were performed. -Animal Studies: All experiments with AAV8-gRNAs (Flcn/Fnip1/Fnip2) on high fat diets for up to 13 weeks, studies in primary human hepatocytes (Fnip1) and studies with AAV-shRNAs (Hpn) were independently performed at least twice on cohorts from independent litters. All attempts at replication were successful. Studies with long-term (30 weeks) inhibition of Flcn/Fnip1/Fnip2, as well as studies on chow, at thermoneutrality or using GalNac-conjugated siRNAs were performed once, but yielded results that were similar to the AAV8-gRNA studies on high fat diets. In some studies, independent methods were used to analyze samples and verify results (e.g. DNA editing, mRNA and protein analysis to confirm gene perturbation).
Randomization	-Animal Studies: For all animal studies, mice were allocated into experimental groups based on matching body weight, body composition and lipid levels. The goal was to balance weight, body composition and lipid levels at baseline (similar values between groups), to allow for evaluation of treatment efficacy following intervention. -Cell culture studies: cells were randomly assigned to experimental groups.
Blinding	-Animal Studies: Investigators were blinded to group allocation during serum chemistry analysis, liver lipid analysis, protein analysis and DNA editing analysis. Blinding was not possible during the in-live part of the mouse studies due to the need to treat mice with different reagents according to experimental designs. In cell culture studies, blinding was not relevant, as the groups needed to be allocated to appropriate experimental and control groups.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	FNIP1 (clone E9Y5U) Rabbit Monoclonal Antibody #36892, Cell Signaling Technologies HSP90 (clone C45G5) Rabbit Monoclonal Antibody #4877 , Cell Signaling Technologies
Validation	The manufacturer provides validation data on the website, evaluating FNIP1 antibody by Western Blot in human cell lines that express or do not express human FNIP1. From the website: "Western Blot analysis of extracts from various cell lines using FNIP1 (E9Y5U) Rabbit mAb or anti-Actinin (D6F6) XP Rabbit mAb #6487. Negative expression of FNIP1 protein in U-937 cells is consistent with the predicted expression pattern." (https://www.cellsignal.com/products/primary-antibodies/fnip1-e9y5u-rabbit-monoclonal-antibody/36892) The HSP90 antibody has been extensively used in publications, with the manufacturer website listing 686 citations (https://www.cellsignal.com/products/primary-antibodies/hsp90-c45g5-rabbit-monoclonal-antibody/4877)

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Cryopreserved Primary Human Hepatocytes from ThermoFisher (Catalog # HMCPTS): Donor HU8450-Male, 16yo, Caucasian
Authentication	Cells were assessed for Phase I enzyme activities, transporter activity and bile canaliculi formation by the manufacturer
Mycoplasma contamination	Cells were not tested for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	The studies involved C57BL/6NTac (WT) mice from Taconic, and Cas9-Ready (ver. 2.5, ID #2673) mice, expressing Cas9 transgene under the CAG promoter (previously reported in Smagris et. al., PLoS Genetics 2024 PMID: 38437227). Mice were between 6-11 weeks old at the beginning of the studies. Mice were housed on a 12h light/dark cycle at 22 ± 1°C or 30°C (for experiments at thermoneutral conditions), at humidity-controlled conditions (30-70% humidity), in static cages (≤5/cage) with free access to food and water.
Wild animals	The study did not involve wild animals.
Reporting on sex	The metabolic studies involved male mice only, which historically have shown more robust metabolic disease phenotypes than females.
Field-collected samples	The study did not involve field-collected samples.
Ethics oversight	All animal procedures were conducted in compliance with protocols approved by the Regeneron Pharmaceuticals Institutional Animal Care and Use Committee.

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*