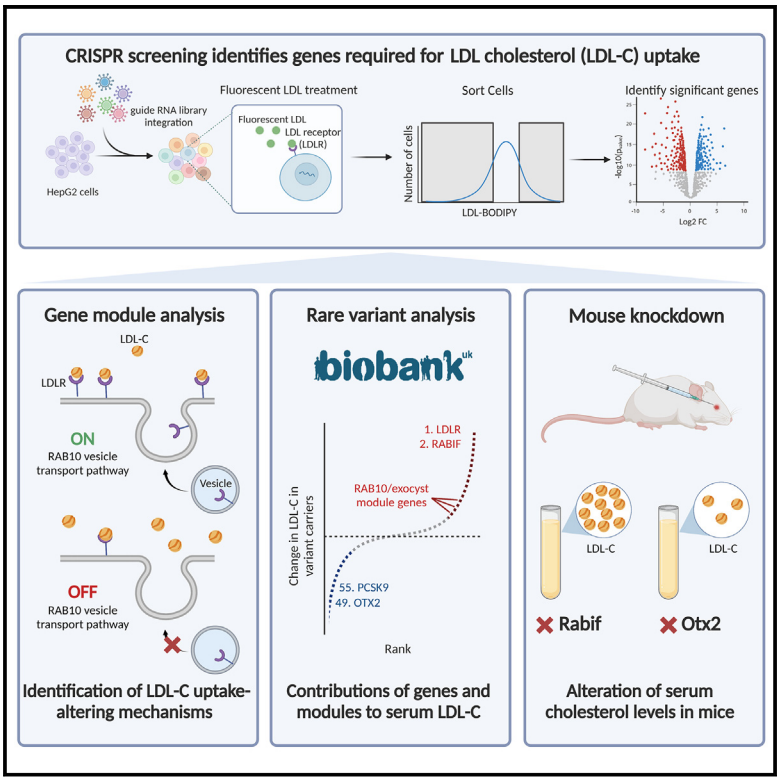
Cell Genomics

Article

Systematic elucidation of genetic mechanisms underlying cholesterol uptake

Graphical abstract



Highlights

- Genome-scale CRISPR screens identify 490 genes that alter liver cell LDL-C uptake
- 21 LDL-C uptake-altering genes enriched in rare coding variants in UK Biobank exomes
- Dysfunction in the RAB10 exocytosis pathway raises LDL-C levels in mice and humans
- *OTX2* loss of function robustly reduces serum LDL-C in mice and humans

Authors

Marisa C. Hamilton, James D. Fife, Ersin Akinci, ..., Haojie Yu, Christopher A. Cassa, Richard I. Sherwood

Correspondence

bchhaoy@nus.edu.sg (H.Y.), ccassa@bwh.harvard.edu (C.A.C.), rsherwood@bwh.harvard.edu (R.I.S.)

In brief

By combining analysis of rare coding variants from the UK Biobank and genome-scale CRISPR-Cas9 knockout and activation screening, Hamilton et al. improve the identification of genes whose disruption alters serum LDL-C levels. They show that RAB10 vesicle transport pathway dysfunction leads to hypercholesterolemia and that *OTX2* disruption reduces serum LDL-C levels.

Hamilton et al., 2023, Cell Genomics *3*, 100304
May 10, 2023
https://doi.org/10.1016/j.xgen.2023.100304

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Systematic elucidation of genetic mechanisms underlying cholesterol uptake

Marisa C. Hamilton,[1] James D. Fife,[1] Ersin Akinci,[1] Tian Yu,[1] Benyapa Khowpinitchai,[1] Minsun Cha,[1] Sammy Barkal,[1] Thi Tun Thi,[2] Grace H.T. Yeo,[3,4] Juan Pablo Ramos Barroso,[1] Matthew Jake Francoeur,[1] Minja Velimirovic,[1] David K. Gifford,[3,4] Guillaume Lettre,[5,6] Haojie Yu,[2,*] Christopher A. Cassa,[1,*] and Richard I. Sherwood[1,7,*]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[2]Precision Medicine Research Programme, Cardiovascular Disease Research Programme, and Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Department of Biological Engineering, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Montreal Heart Institute, Montréal, QC H1T 1C8, Canada
[6]Faculté de Médecine, Université de Montréal, Montréal, QC H3T 1J4, Canada
[7]Lead contact
*Correspondence: bchhaoy@nus.edu.sg (H.Y.), ccassa@bwh.harvard.edu (C.A.C.), rsherwood@bwh.harvard.edu (R.I.S.)
https://doi.org/10.1016/j.xgen.2023.100304

SUMMARY

Genetic variation contributes greatly to LDL cholesterol (LDL-C) levels and coronary artery disease risk. By combining analysis of rare coding variants from the UK Biobank and genome-scale CRISPR-Cas9 knockout and activation screening, we substantially improve the identification of genes whose disruption alters serum LDL-C levels. We identify 21 genes in which rare coding variants significantly alter LDL-C levels at least partially through altered LDL-C uptake. We use co-essentiality-based gene module analysis to show that dysfunction of the RAB10 vesicle transport pathway leads to hypercholesterolemia in humans and mice by impairing surface LDL receptor levels. Further, we demonstrate that loss of function of *OTX2* leads to robust reduction in serum LDL-C levels in mice and humans by increasing cellular LDL-C uptake. Altogether, we present an integrated approach that improves our understanding of the genetic regulators of LDL-C levels and provides a roadmap for further efforts to dissect complex human disease genetics.

INTRODUCTION

Coronary artery disease (CAD) is a common cause of death and disability worldwide.[1] Genetic differences in cholesterol metabolism contribute substantially to CAD risk, along with environmental and clinical risk factors. Serum cholesterol measurements are quantitative and nearly uniformly measured across biobanks, providing human phenotypic data that can aid in the identification of genetic risk factors for CAD. Rare monogenic variants can alter low-density lipoprotein cholesterol (LDL-C) uptake in genes such as *LDLR*, *APOB*, and *PCSK9*.[2] Common, lower-effect-size genetic variants that affect LDL-C levels and CAD risk have also been identified through genome-wide association screens (GWASs).[3–7]

Genetic findings from variants with effects on cholesterol levels have led to the development of FDA-approved drugs targeting PCSK9 and a clinical pipeline targeting ANGPTL3, APOC3, and LPA.[8,9] Nonetheless, the genetics of cholesterol levels are far from completely understood. A trans-ancestry GWAS meta-analysis from the Global Lipids Genetics Consortium (GLGC), using data from 1.65 million subjects, has identified >900 genome-wide significant loci associated with blood lipid levels, including >400 loci associated with LDL-C.[10] Yet, the majority of genes or non-coding RNAs that modulate these loci have not been identified. Associations between deleterious coding variants and cholesterol levels from exome sequencing cohorts, including over 400,000 UK Biobank (UKB) individuals, identified 14 genes with coding variations significantly associated with LDL-C levels.[11–14] Thus, a major disconnect currently remains between the hundreds of loci flagged in GWASs and the paucity of genes shown to drive these associations.

Organismal regulation of LDL-C is accomplished through a complex set of mechanisms. Liver cell LDL uptake is a dominant mechanism controlling serum LDL-C levels, being the primary deficiency in familial hypercholesterolemia patients[15] and the primary target of statins and PCSK9 inhibitors.[16] Cell culture models have been instrumental in dissecting the roles of the sterol-responsive SREBP and LXR pathways,[17,18] which regulate cholesterol biosynthesis and the LDL-C uptake machinery. Cellular uptake of fluorescent LDL-C can identify genetic regulators of cholesterol metabolism in the context of RNAi and CRISPR-Cas9 screening.[19–22] In particular, a genome-wide CRISPR-Cas9 screen in hepatic cell lines identified 163 genes with mutations that alter LDL-C uptake.[22] Yet, how these *in vitro* LDL-C uptake regulators are functionally related in biological pathways, and whether variation in these

genes plays a role in controlling serum LDL-C levels, is poorly understood.

In this work, we combine human biobank coding rare variant analysis, CRISPR screening in human cell models, gene network analysis, and mouse knockdown analysis to dissect the genetics of LDL-C uptake. Using this approach, we find 21 genes that alter LDL-C levels, at least partially, through altered LDL-C uptake. Genes associated with the GTPase RAB10 and exocytosis machinery greatly influence LDL-C levels, defining a novel hypercholesterolemia pathway. In addition, our integrated pipeline identifies genes—including *OTX2*—that decrease serum LDL-C levels upon disruption at least in part through increased LDL-C uptake. In sum, our integrated approach evaluating the results of deleterious coding variants in human cell models through CRISPR screening and in the human population through exome analysis reveals novel genetic contributors to LDL-C levels and provides a roadmap to improved understanding of quantitative human trait genetics.

RESULTS

Genome-scale CRISPR-Cas9-nuclease screening identifies hepatic cell LDL-C uptake-altering genes

Uptake of fluorescently labeled LDL-C by liver cells can assay the genetic contributors to LDL-C levels.[19–23] We used a CRISPR-Cas9 screening platform to determine which human genes govern the uptake of LDL-C in HepG2 hepatocellular carcinoma cells,[24] a model of liver cholesterol metabolism that has a relatively normal karyotype[25] and has been the subject of extensive epigenetic profiling.[26]

To assess fluorescent LDL-C uptake after mutation of known LDL-C uptake-altering genes, we lentivirally infected HepG2 with guide RNAs (gRNAs) targeting *LDLR* and the LDLR-degrading E3 ubiquitin-protein ligase *MYLIP* (a.k.a. IDOL).[27] We assessed LDL-C uptake under serum-starved (SS) and serum-containing (SER) conditions, and we found that Cas9-nuclease mutation of *LDLR* leads to a 3- to 4-fold reduction in LDL-C uptake, while mutation of *MYLIP* leads to up to 1.5- to 1.7-fold increase in LDL-C uptake (Figures 1A and S1A).

We performed a high-throughput CRISPR-Cas9-nuclease screen targeting genes adjacent to variants associated with serum LDL-C from a multi-ancestry GWAS from the Million Veteran Program (MVP) cohort.[3] We targeted 2,634 genes within 1 Mb of the 293 lead genome-wide significant MVP serum LDL-C variants using a gRNA library (Knockout [KO]-Library 1), including four gRNAs per gene predicted by the Cas9 indel prediction algorithm inDelphi[28] to induce frameshifts in >80% of edited alleles, as well as 200 non-targeting control gRNAs.

We performed lentiviral CRISPR-Cas9-nuclease screens using KO-Library 1 in four biological replicates, ensuring >500 cells per gRNA at all stages. We performed screens in SER and SS conditions, as SS HepG2 showed upregulation of a gene set associated with lipid biosynthesis (Table S1).[29] We flow cytometrically isolated four populations per replicate with the lowest 12.5%, next lowest 25%, highest 12.5%, and next highest 25% LDL uptake (Figure 1B), using next-generation sequencing (NGS) on all sorted populations, followed by the MAGeCK RRA pipeline,[30] to identify genes with significant effects on LDL-C uptake.

We found strong replicability in pre-sort gRNA counts (Spearman rank correlation (R) = 0.79–0.91; Figure S1B), appreciable replicability in the ratio of gRNA representation in cells with the highest vs. lowest 37.5% LDL uptake (R = 0.2–0.21; Figures 1C and 1D), and strong replicability among the 100 genes with the strongest effect on LDL-C uptake (R = 0.77–0.79; Figure 1E). MAGeCK analysis identified 37 genes whose disruption significantly alters LDL-C uptake (false discovery rate [FDR] <0.1) under either SS or SER conditions (Figure S1C; Table S2).

To ask whether the resolution to identify significant LDL-C uptake-altering genes is improved by increasing the number of cells screened per gRNA, we designed a smaller gRNA library targeting the 360 genes with the most robust effect on LDL-C uptake in our original screen according to MAGeCK (KO-Library 2). Screening KO-Library 2 using the same protocol as for KO-Library 1 identified 152 genes whose disruption significantly alters LDL-C uptake (FDR <0.1, Figure S1D; Table S2) under either SS or SER conditions and yielded results that are concordant with the larger screen.

Given the added power to identify LDL-C-uptake-altering genes using smaller gRNA libraries, we performed CRISPR-Cas9 screening using two additional gRNA libraries. One library targets 1,803 annotated human transcription factors[31,32] (KO-Library 3) to increase the resolution to identify transcriptional regulators of LDL-C uptake (Figure S1E; Table S2). The final library (KO-Library 4) targets 522 genes with significant or near-significant effects on LDL-C uptake from the three previous screens as well as from a published genome-wide LDL-C uptake CRISPR-Cas9 screen performed in Huh7 cells.[22] Altogether, the four screens identify a total of 490 genes whose knockout alters LDL-C uptake under either SER or SS conditions (299 decrease uptake, 180 increase uptake, 11 do both at FDR <0.1; Figure 1F; Table S2), increasing the number of such genes compared with previously published screens.[21,22,33,34] We observed a positive correlation in the gene-level phenotype across our screens (R = 0.36–0.53 for 229 union gene set; Figure S1F). We find robust but incomplete correlation in the gene-level log2 fold change in screens of SER vs. SS conditions (R = 0.43 for 1,800 genes in KO-Library 3; Figure S1G) and in our HepG2 SS screens vs. previous screens in SS HuH7 cells (R = 0.50 for 203 genes in KO-Library 4; Figure S1H), suggesting that environmental conditions and cell line influence the genetic control of LDL-C uptake.

The screens identify known players in cholesterol uptake and metabolism and flag candidate genes. Mutating known LDL-C uptake-implicated genes such as *LDLR* (the most robust hit), *SREBF2*,[17,35] *HNF4A*,[36] and *SCAP*[17,35] impairs LDL uptake, while mutating *MYLIP*[27] and the cholesterol biosynthesis and processing enzymes *ACAT2* and *SQLE*[35,37] increases LDL uptake (Figure 1F; Table S2). The majority of the significant genes have not been implicated in cholesterol homeostasis, suggesting this dataset can identify novel genes and pathways involved in LDL-C uptake. To address the accuracy of these screening data, we cloned 20 individual gRNAs targeting genes found to affect LDL-C uptake and performed fluorescent LDL-C uptake assays on all 20 knockouts in triplicate. Correlations between changes in LDL-C uptake in our CRISPR screen and individual

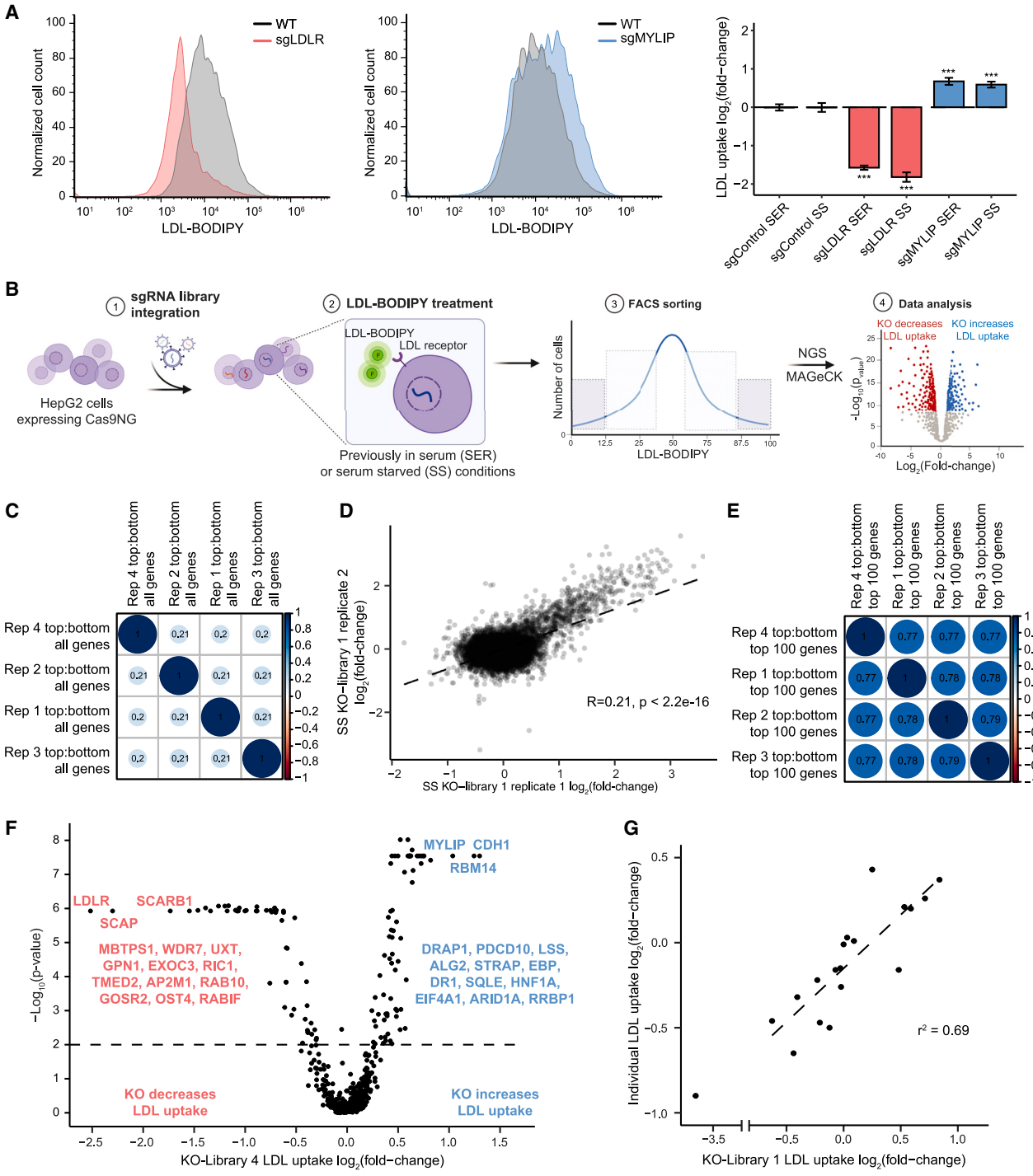


Figure 1. CRISPR screening identifies genes required for LDL-C uptake

(A) Flow cytometric analysis of LDL-C uptake in HepG2 cells with CRISPR-Cas9 knockout of LDLR (red) and MYLIP (blue) under SER and SS conditions; n = 10. Two-tailed t test (***p $\leq$ 0.001).

(B) Schematic of LDL-C uptake CRISPR-Cas9 knockout screening.

(C–E) Spearman correlation and p value of the ratio of representation of all gRNAs (C and D) or gRNAs in the 100 genes with strongest effects on LDL-C uptake (E) in cells with top 37.5% vs. bottom 37.5% LDL-C uptake in KO-Library 1 under SS conditions.

(F) Volcano plot showing MAGeCK LDL-C uptake log2 fold change (x axis) and minimum MAGeCK −log10 p value (y axis) for the 522 genes in KO-Library 4 under SS conditions. Several significant genes are listed; n of permutations = 100,000.

(G) Comparison of LDL-C uptake log2 fold change for 20 genes between KO-Library 1 (x axis) and individual gRNA knockout testing (y axis).

validation (R^2 = 0.69; Figures 1G and S1I) confirm the accuracy of the CRISPR screening results.

CRISPR activation screening reveals tunable regulators of LDL-C uptake

To determine which genes alter LDL-C uptake upon overexpression, we performed systematic upregulation of 200 LDL-C uptake-altering genes using CRISPR activation (CRISPRa) screening. To maximize CRISPR-based gene upregulation, we employed the SunTag system,[38] fusing ScFvGcn4 with a combination of human activation domains. We also recruited two transactivators, p65 and HSF1, to the target site through fusion with the MS2 coat protein (MCP)[39] (Figure S2A). This combined CRISPRa platform leads to significantly stronger gene activation compared with the individual SunTag and MCP components (Figure S2B).

In addition to maximizing the potency of the CRISPRa platform, we characterized the reliability of CRISPRa screening. To measure the variable efficiency of gRNAs in CRISPRa screening, we developed and employed the CRISPRa Outcome and Phenotype (COP) platform. We designed a high-throughput pooled lentiviral CRISPRa library in which each promoter-targeted gRNA has two targets: a target in an adjacent barcoded reporter construct containing a 205-nt proximal promoter sequence and a native genomic target in the promoter of a target gene (Figure 2A). For each of 200 target genes, we cloned eight on-target gRNAs predicted to have high activity,[40] as well as two non-targeting control gRNAs. After transduction of HepG2 cells stably expressing dCas9-10xGcn4 and ScFvGcn4-SBNO1-NFE2L1-KRT40 in three biological replicates, we sorted the cells with the top and bottom 30% cutoffs for LDL-C uptake. In addition, we collected RNA to assess the magnitude of promoter reporter activation upon gRNA treatment through the transcribed barcode (Figure S2A).

The majority of gRNAs were active in CRISPRa, with on-target gRNAs inducing significantly higher reporter activity than control gRNAs (Figure 2B; Table S3). Furthermore, 75% of on-target gRNAs produced stronger reporter activity than the median promoter-matched control gRNA (Figure 2B; Table S3). gRNA cleavage efficiency predictors,[41,42] gRNA binding strand, and gRNA position showed significant but modest correlations with CRISPRa activity (Figure S2C).

We evaluated the endogenous activation strength of 15 gRNAs targeting four cell-surface proteins that were also evaluated in the COP screen, using flow cytometric antibody staining to measure CRISPRa potency. The correlation between COP-measured activation and endogenous activation was high (R = 0.81; Figures 2C and S2D). In addition, reporter activity was significantly, albeit incompletely, correlated with phenotypic impact on LDL uptake for on-target gRNAs targeting the 10 genes with strongest phenotypes (R = 0.38; Figure S2E).

We observed that the ratio of gRNA representation in top- vs. bottom-sorted cells was replicate-consistent (R = 0.43–0.69; Figure S2F). MAGeCK analysis showed that upregulation of 32 genes significantly altered LDL-C uptake, with *LDLR* upregulation most robustly increasing and *MYLIP* upregulation most robustly decreasing uptake (Figure 2D; Table S4). Comparing the effects of gene knockout with upregulation, significant anti-correlation (R = −0.31; Figure 2E) suggests that LDL-C uptake is a tunable process controlled by a set of core regulators.

Gene module analysis elucidates LDL-C uptake-altering mechanisms

We employed co-essentiality analysis to group genes into functionally coherent modules. Utilizing genome-wide CRISPR cell essentiality screen datasets (DepMap),[43] genes with similar functions can be clustered into modules, as genes with similar cellular functions will affect fitness in the same cell lines.[44–48] We combined features of co-essentiality pipelines,[47,48] using ClusterOne[49] (d = 0.2) to group nearly twice as many genes into functional modules as Gene Ontology (GO) and search tool for the retrieval of interacting genes/proteins (STRING). We then annotated each module using overlapping GO modules for interpretability, finding that 85% of modules have a significantly enriched GO term (STAR Methods; Table S5).

Employing co-essential module enrichment on the 490 significant LDL uptake-altering genes, we identified 33 modules significantly associated with LDL-C uptake (STAR Methods; Table S5). The most significantly associated modules contain genes associated with vesicle-mediated transport, protein N-linked glycosylation, lipid biosynthesis, tRNA processing, and translation initiation (Figures 3A and S3A). Most modules comprise a mix of genes annotated to be involved with a cellular function and unannotated genes.[48] To evaluate whether unannotated genes contribute to the annotated cellular function, we explored module 146, whose most significant GO association is in protein N-linked glycosylation. We measured binding of two N-glycan-binding lectins, WGA and ECL, on the surface of HepG2 cells through flow cytometry. We confirmed significantly reduced lectin binding upon knockout of two GO-annotated N-glycosylation genes, *OST4* and *STT3A*, as well as four of five genes that are not annotated as N-glycosylation genes by GO (Figure S3B). RNA sequencing (RNA-seq) of HepG2 cells with knockout of one such unannotated gene, *HYOU1*, reveals compensatory upregulation of co-essential modules involved in N-glycosylation (Figure S3C; Table S1).

We also identified 13 significantly associated modules related to vesicle transport (Figures 3B, 3C, and S3D). The most significant, module 255, for which 11 of the 14 included genes significantly alter LDL-C uptake, contains RAB10 GTPase and exocytosis machinery (Figures 3B and 3C). RAB10 is a GTPase implicated in numerous trafficking processes, including Golgi-to-membrane and endosomal transport;[50] RABIF is a RAB10 holdase/chaperone that is required for RAB10 stability;[51] DENND4C is a RAB10 guanine exchange factor (GEF);[52] and RALGAPB is a RAB10 GTPase-activating protein (GAP). Module 255 also contains seven members of the exocyst complex, which is required for RAB10-mediated vesicle fusion with the plasma membrane,[53] as well as three SNARE complex proteins known to be involved in exocytosis.[53] Altogether, this cluster implicates exocytosis in LDL-C uptake, in line with previous findings.[22,54]

Beyond this exocytosis module, co-essential clustering allows dissection of the multiple roles of vesicle trafficking in LDL-C uptake (Figure S3D),[55] highlighting the multivesicular body (module 514), the WASH and CCC endosome sorting complexes[56] (modules 11 and 277), endosome recycling (modules 476 and 165),

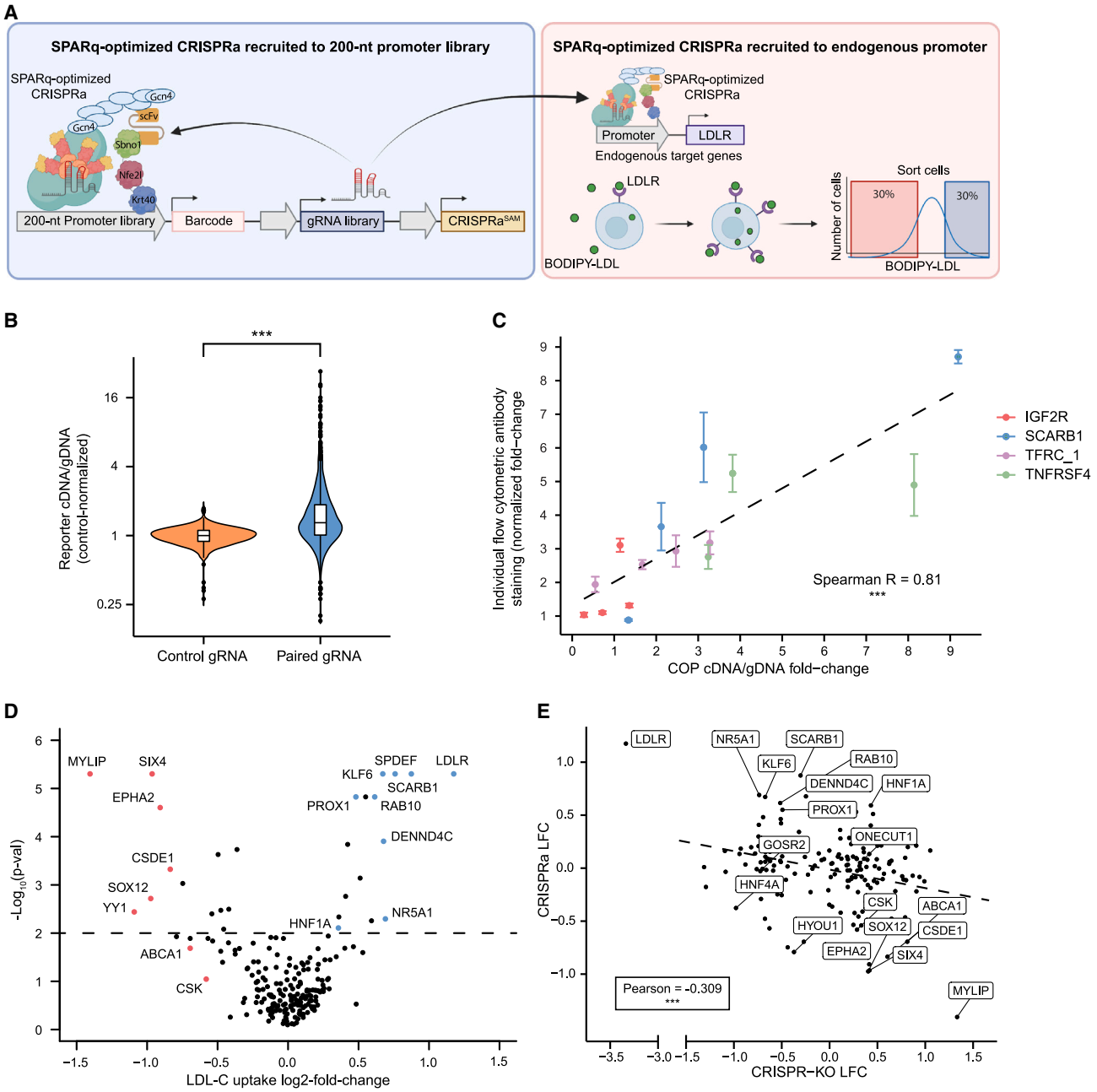


Figure 2. CRISPR activation screening reveals tunable regulators of LDL-C uptake

(A) Schematic of CRISPRa Outcome and Phenotype (COP) screening approach.

(B) Reporter cDNA:gDNA ratio of all promoter-paired gRNAs vs. control gRNAs, normalized by the average cDNA:gDNA ratio of control gRNAs for each gene; Mann-Whitney U test.

(C) Spearman correlation and p value of individual flow cytometric cell surface antibody staining and reporter cDNA:gDNA ratio; n = 2–4.

(D) Volcano plot showing MAGeCK LDL-C uptake log2 fold change (x axis) and minimum MAGeCK −log10 p value (y axis) for the 200 genes in the CRISPRa library. Several significant genes are listed.

(E) Comparison of LDL-C uptake log2 fold change for 200 genes targeted by CRISPR-KO (x axis) and CRISPRa (y axis); two-tailed t test (***p $\leq$ 0.001).

vesicle ATPase acidification (modules 229 and 514), and ER-Golgi trafficking (modules 98, 372, and 461), in addition to the exocyst complex and polarized transport machinery (modules 255 and 69).[22,57]

To assess the roles of co-essential modules on LDL-C uptake, we identified 10 small molecules that modulate the function of significant genes and pathways from our screens (Table S22). Eight of the 10 molecules significantly alter LDL-C uptake in

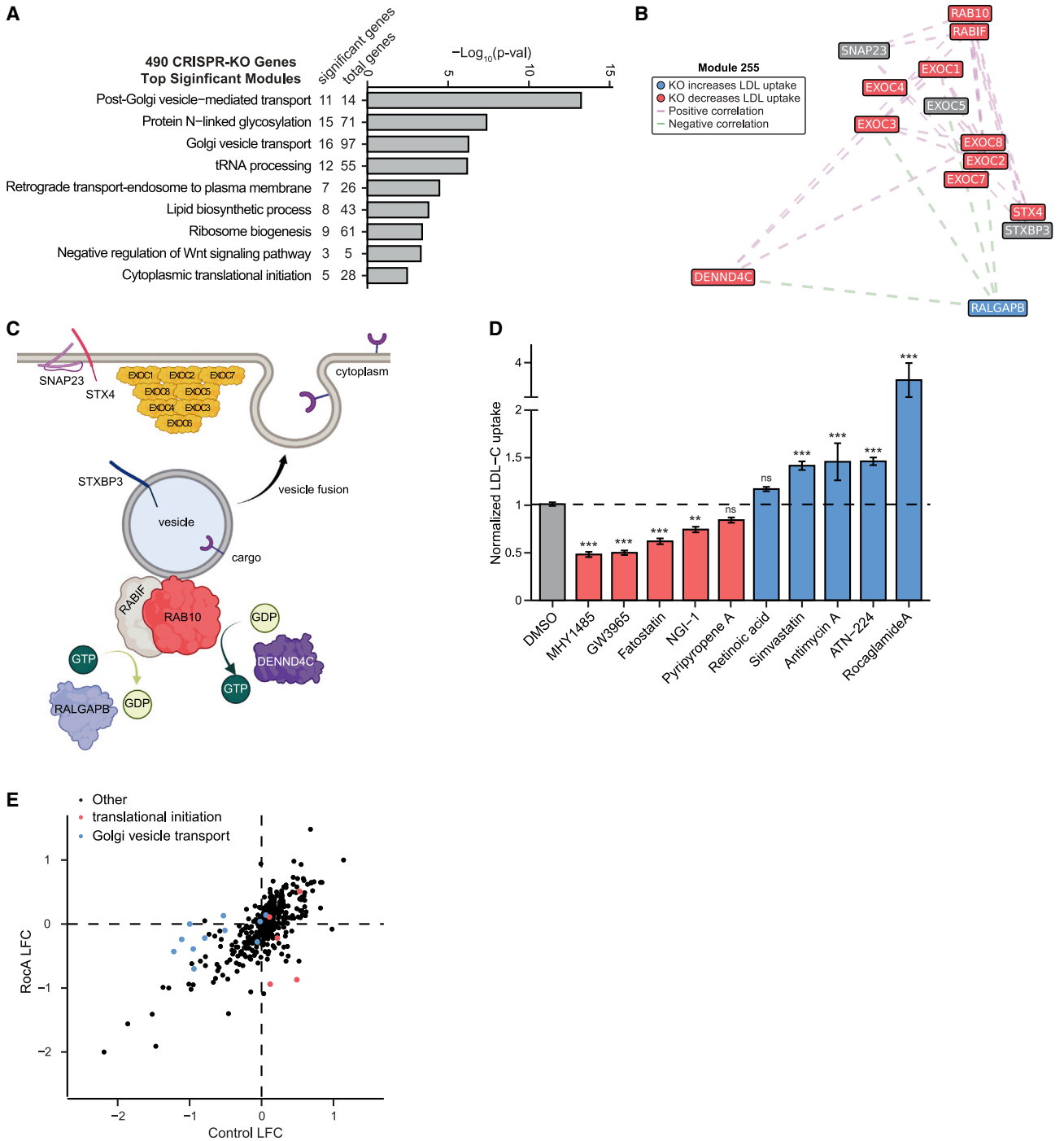


Figure 3. Gene module analysis elucidates LDL-C uptake-altering mechanisms

(A) Co-essential modules with the most significant enrichment of CRISPR-KO genes (hypergeometric p value) labeled with the most significantly associated GO term for each module.

(B) A t-distributed stochastic neighbor embedding (t-SNE) plot based on principal component analysis (PCA) of module 255. CRISPR-KO genes that decrease LDL uptake are labeled red, those that increase LDL uptake are labeled blue. Genes with pairwise Pearson correlation of >0.2 (purple) or <−0.2 (green) are connected with dashed lines.

(C) Schematic of known functions and complexes of module 255, including RAB10 GTPase and exocyst complex components.

(legend continued on next page)

the direction expected based on the CRISPR screening data (Figure 3D), compared with 6 of 14 control molecules chosen for their effect on DNA repair[58] (Figure S3E).

Reasoning that gene knockouts may be redundant or synergistic with small molecules that target the same cellular pathway,[59] we performed CRISPR screening using KO-Library 4 in the presence of eight small molecules as well as a DMSO control, each in three replicates. We observed evidence of gene-small molecule interactions, including *EIF4A1* knockouts increasing LDL-C uptake in cells treated with DMSO or seven unrelated small molecules and decreasing LDL-C uptake in cells treated with the EIF4A1 inhibitor Rocaglamide A (Figure 3E; Table S6). Similarly, the LDL-C uptake-increasing effects of knockout of the lipid biosynthesis pathway members *LSS* and *SQLE* were less pronounced in cells receiving the HMGCR inhibitor simvastatin, presumably because simvastatin acts through the same LDL-C uptake-increasing mechanism (Figure S3F; Table S6). Co-essential modules with differential effects in the presence of each drug compared with DMSO-treated cells supported these results. We identified expected effects (the "translational initiation" module was strongly depleted in Rocaglamide A-treated cells, while the "lipid biosynthetic process" module was strongly depleted in simvastatin-treated cells [Figures 3E and S3F; Table S7]) as well as unexpected molecule-module interactions (e.g., the "Golgi vesicle transport" module was mitigated in Rocaglamide A-treated cells [Figures 3E and S3F; Table S7]).

We built a correlation matrix of the differential effects of each co-essential module on LDL-C uptake in the presence of the eight small molecules. Certain co-essential modules cluster in their interaction with these small molecules, such as modules related to vesicle transport and N-linked glycosylation and modules related to translation, mitochondrial function, and lipid biosynthesis functions (Figure S3G; Table S8). Thus, analysis of gene-small molecule interactions identifies cellular functions that act in coordinated fashion to alter LDL-C uptake.

Rare variant analysis reveals monogenic contributions to serum LDL-C levels

We examined the effects of rare coding variants on serum LDL-C levels using exome sequencing data and patient LDL-C values from the UKB.[60] First, we analyzed the coding burden of rare putative loss of function (pLOF) and computationally predicted damaging missense variants from an analysis of 454,787 individuals.[14] Among the 459 CRISPR-KO significant genes with qualifying variants in this dataset, we identified 11 genes with LDL-C burden at a 10% FDR (Figures 4A and S4A; Tables S9 and 10). These genes include LDL-C uptake regulators such as *LDLR*, *MYLIP*, and *HNF1A*, as well as genes such as the RAB10 GEF *DENND4C* (carriers have increased LDL-C). We found that CRISPR-KO genes are enriched in genes with LDL-C burden in this analysis Figure S4B), supporting the idea that HepG2 LDL-C uptake CRISPR-KO screening is a valuable approach to improve identification of genes associated with serum LDL-C levels. We note that KO-Library 1, which contributed to the identification of a subset of the CRISPR-KO significant genes, is enriched in genes with LDL-C burden, and thus we cannot rule out that some enrichment in CRISPR-KO hits with coding burden could derive from this non-uniform starting gene set. Whole-genome LDL-C burden analysis using this approach revealed 51 significant genes (Tables S9 and 10).

The pLOF-focused coding burden analysis is restricted to variants that are predicted to be highly damaging[62]; however, this threshold diminishes the ability to identify associations in genes that have low allele counts, due to either strong negative selection of damaging variants or sequence length and context. We applied a more inclusive rare variant analysis pipeline, including a larger set of rare coding variants. We evaluated 12 computational models that predict variant deleteriousness (e.g., CADD,[63] VEST4[64]) and evolutionary conservation scores (e.g., GERP,[65] PhyloP[66]). We found that VEST4 outperforms other methods at classifying pathogenic and benign variant annotations from ClinVar[67] (area under the receiver operating characteristics [AUROC] = 0.96) for these variants (Figure S4C). We also found that VEST4 showed the highest mean correlation rank between deleteriousness scores and serum LDL-C levels for 8,344 UKB missense variants in 33 genes with significant rare variant enrichment,[14] adjusting serum LDL-C levels for age, sex, polygenic risk score (PRS), and statin use[5] (adjusted LDL-C) (Figure S4D). While the correlation between VEST4 score and adjusted LDL-C was modest and variable across these genes (median R = 0.077), the fact that correlations were significantly non-zero in 12/33 genes and robust in genes such as *LDLR*, *ABCA1*, and *ASGR1* (Figures 4B, S4D, and S4E) reinforced the idea that accounting for deleteriousness using VEST4 may improve the estimation of the impact of monogenic coding variants on LDL-C levels.

We assessed whether variants in the significant genes identified in our CRISPR-KO screens altered adjusted LDL-C in the UKB cohort by performing collapsing tests considering all variants in 188,891 individuals with four VEST4 score thresholds (top 25%, 50%, 75%, or 100% most deleterious; Figure 4C). Of 417 CRISPR-KO significant genes with $\geq$10 carriers, we found 13 genes associated with LDL-C at a 10% FDR (Figures 4A, 4D, 4E, and S4F; Tables S9 and S11). Only four genes were identified in both analyses (*ABCA1*, *ABCG5*, *APOE*, and *LDLR*) (Figure 4A), revealing the complementary nature of these approaches. Genes uniquely identified in the deleterious-thresholded approach include the RAB10 holdase/chaperone *RABIF* (carriers have increased LDL-C) and the transcription factor *OTX2* (carriers have decreased LDL-C). Whole-genome LDL-C rare variant analysis using the deleterious-thresholded approach revealed 42 significant genes (Tables S9 and S11). VEST4 scores correlated reasonably well with UKB carrier LDL-C levels in genes with significant

(D) Mean control-normalized LDL-C uptake fold change of HepG2 cells treated with DMSO (control) or one of 10 small molecules. Dunnett's multiple comparisons test; n = 6–26 (**p $\leq$ 0.01, ***p $\leq$ 0.001).

(E) Comparison of LDL-C uptake log2 fold change for 522 genes targeted by CRISPR-KO in the presence of DMSO (x axis) and Rocaglamide A (y axis). Genes in denoted co-essential modules are highlighted.

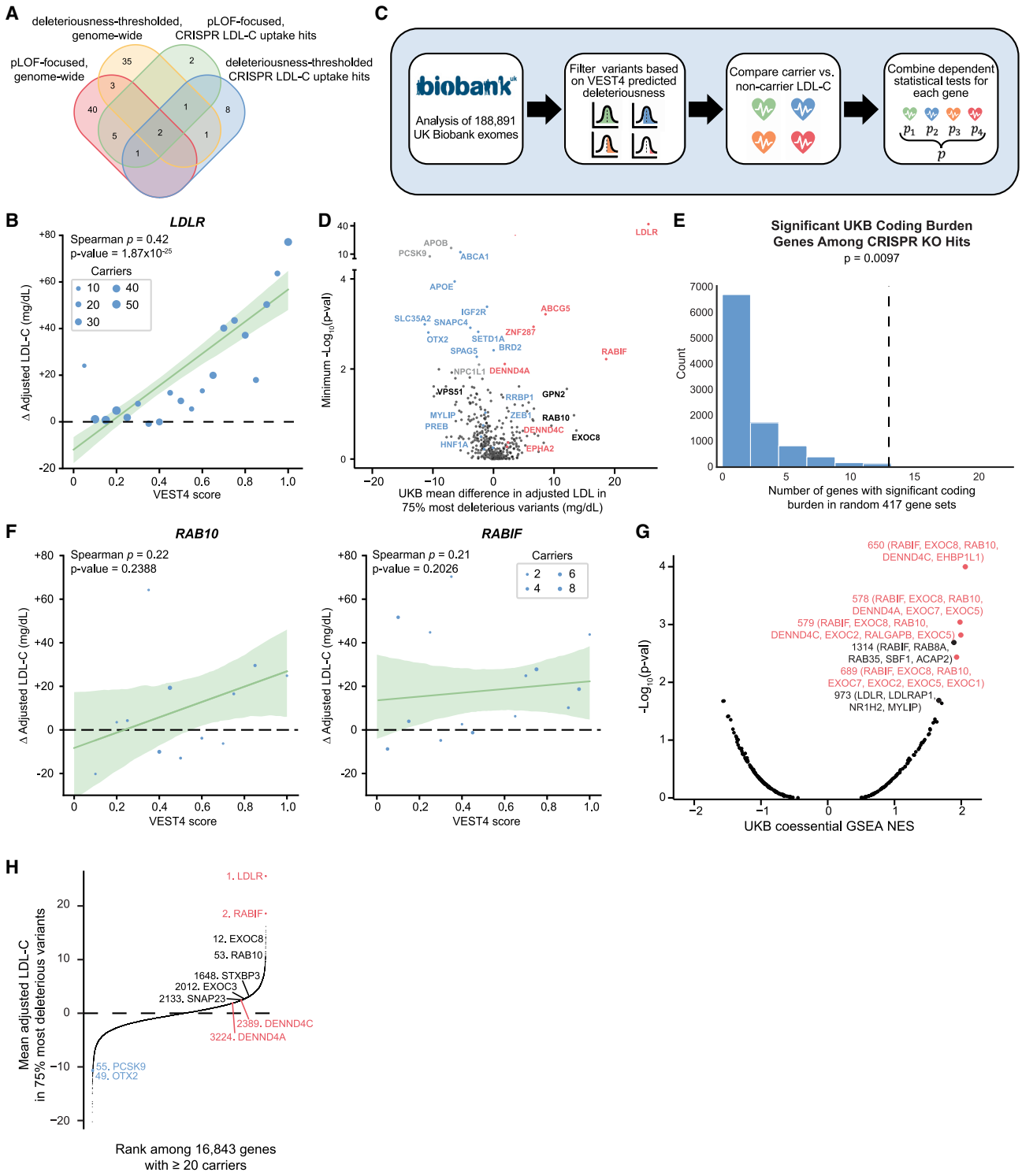


Figure 4. Coding rare variant analysis reveals contributions of genes and modules to serum LDL-C

(A) Venn diagram showing genes with significant enrichment of coding variants in the UKB cohort using two analysis methods.

(B) Spearman correlation and p value within UKB *LDLR* variant carriers between VEST4 score (5% bins) and change in adjusted LDL-C compared with non-carriers. Bubble size represents the number of carriers; 95% confidence interval around the linear regression line is shown.

(C) Schematic of a collapsing burden model that accounts for variant deleteriousness to identify genes associated with altered serum LDL-C in the UKB exome cohort.

(legend continued on next page)

coding burden, such as *OTX2* and *RABIF*, as well as in other CRISPR-KO hit genes (Figures 4F and S4G), adding further evidence that deleterious variants in these genes lead to altered serum LDL-C levels.

We also examined rare variant associations with CAD among 469,828 UKB exome sequencing participants in 21 genes. Using a Cox proportional-hazards regression, we found CAD to be significantly associated with LOF variants in *LDLR* (hazard ratio [HR] = 6.81 [10.16, 4.56], $p = 5.94 \times 10^{-21}$) and *SPAG5* (HR = 2.08 [1.47, 2.96], $p = 4.39 \times 10^{-5}$) and missense variants in *LDLR* (HR = 1.28 [1.16, 1.42], $p = 2.29 \times 10^{-6}$) (Figure S4H; Table S12). These findings may be limited by statistical power given the low numbers of rare LOF variant carriers, the number of CAD cases among variant carriers, or the broad spectrum of functional effects of missense variants. For example, *PCSK9* LOF carriers lack a significant decrease in CAD risk (p = 0.13; Figure S4H; Table S12), at odds with previous reports.[68]

Among all genes with significance in the CRISPR-KO screens, there is a significant negative correlation between their effect on LDL-C uptake and the effect of deleterious coding variants on serum LDL-C levels (Pearson correlation [r] = -0.20; Figure S4I). Restricting to genes prioritized by both the CRISPR-KO LDL-C uptake phenotype and the coding rare variant analysis approaches, there is a significant and more robust negative correlation between their effect on LDL-C uptake and their effect on serum LDL-C levels ($r = -0.79$; Figure S4J).

We asked whether co-essential modules showed evidence of coherent association with serum LDL-C levels through gene set enrichment analysis (GSEA)[69] using the ranked list of mean carrier serum LDL-C among variants in the top 75% of VEST4 deleteriousness for 16,843 genes with 20 or more qualifying variants. Among 241 d = 0.8 modules, we found four significant modules, all associated with increased LDL-C levels, comprising a total of 11 genes (Figure 4G; Table S13). These genes can primarily be functionally designated as RAB10 GTPase components and members of the exocyst complex (Figure 3C). Focused gene-level rare variant analysis on these 11 genes reveals that, among them, *DENND4A*, *DENND4C*, and *RABIF* have LDL-C burden (Figure 4A; Table S9). Similarly, among 291 d = 0.2 modules, the module with the strongest GSEA normalized enrichment score and the lowest p value (Figure S4K; Table S14) was module 255, which also comprises RAB10-related and exocyst pathway genes and scores as the most enriched module in the CRISPR-KO screen (Figures 3B and 3C).

These analyses indicate that RAB10 GTPase dysfunction may contribute to hypercholesterolemia through impaired LDL-C uptake. Of the 16,849 genes assessed in the UKB exome cohort, deleterious variant carriers in *RABIF* have the second highest mean adjusted LDL-C (behind only *LDLR*), and carriers of *RAB10* variants have the 53rd highest mean adjusted LDL-C (Figure 4H). RAB10 GEFs *DENND4C* (2,389th of 16,849) and *DENND4A* (3,224th)[52] are both significantly associated with serum LDL-C, albeit with weaker magnitude, and carriers of variants in other module 255 genes show elevated LDL-C (Figure 4H). Altogether, these analyses suggest that disruption of RAB10 vesicle-mediated processes, including exocytosis mediated by SNARE proteins and the exocyst complex, drives higher LDL-C levels in the human population.

Disrupting the RAB10 vesicle transport pathway impedes LDLR membrane trafficking

Flow cytometry with an anti-LDLR antibody indicated that HepG2 cells with CRISPR knockout of *RAB10* and *DENND4C* have decreased surface LDLR levels (Figure 5A). We repeated these knockouts in a doxycycline-inducible LDLR-mCherry HepG2 cell line. Through time-course fluorescence imaging, we found that LDLR is never efficiently trafficked to the cell membrane in *RAB10*-knockout cells (Figures 5B and S5A); instead, LDLR has a vesicular localization. In addition to altered localization, flow cytometric quantification revealed that *RAB10*- and *DENND4C*-knockout cells have significantly less LDLR-mCherry, indicating decreased LDLR stability (Figures 5C and S5B). Finally, RNA-seq on *DENND4C*-knockout HepG2 vs. control cells revealed enrichment in the expression of a co-essential module related to vesicle-mediated transport (Figure 5D; Table S1); RNA-seq analysis of RAB10 knockout showed an elevated cellular stress signature and was thus excluded from analysis. Altogether, these data suggest that deficiencies in the RAB10 pathway lead to decreased membrane LDLR levels through decreased membrane trafficking and consequent decreased LDLR stability.

Mouse knockdown of Rabif, Otx2, and Csk alters serum cholesterol levels

We tested whether the liver-specific knockdown of three LDL-C uptake-altering genes affects mouse serum cholesterol levels. We chose two genes with robust effects in our CRISPR-KO and CRISPRa screens as well as robust and statistically significant association with LDL-C in the UKB rare variant analysis, *RABIF* and *OTX2* (Figure 4H), and a third gene with robust *in vitro* phenotypes but without UKB burden, *CSK*. Individual knockout of each gene significantly altered HepG2 LDL-C uptake, with ectopic expression of each gene's open reading frame significantly altering LDL uptake in the opposite direction (Figure 6A).

(D) Volcano plot showing mean adjusted LDL-C difference (x axis, mg/dL) and two-sided −log10 Mann-Whitney U p value (y axis) in the VEST4 burden analysis for 417 CRISPR-KO significant genes and 3 additional well-known LDL-C burden genes (gray). Genes significantly associated with increased (red) and decreased (blue) LDL-C levels in burden analyses and selected non-significant genes (black) are highlighted.
(E) Number of genes significantly (two-sided Mann-Whitney U p value, Benjamini-Hochberg FDR = 0.1) associated with serum LDL-C in 10,000 random selections of 417-gene sets.
(F) Correlation within UKB *RAB10* and *RABIF* variant carriers between VEST4 score and change in adjusted LDL-C compared with non-carriers. Plot formatting as in (B).
(G) Volcano plot showing UKB LDL-C rank co-essential GSEA normalized enrichment score (NES) (x axis) and GSEA −log10 p-value[61] for 241 d = 0.8 co-essential modules. The four significant modules (red) (Benjamini Hochberg FDR = 0.1) and several enriched non-significant modules are highlighted.
(H) UKB mean carrier-adjusted LDL-C rank for the 75% most deleterious variants for 16,843 genes. Genes significantly associated with increased (red) and decreased (blue) LDL-C levels (Benjamini-Hochberg FDR = 0.1) and selected RAB10-associated non-significant genes (black) are highlighted.

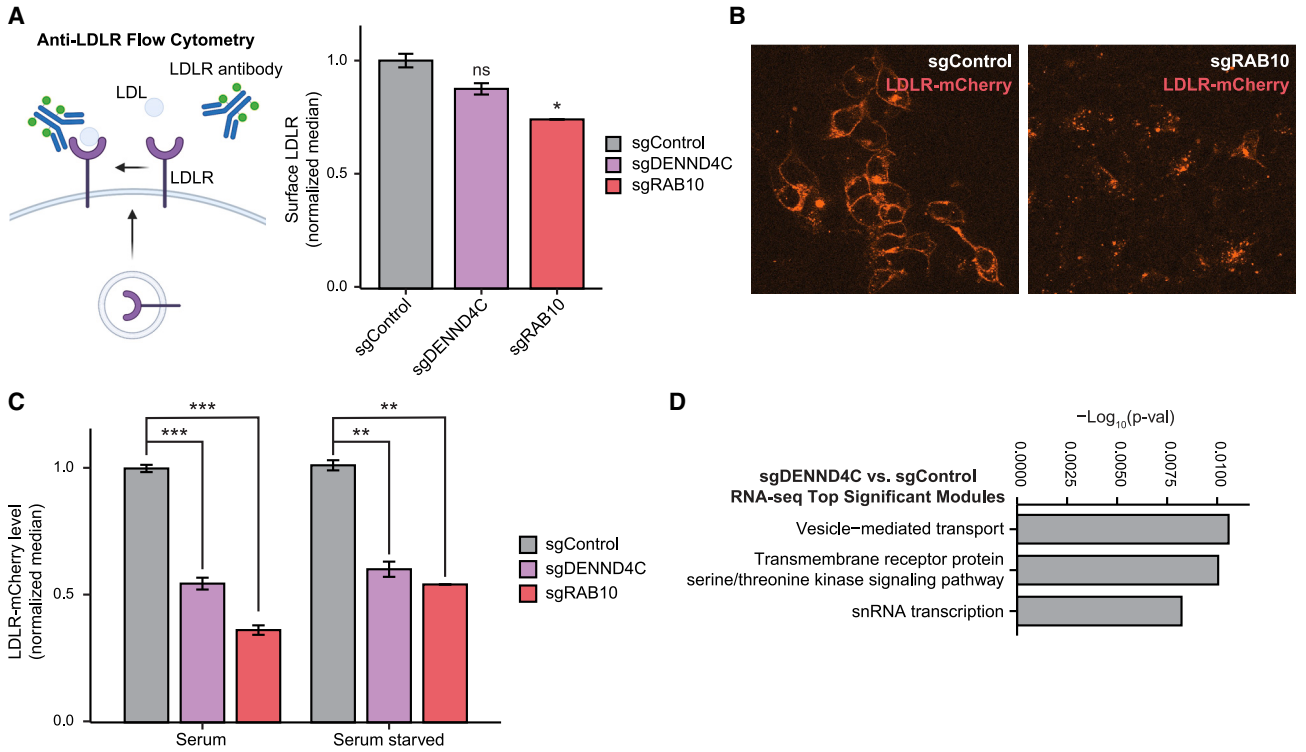


Figure 5. Disrupting the RAB10 vesicle transport pathway impedes LDLR membrane trafficking

(A) Median anti-LDLR flow cytometric staining intensity for HepG2 cells with gene KO as indicated. Two-tailed t test; n = 2.

(B) Fluorescence microscopic images of HepG2 cells with inducible LDLR-mCherry expression and gene KO as indicated. LDLR-mCherry expression was induced 6 h prior to imaging.

(C) Flow cytometric expression of LDLR-mCherry in HepG2 cells with inducible LDLR-mCherry expression and gene KO as indicated. LDLR-mCherry expression was induced 24 h prior to measurement in the indicated medium conditions; n = 2–4. Two-tailed t test (*$p < 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).

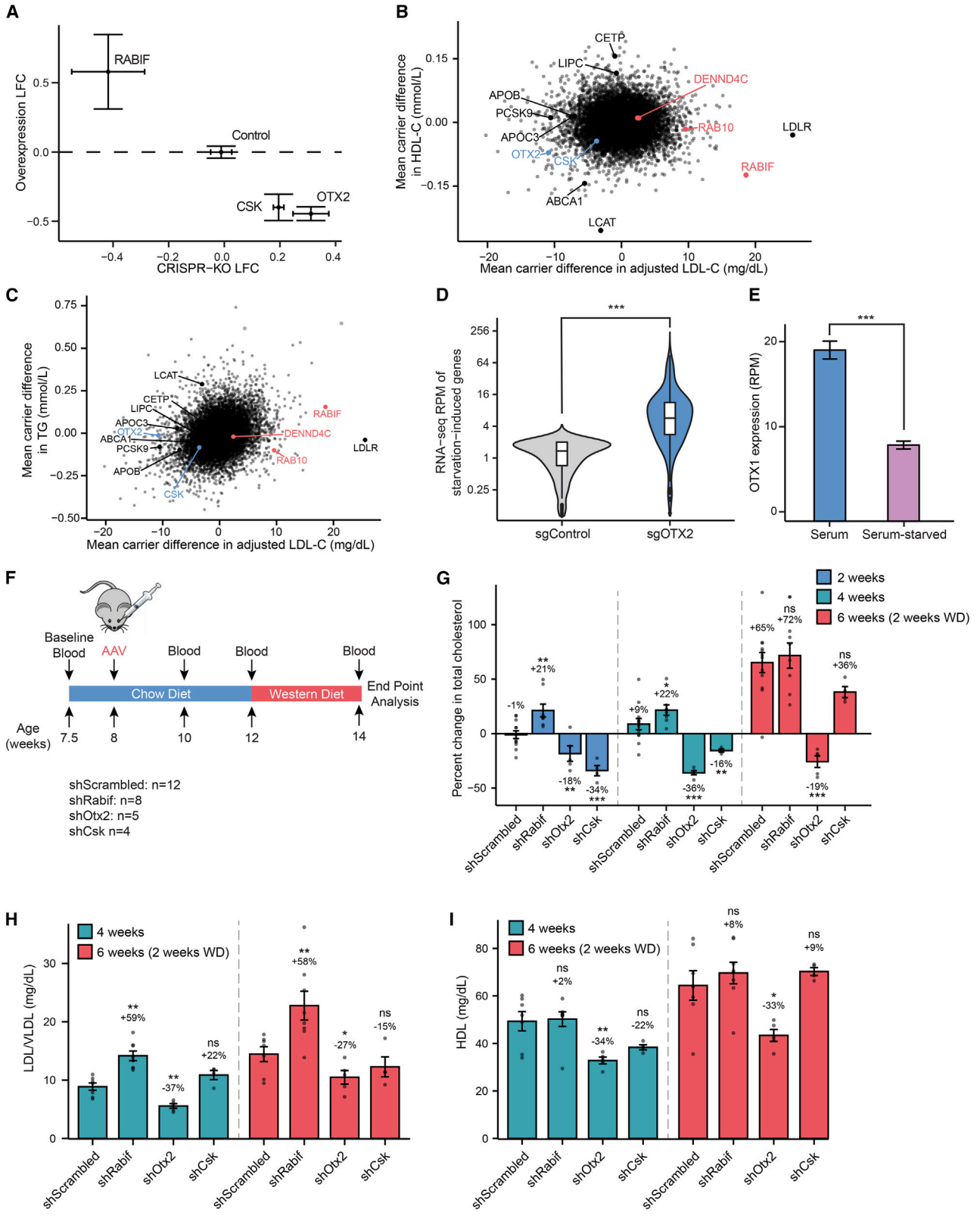
(D) Co-essential module enrichment among sgDENND4C RNA-seq upregulated genes. Hypergeometric p values are shown.

For *RABIF*, this was done through coordinated overexpression of *RABIF* and *RAB10*, as these proteins are coordinately required for proper function.

We also explored whether UKB carriers of deleterious coding variants in these genes show alterations in other serum lipids. *RABIF* carriers show low HDL-C and high triglyceride (TG) levels, although these trends do not rise to statistical significance (Figures 6B, 6C, and S6A). *OTX2* and *CSK* variant carriers show low HDL-C and middling TGs, although also not significant. Associations among these genes and lipoprotein A and apolipoprotein A levels are also predominantly non-significant.

We performed RNA-seq of knockout HepG2 cells to explore mechanisms underlying the LDL-C uptake-altering effects of *OTX2*. *OTX2* knockout induces upregulation of a serum starvation response gene program in non-starved HepG2 (Figure 6D; Table S1). While *OTX2* expression is too low to reliably measure in HepG2 RNA-seq, its paralog *OTX1* is among the most highly downregulated genes upon HepG2 serum starvation (2.4-fold; Figure 6E; Table S1). *OTX1* also significantly increases LDL-C uptake in our CRISPR screen, suggesting these genes may play overlapping roles. Co-essential GSEA on genes showing differential expression in SS *OTX2*-knockout HepG2 cells showed upregulation of the "lipid metabolic process" module (module 660) and "cholesterol esterification" module (module 335) and downregulation of the "TG catabolic process" module (module 324) (Table S15). We analyzed genome-wide Perturb-seq data in which RNA-seq profiling was performed after CRISPRi knockdown of 9,858 genes in the myelogenous leukemia cell line K562.[70] While knockdown of *OTX1* and *OTX2* was performed in this study, the *OTX2* knockdown experiment failed quality control. Co-essential GSEA showed that the module with the most upregulated expression upon *OTX1* knockdown was the "regulation of cholesterol biosynthetic process" module (module 82; Table S16). The most downregulated module (module 160; Table S16) comprises genes involved in "signal-recognition particle (SRP)-dependent co-translational protein targeting to membrane," or the ER targeting of proteins destined for secretion or membrane deposition.

To begin to determine how OTX genes control these gene expression changes, we analyzed OTX1 chromatin immunoprecipitation (ChIP)-seq data from K562 cells.[71] These data identify 10,690 OTX1 binding sites in the K562 genome (Table S16). We used GREAT[72] to identify 3,549 human genes with OTX1 promoter binding (+5 to −1 kb of the gene's transcription start site [TSS]). We found that the 10% of genes most upregulated upon *OTX1* knockdown were significantly more likely than other genes to have OTX1 promoter binding (36% vs. 31%, $p < 0.01$), while genes downregulated upon *OTX1* knockdown had



(legend on next page)

marginally decreased likelihood of OTX1 promoter binding (28% vs. 31%, $p < 0.05$; Table S16; STAR Methods). Further, genes with OTX1 promoter binding have significantly higher expression upon *OTX1* knockdown than genes without such binding (median log2 fold change [LFC] 0.019 vs. 0.012, $p < 0.0001$; Table S16; STAR Methods). These data provide supportive evidence that OTX1 acts more often as a transcriptional repressor than an activator. OTX1 promoter binding is significantly enriched at promoters of genes in module 160 (51% vs. 31%, $p < 0.01$; Table S16) and not significantly different at module 82 promoters, suggesting that these module 160 genes involved in co-translational ER targeting may represent key *cis*-regulatory targets of OTX1. Taken together, our data point to a role for *OTX* genes in regulating gene programs related to lipid metabolism and cholesterol biosynthesis across several cell states. *CSK* knockout induces the expression of co-essential modules related to cell motility and cytoskeleton organization (Figure S6B).

To evaluate the effects of these genes in mice, we performed adeno-associated virus 8 (AAV8) shRNA-mediated gene knockdown of all three genes using retro-orbital injection, which yields efficient liver delivery.[73–75] We measured serum cholesterol levels prior to knockdown and then every other week for 6 weeks, with the first 4 weeks on standard chow and the final 2 on Western diet (WD) (Figure 6F). At the 6-week endpoint, we confirmed that shRNA treatment induced robust decreases in target transcript levels in liver (on average 88% decrease for *Rabif*, 60% decrease for *Csk*, and *Otx2* transcripts were undetectably low; Figure S6C).

We observed rapid, robust, and durable shRNA-mediated alteration of serum cholesterol compared with pre-treatment baseline levels. *Rabif* knockdown increased total cholesterol levels by an average 21% compared with pre-treatment levels within 2 weeks, significantly more than the 1% decrease observed in scrambled shRNA control mice (Figure 6G). This serum cholesterol increase was maintained at 4 weeks (22% vs. 9% in shScrambled) and expanded at 6 weeks after 2 weeks on WD (72% vs. 65% in shScrambled) (Figure 6G). The increased total cholesterol in sh*Rabif*-treated mice was primarily driven by changes in the VLDL/LDL fraction, which was 59% higher than in shScrambled-treated mice at week 4 and 58% higher at week 6 (Figure 6H), with only minimal changes in HDL ($n = 8$, non-significant; Figure 6I). *Rabif* knockdown also increased serum TG levels by 13% compared with control knockdown after 4 weeks, although this change was not significant (Figure S6D). Altogether, mouse liver-targeted knockdown of *Rabif* led to markedly increased serum LDL-C levels, in line with the higher serum LDL-C observed in individuals with deleterious *RABIF* variants and the decrease in liver cell LDL-C uptake after *RABIF* knockout.

On the other hand, knockdown of *Otx2* led to robustly decreased serum cholesterol as predicted from biobank and cell line analysis. *Otx2* knockdown decreased cholesterol levels by 18% and 36% at 2 and 4 weeks, respectively (Figure 6G), comparable to decreases seen after liver-targeted siRNA-mediated knockdown and CRISPR-Cas9 mutation of *Pcsk9* in chow-fed wild-type mice (30%–40%).[76,77] *Csk* knockdown decreased total cholesterol levels compared with pre-treatment levels by an average of 34% and 16% at those time points. This decrease in cholesterol was maintained in sh*Otx2*-treated mice at week 6 after 2 weeks of WD (19% decrease), but not in sh*Csk*-treated mice (36% increase; Figure 6G). *Otx2* knockdown decreased both VLDL/LDL and HDL levels significantly compared with control knockdown at 4 weeks (by 37% and 34%) and 6 weeks (by 27% and 33%) (Figures 6H and 6I) and also robustly decreased TG levels compared with control knockdown by 26% at 4 weeks (Figure S6D). *Csk* knockdown did not induce significant changes in separated VLDL/LDL or HDL levels at either 4 or 6 weeks, nor did it alter TG levels (Figures 6H, 6I, and S6D).

Some genetic mutations that lower circulating LDL-C levels, such as those in *APOB*, cause increased hepatic lipid and cholesterol levels, while mutations in *PCSK9* do not.[78] We found that mice with *Otx2* knockdown show on average 18% reduction in liver total cholesterol and 24% reduction in liver TG levels (Figure S6E). These data demonstrate that *Otx2* reduces lipid and cholesterol content both in the blood and in the liver, suggesting that *Otx2* loss is unlikely to reduce circulating LDL-C by impairing VLDL secretion. We conclude that liver-targeted knockdown of *Otx2* leads to robust decrease in serum LDL-C levels and makes mice refractory to WD-induced increases in cholesterol level.

LDL uptake-altering genes may underlie GWAS loci

We asked whether LDL uptake-altering genes that we identified could potentially underlie loci associated with LDL-C in a recent trans-ancestry GWAS.[10] Examining the RAB10 co-essential module, we noted GWAS-identified loci significantly associated with LDL-C adjacent to *RAB10* and *DENND4C*.[10] Bayesian fine-mapping[79] of the *RAB10* locus suggests a single likely causal variant (rs142787485, posterior inclusion probability

Figure 6. Mouse knockdown of Rabif, Csk, and Otx2 alters serum cholesterol levels

(A) Comparison of the effects of CRISPR-Cas9 gene KO (x axis) and ectopic open reading frame (ORF) expression (y axis) for *CSK*, *OTX2*, *RABIF*, and non-targeting control. Error bars represent standard error.

(B and C) UKB mean carrier adjusted LDL-C (x axis) vs. mean carrier HDL-C (B, y axis) and mean carrier triglycerides (C, y axis) for all genes. Benchmark genes and novel genes are highlighted.

(D) HepG2 control or *OTX2* KO RNA-seq reads per million (RPM) for the 100 genes that are most robustly upregulated in HepG2 cells upon serum starvation. Paired Mann-Whitney U test.

(E) RNA-seq RPM for *OTX1* in HepG2 cultured under serum-containing or serum-starved conditions; $n = 46$. Two-tailed t test.

(F) Schematic of mouse AAV shRNA experiment. Cohort sizes for each treatment group are listed.

(G) Average percentage change in total cholesterol from pre-treatment measurement in mice treated with the designated AAV8 shRNA at the designated time points. Two-tailed t test.

(H and I) Average LDL/VLDL (H) and HDL (I) in mice treated with the designated AAV8 shRNA at the designated time points. Average difference compared with shScrambled and p value are noted. Two-tailed t test (*$p < 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).

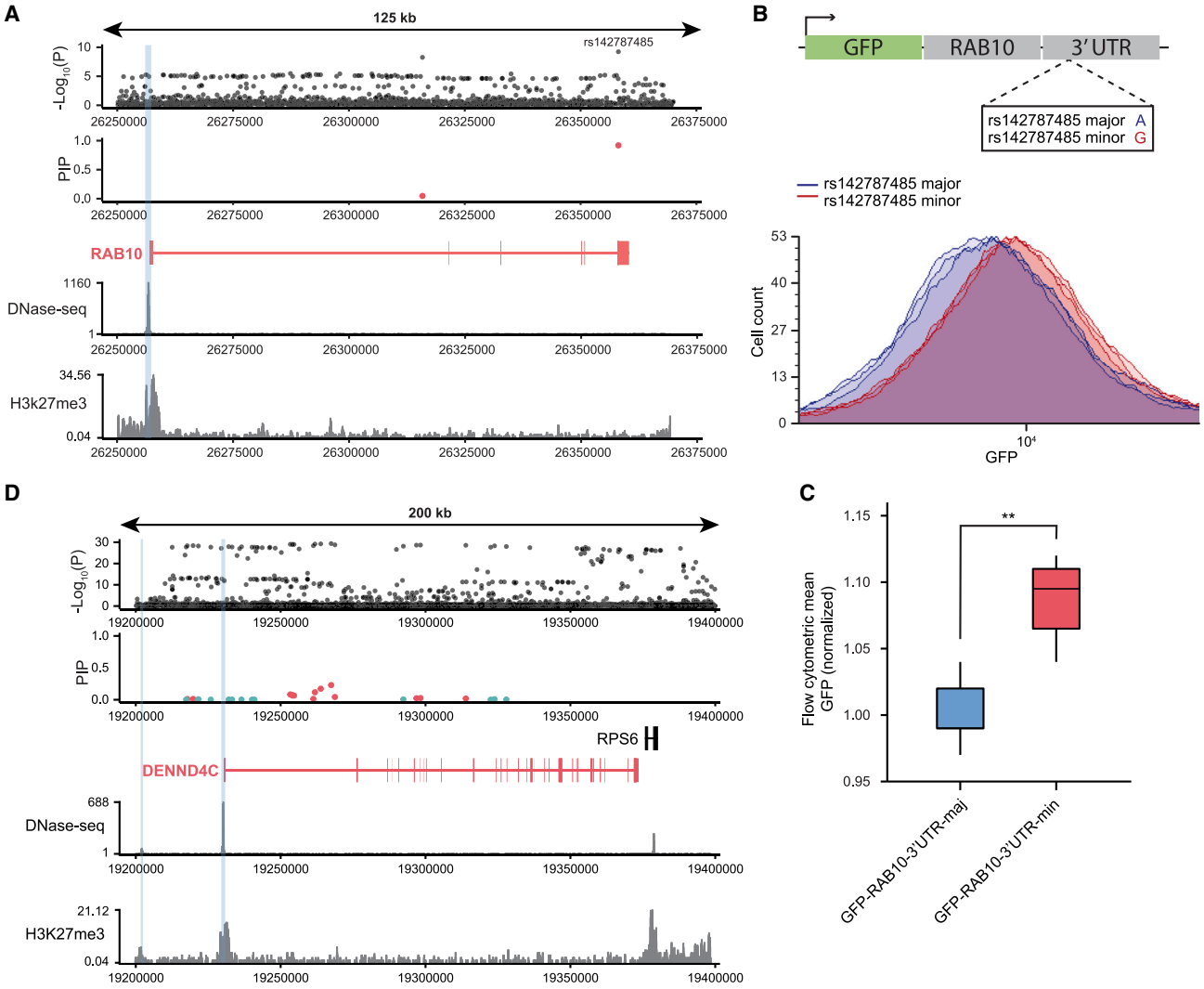


Figure 7. LDL uptake-altering genes may underlie GWAS loci

(A) Fine mapping of the GWAS signal at the *RAB10* locus (black, LDL-GWAS variants; red, 95% confidence set; teal, 99% confidence set). Signal tracks represent HepG2 DNase-seq and H3K27me3 ChIP-seq. Blue bars represent ABC-predicted enhancer connections with *RAB10*.

(B) Flow cytometric analysis of HepG2 cells with expression of a GFP-RAB10 fusion protein expression construct containing the full RAB10 3′ UTR with rs142787485 major (blue) or minor (red) allele.

(C) GFP-RAB10 3′ UTR mean GFP flow cytometry; n = 6. Two-tailed t test (**$p \leq 0.01$).

(D) Fine-mapping of the GWAS signal at the *DENND4C* locus. Plot formatting as in (A).

[PIP] = 0.92) in a conserved region of the 3′ UTR of *RAB10* (Figure 7A). The minor allele of rs142787485, which is associated with lower LDL-C levels in carriers,[10] disrupts predicted binding sites of two microRNA species (Figure S7A),[80] suggesting that the minor allele may possess increased *RAB10* transcript stability. To evaluate the impact of this variant, we cloned a GFP-RAB10 fusion protein expression construct containing the full RAB10 3′ UTR with rs142787485 major or minor allele. We found that, upon stable lentiviral transduction into HepG2 cells, GFP-RAB10 with a 3′ UTR with rs142787485^{min} led to 9% higher protein levels than rs142787485^{maj} (Figures 7B and 7C). Altogether, these data suggest that a common variant that increases RAB10 protein levels is associated with reduced LDL-C levels in the population. Fine mapping of the *DENND4C* locus yields seven intronic variants with PIP between 5% and 25% (Figure 7D), complicating evaluation of regulatory mechanisms.

Other robustly LDL uptake-altering genes that are not classically recognized cholesterol modulators but which are near LDL-C GWAS loci include *CSK*, *ARID1A*, and *RRBP1* (Figures S7B–S7D). In particular, the *ARID1A* locus contains a single credible fine-mapped variant in the *ARID1A* promoter (rs114165349, PIP = 0.99), which is predicted to regulate its expression by the ABC model (Figure S7B). ARID1A is a core epigenetic regulator whose loss in mouse liver leads to

pleiotropic changes, including elevated cholesterol.[81] In addition, the GWAS signal at the *RRBP1* locus can be localized to a single fine-mapped variant (rs2618566, PIP = 1.00) in a distal regulatory region that is a liver expression quantitative trait loci (eQTL)[82] (Figure S7C). Our exome rare variant analysis identifies significantly decreased LDL-C in carriers of rare *ARID1A* and *RRBP1* coding variants (Figure S7E), and mouse *RRBP1* knockdown has been shown to disrupt lipid homeostasis and lead to lower circulating LDL-C levels.[83] Nonetheless, further experiments are required to determine whether these non-coding GWAS variants alter expression or function of these LDL uptake-altering genes.

DISCUSSION

Cholesterol has been recognized for over 100 years as a risk factor for CAD.[84,85] Currently, genetic contributors to LDL-C levels remain incompletely understood, as evidenced by GWAS-associated loci lacking known LDL-C regulatory genes. We employed a pipeline combining human cell line CRISPR screening, co-essentiality-based gene module analysis, human genetic association, and mouse knockdown to expand the understanding of genes and cellular mechanisms whose dysfunction alters serum LDL-C levels. Overall, combining genes identified here with those from published genome-wide analyses,[11–14] we expand the list of genes significantly associated with LDL-C levels (Table S9). Of these, 21 function at least partially by altering liver cellular LDL-C uptake. This gene set is thus a valuable resource to understand pathways and mechanisms altering cholesterol metabolism, to evaluate therapeutic targets, and to connect GWAS-identified variants with target genes.

Combining human genetic association and cellular genetic screening compensates for shortfalls in each approach. While genetic association cohort sizes have increased,[14] the rarity of deleterious coding variants and the genetic and environmental heterogeneity among carriers complicate the identification of disease-associated genes from human cohorts alone. Conversely, CRISPR screening allows scalable and well-controlled loss-of-function screening; however, variability in cellular LDL-C uptake is not a perfect surrogate of serum LDL-C levels. Focusing rare variant analysis specifically on genes with *in vitro* LDL-C uptake phenotypes increases the power to identify significant associations. This is not simply due to decreasing the number of tested hypotheses; LDL-C uptake-impairing genes are highly enriched in genes with LDL-C burden (Figures 4E and S4E).

These previously unappreciated genes often show robust effects on LDL-C levels in deleterious variant carriers, as with *RABIF* and *RAB10*, for which carriers have the 2nd and 53rd highest mean LDL-C among 16,849 evaluated genes, and *OTX2*, for which carriers have the 49th lowest mean LDL-C, on par with *PCSK9* (Figure 4G). We validated in mice that the effects of *RABIF* and *OTX2* knockdown on LDL-C levels are similar in magnitude to those of known monogenic hyper-/hypocholesterolemia genes. These genes have likely eluded prior identification due to the relative rarity of deleterious variant carriers. Classical hyper-/hypocholesterolemia genes have been identified through characterization of human loss-of-function alleles and knockout mice.[15] However, germline knockouts of *Otx2*, *Rab10*, and *Dennd4c*, along with many other genes we implicate in LDL-C uptake, are embryonic lethal,[86–89] and humans with biallelic disruption of these genes have not been reported. In addition, genes such as *RAB10* and *RABIF* are small (200 and 123 amino acids, respectively), making them inherently less likely to accumulate mutations. Our use of cell-type-specific CRISPR screening and analysis of monoallelic partial loss-of-function phenotypes from exome biobank data provide innovative avenues to identify disease-relevant genes.

In addition, we find that co-essentiality provides a powerful approach to grouping genes into functional modules. Co-essentiality annotates a larger number of genes than GO and PPI database analysis[48] and, because it is based on functional impacts of gene knockout, is well calibrated for pathway analysis of CRISPR screening data. Of note, we define a pathway contributing to genetic propensity toward hypercholesterolemia. We find that the RAB10-driven exocytosis pathway shows significant LDL-C burden. A recent GWAS also identifies common variants near *RAB10* and *DENND4C* that are associated with LDL-C levels, and we validated that the RAB10 3′ UTR variant that is associated with decreased serum LDL-C levels increases RAB10 protein levels by 9% in liver cells.

RAB10 is a multi-functional GTPase that has been implicated in a diverse array of vesicle-mediated processes, including endocytosis and exocytosis[50,54]; however, the co-essentiality data suggest stronger shared function with the exocyst complex than other vesicle-mediated processes. The exocyst complex has been previously implicated in polarized transport of LDLR to basolateral membranes,[57] and mutations in exocyst genes impair LDLR uptake while increasing transferrin uptake.[22] In our time-course microscopy data, we are unable to detect substantial surface LDLR expression in RAB10-mutant cells at its earliest detectable time point, suggesting that impaired initial membrane trafficking contributes to the phenotype. Recent work also indicates that RAB10-deficient cells show defective LDLR endosomal recycling as well.[54] In addition, RAB10 and exocyst knockout impairs LDL-C uptake in LDLR knockout cells,[22,54] and RAB10 has been implicated in lipophagy,[90] suggesting that this pathway's roles in lipid uptake and trafficking are multi-faceted.

Further, we show that decreased function of *OTX2* robustly lowers LDL-C levels in human cohorts and mice through increased LDL-C uptake. Deleterious *OTX2* alleles are too rare to ascertain effects on CAD. *Otx2* is essential for gastrulation and early brain and eye development[86]; its roles outside of the nervous system are poorly understood. *OTX2* RNA expression is undetectably low in human and mouse liver tissue and in HepG2 cells, yet its downregulation in HepG2 and mouse liver yields a strong transcriptional and cholesterol phenotype. Knockout of the orthologous and partially functionally redundant *OTX1* gene[91] in HepG2 yields similar increase in HepG2 LDL-C uptake, and individuals with deleterious *OTX1* variants have lower LDL-C levels, albeit not as robustly as for *OTX2*. Whether dual inhibition of *OTX1* and *OTX2* more dramatically lowers LDL-C levels and whether hepatic *OTX2* inhibition has side effects beyond its effects on serum lipids are compelling areas for further research.

Limitations of the study

Here we focus on genetic control of liver cell LDL-C uptake, which is but one of many points of control of serum LDL-C levels. As such, we posit that many additional genes are likely to be *bona fide* LDL-C regulators with distinct phenotypes that contribute to organismal LDL-C levels. We also note that our CRISPR screens were performed in a liver cancer cell line, which may not fully reflect genetic programs involved in hepatocyte LDL-C uptake.

In addition, the rare variant analysis was designed to be an expansive screen for discovery, given two additional sources of information: functional screening data and gene families derived from co-essentiality screening. The Mann-Whitney U test is a weaker test statistic than those used by traditional rare variant association methods.[92] Methods to correct for population of origin in rare variant analysis could further refine this method; adding cohorts from diverse ancestries could improve understanding of drivers of cholesterol metabolism across genetic backgrounds. Additionally, increases in cohort size should expand the gene sets associated with LDL-C levels. For example, genes in the lipid biosynthesis co-essential module associate with lower LDL-C levels (Figure S7F; Table S11), yet only the statin target *HMGCR* rises to significance (Table S9).

Altogether, the integrated approach combining cellular genetic screening, biobank genetic association, and gene network analysis defined here should provide a roadmap to guide efforts to dissect complex trait genetics.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - Animal models
- METHOD DETAILS
  - CRISPR-KO screening
  - Cloning and testing of individual gRNAs
  - Small molecule analysis
  - Lectin flow cytometry analysis
  - RNA-seq
  - CRISPRa Outcome and Phenotype (COP) screen design and execution
  - COP data analysis
  - Cloning and testing of individual COP gRNAs
  - Co-essentiality modules and enrichment
  - Exome rare variant analysis
  - Gene overexpression and RAB10 pathway imaging analysis
  - RAB10 3' UTR evaluation
  - Mouse AAV8 shRNA experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100304.

ACKNOWLEDGMENTS

The authors thank Grigoriy Losyev, Benjamin Holmes, Mandana Arbab, Allison James, Benjamin Angulo, Sumaiya Nazeen, and Shamil Sunyaev for technical assistance and are grateful for funding from 1R01HG008754 (D.K.G., R.I.S.), 1R21HG010391 (R.I.S., C.A.C.), R01HG010372 (J.D.F. and C.A.C.), the American Cancer Society (R.I.S.), the American Heart Association (R.I.S.), the National Organization for Rare Disorders (R.I.S.), the Qatar Biomedical Research Institute (R.I.S.), and a National University of Singapore start-up grant (H.Y.). We are indebted to the UK Biobank and its participants (UK Biobank application 41250 and IRB protocol 2020P002093).

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, writing – original draft, and writing – reviewing and editing, R.I.S., C.A.C., H.Y., M.C.H., J.D.F., E.A., T.Y., D.K.G., and G.L.; investigation and validation, R.I.S., H.Y., M.C.H., E.A., B.K., M.C., T.T.T., M.V., M.J.F., and S.B.; software, formal analysis and visualization, R.I.S., C.A.C., H.Y., M.C.H., J.D.F., E.A., T.Y., B.K., G.H.T.Y., J.P.R.B., and G.L.; funding acquisition and supervision, R.I.S., C.A.C., H.Y., and D.K.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 27, 2022
Revised: December 2, 2022
Accepted: March 24, 2023
Published: April 21, 2023

REFERENCES

1. GBD 2015 Mortality and Causes of Death Collaborators (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet *388*, 1459–1544. https://doi.org/10.1016/S0140-6736(16)31012-1.
2. Mundal, L.J., Igland, J., Veierød, M.B., Holven, K.B., Ose, L., Selmer, R.M., Wisloff, T., Kristiansen, I.S., Tell, G.S., Leren, T.P., and Retterstøl, K. (2018). Impact of age on excess risk of coronary heart disease in patients with familial hypercholesterolaemia. Heart *104*, 1600–1607. https://doi.org/10.1136/heartjnl-2017-312706.
3. Global Lipids Genetics Consortium, Myocardial Infarction Genetics (MIGen) Consortium, The Geisinger-Regeneron DiscovEHR Collaboration, The VA Million Veteran Program, Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., et al. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. *50*, 1514–1523. https://doi.org/10.1038/s41588-018-0222-9.
4. Klarin, D., Zhu, Q.M., Emdin, C.A., Chaffin, M., Horner, S., McMillan, B.J., Leed, A., Weale, M.E., Spencer, C.C.A., Aguet, F., et al. (2017). Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. Nat. Genet. *49*, 1392–1397. https://doi.org/10.1038/ng.3914.
5. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013).

Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283. https://doi.org/10.1038/ng.2797.

6. Howson, J.M.M., Zhao, W., Barnes, D.R., Ho, W.-K., Young, R., Paul, D.S., Waite, L.L., Freitag, D.F., Fauman, E.B., Salfati, E.L., et al. (2017). Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. Nat. Genet. *49*, 1113–1119. https://doi.org/10.1038/ng.3874.

7. Nikpay, M., Goel, A., Won, H.-H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. *47*, 1121–1130. https://doi.org/10.1038/ng.3396.

8. Viney, N.J., van Capelleveen, J.C., Geary, R.S., Xia, S., Tami, J.A., Yu, R.Z., Marcovina, S.M., Hughes, S.G., Graham, M.J., Crooke, R.M., et al. (2016). Antisense oligonucleotides targeting apolipoprotein(a) in people with raised lipoprotein(a): two randomised, double-blind, placebo-controlled, dose-ranging trials. Lancet *388*, 2239–2253. https://doi.org/10.1016/S0140-6736(16)31009-1.

9. Musunuru, K., and Kathiresan, S. (2019). Genetics of common, complex coronary artery disease. Cell *177*, 132–145. https://doi.org/10.1016/j.cell.2019.02.015.

10. Graham, S.E., Clarke, S.L., Wu, K.-H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. Nature *600*, 675–679. https://doi.org/10.1038/s41586-021-04064-3.

11. Jurgens, S.J., Choi, S.H., Morrill, V.N., Chaffin, M., Pirruccello, J.P., Halford, J.L., Weng, L.-C., Nauffal, V., Roselli, C., Hall, A.W., et al. (2022). Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. Nat. Genet. *54*, 240–250. https://doi.org/10.1038/s41588-021-01011-w.

12. Wang, Q., Dhindsa, R.S., Carss, K., Harper, A.R., Nag, A., Tachmazidou, I., Vitsios, D., Deevi, S.V.V., Mackay, A., Muthas, D., et al. (2021). Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature *597*, 527–532. https://doi.org/10.1038/s41586-021-03855-y.

13. Hindy, G., Dornbos, P., Chaffin, M.D., Liu, D.J., Wang, M., Selvaraj, M.S., Zhang, D., Park, J., Aguilar-Salinas, C.A., Antonacci-Fulton, L., et al. (2022). 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. https://doi.org/10.1016/j.ajhg.2021.11.021.

14. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. Nature *599*, 628–634. https://doi.org/10.1038/s41586-021-04103-z.

15. Berberich, A.J., and Hegele, R.A. (2019). The complex molecular genetics of familial hypercholesterolaemia. Nat. Rev. Cardiol. *16*, 9–20. https://doi.org/10.1038/s41569-018-0052-6.

16. Young, S.G., and Fong, L.G. (2012). Lowering plasma cholesterol by raising LDL receptors—revisited. N. Engl. J. Med. *366*, 1154–1155. https://doi.org/10.1056/NEJMe1202168.

17. Horton, J.D., Goldstein, J.L., and Brown, M.S. (2002). SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J. Clin. Invest. *109*, 1125–1131. https://doi.org/10.1172/JCI15593.

18. Wang, B., and Tontonoz, P. (2018). Liver X receptors in lipid signalling and membrane homeostasis. Nat. Rev. Endocrinol. *14*, 452–463. https://doi.org/10.1038/s41574-018-0037-x.

19. Bartz, F., Kern, L., Erz, D., Zhu, M., Gilbert, D., Meinhof, T., Wirkner, U., Erfle, H., Muckenthaler, M., Pepperkok, R., and Runz, H. (2009). Identification of cholesterol-regulating genes by targeted RNAi screening. Cell Metab. *10*, 63–75. https://doi.org/10.1016/j.cmet.2009.05.009.

20. Blattmann, P., Schuberth, C., Pepperkok, R., and Runz, H. (2013). RNAi–based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol-regulatory function. PLoS Genet. *9*, e1003338. https://doi.org/10.1371/journal.pgen.1003338.

21. Kraehling, J.R., Chidlow, J.H., Rajagopal, C., Sugiyama, M.G., Fowler, J.W., Lee, M.Y., Zhang, X., Ramírez, C.M., Park, E.J., Tao, B., et al. (2016). Genome-wide RNAi screen reveals ALK1 mediates LDL uptake and transcytosis in endothelial cells. Nat. Commun. *7*, 13516. https://doi.org/10.1038/ncomms13516.

22. Emmer, B.T., Sherman, E.J., Lascuna, P.J., Graham, S.E., Willer, C.J., and Ginsburg, D. (2021). Genome-scale CRISPR screening for modifiers of cellular LDL uptake. PLoS Genet. *17*, e1009285. https://doi.org/10.1371/journal.pgen.1009285.

23. Thormaehlen, A.S., Schuberth, C., Won, H.-H., Blattmann, P., Joggerst-Thomalla, B., Theiss, S., Asselta, R., Duga, S., Merlini, P.A., Ardissino, D., et al. (2015). Systematic cell-based phenotyping of missense alleles empowers rare variant association studies: a case for LDLR and myocardial infarction. PLoS Genet. *11*, e1004855. https://doi.org/10.1371/journal.pgen.1004855.

24. Aden, D.P., Fogel, A., Plotkin, S., Damjanov, I., and Knowles, B.B. (1979). Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. Nature *282*, 615–616. https://doi.org/10.1038/282615a0.

25. Zhou, B., Ho, S.S., Greer, S.U., Spies, N., Bell, J.M., Zhang, X., Zhu, X., Arthur, J.G., Byeon, S., Pattni, R., et al. (2019). Haplotype-resolved and integrated genome analysis of the cancer cell line HepG2. Nucleic Acids Res. *47*, 3846–3861. https://doi.org/10.1093/nar/gkz169.

26. Partridge, E.C., Chhetri, S.B., Prokop, J.W., Ramaker, R.C., Jansen, C.S., Goh, S.-T., Mackiewicz, M., Newberry, K.M., Brandsmeier, L.A., Meadows, S.K., et al. (2020). Occupancy maps of 208 chromatin-associated proteins in one human cell type. Nature *583*, 720–728. https://doi.org/10.1038/s41586-020-2023-4.

27. Zelcer, N., Hong, C., Boyadjian, R., and Tontonoz, P. (2009). LXR regulates cholesterol uptake through Idol-dependent ubiquitination of the LDL receptor. Science *325*, 100–104. https://doi.org/10.1126/science.1168974.

28. Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K., and Sherwood, R.I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. Nature *563*, 646–651. https://doi.org/10.1038/s41586-018-0686-x.

29. Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863. https://doi.org/10.1038/nprot.2017.016.

30. Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.-H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., et al. (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nat. Protoc. *14*, 756–780. https://doi.org/10.1038/s41596-018-0113-7.

31. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

32. Lin, L., Holmes, B., Shen, M.W., Kammeron, D., Geijsen, N., Gifford, D.K., and Sherwood, R.I. (2020). Comprehensive mapping of key regulatory networks that drive oncogene expression. Cell Rep. *33*, 108426. https://doi.org/10.1016/j.celrep.2020.108426.

33. Lucero, D., Dikilitas, O., Mendelson, M.M., Aligabi, Z., Islam, P., Neufeld, E.B., Bansal, A.T., Freeman, L.A., Vaisman, B., Tang, J., et al. (2022). Transgelin: a new gene involved in LDL endocytosis identified by a genome-wide CRISPR-Cas9 screen. J. Lipid Res. *63*, 100160. https://doi.org/10.1016/j.jlr.2021.100160.

34. Smith, G.A., Padmanabhan, A., Lau, B.H., Pampana, A., Li, L., Clara Lee, Y., Pelonero, A., Nishino, T., Sadagopan, N., Jain, R., et al. (2021). CSDE1

is a post-transcriptional regulator of the LDL receptor. Preprint at bioRxiv. https://doi.org/10.1101/2020.08.03.235028.

35. Ye, J., and DeBose-Boyd, R.A. (2011). Regulation of cholesterol and fatty acid synthesis. Cold Spring Harb. Perspect. Biol. *3*, a004754. https://doi.org/10.1101/cshperspect.a004754.

36. Yin, L., Ma, H., Ge, X., Edwards, P.A., and Zhang, Y. (2011). Hepatic HNF4α is essential for maintaining triglyceride and cholesterol homeostasis. Arterioscler. Thromb. Vasc. Biol. *31*, 328–336. https://doi.org/10.1161/ATVBAHA.110.217828.

37. Luo, J., Yang, H., and Song, B.-L. (2020). Mechanisms and regulation of cholesterol homeostasis. Nat. Rev. Mol. Cell Biol. *21*, 225–245. https://doi.org/10.1038/s41580-019-0190-7.

38. Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646. https://doi.org/10.1016/j.cell.2014.09.039.

39. Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588. https://doi.org/10.1038/nature14136.

40. Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W., Goodale, A., Root, D.E., Piccioni, F., and Doench, J.G. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat. Commun. *9*, 5416. https://doi.org/10.1038/s41467-018-07901-8.

41. Kim, H.K., Kim, Y., Lee, S., Min, S., Bae, J.Y., Choi, J.W., Park, J., Jung, D., Yoon, S., and Kim, H.H. (2019). SpCas9 activity prediction by DeepSpCas9, a deep learning–based model with high generalization performance. Sci. Adv. *5*, eaax9249. https://doi.org/10.1126/sciadv.aax9249.

42. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191. https://doi.org/10.1038/nbt.3437.

43. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency Map. Cell *170*, 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.

44. Boyle, E.A., Pritchard, J.K., and Greenleaf, W.J. (2018). High-resolution mapping of cancer cell networks using co-functional interactions. Mol. Syst. Biol. *14*, e8594. https://doi.org/10.15252/msb.20188594.

45. Kim, E., Dede, M., Lenoir, W.F., Wang, G., Srinivasan, S., Colic, M., and Hart, T. (2019). A network of human functional gene interactions from knockout fitness screens in cancer cells. Life Sci. Alliance *2*, e201800278. https://doi.org/10.26508/lsa.201800278.

46. Wang, T.P., Doss, M., Tokar, J.L., Reddy, S., Barta, S.K., Yu, J.Q., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras. Cell *42*, 890–892.e15. https://doi.org/10.1016/j.cell.2017.01.013.

47. Amici, D.R., Jackson, J.M., Truica, M.I., Smith, R.S., Abdulkadir, S.A., and Mendillo, M.L. (2021). FIREWORKS: a bottom-up approach to integrative coessentiality network analysis. Life Sci. Alliance *4*, e202000882. https://doi.org/10.26508/lsa.202000882.

48. Wainberg, M., Kamber, R.A., Balsubramani, A., Meyers, R.M., Sinnott-Armstrong, N., Hornburg, D., Jiang, L., Chan, J., Jian, R., Gu, M., et al. (2021). A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. Nat. Genet. *53*, 638–649. https://doi.org/10.1038/s41588-021-00840-z.

49. Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nat. Methods *9*, 471–472. https://doi.org/10.1038/nmeth.1938.

50. Chua, C.E.L., and Tang, B.L. (2018). Rab 10-a traffic controller in multiple cellular pathways and locations. J. Cell. Physiol. *233*, 6483–6494. https://doi.org/10.1002/jcp.26503.

51. Gulbranson, D.R., Davis, E.M., Demmitt, B.A., Ouyang, Y., Ye, Y., Yu, H., and Shen, J. (2017). RABIF/MSS4 is a Rab-stabilizing holdase chaperone required for GLUT4 exocytosis. Proc. Natl. Acad. Sci. *114*, E8224–E8233. https://doi.org/10.1073/pnas.1712176114.

52. Sano, H., Peck, G.R., Kettenbach, A.N., Gerber, S.A., and Lienhard, G.E. (2011). Insulin-stimulated GLUT4 protein translocation in adipocytes requires the Rab10 guanine nucleotide exchange factor Dennd4C. J. Biol. Chem. *286*, 16541–16545. https://doi.org/10.1074/jbc.C111.228908.

53. Zeng, J., Feng, S., Wu, B., and Guo, W. (2017). Polarized exocytosis. Cold Spring Harb. Perspect. Biol. *9*, a027870. https://doi.org/10.1101/cshperspect.a027870.

54. Khan, T.G., Ginsburg, D., and Emmer, B.T. (2022). The small GTPase RAB10 regulates endosomal recycling of the LDL receptor and transferrin receptor in hepatocytes. J. Lipid Res. *63*, 100248. https://doi.org/10.1016/j.jlr.2022.100248.

55. Islam, M.M., Hlushchenko, I., and Pfisterer, S.G. (2022). Low-density lipoprotein internalization, degradation and receptor recycling along membrane contact sites. Front. Cell Dev. Biol. *10*, 826379.

56. Fedoseienko, A., Wijers, M., Wolters, J.C., Dekker, D., Smit, M., Huijkman, N., Kloosterhuis, N., Klug, H., Schepers, A., Willems van Dijk, K., et al. (2018). The COMMD family regulates plasma LDL levels and attenuates atherosclerosis through stabilizing the CCC complex in endosomal LDLR trafficking. Circ. Res. *122*, 1648–1660. https://doi.org/10.1161/CIRCRESAHA.117.312004.

57. Grindstaff, K.K., Yeaman, C., Anandasabapathy, N., Hsu, S.C., Rodriguez-Boulan, E., Scheller, R.H., and Nelson, W.J. (1998). Sec6/8 complex is recruited to cell-cell contacts and specifies transport vesicle delivery to the basal-lateral membrane in epithelial cells. Cell *93*, 731–740. https://doi.org/10.1016/s0092-8674(00)81435-x.

58. Bermudez-Cabrera, H.C., Culbertson, S., Barkal, S., Holmes, B., Shen, M.W., Zhang, S., Gifford, D.K., and Sherwood, R.I. (2021). Small molecule inhibition of ATM kinase increases CRISPR-Cas9 1-bp insertion frequency. Nat. Commun. *12*, 5111. https://doi.org/10.1038/s41467-021-25415-8.

59. McDermott, U. (2019). Large-scale compound screens and pharmacogenomic interactions in cancer. Curr. Opin. Genet. Dev. *54*, 12–16. https://doi.org/10.1016/j.gde.2019.02.002.

60. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.

61. Hung, J.-H., Yang, T.-H., Hu, Z., Weng, Z., and DeLisi, C. (2012). Gene set enrichment analysis: performance evaluation and usage guidelines. Brief. Bioinform. *13*, 281–291. https://doi.org/10.1093/bib/bbr049.

62. Povysil, G., Petrovski, S., Hostyk, J., Aggarwal, V., Allen, A.S., and Goldstein, D.B. (2019). Rare-variant collapsing analyses for complex traits: guidelines and applications. Nat. Rev. Genet. *20*, 747–759. https://doi.org/10.1038/s41576-019-0177-4.

63. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315. https://doi.org/10.1038/ng.2892.

64. Carter, H., Douville, C., Stenson, P.D., Cooper, D.N., and Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genom. *14*, S3. https://doi.org/10.1186/1471-2164-14-S3-S3.

65. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. *6*, e1001025. https://doi.org/10.1371/journal.pcbi.1001025.

66. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121. https://doi.org/10.1101/gr.097857.109.

67. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980–D985. https://doi.org/10.1093/nar/gkt1113.

68. Cohen, J.C., Boerwinkle, E., Mosley, T.H., and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med. *354*, 1264–1272. https://doi.org/10.1056/NEJMoa054013.

69. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

70. Replogle, J.M., Saunders, R.A., Pogson, A.N., Hussmann, J.A., Lenail, A., Guna, A., Mascibroda, L., Wagner, E.J., Adelman, K., Lithwick-Yanai, G., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell *185*, 2559–2575.e28. https://doi.org/10.1016/j.cell.2022.05.013.

71. ENCODE Project Consortium, Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

72. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501. https://doi.org/10.1038/nbt.1630.

73. Yu, H., Rimbert, A., Palmer, A.E., Toyohara, T., Xia, Y., Xia, F., Ferreira, L.M.R., Chen, Z., Chen, T., Loaiza, N., et al. (2019). GPR146 deficiency protects against hypercholesterolemia and atherosclerosis. Cell *179*, 1276–1288.e14. https://doi.org/10.1016/j.cell.2019.10.034.

74. Koornneef, A., Maczuga, P., van Logtenstein, R., Borel, F., Blits, B., Ritsema, T., van Deventer, S., Petry, H., and Konstantinova, P. (2011). Apolipoprotein B knockdown by AAV-delivered shRNA lowers plasma cholesterol in mice. Mol. Ther. *19*, 731–740. https://doi.org/10.1038/mt.2011.6.

75. Lei, S., Sun, R.-Z., Wang, D., Gong, M.-Z., Su, X.-P., Yi, F., and Peng, Z.-W. (2016). Increased hepatic fatty acids uptake and oxidation by LRPPRC-driven oxidative phosphorylation reduces blood lipid levels. Front. Physiol. *7*, 270. https://doi.org/10.3389/fphys.2016.00270.

76. Ding, Q., Strong, A., Patel, K.M., Ng, S.-L., Gosis, B.S., Regan, S.N., Cowan, C.A., Rader, D.J., and Musunuru, K. (2014). Permanent alteration of PCSK9 with in vivo CRISPR-cas9 genome editing. Circ. Res. *115*, 488–492. https://doi.org/10.1161/CIRCRESAHA.115.304351.

77. Frank-Kamenetsky, M., Grefhorst, A., Anderson, N.N., Racie, T.S., Bramlage, B., Akinc, A., Butler, D., Charisse, K., Dorkin, R., Fan, Y., et al. (2008). Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. Proc. Natl. Acad. Sci. USA *105*, 11915–11920. https://doi.org/10.1073/pnas.0805434105.

78. Welty, F.K. (2020). Hypobetalipoproteinemia and abetalipoproteinemia: liver disease and cardiovascular disease. Curr. Opin. Lipidol. *31*, 49–55. https://doi.org/10.1097/MOL.0000000000000663.

79. Wellcome Trust Case Control Consortium, Maller, J.B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J.M.M., Auton, A., Myers, S., et al. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. *44*, 1294–1301. https://doi.org/10.1038/ng.2435.

80. Chen, Y., and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. *48*, D127–D131. https://doi.org/10.1093/nar/gkz757.

81. Moore, A., Wu, L., Chuang, J.-C., Sun, X., Luo, X., Gopal, P., Li, L., Celen, C., Zimmer, M., and Zhu, H. (2019). Arid1a loss drives nonalcoholic steatohepatitis in mice through epigenetic dysregulation of hepatic lipogenesis and fatty acid oxidation. Hepatology *69*, 1931–1945. https://doi.org/10.1002/hep.30487.

82. Ramdas, S., Judd, J., Graham, S.E., Kanoni, S., Wang, Y., Surakka, I., Wenz, B., Clarke, S.L., Chesi, A., Wells, A., et al. (2021). A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids. Genomics *109*, 1366–1387. https://doi.org/10.1101/2021.12.07.470215.

83. Anastasia, I., Ilacqua, N., Raimondi, A., Lemieux, P., Ghandehari-Alavijeh, R., Faure, G., Mekhedov, S.L., Williams, K.J., Caicci, F., Valle, G., et al. (2021). Mitochondria-rough-ER contacts in the liver regulate systemic lipid homeostasis. Cell Rep. *34*, 108873. https://doi.org/10.1016/j.celrep.2021.108873.

84. MÜLLER, C. (1939). Angina pectoris in hereditary xanthomatosis. Arch. Intern. Med. *64*, 675–700. https://doi.org/10.1001/archinte.1939.00190040016002.

85. Wilson, P.W.F. (2013). Lipids and vascular disease: a framingham perspective. Glob. Heart *8*, 25–33. https://doi.org/10.1016/j.gheart.2012.12.009.

86. Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A., and Brûlet, P. (1995). Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. Dev. Camb. Engl. *121*, 3279–3290.

87. Nada, S., Yagi, T., Takeda, H., Tokunaga, T., Nakagawa, H., Ikawa, Y., Okada, M., and Aizawa, S. (1993). Constitutive activation of Src family kinases in mouse embryos that lack Csk. Cell *73*, 1125–1135. https://doi.org/10.1016/0092-8674(93)90642-4.

88. Lv, P., Sheng, Y., Zhao, Z., Zhao, W., Gu, L., Xu, T., and Song, E. (2015). Targeted disruption of Rab10 causes early embryonic lethality. Protein Cell *6*, 463–467. https://doi.org/10.1007/s13238-015-0150-8.

89. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508–514. https://doi.org/10.1038/nature19356.

90. Li, Z., Schulze, R.J., Weller, S.G., Krueger, E.W., Schott, M.B., Zhang, X., Casey, C.A., Liu, J., Stöckli, J., James, D.E., and McNiven, M.A. (2016). A novel Rab10-EHBP1-EHD2 complex essential for the autophagic engulfment of lipid droplets. Sci. Adv. *2*, e1601470. https://doi.org/10.1126/sciadv.1601470.

91. Suda, Y., Nakabayashi, J., Matsuo, I., and Aizawa, S. (1999). Functional equivalency between Otx2 and Otx1 in development of the rostral head. Development *126*, 743–757. https://doi.org/10.1242/dev.126.4.743.

92. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., NHLBI GO Exome Sequencing Project—ESP Lung Project Team; Christiani, D.C., Wurfel, M.M., and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am. J. Hum. Genet. *91*, 224–237. https://doi.org/10.1016/j.ajhg.2012.06.007.

93. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784. https://doi.org/10.1038/nmeth.3047.

94. Jeng, E.E., Bhadkamkar, V., Ibe, N.U., Gause, H., Jiang, L., Chan, J., Jian, R., Jimenez-Morales, D., Stevenson, E., Krogan, N.J., et al. (2019). Systematic Identification of Host Cell Regulators of Legionella pneumophila Pathogenesis Using a Genome-wide CRISPR Screen. Cell Host & Microbe, ISSN 1931-3128 *26*, 551–563.e6. https://doi.org/10.1016/j.chom.2019.08.017.

95. Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554. https://doi.org/10.1186/s13059-014-0554-4.

96. Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. G3 *7*, 2719–2727. https://doi.org/10.1534/g3.117.041277.

97. Erard, N., Knott, S.R.V., and Hannon, G.J. (2017). A CRISPR resource for individual, combinatorial, or Multiplexed gene knockout. Mol. Cell *67*, 348–354.e4. https://doi.org/10.1016/j.molcel.2017.06.030.

98. Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491. https://doi.org/10.1016/j.cell.2013.12.001.

99. Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge – accurate paired shotgun read merging via overlap. PLoS One *12*, e0185056. https://doi.org/10.1371/journal.pone.0185056.

100. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

101. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinforma. Oxf. Engl. *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

102. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

103. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

104. Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47*, D419–D426. https://doi.org/10.1093/nar/gky1038.

105. van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

106. Hail Team (2022). Hail. https://doi.org/10.5281/ZENODO.6502316.

107. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

108. Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R., et al. (2022). Ensembl 2022. Nucleic Acids Res. *50*, D988–D995. https://doi.org/10.1093/nar/gkab1049.

109. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

110. Liu, X., Li, C., Mou, C., Dong, Y., and Tu, Y. (2020). dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med. *12*, 103. https://doi.org/10.1186/s13073-020-00803-9.

111. Fife, J.D., Tran, T., Bernatchez, J.R., Shepard, K.E., Koch, C., Patel, A.P., Fahed, A.C., Krishnamurthy, S., Regeneron Genetics Center, DiscovEHR Collaboration, et al. (2021). A framework for integrated clinical risk assessment using population sequencing data (Genetic and Genomic Medicine). https://doi.org/10.1101/2021.08.12.21261563.

112. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

113. Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. J. Open Source Softw. *4*, 1317. https://doi.org/10.21105/joss.01317.

114. Szustakowski, J.D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P.G., Sasson, A., Wong, E., Liu, D., Wade Davis, J., Haefliger, C., et al. (2021). Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. Nat. Genet. *53*, 942–948. https://doi.org/10.1038/s41588-021-00885-0.

115. Patel, A.P., Wang, M., Fahed, A.C., Mason-Suares, H., Brockman, D., Pelletier, R., Amr, S., Machini, K., Hawley, M., Witkowski, L., et al. (2020). Association of rare pathogenic DNA variants for familial hypercholesterolemia, hereditary breast and ovarian cancer syndrome, and lynch syndrome with disease risk in adults according to family history. JAMA Netw. Open *3*, e203959. https://doi.org/10.1001/jamanetworkopen.2020.3959.

116. Lee, J.H., Giannikopoulos, P., Duncan, S.A., Wang, J., Johansen, C.T., Brown, J.D., Plutzky, J., Hegele, R.A., Glimcher, L.H., and Lee, A.-H. (2011). The transcription factor cyclic AMP–responsive element–binding protein H regulates triglyceride metabolism. Nat. Med. *17*, 812–815. https://doi.org/10.1038/nm.2347.

117. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

118. Sarkar, S.K., and Chang, C.-K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. J. Am. Stat. Assoc. *92*, 1601–1608. https://doi.org/10.1080/01621459.1997.10473682.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Purified Mouse Anti-Human LDLR	BD Biosciences	Cat#565641; RRID:AB_2739317
Donkey anti-mouse AlexaFluor647	Jackson ImmunoResearch	Cat#715-605-151; RRID:AB_2340863
APC anti-human CD36L1	BioLegend	Cat#363207; RRID:AB_2721444
APC anti-human CD132	BioLegend	Cat#350007; RRID:AB_10722613
APC anti-human CD71	BioLegend	Cat#334107; RRID:AB_10916388
APC anti-human CD222	BioLegend	Cat#364205; RRID:AB_2904389
Bacterial and virus strains		
NEB Stable Competent *E. coli* (High Efficiency)	New England Biolabs	Cat#C3040H
AAV8-scramble-shRNA	Vector Biolabs	N/A
AAV8-*Rabif*-shRNA	Vector Biolabs	N/A
AAV8-Otx2-shRNA	Vector Biolabs	N/A
AAV8-*Csk*-shRNA	Vector Biolabs	N/A
Chemicals, peptides, and recombinant proteins		
Bodipy-LDL	Thermo Fisher Scientific	Cat#L3483
NGI-1	Cayman	Cat#31519
ATN-224	Cayman	Cat#23553
Pyripyropene A	Cayman	Cat#11896
Fatostatin	Cayman	Cat#13562
KX2-391	Cayman	Cat#21429
MHY1485	Cayman	Cat#18008
GW3965	Cayman	Cat#10054
Rocaglamide A	MedChemExpress	Cat#HY-19356
Rho Activator II	Cytoskeleton.com	Cat#CN03-A
Tempol	Cayman	Cat#27051
Simvastatin	Cayman	Cat#10010344
Puromycin Dihydrochloride	Gibco	Cat#A1113803
NEBuilder HiFi DNA Assembly Master Mix	New England Biolabs	Cat#E2621L
DMEM, high glucose, pyruvate	Gibco	Cat#11995073
Fetal Bovine Serum	Thermo Fisher Scientific	Cat#10437028
Penicillin-Streptomycin (10,000 U/mL)	Gibco	Cat#15140122
Opti-MEM Reduced Serum Medium	Gibco	Cat#11058021
Trypsin-EDTA (0.25%), phenol red	Gibco	Cat#25200072
PureLink™ Genomic DNA Mini Kit	Invitrogen	Cat#K182002
DPBS, no calcium, no magnesium	Gibco	Cat# 14190250
QIAquick PCR Purification Kit	Qiagen	Cat#28106
QIAquick Gel Extraction Kit Qiagen	Qiagen	Cat#28706
QIAprep Spin Miniprep Kit Qiagen	Qiagen	Cat#27106
SYBR Safe DNA Gel Stain	Invitrogen	Cat#S33102
BsmBI-v2	New England Biolabs	Cat#R0739L
1 Kb Plus DNA Ladder	Invitrogen	Cat#10787026
UltraPure™ Agarose	Invitrogen	Cat#16500100
NEBNext Ultra II Q5 Master Mix	New England Biolabs	Cat#M0544L
Dimethyl sulfoxide	Sigma-Aldrich	Cat#D2650
Polybrene infection/Transfection Reagent	Sigma-Aldrich	Cat#TR-1003

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNEasy Plus mini kit	Qiagen	Cat#74134
Zymo Quick-RNA 96 kit	Zymo Research	Cat# R1052
Lexogen QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina	Lexogen	Cat#015.96
*Trans*IT-Lenti Transfection Reagent	Mirus	Cat#MIR6605
Lenti-X Concentrator	Takara	Cat#631232
Erythrina Cristagalli Lectin (ECL), Biotinylated	Vector Biolabs	Cat#B-1145-5
Wheat Germ Agglutinin (WGA), Biotinylated	Vector Biolabs	Cat# B-1025-5
45% fat Kcal diet (0.2% total cholesterol)	Envigo	TD.10885
Deposited data		
RNA-seq data	This paper	Tables S1 and S6 and GEO: GSE229650
CRISPR screen data	This paper	Tables S2, S3, and S4 and SRA: PRJNA943253
Experimental models: Cell lines		
HepG2	ATCC	Cat#HB-8065
HEK293T	Gift from Dr. Steven Elledge	Harvard Medical School, Boston, MA
Experimental models: Organisms/strains		
C57BL/6	*InVivos*	N/A
Oligonucleotides		
gRNA_60bp_fw TAACTTGAAAGTATTTCGATTTCTT GGCTTTATATATCTTGTGGAAAGGACGAAACACCG	This paper	method details
gRNA_60bp_rv GTTGATAACGGACTAGCCTTATTTA AACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC	This paper	method details
U6_Bc_r1seq_halftail (24 distinct versions of this primer with staggered-length in-line barcodes denoted here as NNNNN) 5' ACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNN GGAAAGGACGAAACACCG 3'	This paper	method details
gRNAFE_r2seq_halftail 5' GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGC CTTATTTAAACTTGCTATGCTGT 3'	This paper	method details
r1seq_fulltail 5' AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTC 3'	This paper	method details
r2seq_fulltail (up to 8 distinct indexed versions of this primer were used to maximize pooling) 5' CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACT GGAGTTCAGACGTGTGCT 3'	This paper	method details
See Table S21 for individual gRNA oligos	This paper	method details
Recombinant DNA		
pCMV-VSV-G	Gift from Dr. Ricardo Weinlich	Hospital Israelita Albert Einstein, Brazil
pMDLg-pRRE	Gift from Dr. Ricardo Weinlich	Hospital Israelita Albert Einstein, Brazil
pRSV-Rev	Gift from Dr. Ricardo Weinlich	Hospital Israelita Albert Einstein, Brazil
LentiCRISPR v2	Sanjana et. al., Nat. Methods[93]	Addgene Cat#52961
pCW-Cas9-Blast	Gift from Dr. Mohan Babu	Addgene Cat#83481
GFP-RAB10 lenti	Jeng et al., Cell Host & Microbe[94]	Addgene Cat#130883

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
lenti GFP-RAB10 3'UTRmaj PuroR	This paper	Addgene Cat#186736
lenti GFP-RAB10 3'UTRmin PuroR	This paper	Addgene Cat#186737
pLenti_rtTA-P2A-BFP_HygroR	This paper	Addgene Cat#186738
lentiTet-LDLR BlastR	This paper	Addgene Cat#186739
lentiTet-RAB10-T2A-RABIF-P2A-mCherry BlastR	This paper	Addgene Cat#186740
lentiTet-OTX2-P2A-mCherry BlastR	This paper	Addgene Cat#186741
lentiTet-CSK-P2A-mCherry BlastR	This paper	Addgene Cat#186742
lentiTet-P2A-mCherry BlastR	This paper	Addgene Cat#186743
pLenti_EF1a-scFv-SBNO1-KBP250-NFEL1-GSTGGT-KRT40-T2A-MCP-P65-HSF1-BFP_HygroR	This paper	Addgene Cat#186744
p2T CAG dCas9-10XGcn4-P2A-mCherry BlastR	This paper	Addgene Cat#186745
lentiCRISPR-v2-FE-PuroR	This paper	Addgene Cat#186746
Software and algorithms		
FCS Express 7	*De Novo* Software	https://denovosoftware.com/
MAGeCK	Li et. al., Genome Biology[95]	https://sourceforge.net/projects/mageck/
Excel	Microsoft	https://www.microsoft.com/en-us/microsoft-365/excel
Prism 8.2.0	GraphPad	https://www.graphpad.com
Rstudio	Rstudio	https://rstudio.com
STAR aligner (v2.5.2b)	STAR aligner	https://github.com/alexdobin/STAR
HT-seq (v0.9.1)	HT-seq	https://htseq.readthedocs.io/en/master/
DESeq2 1.26.0	Bioconductor	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Biorender	Biorender	https://biorender.com/
FloJo version 10.6.2	BD Biosciences	https://www.flowjo.com/
Code generated in this manuscript and publicly available data	This paper	Mendeley Data: https://doi.org/10.17632/9wgk5ny69n.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Richard I. Sherwood (rsherwood@bwh.harvard.edu)

Materials availability

Plasmids generated in this study have been deposited to Addgene (Deposit 81242) and will be publicly available as of the date of publication.

Data and code availability

- RNA-seq and CRISPR screening data that support the findings of this study are provided in the supplemental information. Data has been deposited at SRA (Bioproject ID PRJNA943253).
- All original code and publicly available data is available at Mendeley repository https://doi.org/10.17632/9wgk5ny69n.1 and https://github.com/j-fife/hamilton-et-al.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

HepG2 hepatocellular carcinoma cells.

Animal models

C57BL/6 mice.

METHOD DETAILS

CRISPR-KO screening

For KO-Library 1, four gRNAs were designed targeting each of the 2,634 genes within 1 Mb of a lead variant from the MVP multiethnic GWAS cohort[3]. gRNAs were selected from the Brunello, TKOv3, and CROATAN libraries[42,96,97] to maximize inDelphi-predicted frameshift fraction[28]. 200 non-targeting control gRNAs were chosen from the Brunello library. KO-Library 2 included the same four gene-targeting gRNAs from KO-Library 1 for 360 genes that showed strong phenotypic effects as well as the same non-targeting controls. KO-Library 3 contained the 4 Brunello gRNAs for 1,803 transcription factor genes[31,32] as well as Brunello non-targeting controls. KO-Library 4 included four Brunello gRNAs targeting 522 genes with significant or near-significant effects on LDL-C uptake from the three previous screens as well as from a recently published genome-wide LDL-C uptake CRISPR-Cas9 screen performed in Huh7 cells[22] and also included Brunello non-targeting controls.

All oligonucleotide libraries (Tables S17, S18, S19, and S20) were ordered from Twist Biosciences or Agilent Technologies in the following sequence format:

CTTGTGGAAAGGACGAAACACCG [19-20-bp protospacer—remove initial G for any 20-bp protospacer with one natively] GTTTAAGAGCTATGCTGGAAACAGCATAGC

Libraries were amplified by PCR using Q5UltraII mastermix (NEB) using the following primers:

gRNA_60bp_fw TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG

gRNA_60bp_rv GTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC

gRNA libraries were cloned into the lentiCRISPR-v2-FE-PuroR backbone (derived from Addgene plasmid #52961[93] by replacing the gRNA hairpin with the FE gRNA hairpin[98]) through BsmBI vector digest and NEBuilder HiFi DNA assembly, ensuring >100-fold representation of each gRNA.

HepG2 Cas9NG-P2A-mCherry cells were seeded at a density of $4x10^4$ cells/cm^2 on 15 cm plates in four biological replicates with 8 μg/mL of polybrene. The cells were transduced with lentivirus using Trans-IT Lenti (Mirus Bio) at a multiplicity of infection (MOI) of 0.3-0.5 as determined by titration. Two days post-transduction, cells were treated with 500 ng/mL puromycin and were selected for 5-7 days to enrich for infected cells.

After selection, cells were seeded at a density of $1.08x10^5$ cells/cm^2 on 15 cm plates and left to incubate overnight. The next day, the media was replaced with either DMEM + 10% FBS (serum condition) or optiMEM (serum-starved condition) and cells were incubated overnight. Approximately 4-6 hours prior to sorting, cells were treated with 2.5 μg/mL BODIPY™ FL LDL (Thermo Fisher) in optiMEM. Cells were trypsinized and sorted into four bins (bottom 12.5%, bottom 12.5-37.5%, top 37.5-12.5%, and top 12.5%) or two bins as specified in the text using a BD FACSAria flow cytometer. After sorting, genomic DNA was isolated from cells using the Purelink Genomic DNA mini kit (Thermo Fisher), and up to 20 μg of genomic DNA per sample was used to amplify the U6-3' to gRNA hairpin region with different in-line barcodes. PCR2 was performed to add full-length Illumina sequencing adapters using internally ordered primers with equivalent sequences to NEBNext Index Primer Sets 1 and 2 (New England Biolabs). All PCRs were performed using Q5UltraII polymerase (NEB). Pooled samples were sequenced using NextSeq (Illumina), using 75-nt reads and collecting greater than 100 reads per gRNA in the library.

The library prep primers were as follows:

PCR1:

U6_Bc_r1seq_halftail (24 distinct versions of this primer with staggered-length in-line barcodes denoted here as NNNNN were used)

5' ACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNN GGAAAGGACGAAACACCG 3'

gRNAFE_r2seq_halftail

5' GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTTATTTAAACTTGCTATGCTGT 3'

PCR2:

r1seq_fulltail

5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC 3'

r2seq_fulltail (up to 8 distinct indexed versions of this primer were used to maximize pooling)

5' CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCT 3'

CRISPR-Cas9 screening analysis was performed using MAGeCK RRA v0.5.9.2 paired analysis comparing top and bottom sorted bins and normalized to non-targeting control gRNAs present in the library.

For the small molecule CRISPR-Cas9 screens, HepG2 Cas9NG-P2A-mChe cells containing the KO-Library 4 were seeded at a density of $1.08x10^5$ cells/cm^2 on 10 cm plates in 3 replicates per condition and left to incubate overnight. The next day, the media was replaced with optiMEM (serum-starved condition) containing either control (DMSO) or one of eight inhibitors (see below) and cells were incubated overnight. Approximately 4-6 hours prior to sorting, cells were treated with 2.5 μg/mL BODIPY™ FL LDL (Thermo Fisher) in optiMEM. Cells were trypsinized and sorted into two bins (bottom 30% and top 30%) using a BD FACSAria flow cytometer. After sorting, genomic DNA was isolated from cells and PCR and sequencing was performed as described above.

Small molecule CRISPR-Cas9 screening analysis was performed using MAGeCK RRA v0.5.9.2 paired analysis comparing top and bottom sorted bins and normalized to non-targeting control gRNAs present in the library. A default number of permutations were performed except when indicated otherwise in figure legends.

NGI-1 (Cayman) 5μM
ATN-224 (Cayman) 40 μM
KX2-391 (Cayman) 2.5 μM
Pyripyropene A (Cayman) 10 μM
Fatostatin (Cayman) 2.5 μM
MHY1485 (Cayman) 10 μM
GW3965 (Cayman) 0.02 μM
Rocaglamide A (MedChemExpress) 0.01 μM
Simvastatin (Cayman) 0.5 uM

Cloning and testing of individual gRNAs

Oligonucleotides including protospacer sequences (Table S21) were ordered in the following format:

GGAAAGGACGAAACACCG [19-20-bp protospacer—remove initial G for any 20-bp protospacer with one natively] GTTTAAGAGCTATGCTGGAAAC

were amplified by PCR to create homology arms and cloned into a lentiCRISPR-v2-FE-PuroR backbone through NEBuilder HiFi DNA assembly as described above. To make KO cell lines, the gRNA constructs were packaged into lentivirus and transduced into HepG2-Cas9NG-P2A-mChe cells seeded at $4x10^4$ cells/cm^2 on 48-well plates in two replicates with 8 μg/mL of polybrene. Two days post-transduction, cells were treated with 500 ng/mL puromycin and were selected for approximately one week.

HepG2 Cas9NG-P2A-mChe KO cell lines were seeded 1:1 with HepG2 WT cells to achieve a total density of $1.08x10^5$ cells/cm^2 on a 96 well plate in at least two technical replicates of two biological replicates per serum condition and incubated overnight. The next day, the media was replaced with either DMEM + 10% FBS (serum condition) or optiMEM (serum-starved condition) and cells were incubated overnight. Approximately 4-6 hours prior to FACS analysis, cells were treated with 2.5 μg/mL BODIPY^{TM} FL LDL in optiMEM. Cells were trypsinized and analyzed for presence of mCherry and LDL uptake using a Beckman CytoFLEX flow cytometer. LDL uptake of each KO cell line was normalized to the LDL uptake of the WT cells within the same well. Differential LDL uptake between KO and control cells was further normalized using data from the control sgCTRL KO line.

ICE analysis was performed using Sanger sequencing reads and significance cut-off value for decomposition of $P < 0.001$.

Small molecule analysis

For the individual small molecule analysis, HepG2 Cas9NG-P2A-mChe cells were seeded at a density of $1.08x10^5$ cells/cm^2 on a 96-well plate in 3-26 replicates per condition and left to incubate overnight. The next day, the media was replaced with optiMEM (serum-starved condition) containing either control (DMSO) or one of 24 inhibitors (Table S22) and cells were incubated overnight. Approximately 4-6 hours prior to analysis, cells were treated with 2.5 μg/mL BODIPY^{TM} FL LDL in optiMEM. Cells were trypsinized, resuspended in FACS suspension media, and analyzed by flow cytometric analysis using a Beckman CytoFLEX flow cytometer. Differential LDL-C uptake between small molecules and control treatment was further normalized using data from DMSO treated cells.

Lectin flow cytometry analysis

HepG2 individual gene KO cell lines were cultured in DMEM + 10% FBS with or without 5 μM NGI-1 for 24 hours. Cells were then trypsinized and 2.5*10^5 cells were resuspended in FACS suspension media (DMEM, no phenol red (Thermo Fisher 31053028) + 2% FBS + 2mM EDTA) containing streptavidin-Dy488 (2.5 ug/mL) and a biotinylated lectin (ECL (5 ug/mL) or WGA (0.625 ug/mL)). Staining reactions were incubated for 20 minutes, protected from light. Cells were washed twice and resuspended in FACS suspension media followed by flow cytometric analysis using a Beckman CytoFLEX flow cytometer. Lectin binding of each KO cell line was normalized using data from the control sgControl line.

RNA-seq

HepG2 Cas9NG-P2A-mChe KO cell lines were seeded at a density of $2.1x10^5$ cells/cm^2 on a 24 well plate with two biological replicates per serum condition and incubated overnight. The next day, the media was replaced with either DMEM + 10% FBS (serum condition) or optiMEM (serum-starved condition) and cells were incubated overnight. RNA was harvested from cells using the Qiagen RNEasy Plus mini kit or the Zymo Quick-RNA 96 kit, prepared for RNA-seq using the Lexogen QuantSeq-Pool kit, and sequenced using Illumina Nextseq at >1*10^6 reads/sample.

RNA-seq reads were mapped using the Quantseq 3' mRNA mapping pipeline as described by Lexogen to GRCh38.p12. Briefly, reads were first trimmed using bbduk from the bbmap suite (v38.79)[99] trimming for low quality tails, poly-A read-through and adapter contamination using the recommended parameters. Then, reads were mapped using the STAR aligner (v2.5.2b)[100] with the recommended modified-Encode settings. Finally, HT-seq (v0.9.1) count was used to obtain per-gene counts[101].

Within each cell line, we conducted differential expression analysis using DESeq2 1.26.0 to identify significantly differentially expressed genes for each gene KO with respect to the control condition.

OTX1 knockdown analysis was performed using published genome-wide Perturb-seq data[70]. Relative magnitude of change in expression (z-score) of 8,248 genes in K562 cells with *OTX1* knockdown was used as input in co-essential GSEA analysis using gseapy v.0.10.448 with 10,000 iterations and 60 parallel processes as negative control. The modules were analyzed according to their normalized enrichment scores (NES), and GO terms from PANTHER version 14 were linked as described below.

CRISPRa Outcome and Phenotype (COP) screen design and execution

An oligonucleotide library was designed pairing the proximal promoter of each of 200 target genes (194-nt upstream of the annotated TSS and 10-nt downstream) with 10 distinct gRNAs each, including 8 on-target gRNAs and 2 non-targeting gRNAs. On-target gRNAs were chosen as the 8 gRNAs with lowest Combined Rank from the CRISPick CRISPRa mode[40] after filtering gRNAs that (i) are 1-nt offset from a lower-ranked gRNA on the same strand; (ii) contain an MfeI recognition site since MfeI was used in cloning this library. Off-target gRNAs were chosen from the non-targeting gRNAs in the Calabrese library[40].

The COP oligonucleotide library (Table S3) was ordered in the following format:

GGAAAGGACGAAACACCG [19-20-nt spacer] gtttaagagctaggccAACAATTGTCAACAGACCATGCC [205-nt promoter] GCTAG CTTGAAGGGGACG

The oligonucleotide library was amplified by PCR to create homology arms by using the primers; 010415_sgRNA_60bp_fw: TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG, and 031621_CRISPRArepPCR1A_rv: TG GTGAGTAGGAGGAAGAGGAAGCGCTTCATGGTGGCACGTCCGTCCCCTTCAAGCTAGC. In order to introduce 15 nt random barcodes, two separate fragments were also amplified by PCR wherea CD5F2_DPE_CTE_Puro-containing plasmid was used as the template by using the primers; 031621_CRISPRArep_PCR1B_fw: GCTAGCTTGAAGGGGACG and 031621_CRISPRArep_PCR1B_rv: CTAAGACAGGCGCAGCCTCCGAGGGATGTGTACATTTGGATGCAGGTCGAAAGGCNNNNNNNNNNNNNNNNTCAACTGAAGCAG AAGAGGT, 031621_CRISPRArep_PCR1C_fw: GCCTTTCGACCTGCATCC and 031621_CRISPRArep_PCR1C_rv: CGGTGTTTC GTCCTTTCCAC, respectively. These three amplicons were then cloned into a plasmid backbone produced with NheI-HF and AgeI-HF double-digest of pLenti_CD5F2_MCPp65_Puro through NEBuilder HiFi, upscaled to 100ul including 0.05 pmol library, and 0.1 pmol of each PCR amplicon. After treating with Plasmid-Safe (Lucigen), the circularized assembly product was linearized with MfeI-HF digest and then was eluted using QIAquick PCR purification kit (Qiagen). The linearized assembly reaction was further amplified by PCR using the following primers; 072121_CRISPRArep_PCR2A_fwnew: tattacagggacagcagagatccagtttggttaattaaTTCC TTGTCAACAGACCATGCC and 050415_SAMsgRNA_60bp_rv atttaaacttgctaggccCTGCAGACATGGGTGATCCTCA TGTTggcctagctcttaaac and cloned into NheI+AgeI-digested pLenti_CD5F2_MCPp65_Puro backbone to produce a high throughput pooled library bearing proximal promoters of 200 LDL-related genes, paired gRNAs, as well as the CRISPRa SAM component MCP-p65-HSF1 in the same vector.

HepG2 cells used in the COP assay were first co-transfected with a non-autonomous Tol2 vector bearing dCas9-10xGcn4-mCherry-BlastR cassette (p2T CAG dCas9-10XGcn4-P2A-mCherry BlastR) and a Tol2 transposase-expressing vector using lipofectamine 3000 (Thermo Fisher). One day post-transfection, the cells were treated with 5ug/mL blasticidin and were selected for approximately 5 days. The cells were further sorted twice using a BD FACSAria flow cytometer for mCherry expression to produce a highly enriched HepG2$^{dCas910x\text{-}Gcn4\text{-}mChe}$ cell population. This cell line was then transduced with a SunTag transcriptional activator lentiviral vector (pLenti_EF1a-scFv-SBNO1-KBP250-NFE2L1-GSTGGT-KRT40-P2A-BFP_HygroR) to produce a cell line stably expressing CRISPRa; HepG2$^{dCas9\text{-}10xGcn4\text{-}mChe\text{-}scFv\text{-}Sbno1\text{-}Nfe2l1\text{-}Krt40\text{-}BFP}$. These cells were subsequently treated with the lentiviral COP library at a multiplicity of infection (MOI) of ~0.5. Two days post-transduction, cells were treated with 500 ng/mL puromycin and were selected for approximately one week.

To build a library connecting barcodes and gRNAs, genomic DNA was isolated from cells using the Purelink Genomic DNA mini kit (Thermo Fisher), and up to 20 μg of genomic DNA per sample was used to amplify promoter-barcode and barcode-gRNA pairs as dictionary library samples separately in PCR by using different in-line barcodes including primers shown below:

Paired Promoter-Barcode NGS sample

031621_CRISPRApro_r1seq_2-4N: CTTTCCCTACACGACGCTCTTCCGATCT(N)$_{2\text{-}4}$TCCTTGTCAACAGACCATGCC

031621_IntPri_r2seq_1-3N:

GGAGTTCAGACGTGTGCTCTTCCGATCT(N)$_{1\text{-}3}$TTTGGATGCAGGTCGAAAGGC

Paired Barcode-gRNA NGS sample

031621_CRISPRa_CD5F2_r1seq_2-4N: CTTTCCCTACACGACGCTCTTCCGATCT(N)$_{2\text{-}4}$CACCTCTTCTGCTTCAGTTGA

031621_CRISPRA_MS2hp_r2seq_1-3N: GGAGTTCAGACGTGTGCTCTTCCGATCT(N)$_{1\text{-}3}$CATGTTGGCCTAGCTCTTAAAC

PCR2 was performed to add full-length Illumina sequencing adapters using internally ordered primers with equivalent sequences to NEBNext Index Primer Sets 1 and 2 (New England Biolabs). All PCRs were performed using Q5UltraII polymerase (NEB). Pooled samples were sequenced using NextSeq (Illumina), using 40-45-nt reads and collecting greater than 100 reads per construct in the library.

In parallel, total RNA (RNEasy Maxi kit, Qiagen) and genomic DNA (Purelink Genomic DNA mini kit) were isolated from each biological replicate. cDNA samples were synthesized from total RNA samples with reverse transcription by using sequence specific primer, 031621_CTE_RTprimer: CCTCCGAGGGATGTGTACA, targeting downstream of the 15nt barcode site. In order to quantify the transcript abundancy of barcodes which would eventually demonstrate the activity of the promoter library in response to their paired gRNAs, 10ug gDNA and 20ug cDNA were PCR amplified by using 8 different r1, and a constant r2 primer for each samples:

122320_CD5F2_3pr2seq: GGAGTTCAGACGTGTGCTCTTCCGATCTNNNGCCAATGCCATAATCCACCTCT
121016_IntPri_r1seqtail_Bc: CTCTTTCCCTACACGACGCTCTTCCGATCT(N)$_{2\text{-}6}$TTTGGATGCAGGTCGAAAGGC

PCR2 was performed to add full-length Illumina sequencing adapters using internally ordered primers with equivalent sequences to NEBNext Index Primer Sets 1 and 2 (New England Biolabs). All PCRs were performed using Q5UltraII polymerase (NEB). Pooled samples were sequenced using NextSeq (Illumina), using 75-nt reads and collecting greater than 100 reads per gRNA in the library.

For phenotypic LDL-C uptake screening, COP library-containing HepG2$^{dCas9\text{-}10xGcn4\text{-}mChe\text{-}scFv\text{-}Sbno1\text{-}Nfe2l1\text{-}Krt40\text{-}BFP}$ cells were seeded at a density of 1.08x10^5 cells/cm^2 on 10 cm plates and left to incubate overnight. The next day, the media was replaced with optiMEM (serum-starved condition) and cells were incubated overnight. Approximately 4-6 hours prior to FACS sorting, cells were treated with 2.5 μg/mL BODIPY^{TM} FL LDL in optiMEM. Cells were trypsinized and sorted into 2 bins (bottom 30%, top 30%) using a BD FACSAria flow cytometer. To quantify the gRNA abundance in sorted cell populations, genomic DNA was isolated from cells using the Purelink Genomic DNA mini kit (Thermo Fisher), and up to 20 μg of genomic DNA per sample was used to amplify U6-3' to gRNA-FE hairpin region with in-line barcodes including primer pairs, all of which are different for each sample, shown below:

101317_U6PE1_BcXr1: ACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNN GGAAAGGACGAAACACCG
031621_CRISPRA_MS2hp_r2seq_1-3N: GGAGTTCAGACGTGTGCTCTTCCGATC(N)$_{1\text{-}3}$CATGTTGGCCTAGCTCTTAAAC

PCR2 was performed to add full-length index primers, shown below. All PCRs were performed using Q5UltraII polymerase (NEB). Pooled samples were sequenced using NextSeq (Illumina), using 50-nt reads and collecting greater than 100 reads per gRNA in the library.

031317_10xr2seq_24bp_N70X_fw: CAAGCAGAAGACGGCATACGAGATNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCT
110717_PE1_fulltail: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC

COP data analysis

The Promoter-Barcode-gRNA dictionary was built upon two sets of pair-end sequencing datasets. In the first sequencing data, each R1 read contains a promoter sequence (>20-nt) from the 200 genes and the corresponding R2 read contains a 15-nt variable barcode sequence. A barcode-promoter pair was considered valid if the barcode predominantly (in >90% of reads) pairs with a single promoter sequence. In the second sequencing data, each R1 read contains the reverse complement of a barcode sequence, and the corresponding R2 read contains the reverse complement of a 19-20-nt gRNA sequence from the 2000 gRNAs targeting the promoters of the 200 genes. Similarly, a barcode-gRNA pair was considered valid if the barcode predominantly (in >90% of reads) pairs with a single gRNA sequence. The barcode-promoter pairs and the barcode-gRNA pairs were merged and saved to the promoter-barcode-gRNA dictionary if the gRNA and the promoter were from the same gene and the Hamming distance between the barcodes was less than 2.

For the cDNA and gDNA reporter sequencing data, the 15-nt variable barcodes were extracted from R1 reads, and matched to the existing barcodes in the promoter-barcode-gRNA dictionary if the Hamming distance was less than 2. For each gRNA,the total number of reads for every matched barcode in the cDNA and gDNA data was aggregated and normalized to reads per million (RPM). The efficiency of the gRNA was calculated as the log2 fold change (LFC) of the cDNA to gDNA barcode count. Within each gene, the gRNA efficiency was proportionally scaled to the 0-1 range, where the gRNA with the lowest LFC had an efficiency of 0, and the gRNA with the highest LFC had an efficiency of 1.

gRNA and promoter features that were used as predictive variables in the linear regression model for gRNA efficiency include CRISPick onTarget efficacy (continuous), CRISPick onTarget rank (ordinal), CRISPick offTarget rank (ordinal), Deep SpCas9 score (continuous), TATA motif presence (nominal), InR motif presence (nominal), gRNA orientation (nominal), and cutsite transcription start site (TSS) offsite bins (nominal). For each gRNA, the cutsite TSS offset was defined as the distance from the TSS to the cutsite. cutsite TSS offset was further grouped into 10nt bins for cutsite within 80-150nt range downstream of the TSS. To ensure model consistency, the 1600 gRNAs were bootstrapped into 100 subsamples, each containing 1600 gRNAs with replacement. For each of the subsamples, a linear regression model was built using the above mentioned features as predictive variables and the gRNA efficiency as the response variable.

The LDL-C uptake gDNA sequencing data is made up of 3 replicates after quality control curation for technical data quality, and each replicate consists of a "top30" sample and a "bottom30" sample. Sequences were extracted from R1 reads and matched to the 2000 gRNA library. For each gRNA in the library, we recorded the matched read count for the "top30" sample and the "bottom30" sample, and calculated the "Top30/Bottom30" read count ratio. Pearson correlations were calculated for each replication pair on the "Top30/Bottom30" ratio. Read counts were selected as the input to the MAGeCK "test" module to calculate the log2 fold change (LFC) and the associated p-value of LDL-C uptake for each gene. To run the MAGeCK "test" module, we specified "Top30" samples (3 replicates) as treatment samples, "Bottom30" samples (3 replicates) as control samples, and we used control gRNAs as the normalization method.

Cloning and testing of individual COP gRNAs

Oligonucleotides including protospacer sequences (Table S21) were ordered in the following format:

GGAAAGGACGAAACACCG [19-20-bp protospacer—remove initial G for any 20-bp protospacer with one natively] GTTTAAGAGCTAGGCCAACATG

Fifteen gRNAs were amplified by PCR to create homology arms and individually cloned into a pLenti-U6-2xMS2gRNA-MCPp65-PuroR backbone through NEBuilder HiFi DNA assembly as described above. To make CRISPRa cell lines, the gRNA constructs were packaged into lentivirus and transduced into HepG2 cells expressing dCas9-10xGcn4-mChe and scFv-Sbno1-Nfe2l1-Krt40-BFP, seeded at 4x10^4 cells/cm^2 on 6-well plates with 8 μg/mL of polybrene in at least two replicates per gRNA. Two days post-transduction, cells were treated with 500 ng/mL puromycin and were selected for approximately one week.

For the cell-surface protein staining, CRISPRa cells were washed, trypsinized, and 6.4*10^4 cells were resuspended in FACS suspension media containing one of four APC-antibodies:

APC anti-human CD36L1 (Biolegend)
APC anti-human CD132 (Biolegend)
APC anti-human CD71 (Biolegend)
APC anti-human CD222 (Biolegend)

Staining reactions were incubated on ice for 20 minutes and protected from light. Cells were washed twice and resuspended in FACS suspension media followed by flow cytometric analysis using a Beckman CytoFLEX flow cytometer to determine magnitude of activation of each cell surface protein. CRISPRa potency of each cell line was normalized to control wells that received a different gRNA and the same antibody.

Co-essentiality modules and enrichment

Modules were generated using DepMap 21Q2 Project Achilles data[43]. This data was altered to account for gene proximity effects using the method described by FIREWORKS[47]. The proximity corrected data was then converted to modules using an existing method[48]. Module enrichments for gene subsets were calculated from the hypergeometric distribution implemented in scipy v1.6.3[102]. Gene Ontology[103] (GO) terms were generated from PANTHER version 14 along with the tool's implementations of term enrichment and FDR control[104]. All GSEA analysis was performed with gseapy v.0.10.4[69] with 10,000-100,000 iterations as negative control. Networks were constructed for all modules with a hypergeometric test FDR less than 0.01. t-Distributed Stochastic Neighbor Embedding (t-SNE)[105] algorithm was used to reduce the dimensionality of the Achillies proximity corrected data, and the first and second principal components were used as the X and Y coordinates of the genes, respectively.

Exome rare variant analysis

UKB variant extraction is performed using Hail v0.2.66[106]. Variant allele frequencies were estimated from the Genome Aggregation Database (gnomAD v2.1)[107]. Variants were included with population maximum allele frequencies of <= 0.001 (Ensembl gnomAD plugin)[108]. Functional consequences and computational score predictions come from Variant Effect Predictor v99[109] leveraging the dbNSFP v4.0 plugin[110]. Variant functional consequences terms were called deleterious if the VEP consequence indicated frameshift, nonsense, start lost, splice donor/acceptor disruption, or stop lost.

Individuals carrying multiple coding variants in the same gene were assigned the variant with the more impactful functional consequence (Deleterious, Missense). If multiple variants have the same worst functional consequence, the variant with the lowest population maximum allele frequency is selected. Consistent with our prior analysis we filtered the most severe variant using CADD score[111]. Therefore, if the allele frequencies are the same, the variant with the highest CADD[63] score is chosen. Comparisons between carriers and non-carriers are performed with the Mann-Whitney U test utilizing scipy v1.6.3.[102]

Phenotypic adjustments are performed through ten-fold cross validation on a linear regression. Adjustments are made by fitting a regression to predict LDL-C. From the fit model's parameters we subtract feature weights associated with each covariate multiplied by each individual's covariates from their original LDL-C value. Linear regressions were performed using scikit-learn v0.24.0.[112]

Clinvar analysis was restricted to all Pathogenic and Benign coding missense variants as reported in the May 2022 release.

Cox proportional hazard analyses were performed using lifelines v0.27.4[113]. Regressions were performed with presence of a rare missense variant and presence of an LOF variant as binary features. Most severe variant was determined using the same strategy previously mentioned. Exon coordinates were determined for genes of interest using MANE transcripts, with an additional 5nt retained up- and downstream of each coding region to capture splice acceptor and donor region variants. Gene-level VCF files were extracted from the UK Biobank WES joint-called pVCFs using bcftools. The VCF files were then normalized to flatten multiallelic sites and align variants to the GRCh38 reference genome. Variants located in NIST Genome in a Bottle ''difficult regions'' were removed from analysis, as were variants with a minor allele frequency greater than 0.1% in the UK Biobank cohort. Further filtering removed variants where more than 10% of samples were missing genotype calls[114] and variants that did not appear in the UK Biobank cohort. To mitigate differences in sequencing coverage between individuals who were sampled at different phases of the UK Biobank project, variants were only retained in the final set if at least 90% of their called genotypes had a read depth of at least 10. All filtering took place by running bcftools through the Swiss Army Knife app on the UK Biobank Research Analysis Platform. Case definitions for coronary artery disease were defined in the UKB using a combination of self-reported data confirmed by trained healthcare professionals, hospitalization records, and national procedural, cancer, and death registries, previously described at the disorder level, and Estimated untreated levels obtained using adjustments for lipid-lowering therapies were used in analyses, as described in this prior supplement[115].

Gene overexpression and RAB10 pathway imaging analysis

To construct HepG2 cell lines with inducible gene overexpression, we first made a stable HepG2 rtTA expressing cell line by performing lentiviral infection with pLenti_rtTA-P2A-BFP_HygroR and sorting stably expressing cells for BFP expression. We then replaced the insert in the Tet promoter-containing vector pCW-Cas9-Blast (Addgene 83481) for LDLR-mCherry, RAB10-T2A-RABIF-P2A-mCherry, OTX2-P2A-mCherry, CSK-P2A-mCherry, or P2A-mCherry as a control. We made stable lines using lentiviral transduction into HepG2-rtTA-BFP cells using Blasticidin selection to select for cells that received the insert.

For LDL-C uptake experiments, we used the protocol above for individual gRNA testing, comparing mCherry+ gene-overexpressing cells with mCherry- wild-type control cells.

To examine the role of the RAB10 pathway in LDLR expression and LDL-C uptake, we infected HepG2 cells and HepG2-rtTA-BFP + Tet-LDLR-mCherry cells with sgRAB10, sgDENND4C, and sgCTRL gRNA lentivirus and selected for infected cells with Puromycin. To measure surface LDLR expression, we stained HepG2 KO cells with mouse anti-LDLR Clone C7 monoclonal antibody (BD 565641) and donkey anti-mouse DyLight649 (Jackson ImmunoResearch) for 20 min on ice followed by flow cytometric analysis using BD CytoFLEX. To track LDLR localization over time, we induced LDLR-mCherry expression in HepG2-rtTA-BFP + Tet-LDLR-mCherry KO cells using 2 μg/mL Doxycycline for the specified time followed by fluorescent imaging using a Zeiss Apotome microscope at 40x resolution. Images were processed using Zeiss ZEN software and ImageJ. We also performed flow cytometry on HepG2-rtTA-BFP + Tet-LDLR-mCherry KO cells induced to express LDLR-mCherry with 2 μg/mL Doxycycline for the specified time and chased with standard media for the indicated times.

RAB10 3' UTR evaluation

We amplified the entire RAB10 3' UTR sequence from HepG2 genomic DNA and inserted it immediately after the stop codon of GFP-RAB10 lenti (Addgene 130883). We also cloned an identical plasmid containing the minor allele of rs142787485. We performed lentiviral transduction of HepG2 with both plasmids, selecting to completion with Puromycin and performing flow cytometry using a BD CytoFLEX to measure GFP levels.

Mouse AAV8 shRNA experiments

All animal care and experimental procedures used in this study were approved by the Institutional Animal Care and Use Committee at National University of Singapore. Mice were housed under a 12-hour light-dark cycle with free access to water and normal chow diet. Male C57BL/6 littermates were obtained from *InVivos* at 6 weeks of age. For studies with special diet, 12-week-old mice were placed on western diet (TD.10885, Envigo) for a total period of 2 weeks.

To knock down hepatic *Rabif, Otx2 or Csk in vivo*, AAV8-scramble-shRNA (Vector Biolabs), AAV8-*Rabif*-shRNA (Vector Biolabs), AAV8-*Otx2*-shRNA (Vector Biolabs), or AAV8-*Csk*-shRNA (Vector Biolabs) were administered to eight-week-old C57BL/6 male mice via retro-orbital injection, with a dose of 6×10^{11} gc per mouse. Heparinized blood samples were taken at 2, 4 and 6 weeks post-injection for plasma lipid analysis, western diet-feeding started at 4 weeks post-injection. Mice were sacrificed at 6 weeks post-injection and livers and plasma were collected for further analysis.

Sequences of shRNAs used for *in vivo* gene targeting are as follows.

Rabif: GCAGACTTGTTGGTACTAATA

Otx2: GCTGTTACCAGCCATCTCAAT

Csk: GCCTTGAGAGAGAAGAAATTT

Blood was collected from tail tips after overnight fasting, and plasma was further isolated via centrifugation. Triglyceride and total cholesterol levels were measured using Infinity Triglycerides Reagent (Thermo Fisher) and Infinity Cholesterol Reagent (Thermo Fisher) respectively according to the manufacturer's instructions. VLDL/LDL fraction of plasma was separated from HDL fraction using HDL and LDL/VLDL Quantitation Kit (Sigma) according to the manufacturer's instructions.

Total RNA was extracted from mouse livers using Trizol reagent (Invitrogen). 1 μg of RNA was applied for reverse transcription with using Maxima H Minus cDNA synthesis system (Thermo). Quantitative real-time PCR (RT-PCR) was performed on a ViiA™ 7 RT-PCR system (Applied Biosystems Inc.). Melting curve analysis was carried out to confirm the RT-PCR products. Expression data is presented after calculating the relative expression level compared to the housekeeping gene Rplp0.

The protocol for hepatic lipid extraction was adapted from previously published work[116] with slight modifications. Briefly, to measure the lipid (Triglyceride and cholesterol) levels of liver, snap-frozen liver samples were weighed and homogenized in ten volumes of ice-cold PBS. Two-hundred microliters of the homogenate was transferred into 1,200 ul of chloroform:methanol (2:1; v/v) mixture followed by vigorous vortex for 30 seconds. One-hundred microliters of ice-cold PBS was then added into the mixture and mixed vigorously for 15 seconds. The mixture was then centrifuged at 4,200 rpm for 10 minutes at 4 degree. Two-hundred microliters of the organic phase (bottom layer) was transferred into a new tube and evaporated for dryness. Two-hundred microliters of 1% Triton X-100 in ethanol was used to dissolve the dried lipid with constant rotation for 2 hours. Triglyceride and cholesterol content were determined using the Infinity Triglycerides reagent (Thermo Scientific) and Infinity Cholesterol reagent (Thermo Scientific) respectively, based on manufacturer's instructions.

QUANTIFICATION AND STATISTICAL ANALYSIS

Number of replicates can be found in the Figure legends or in the Methods Details of STAR Methods. All measurements were taken from distinct samples. All figures show mean with standard error bars unless specified otherwise. For CRISPR screens, MAGECK-RRA[95] was used to calculate p-values and FDR. For case-control comparisons, two-tailed t-tests and Mann-Whitney U-tests were performed to compare treatment and control groups as indicated in Figure legends. For gene module analysis, hypergeometric tests were used to calculate significance. For exome analysis, significance is reported based on a false discovery rate of 0.1 unless noted otherwise. Tests are conducted using a two-sided Mann Whitney U statistic comparing adjusted LDL-C values between rare-variant carriers and non-carriers. False discovery rate corrections are done using the Benjamini-Hochberg procedure[117] correcting for the number of independent tests performed. For comparisons involving multiple dependent tests (*e.g.* genes at different groupings of variants) we apply Simes method[118] on the group of hypotheses to represent a single independent test. Testing groups include limiting variants with increasing VEST4 predicted impact, taken with no limit, the top 25%, top 50%, and top 75% of VEST4 values. P-values are represented by asterisks (*$p<0.05$, **$p \leq 0.01$, ***$p \leq 0.001$). Statistical analysis and visualization were carried out in Graph Pad Prism version 9.3.1 and R version 3.6.3. Flow cytometric analysis was performed using FCS Express version 7.12 and FloJo version 10.6.2.