MATERIALS AND METHODS

Generation of Adenoviruses

The *Streptococcus pyogenes* CRISPR-Cas9 system and the guide RNA protospacers were inserted into the Adeno-X vector (Clontech) as previously described.[1] The protospacers to specifically target the rs12740374 minor allele sequence (5'- GTGCTTGATTGAGCAACCTC-3') or major allele sequence (5'- GTGCTTGATTGAGCACCCTC-3') were designed by visual inspection. Irrelevant protospacers were used as controls for the primary human hepatocyte experiments (protospacer 5'-TTTTTTGTTTTTTGTTTTTT-3') and the BAC transgenic mouse experiments (5'- GGTGCTAGCCTTGCGTTCCG-3'). The Penn Vector Core at the University of Pennsylvania used these vectors to generate recombinant adenoviral particles (designated CRISPR-SNP, CRISPR-major, or CRISPR-control). The Penn Vector Core also provided pre-made adenovirus encoding GFP.

Primary Human Hepatocyte Experiments

Single vials containing 5 to 15 million viable primary human hepatocytes each from lots derived from 20 individuals were obtained from Thermo Fisher Scientific, representing a combination of: Human Plateable Hepatocytes, Induction Qualified; Human Plateable Hepatocytes, Metabolism Qualified; and Human Plateable Hepatocytes, Transporter Qualified. All were rated to survive in culture for at least three days after plating. The cells from each vial were thawed into Cryopreserved Hepatocyte Recovery Medium (Gibco), gently spun down, and then resuspended in William's Medium E with added Hepatocyte Thawing & Plating Supplement (fetal bovine serum, dexamethasone, and a cocktail solution of penicillin-streptomycin, bovine insulin, GlutaMAX, and HEPES) (Thermo Fisher Scientific) and plated at 400,000 cells/well in collagen-coated 24-well plates, according to the manufacturer's instructions. After 6 hours, each well was refed with 500 µL maintenance medium [William's Medium E with added Hepatocyte Maintenance Supplement (dexamethasone and a cocktail solution of penicillin-streptomycin, insulin, transferrin, selenium complex, BSA, linoleic acid, GlutaMAX and HEPES)] (Thermo Fisher Scientific). Each well was refed with 500 µL maintenance medium daily for the duration of the experiments.

Residual cells left over in the thawed vials were washed out with PBS and collected, and genomic DNA was isolated using the DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturer's instructions. A 423-bp region surrounding the rs12740374 SNP was PCR amplified using the following primers: F: 5'-AGGAACTGGAAAAGCCCTGT-3' and R: 5'-GAGGCCACAGCAGGTTAGAC-3'. PCR amplicons were subjected to Sanger sequencing to rapidly determine the rs12740374 genotype for each lot while the experiments were ongoing.

One day after plating, one to four pairs of wells for each lot of heterozygous cells (depending on availability of cells) were treated with either CRISPR-SNP adenovirus or CRISPR-control adenovirus. One set of wells was treated with GFP adenovirus to assess the efficiency of infection. Other wells of the heterozygous cells and the homozygous major cells were not treated with virus and were used to generate the data in Figure 1. Each virus-treated well received 1.5×10^7 viral particles mixed into 250 µL maintenance medium, for an estimated MOI of ~50. The virus was aspirated after four hours and replaced with 500 µL maintenance medium. Two days later, some of the virus-treated wells were used for DNA isolation with QuickExtract DNA Extraction Solution (Epicentre) and PCR amplification of the rs12740374 SNP sequence (as described above) for deep sequencing (as described below). The other virus-treated wells as well as non-virus-treated wells were used for quantitative reverse

transcriptase-polymerase chain reaction (qRT-PCR) gene expression studies (as described below). Of note, the non-virus-treated wells were assessed at this timepoint, rather than immediately upon thawing or one day after plating, in order to serve as a baseline for the genome-edited cells.

Induced Pluripotent Stem Cell-derived Hepatocyte-like Cell Experiments

These experiments were performed as previously described.[2] Briefly, a multi-ethnic population cohort of iPSC lines were generated from healthy human subjects as approved by the University of Pennsylvania Human Subjects Research Institutional Review Board. rs12740374 genotypes were available from use of the Infinium Human CoreExome-24 BeadChip (Illumina). One iPSC line from each of 86 participants was differentiated into HLCs using a feeder-free protocol.[2] In brief, we incubated cells in (1) RPMI-B27 (RPMI-1640 from Thermo Fisher Scientific; 2% B-27 Supplement Minus Insulin from Thermo Fisher Scientific) medium supplemented with recombinant human/mouse/rat activin A (100 ng/mL; R&D Systems), recombinant human BMP-4 (10 ng/mL; Peprotech), and recombinant human FGF basic (20 ng/mL; R&D Systems) for 2 days, followed by 3 days of incubation in RPMI-B27 (minus insulin) supplemented with activin A alone, in ambient oxygen/5% CO_2, yielding definitive endoderm; (2) RPMI-B27 (with insulin, i.e., made with 2% B-27 Supplement) supplemented with BMP-4 (20 ng/mL; PeproTech) and FGF basic (10 ng/mL) for 5 days in 5% oxygen/5% CO_2, yielding hepatic progenitor cells; (3) RPMI-B-27 (with insulin) supplemented with recombinant human HGF (20 ng/mL; PeproTech) for 5 days in 5% oxygen/5% CO_2, yielding immature HLCs; and (4) HCM Hepatocyte Culture Medium (Lonza) without EGF and supplemented with recombinant human oncostatin M (20 ng/mL; R&D Systems) for 5 days in ambient oxygen/5% CO_2. The quality of the HLCs was assessed as described.[2] RNA was extracted from the HLC samples and used for RNA-seq analysis[2] and for specific qRT-PCR studies for *SORT1* and *PSRC1* (as described below). In parallel, we used RNA-seq data from 96 whole-liver samples in the GTEx project (dbGaP: phs000424.v6.p1) to determine *SORT1* expression levels.[2]

Generation of BAC Transgenic Mice and Somatic In Vivo Genome Editing Experiments

All procedures used in animal studies were approved by the pertinent Institutional Animal Care and Use Committees at Harvard University and the University of Pennsylvania and were consistent with local, state, and federal regulations as applicable. BAC clone RP11-463O24 (with the insert hg18/chr1:109541561-109744884, containing the minor allele of rs12740374) from the human RPCI-11 Human Male BAC Library (BACPAC Resource Library, Children's Hospital Oakland Research Institute) was grown in DH10B *E. coli* cells and purified with the NucleoBond BAC 100 kit (MACHEREY-NAGEL). The BAC DNA was used for pronuclear injection into C57BL/6J embryos at the Harvard University Genome Modification Facility. Ten founder mice that were positive for the BAC transgene by PCR analysis were obtained. In each of these mice, the integrity of the transgene was tested with PCR of amplicons distributed throughout the BAC insert sequence, and quantitative PCR with these amplicons was used to assess relative copy number among the mice that were positive for all of the amplicons. The founder mouse with the lowest copy number of the complete transgene and its descendants were bred with wild-type C57BL/6J mice to the F2 generation, in which roughly 50% of the offspring were positive for the complete transgene, consistent with Mendelian transmission of a single transgene insertion site (data not shown). Quantitative PCR with TaqMan SNP Genotyping Assay, Human SM, Assay ID C__25753757_20 (Applied Biosystems) confirmed a consistent copy number across all of the positive F2 mice.

Mice that were three to four months of age were used for experiments. Mice were given CRISPR-SNP adenovirus, CRISPR-major adenovirus, or CRISPR-control adenovirus. The mice were administered 1×10^{11} viral particles each via retro-orbital injection; the mice were anesthetized with 1%-2% inhaled isoflurane during injections. As much as possible, the mice in the groups were matched with respect to age and sex but were otherwise assigned to the groups randomly. After four days or two weeks, the mice were euthanized by carbon dioxide asphyxiation, and whole liver samples were harvested and snap-frozen in liquid nitrogen. Liver genomic DNA was subsequently isolated using the DNeasy Blood & Tissue Kit. A 423-bp region surrounding the rs12740374 SNP was PCR amplified (as described above). PCR products were purified, analyzed using the Surveyor Mutation Detection Kit (Integrated DNA Technologies) according to the manufacturer's instructions, and resolved on 2.0% agarose gels.

Quantitative Reverse Transcriptase–Polymerase Chain Reaction Gene Expression Studies

Wells of primary human hepatocytes were lysed directly in TRIzol Reagent (Thermo Fisher Scientific), and small pieces of liver from the virus-treated BAC transgenic mice were homogenized in TRIzol Reagent. Wells of HLCs were processed with the miRNeasy Mini Kit (QIAGEN). This was followed by RNA isolation according to the manufacturer's instructions and reverse-transcription using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) with an equimolar mixture of random hexamers and oligo-dT. Gene expression was measured using the following TaqMan Gene Expression Assays along with TaqMan Gene Expression Master Mix (Thermo Fisher Scientific): *SORT1*, Hs00361760_m1; *PSRC1*, Hs00934024_g1; *SARS*, Hs00197856_m1; *PSMA5*, Hs00936004_m1. Human *ACTB* (4310881E) or *B2M* (4326319E) was used to measure the reference gene for human cells. Mouse *Actb* (4326319E) was used to measure the reference gene for the mouse liver samples. Reactions were plated in technical duplicate and carried out on a ViiA 7 Real-Time PCR system (Applied Biosystems). Relative expression levels were quantitated by the $\Delta\Delta C_t$ method.

Western Blot Analysis

For Western blot analysis, liver samples were lysed in a homogenization buffer (25 mM Tris-HCl, 10 mM $Na_4P_2O_7$, 1% NP40, 10 mM NaF, 1 mM EGTA, and 1 mM EDTA) containing protease inhibitor cocktail (Sigma-Aldrich, 11697498001). Antibodies against sortilin (Abcam, ab16640) and β-actin (Cell Signaling Technology, 4967) were used.

Deep Sequencing

PCR amplicons from virus-treated primary human hepatocytes or BAC transgenic mice were subjected to next-generation DNA sequencing at the Massachusetts General Hospital CCIB DNA Core (CRISPR Sequencing service; https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/crispr_sequencing_main.jsp). Mutagenesis rates were determined as previously described.[3] In brief, sequencing data were processed according to standard Illumina sequencing analysis procedures. Processed reads were mapped to the expected PCR amplicon (with the rs12740374 minor allele) as the reference sequence using a custom script; reads that did not map to reference were discarded. Indel frequencies were determined as follows. The reads were analyzed using a custom script to identify indels by matching reads against reference, with indels involving any portion of the sequence within 15 nt upstream or downstream of the predicted CRISPR-Cas9 cleavage site (3 nt upstream of the 3' end of the protospacer) considered to be possible CRISPR-Cas9-induced mutations. Reads for which there was any 18-nt sequence with more than 2 mismatches with the corresponding 18-nt

portion of the reference sequence, either upstream or downstream of a candidate indel, were discarded as errors. For reads from CRISPR-SNP-treated primary human hepatocytes, a custom script was used to discriminate whenever possible between indels on minor allele-bearing chromosomes and indels on major allele-bearing chromosomes.

Statistical Analyses

For untreated primary human hepatocytes, triplicate wells were analyzed for gene expression from each of the 20 lots. The average gene expression levels for each lot were compared between the rs12740374 homozygous major lots (N = 14) and the heterozygous lots (N = 6) using the Mann-Whitney U test. For virus-treated primary human hepatocytes, there were a total of 19 pairs of wells (one well treated with CRISPR-control, one well treated with CRISPR-SNP) representing the six heterozygous lots (there were differing numbers of pairs of wells due to varying availability of cells from the original vials to plate at 400,000 cells/well). Although Figures 2C and 2D show aggregated gene expression data for each of the six lots, the 19 pairs of wells were treated individually for statistical analyses with the Wilcoxon signed-rank test (N = 19 paired sets) as shown in Figure 2B. For the BAC transgenic mice, the average gene expression levels were compared between the CRISPR-control-treated mice and either CRISPR-SNP-treated mice or CRISPR-major-treated mice using the Mann-Whitney U test. All statistical analyses were performed using GraphPad Prism 6 for Mac OS X.

Supplemental References

1. Ding Q, Strong A, Patel KM, Ng SL, Gosis BS, Regan SN, Cowan CA, Rader DJ, Musunuru K. Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circ Res*. 2014;115:488–492.

2. Pashos EE, Park Y, Wang X, et al. Large, diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. *Cell Stem Cell*. 2017;20:558–570.

3. Yang Y, Wang L, Bell P, McMenamin D, He Z, White J, Yu H, Xu C, Morizono H, Musunuru K, Batshaw ML, Wilson JM. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat Biotechnol*. 2016;34:334–338.