SUPPLEMENTAL MATERIAL to

Posttranscriptional regulation of the LDL Receptor in humans by the U2-spliceosome and its interactors

Paolo Zanoni[1,2,*], Grigorios Panteloglou[1,2,*], Alaa Othman[3], Joel T. Haas[4], Roger Meier[5],

Antoine Rimbert[6], Marta Futema[7], Yara Abou Khalil[8,9], Simon F. Norrelykke[5],

Andrzej J. Rzepiela[5], Szymon Stoma[5], Michael Stebler[5], Freerk van Dijk[10], Melinde Wijers[6],

Justina C. Wolters[6], Nawar Dalila[11], Nicolette C. A. Huijkman[6], Marieke Smit[6],

Antonio Gallo[12], Valérie Carreau[12], Anne Philippi[13], Jean-Pierre Rabès[8,14,15],

Catherine Boileau[8,16], Michele Visentin[17], Luisa Vonghia[18,19], Jonas Weyler[18,19],

Sven Francque[18,19], An Verrijken[19,20], Ann Verhaegen[19,20], Luc Van Gaal[19,20],

Adriaan van der Graaf[10], Belle V. van Rosmalen[21], Jerome Robert[1], Srividya Velagapudi[1,2],

Mustafa Yalcinkaya[1,2], Michaela Keel[1,2], Silvija Radosavljevic[1,2], Andreas Geier[22],

Anne Tybjaerg-Hansen[11], Mathilde Varret[8], Lucia Rohrer[1,2], Steve E. Humphries[23],

Bart Staels[4], Bart van de Sluis[6], Jan Albert Kuivenhoven[6], and Arnold von Eckardstein[#,1,2]

* These authors contributed equally to this work

1. Supplemental Material and Methods

Material

siRNAs, antibodies, cells, as well as sources of data are described in the Major Resource Table. Sources of all other materials are described in the running text

Cell culture

Huh7, HepG2 and HEK293T cells were cultured in DMEM (cat. D6546, Sigma Aldrich, Buchs, Switzerland) complemented with 10% fetal bovine serum (cat. 10500056, Thermo Fischer Scientific, Reinach, Switzerland) and penicillin/streptomycin (100U/ml each, cat. 15140122, Thermo Fischer Scientific).

Real-time PCR

Total RNA was extracted using Tri-reagent (cat. T9424, Sigma Aldrich) and treated with DNAse I (cat. 04 716 728 001, Roche, Switzerland) according to manufacturer`s instructions. cDNAs were generated using the RevertAid First Strand Synthesis kit (cat. K1621, Thermo Fischer Scientific). RT-PCR reactions were performed on a Roche Light Cycler 480-II (cat. 05015243001, Roche, Switzerland) using the LightCycler ® 480 SYBR Green I Master (cat. 04887352001, Roche, Switzerland) according to the manufacturer's instructions. Data were analyzed by performing relative quantification based on Cp values for the reference gene and the gene of interest. A standard curve for each primer couple was generated for each experiment.

Isolation and labeling of plasma lipoproteins

LDL ($1.019<d<1.063$ g/ml) and HDL ($1.063<d<1.021$ g/ml) were isolated from frozen human normolipidemic plasma of blood donors by sequential ultracentrifugation as described

previously. For the genome-wide siRNA screening, LDL and HDL were labeled with Atto594 and Atto655, respectively (AD594-35 and AD655-35, Atto-Tec, Siegen, Germany) - as published before[30]. For validation experiments, we labeled both LDL and HDL with ^{125}I according to the McFarlane method[30,31].

siRNA genome-wide screening

The genome wide siRNA screening of genes limiting the uptake of fluorescently labeled LDL was performed in the Scientific Center for Optical and Electron Microscopy (ScopeM) of ETH Zurich: https://scopem.ethz.ch/. The Ambion Silencer Select Human Genome siRNA library V4 containing three unique, non-overlapping siRNAs for each of the 21,584 human genes, was diluted in sterile nuclease-free water and re-plated in 192 BD-Falcon clear bottom 384 wells assay plates (cat. 08-772-151, Thermo Fischer Scientific). 5µl of an 80nM solution of each siRNA was plated in each well of the assay plates maintaining the same layout as in the master library. As internal controls, each of the following Ambion Silencer Select siRNA oligonucleotides (Thermo Fischer Scientific) were plated in four replicate wells in each of the 192 assay plates: anti-*PLK1*, anti-*LDLR*, and the Silencer™ Select Negative Control No. 1 siRNA. The assay plates were sealed and maintained at -20 °C until use. One copy of the entire library was used to perform our screening. Libraries were plated by using a Tecan Freedom Evo automated liquid handling robot (Tecan, Männedorf, Switzerland). Liquid handling for the screening was performed by an EL406 liquid handling robot (Bio Tek Instruments, Luzern, Switzerland).

On Day 1, the assay plates were thawed and spun at 1000g for 3 minutes at room temperature. RNAiMax (cat. 13778150, Thermo Fischer Scientific) was diluted in DMEM (0.075ul of RNAiMax for each 10ul of DMEM) and 10ul of diluted RNAiMax were added to each well. The plates were incubated under these conditions for 1 hour at room temperature.

3

In the meantime, Huh-7 cells were detached by trypsin treatment for 15 minutes at 37°C. After trypsinization, the cells were re-suspended in DMEM 12.3% FBS and counted using a Neubauer chamber. The cells were then diluted to a final concentration of 26154 cells/ml (corresponding to a final concentration of 1700 cells/well) in DMEM 12.3% FBS in a 5l Double Sidearm CellStar® Spinner Flasks (Wheaton, Millville, NJ, USA) in a water bath set at 37°C. After 5-10 minutes of gentle stirring, 65ul of cell suspension was added to each well and the plates were immediately placed on pre-heated metal blocks in a cell culture incubator set at 37°C, 5% CO_2. The cells were maintained under these conditions for 72h. Final siRNA concentration was 5nM. We will refer to the abovementioned transfection protocol as "reverse transfection" in this manuscript, to indicate that the cell suspension was added at the end of the transfection process.

On Day 4, 60μl of medium was aspirated from each well, followed by the addition of 20μl of a solution with 66μg each of Atto594-LDL and Atto655-HDL protein/ml in DMEM, for a final concentration of 33μg/ml. After incubation for 4h at 37°C, 5% CO_2, the cells of each well were washed six times with PBS, followed by the addition of 50μl of an isotonic 2% paraformaldehyde solution containing 20ug/ml of Hoechst 33258 (cat. 861405, Sigma Aldrich). After another 15 minutes at room temperature, the cells were washed again six times with PBS to eliminate the paraformaldehyde. After the last wash, PBS was aspirated and substituted with a 0.05% solution of sodium azide in water. The plates were then sealed with adhesive aluminum foil and kept at 4°C protected from light until imaging.

Imaging was performed by two twin ImageXpress micro HCS microscopes (Molecular Devices, Sunnyvale, CA, USA) fed by a robotic arm and equipped with Photometrics CoolSNAP HQ[2] CCD cameras (Photometrics, Tucson, AZ, USA) and a Lumencor Spectra X solid-state light engine (Lumencor, Beaverton, OR, USA). For each assay plate, two datasets were collected. First, for cell counting, the whole surface of each well was captured at 4x

4

magnification in the DAPI channel. Second, 9 tiled sites in each well were then acquired with a 20x objective in the DAPI, RFP and CY5 channel to allow for the final image analysis. After acquisition, image segmentation and the subsequent image analysis were performed using Cell Profiler (http://cellprofiler.org/, The Broad Institute, Cambridge, MA, USA)[32]. The data were recorded at first for more than 100 different assay features. Based on their close relationship with the phenotype of interest, the number of assay features taken into consideration for further analysis was restricted subsequently to the following five: foci count per cell, foci mean intensity, cytoplasm granularity 1 and 2, cytoplasm median intensity. Furthermore, the "nuclei count" feature was used to measure toxicity as well as to determine transfection efficiency by measuring the extent of cell death induced by knockdown of the essential gene *PLK1*.

Downstream data analysis and hit gene identification was performed using the R statistical software (R-project.org) and applying the procedures described in the statistics section. Hits coming from the 'median cytoplasm intensity' assay feature were selected for validation as this feature displayed the highest Z'-factor values when compared to other assay features. A previous RNA Sequencing experiment conducted in similar conditions (see below) was used to exclude hit genes that are not expressed in Huh-7 cells or for whom no expression data could be generated, for example pseudogenes or uncharacterized loci. This led to the exclusion of 4 ribosomal pseudogenes (namely RPL13AP20, RPL34P34, RPL18AP16 and RPL21P20) from the hits that led to a decrease in LDL uptake upon knockdown. We also excluded 6 uncharacterized loci, 1 pseudogene and 10 non expressed genes from the list of genes that increased uptake upon knockdown.

Validation of the screening hits

The validation experiments in Huh-7 and HepG2 cells were performed in 24 wells plates.

Four pooled siRNA oligonucleotides (Major Resource Table), against each of the top hits

were reverse transfected (see screening methods above) using Lipofectamine RNAiMax

(Thermo Fischer Scientific) according to the manufacturer's instructions. This pooling

strategy has been shown to target multiple transcripts as well as multiple positions within the

same transcript at the same time, thus maximizing knockdown efficiency[33]. The final siRNA

concentration after the addition of the cell suspension was 10nM. As the positive control,

LDLR was knocked down with siRNA oligonucleotide from Ambion Silencer Select. For the

validation of RBM25, siRNAs from Dharmacon and Sigma were used, because of potential

off-target effects of the Dharmacon anti-RBM25 siRNA (see Results section). Knockdown

efficiency was determined by RT-PCR using GAPDH expression levels as the reference

(primers sequences in Supplemental Table IX). Cell association of 125I-LDL or 125I-HDL

(33.3μg/ml, 2h incubation at 37°C) was recorded 72h after transfection as described

previously[30]. In brief, the cells were incubated with the respective iodinated lipoprotein

without (total) or with 100 times excess of the respective non-labeled lipoprotein. Specific

cellular association was calculated by subtracting the values obtained in the presence of

excess unlabeled lipoprotein (unspecific) from those obtained in the absence of unlabeled

lipoprotein (total).

RNA Sequencing in Huh-7 cells

In order to identify genes differentially spliced upon knock-down of U2 spliceosome

components, Huh-7 cells were first transfected with Dharmacon siRNA pools (see Major

Resource Table) in 3 separate replicate experiments according to the same protocol used for

cell association studies. 72h after transfection, the cells were washed once with PBS. Total

6

RNA was extracted with the Genelute Mammalian Total RNA Miniprep Kit (cat. RTN350, Sigma-Aldrich, Switzerland). Genomic DNA was eliminated by on-column DNase digestion (cat. D2816, Sigma-Aldrich, Switzerland). RNA was then sequenced at the Functional Genomics Center Zurich (https://fgcz.ch/) after poly(A) selection with the TruSeq RNA Library Prep Kit (Illumina, San Diego, CA, USA) on an Illumina HiSeq4000 automated sequencer (Illumina, San Diego, CA, USA) to generate single-end reads of length 126nt. The average read depth was $3.13*10^7$ reads/sample. The raw reads were first cleaned by removing adapter sequences, trimming low quality ends, and filtering reads with low quality (phred quality <20) using Trimmomatic[34]. Sequence alignment of the resulting high-quality reads to the Homo Sapiens reference genome (build GRCh38) was carried out using STAR (Version 2.5.1b)[35]. Basal gene expression was determined by performing the same experiment on non-transfected Huh-7 cells. Gene expression values were computed with the function featureCounts from the Bioconductor package Rsubread[36]. To detect candidates for differential exon usage we used DEXSeq[37].

Measurement of intron 3 retention by RT-PCR

Alternative splicing of the segment between exons 3 and 4 of *LDLR* upon knockdown of spliceosome gene hits in Huh-7 cells as well as in human liver samples (see below) was studied by RT-PCR using previously published[9] primer couples annealing in exon 3, intron 3 and exon 4 (see Supplemental Table IX for primer sequences).

Molecular cloning and LDLR-ret fragment overexpression

The coding sequence of the LDLR fragment formed after intron 3 retention (LDLRret) was obtained by PCR amplification of cDNA after *SF3B1* knockdown using the primers listed in Supplemental Table IX and subcloned in pCR-Blunt (Thermo Fischer Scientific, The

Netherlands) according to the manufacturer's instructions. A synthetic oligonucleotide encoding a version of LDLRret carboxyterminally tagged with hemagglutinin (HA) was obtained through Invitrogen GeneArtTM StringsTM DNA Fragments (Thermo Fischer Scientific) and subcloned in pCR-Blunt as well. Both the untagged and the HA-tagged versions of the LDLRret fragment were subsequently subcloned into the pcDNA3.1 Mammalian Expression Vector (cat. V79020, Thermofisher Scientific) under the control of a CMV promoter. The coding sequences of either construct were confirmed by Sanger sequencing. In the overexpression experiments, HEK293T cells were seeded first in 6 wells plates, $1*10^6$ cells/well in complete medium. 16h after seeding, 2.5 µg per well of pcDNA3.1 encoding for the LDLRret fragment were transfected using Lipofectamine 3000 (cat. L3000015, Invitrogen) according to the manufacturer's instructions (7µl of Lipofectamine 3000 and 5ul of P3000 reagent per well). The same amount of an empty pcDNA3.1 vector was used as the control. For proteasomal inhibition studies, the cells were exposed to the MG-132 (Merck, cat. SML1135) proteasomal inhibitor at a concentration of 5µM in DMSO as described previously[38]. Protein lysates and media were collected 48h after transfection. Media were spun at 1500g for 5' and the supernatant was used to detect the fragment.

Molecular cloning of RBM25 and its mutants

The coding sequences of the wild type and the mutant RBM25 were designed based on NCBI's reference sequence NM_021239.3 and were synthesized and cloned by Proteogenix (Schiltigheim, France) into pLVXpuro vector (cat. 632164, Takara Bio/ Clontech). The coding sequences of either construct were confirmed by Sanger sequencing. The coding sequences were thereafter subcloned into pcDNA3.1 vector, by digesting the original pLVX-puro constructs with BstBI (cat. R0519S, NEB), Klenow (M0210S, NEB) and XbaI (cat. FD0685, Thermofisher scientific) and the pcDNA3.1 empty vector with EcoRV (FD0303,

8

Thermofisher scientific) and XbaI, respectively, according to the manufacturer's instructions. After gel purification with Zymoclean Gel DNA Recovery Kit (D4001, Zymo Research), the digested material was ligated using T4 DNA ligase (M0202S, NEB). The correct sequences of the generated plasmids were confirmed by both restriction enzyme digestion and Sanger Sequencing.

For the overexpression experiments, $4*10^5$ Huh7 cells were seeded onto 6-well plates and, after 24 hours culturing, transfected with 4 μg from each plasmid using Lipofectamine 2000 (Cat. 11668019, Invitrogen) according to the manufacturer's instructions (6 μl lipofectamine 2,000/well). The same amount of an empty pcDNA3.1 vector was used as the control. 24 hours after transfection the cells were treated with complete medium containing 750 μg/ml G418 (Cat. 10131-027, Gibco) for an additional 48h before they were harvested

Western blotting

Protein levels of LDLR and RBM25 in the whole cell lysate 72 hours after transfection were determined by Western blotting. Firstly, proteins were separated on an 8% SDS-PAGE gel and then transferred to an Amersham Hybond P 0.45 PVDF membrane (GE Healthcare Europe, Glattburg, Switzerland). All antibodies (see Major Resource Table) were diluted in a 0.01% PBST solution containing 5% skim milk. Primary antibodies were incubated overnight at 4° C. Secondary antibodies were incubated for 2 hours at room temperature. Detection was carried out using Thermofisher Scientific SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Fischer Scientific) on a Fusion FX (Vilber Lourmat, Marne-la-Vallée, France) imaging system. Densitometry was performed using either the Image Studio Lite software (LI-COR Biotechnology, Lincoln, NE, USA) (Figure 4C,4D) or ImageJ (https://imagej.nih.gov/ij/; Supplementary Figures IIIc, IIIe).

For Western blotting experiments involving cells overexpressing the LDLRret fragment, proteins were first separated on a 15% SDS-PAGE gel and then transferred to an Amersham Hybond P 0.45 PVDF membrane. The HA-tagged form of the fragment was detected with an HRP-conjugated anti-HA antibody (monoclonal Ha-HRP, clone HA-7, H6533, Sigma Aldrich). Beta actin (SC-1616, Santa Cruz) was used as loading control.

Flow cytometry based analysis of LDLR cell suface expression

LDLR cell surface levels were determined by flow cytometry on live Huh-7 cells. 72 hours after transfection, the medium was aspirated and the cells were washed twice with PBS and detached using an Accutase solution (Sigma-Aldrich, A6964) for 5 min at 37°C. The cells were then collected in complete medium and counted using a Beckman Coulter Z2 cell counter (Beckman Coulter, Nyon, Switzerland). After counting, the cells were washed in ice cold PBS and then incubated in Blocking buffer (PBS containing 0.5% BSA and 2% FBS) for 30 min on ice. After blocking, the cells were incubated with a primary anti-LDLR antibody (Santa Cruz Biotechnology, see Major Resource Table) diluted to a final concentration of 2μg/ml in FACS buffer (PBS containing 0.5% BSA and 0.05% NaN_3) for 1 hour on ice. After washing twice with FACS buffer, the cells were incubated with either a chicken anti-Mouse IgG (H+L) Cross-Adsorbed AlexaFluor 488-conjugated Secondary Antibody (cat. A-21200, Thermo Fischer Scientific) diluted to a final concentration of 4ug/ml in FACS buffer for 1h on ice in the dark (figure 4D) or anti-Mouse IgG (H+L) Cross-Adsorbed AlexaFluor 647-conjugated Secondary Antibody (Thermo Fischer Scientific cat. A-21236) (Supplemental Figures IIIg, Xe and XId). Finally, the cells were washed twice with FACS buffer and re-suspended in ice cold FACS buffer for the subsequent acquisition. In Figure 4D, cells that were not incubated with any antibody, and cells that were incubated with the secondary antibody only were used as the negative controls to determine the signal-to-noise ratio.

10

Sample acquisition was carried out on a BD LSR II Fortessa (BD-Biosciences, Allschwil, Switzerland) and using BD FACSDIVA™ software. Data analysis was carried out using FlowJo version 10 (FlowJO LLC, Ashland, OR, USA). In all experiments, approximately 10^4 events per condition recorded at the final gate were used for analysis. For data in Figure 4D, dead cells were excluded by gating. For data show in Supplemental Figures IIIg, Xe and XId, life and dead cells were discriminated by staining with Propidium Iodide (PI) (cat. 81845, Fluka) and FITC-Annexin V (cat. 640945, Biolegend, London, UK) at final concentrations of 1 ug/ml and 2 ug/ml, respectively. Doublets were excluded through gating.

Measurement of Atto655-LDL via flow cytometry

Huh-7 cells were incubated with Atto655-LDL to measure the LDL uptake. Briefly, 72 hours post-transfection, the cells were washed with PBS and they were incubated with 10 µg/ml Atto655-LDL contained in Assay Medium (DMEM supplemented with 0.2% BSA), in the absence or presence of 100x (1 mg/ml) non-labeled LDL, for 2 hours at 37°C. At the end of the incubation, the cells were washed with PBS and detached using Accutase® solution (cat A6964, Sigma Aldrich, Switzerland). Prior acquisition, and in order to exclude signal originating from dead cells, the cells were treated with PI and FITC-Annexin V, as described above. Sample acquisition was carried out on a BD LSR II Fortessa (BD-Biosciences, Allschwil, Switzerland) and using BD FACSDIVA™ software. Data analysis was carried out using FlowJo version 10 (FlowJO LLC, Ashland, OR, USA). Specific cellular uptake of LDL was calculated by subtracting the Median Fluorescence Intensity (MFI) value obtained for each sample in the presence of excess unlabeled LDL (unspecific) from the respective MFI value obtained in the absence of unlabeled LDL (total) of the respective samples.

Analysis of LDLR expression by targeted proteomics

To confirm LDLRret fragment overexpression by targeted proteomics, cell lysates from transfected HEK293T were first separated on a SDS-PAGE (4-12% Bis-Tris gel, Expedeon) and the bands corresponding to the full size LDLR protein and to the LDLRret fragment were excised and subjected to in-gel trypsin digestion. Two peptides were used for quantification of the LDLR protein: peptide 1 (CIPQFWR) being located in the overlapping part of the protein and peptide 2 (NVVALDTEVASNR) representing a portion of LDLR that is specific for the full-length protein. Isotopically labeled standard peptides (CIPQFWR containing $^{13}C^{15}N$-labeled arginine, PEPotec grade 2 Thermo Scientific; NVVALDTEVASNR containing $^{13}C^{15}N$-labeled arginine derived from a synthetic protein concatemer, PolyQuant GmbH, Regensburg, Germany) were added to the trypsinized samples for relative comparison of the different samples. The peptides were targeted and analyzed by a triple quadrupole mass spectrometer (MS) equipped with a nano-electrospray ion source (TSQ Vantage, Thermo Scientific) coupled to liquid chromatography (Ultimate 3000 nano-UHPLC system, Dionex) as described previously[38]. The MS traces were manually curated using the Skyline software[39] and the summarized peak areas for the endogenous peptide and isotopically labeled standard peptides were plotted. The endogenous amount was estimated by reference to the known amount of the isotopically labelled standard peptides that was injected in the LC-MS measurement.

Minigene experiments

Both the short (chR19:11,102,285-11,102,921) and the long (chR19:11,102,283-11,105,702) LDLR genomic sequences of interest were amplified by PCR from HEK293T genomic DNA using the primer couples reported in Supplemental Method Table S1, cloned first in pCR-Blunt and then subcloned in the pSPL3 exon trapping vector[40] for expression. Reaction

12

conditions for PCR cloning of the short construct: Step 1: 95ºC, 2 minutes; Step 2: 95ºC, 30 seconds; Step 3: 66ºC, 30 seconds; Step 4: 72ºC, 70.0 seconds; Step 5: Repeat steps 2-4 29 more times; Step 6: 72ºC, 5 minutes; Step 7: 4ºC, forever. Reaction conditions for PCR cloning of the long construct: Step 1: 95ºC, 2 minutes; Step 2: 95ºC, 30 seconds; Step 3: 66ºC, 30 seconds; Step 4: 72ºC, 6minutes; Step 5: Repeat steps 2-4 29 more times; Step 6: 72ºC, 5 minutes; Step 7: 4ºC, forever.

To determine the splicing patterns of exons 3 and 4 in *LDLR*, the two *LDLR* minigenes as well as an empty pSPL3 vector were transfected in HEK293T cells according to the same protocol described above for the LDLRret fragment overexpression studies. 48 hours after transfection, the cells were lysed in Trizol Reagent (Thermo Fischer Scientific) and the lysates were used for total RNA extraction. RNA sequencing on poly-A selected transcripts was performed as described above and sequencing reads were aligned to the reference sequence of each minigene construct. The coverage data shown in this manuscript were obtained from .bam files using the backbone of the long *LDLR* minigene as reference for alignment and were normalized to the average coverage in Exon 3 of *LDLR*.

Expression of spliceosome genes and *LDLR* transcripts in human liver

For the analysis of LDLR isoform-level expression data in healthy human liver samples obtained from bulk tissue RNA-Seq data, 14 healthy untreated liver samples were chosen for analysis (Gene Expression Omnibus, accession number GSE126848) [12].Upon a detailed examination of the quality of each sample, one out of fourteen samples (SRR8601557) was excluded from the generated plots due to lower quality of the reads. Additionally, sample SRR8601570 was observed to have 2-3x lower number of mapping reads compared to the rest of the samples but was retained for analysis due to the small size of the group.

13

Human liver biopsies were obtained at the Obesity Clinic of the Antwerp University Hospital from 155 non-diabetic, obese (BMI > 35) patients as described previously[13]. Briefly, patients suspected for NAFLD were biopsied at baseline and referred for intensive dietary intervention or bariatric surgery depending on presence of comorbidities. For gene expression analysis using the Affymetrix Human Gene 2.0 ST array, data was normalized by Robust Multichip Averaging (RMA) at the transcript level using the oligo package in R (R-project.org). To assess *LDLR* exon usage, RMA was applied at the probe level and each probe's intensity was normalized by the total *LDLR* expression for that patient as calculated by the transcript level summarization. Normalized probe intensities are presented as the $\log_2$ ratio of each probe to the total LDLR transcript expression in each patient.

Healthy human liver tissue was obtained from 17 individuals undergoing partial hepatectomy because of either focal nodular hyperplasia or hepatocellular adenomas at the Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. Spare normal tissue, defined as based on portal tract and central vein architecture, less than 5% steatosis, no inflammation, cholestasis and/or fibrosis, was used for RNA extraction and RT-PCR analysis. Nine additional liver biopsies were obtained in the Division of Hepatology, University Hospital Würzburg, Germany from patients suspected to have NAFLD and included into the RT-PCR analysis as well.

LDLR transcript analysis in peripheral blood cells

Whole blood samples of 4,000 unrelated individuals were obtained from the Dutch BIOS consortium[14]. Paired-end RNA-sequencing data were generated. A reference sequence using the human genome build hg19 and Ensembl annotation version 75 was generated using the Kallisto[41] (v 0.42.2.1) index. Subsequently the paired-end fastq files were used as input for Kallisto quant, which was executed with the aforementioned reference sequence and default

settings to quantify transcript abundances per sample. For the *LDLR* transcript analysis, the available phenotype data were first filtered for missing values. After excluding samples of individuals without data on LDL cholesterol, 2,462 samples were left. Samples were then subdivided by age and LDL cholesterol quartiles and statistically analyzed as described in the statistics section below.

Analysis of the UK Biobank data

The raw GWAS data for variants and phenotypes were downloaded from http://www.nealelab.is/uk-biobank (Version 2). The data included associations of 13,791,467 detected and imputed SNPs from 361,194 samples (194,174 females and 167,020 males) from the UK biobank. Regression models were adjusted for age, age^2, inferred_sex, age * inferred_sex, age^2 * inferred_sex, and principal components (PCs) 1 to 20. Variants were assigned to each gene by using the UCSC Genome Browser Variant Annotation Integrator (https://genome.ucsc.edu/cgi-bin/hgVai), build GRCh37/h19. Only variants between the transcription start and end were considered as shown in Supplemental Table X. P-values for association were corrected using the Bonferroni correction method based on the total number of variants (1360) in the selected genes. Association between exome variants in the elected genes were downloaded from the Amazon cloud server s3://helix-researchpublic/ukbb_exome_analysis_results/V1.3/, as provided by Helix Research and UK Biobank[15,42]. The latest version of the analysis (1.3) was used. Data included 984,819 variants in 15,474 genes from 40,468 exomes. Both loss of function and coding-based models were considered for European and all ethnicities. Data analysis and visualization was done using R (R-project.org).

15

Analysis of *RBM25* gene variants in FH patients

Whole exome sequencing (WES) data of 71 FH probands negative for mutations in the known major FH-causing genes (*LDLR*, *APOB*, and *PCSK9*) were generated as part of the UK10K project[17]. A previous analysis suggested an enrichment of rare (minor allele frequency [MAF] <0.005) *RBM25* gene variants in the FH cohort in comparison to 1926 controls from the UK10K study[17]. The gene burden test was re-analyzed using a much larger control cohort of 56,352 European data provided by the gnomAD study[16]. Variants identified by WES in both FH cases and gnomAD, were filtered to select those with MAF<0.0001 that were annotated as missense or loss-of-function (LOF). The MAF cut-off of 0.0001 was based on the analysis of another genetically heterogeneous dominant disease, hypertrophic cardiomyopathy, as previously demonstrated by Whiffin et al[43].

Statistics

The RNAi screening assay feature data were analyzed as follows: data were first normalized by the median value of each batch, microscope, plate and well and were finally expressed as robust Z-score[44] normalized values. The Redundant siRNA Activity (RSA) analysis was performed for each assay feature on the normalized data to rank the genes and detect the top hits[45], defined here as the genes with an RSA p-value of less than 0.001. This p-value cutoff was dictated by our ability to verify the results *in vitro*.

Transfection efficiency and the dynamic range of the screening assay were determined by calculating the Z'-factor between positive and negative transfection and assay controls as published before[44]. Dimensionality reduction across the five main assay features mentioned above was performed using the Locally-Linear-Embedding[46] method on $\log_2$-transformed data through the *sklearn* Python implementation of the method (http://scikit-learn.org/).

16

Top hits were clustered by function using the online String tool (https://string-db.org/) according to its default parameters. Gene Ontology (GO) analysis for statistical overrepresentation of GO-Slim biological process terms was performed using the Panther Gene List Analysis tool (http://www.pantherdb.org). As by default settings, enrichment was calculated by Fisher's exact test and corroborated by the calculation of the false discovery rate (FDR) according to the Benjamini-Hochberg procedure. GO terms with FDR <0.05 were reported as significant in this manuscript.

For the experiments in Figures 1C, 4C, and 4D as well as in supplemental Figure II the data were normalized to the non-targeting siRNA condition, while for the experiments in supplemental Figure XI the data were normalized to the overexpression of wild type RBM25 condition. In all cases, significance was calculated by one-way ANOVA followed by Dunnet`s post-hoc test between the respective conditions and the respective control, as indicated in the figure legend. For the experiments in Figures 1D and in supplemental Figure III and V, the data from each vendor's targeting siRNA were normalized to the respective vendor's non targeting siRNA. Respectively, for the experiments in supplemental Figure X the data were normalized to the pcDNA3.1 empty vector control. In all cases, the significance was calculated by t-test between the two reported conditions in each experiment. The number of replicates for each experiment is reported in the corresponding figure legend. The regression analysis in Figure 2C was performed using the linear regression function of GraphPad Prism, ver. 5, without constraints and according with its default parameters. The difference between expression data in Figure 5B was tested by Mann-Whitney test. For Figures 5C and 5D as well as supplemental figures VIa, VIb, and VIc, correlation analyses were performed according to Pearson.

To measure the correlation between *LDLR* transcript abundances in peripheral blood samples from the BIOS population, age and LDL cholesterol a Spearman correlation test was

17

performed between the Transcripts Per Million (TPM) normalized Kallisto values (see above) and the phenotypes of interest using the R cor.test function.

To examine the enrichment of *RBM25* gene variants in FH in comparison to gnomAD, a binomial test was applied.

Study approval

Human studies were approved by institutional boards of each institution involved: the ethics committee of the University Medical Centre Groningen, the Ethical Committee of the Antwerp University Hospital (file 6/25/125) and the Ethics Committee of the University Hospital Würzburg (AZ188/17 and AZ96/12). For the AMC liver samples, the local medical ethics committee "Medisch Ethische Toetsings Commissie van het Amsterdam UMC, locatie AMC" approved the protocol for this study before the General Data Protection Regulation (GDPR) law came into force, waiving the need of inform consent as the study employed spare biological material from hepatectomies with therapeutic purpose.

For the UK10K FH sequencing data, all consents and local review board approvals were in accordance with the UK10K project ethical framework. All participants provided written informed consent.
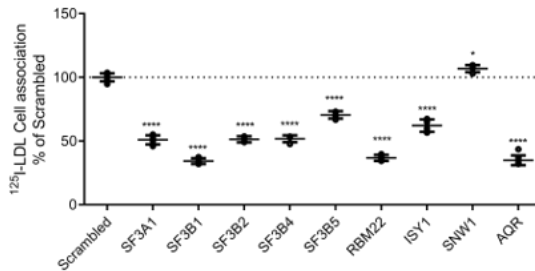
2. Supplemental Figures
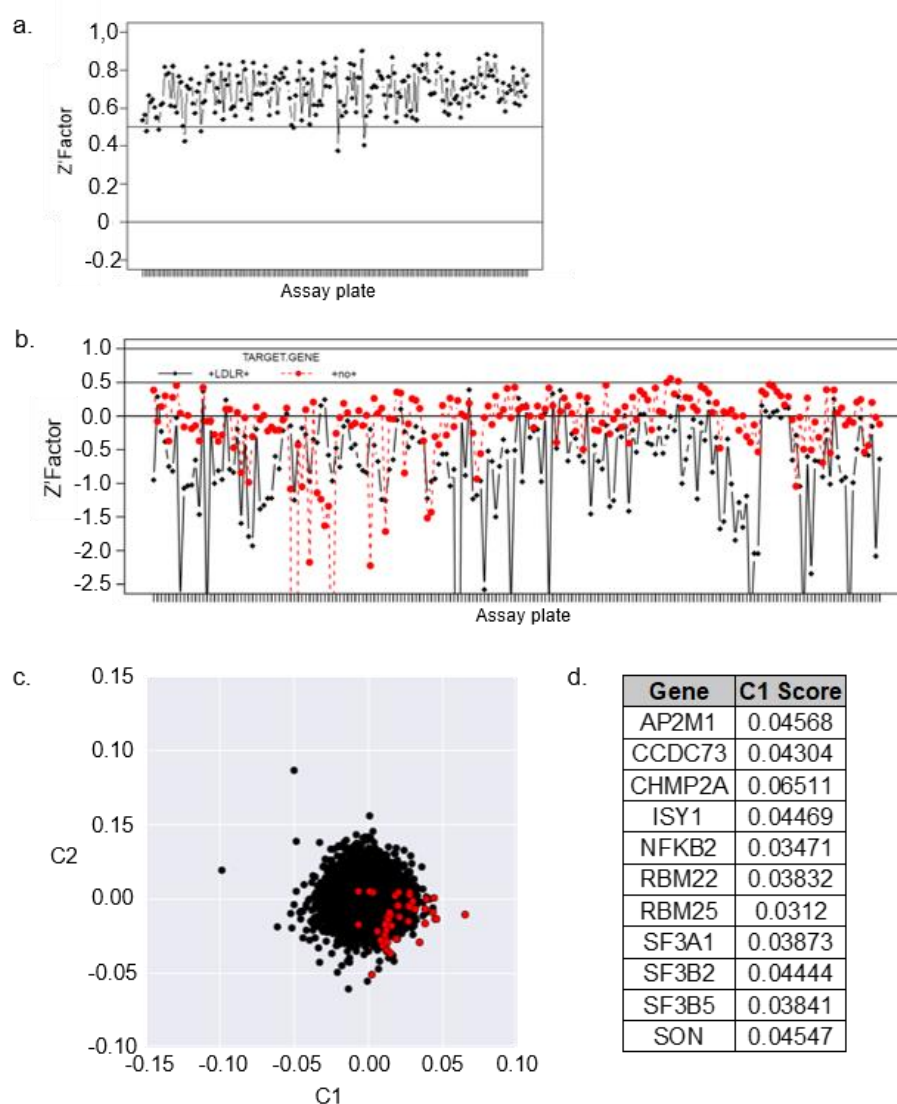
Supplemental Figure I. Quality control (QC) of the screening and non-linear dimensionality reduction of the dataset. a). QC for transfection efficiency. The effect size on nuclei counts of an siRNA against *PLK1*, an essential kinase that upon knockdown results in cell death, was measured in comparison to control wells that received a non-targeting siRNA pool and is expressed here as Z'-factor. Each dot represents one of the 192 screening plates. Z'-factors between 0.5 and 1 are considered excellent. b). Overview on negative and positive controls. This figure depicts the Z'-factors for median cytoplasm intensity in the LDL channel for negative control wells that did not receive any fl-LDL (in red) as well as for positive control wells that received an siRNA against *LDLR* (black). For either control, the Z'-factor was calculated in comparison to wells that received a non-targeting control siRNA and were subsequently incubated with fl-LDL. Z'-factors between 0.5 and 1 are considered excellent. c). Locally linear embedding (LLE) analysis. The graph shows the first two components of the screening dataset after LLE dimensionality reduction was applied. As LLE generates slightly different results at each iteration, a representative LLE outcome is reported in this figure. Each dot represents one gene. Red dots represent our top hit genes limiting LDL uptake after RSA analysis of the median cytoplasm intensity feature. Note how most of the red dots map at the extreme right of the C1 component. d). Top hit genes after LLE. This list contains all genes with a value over 0.03 on the C1 axis after LLE and corresponds to the iteration of the analysis shown on the left.
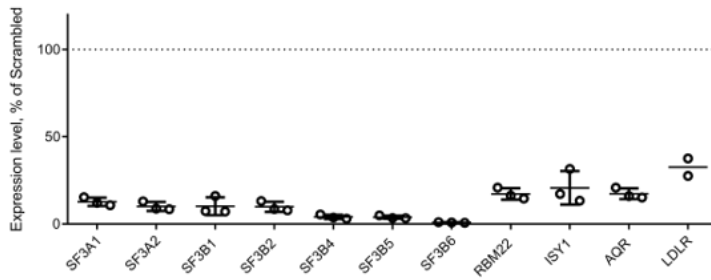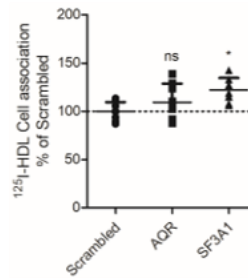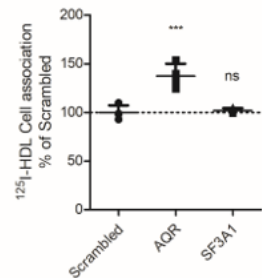
Supplemental Figure II. Additional in vitro validation experiments and determination of knockdown efficiency. a) Knockdown efficiency of siRNA pools in Huh-7 cells was measured by real-time PCR 72 hours after siRNA transfection and is expressed as percentage of expression compared to a non-silencing control (scrambled). With the exception of LDLR (from Ambion Silencer Select) all siRNAs were ON-TARGETplus Smart Pool siRNAs from Dharmacon. Data are shown as means±SD. **b)** ^{125}I-LDL cell association in HepG2 cells. **c)** ^{125}I-HDL cell association in Huh-7 cells. **d)** ^{125}I-HDL cell association in HepG2 cells. Cell association experiments were performed 72 hours after transfection of pooled siRNAs by incubating the cells for 2 hours at 37°C in the presence of 33.3ug/ml of radioiodinated lipoproteins or 40fold excess of unlabeled LDL. Data are shown as means±SD. Significance was calculated by one-way ANOVA followed by Dunnet's post-hoc test between all targeting siRNAs and the non-targeting siRNA (Scrambled). ns = not significant, * = p<0.05, ** = p<0.01, *** = p<0.001, ****= p<0.0001.
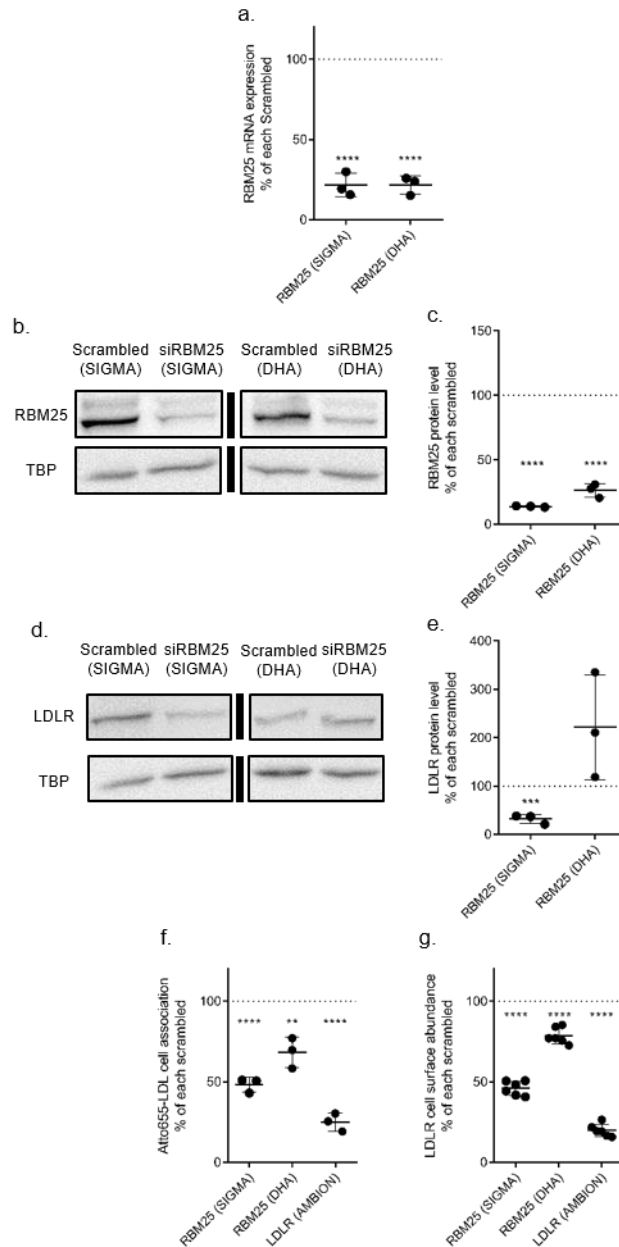
Supplemental Figure III. Validation of RBM25. a). Knockdown efficiency of siRNAs obtained from different vendors (Sigma or Dharmacon) in Huh-7 cells was measured by real-time PCR 72 hours after siRNA transfection and is expressed as percentage of expression compared to the respective vendor's non-silencing control (scrambled). Data are shown as means ±SD of three replicate experiments. **b) and c) Effect of RBM25 knockdown on RBM25 protein levels.** RBM25 protein levels in Huh-7 cells 72 hours after transfection with the indicated siRNAs were measured by western blot. A representative blot is shown in **b). c)** shows the relative density of RBM25 bands after knockdown of RBM25 with the indicated siRNAs, relative to the respective scrambled control. TATA-binding-protein (TBP) was used as the loading control. Bars represent means ±SD of three replicate experiments. **d) and e). Effect of RBM25 knockdown on LDLR protein levels.** LDLR protein levels in Huh-7 cells 72 hours after transfection with the indicated siRNAs were measured by western blot. A representative blot is shown in **d). e)** shows the relative density of LDLR bands after knockdown of RBM25 with the indicated siRNAs, relative to the respective scrambled

control. TATA-binding-protein (TBP) was used as the loading control. Bars represent means ±SD of three replicate experiments. **f) Effect of RBM25 knockdown on Atto655-LDL uptake**. Atto655-LDL uptake was measured 72 hours after transfection with siRNAs against RBM25 or non targeting controls (scrambled) obtained from different vendors by incubating the cells for 2 hours at 37°C with 10 ug/ml Atto655-LDL in the presence or absence of 100fold excess of unlabeled LDL. Each point represents one of three (**a through f**) or six (**g**) experiments. Data are shown as means ±SD. **g) Effect of RBM25 knockdown on LDLR cell surface levels**. LDLR cell surface levels in live Huh-7 cells were measured by flow cytometry 72 hours after knockdown of RBM25 with the indicated siRNAs. The siRNA against LDLR was used as the positive control. The data are normalized to the respective vendor's non-silencing control. Each point represents one of three identical experiments. Data are shown as means ±SD. Significance was calculated by t-test between each targeting siRNA and the respective non-targeting control (scrambled) of each vendor. ns = not significant, * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, ****= $p<0.0001$
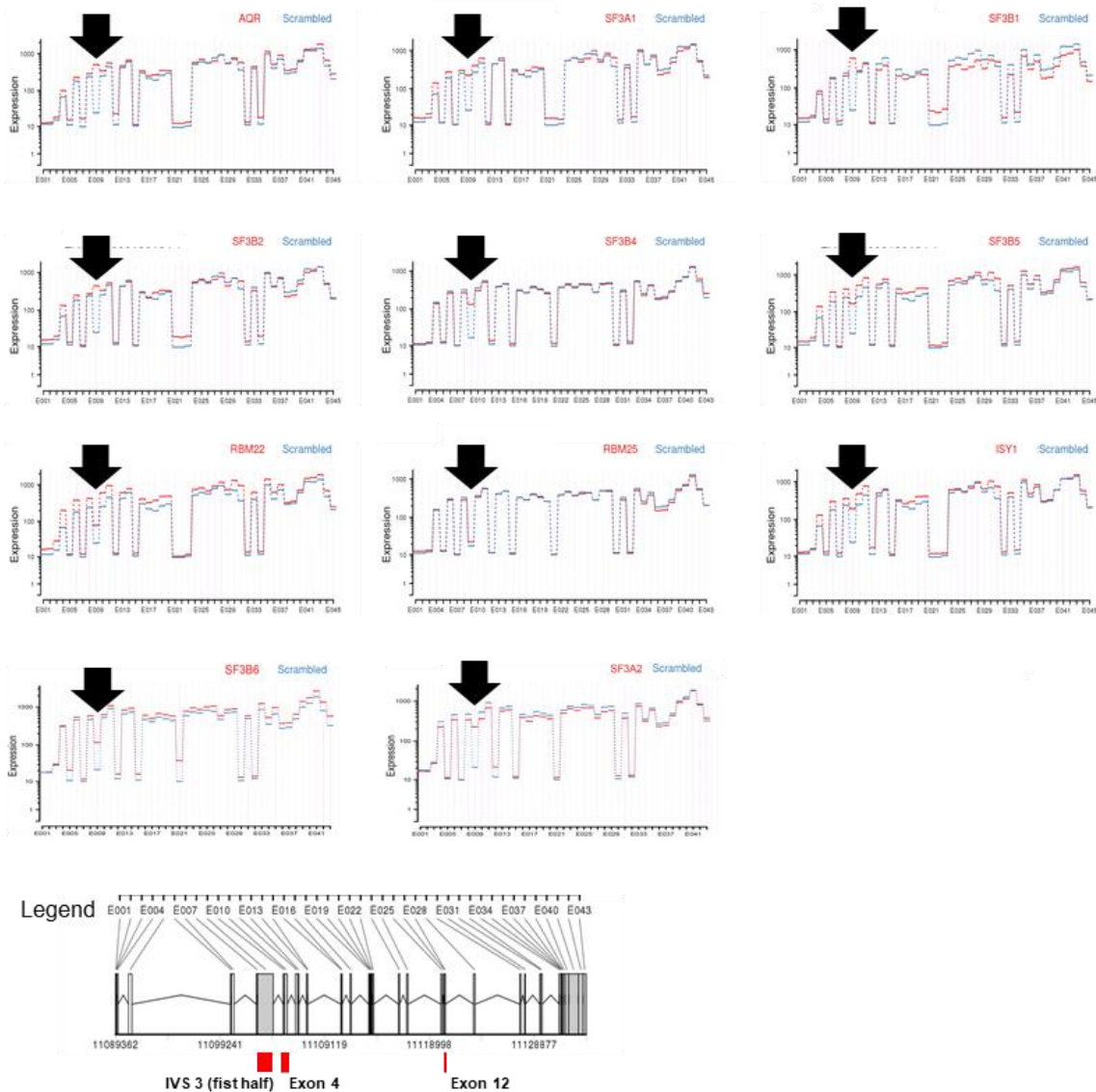
Supplemental Figure IV. *LDLR* **Exon level expression after knockdown of spliceosome genes**. Expression of the *LDLR* exons was recorded by RNA sequencing of Huh-7 cells 72 hours after knockdown of each spliceosomal hit gene. Segments represent differential exon usage in each sector of the *LDLR* genomic sequence as identified by the DEXSeq algorithm. Canonical exons within the ENSG00000130164 genomic reference are depicted in the legend. Normalized read counts are reported on the y axis. The black arrow indicates the location of ENSG00000130164:E009, corresponding to the first half of intron 3. Data represent the average of 3 replicate experiments
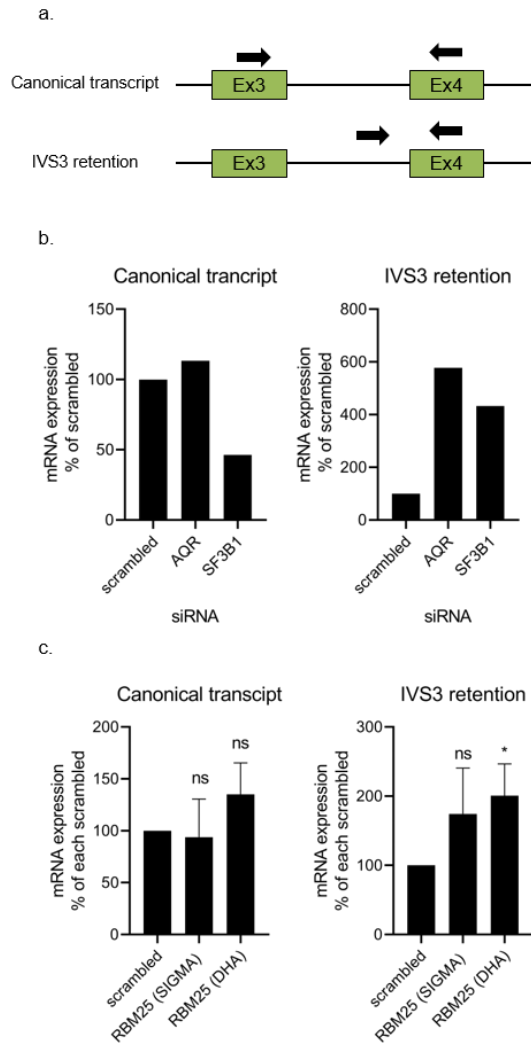
Supplemental Figure V. Confirmation of *LDLR* intron 3 retention by RT-PCR. a). RT-PCR setup. The approximate location of each primer within the region of *LDLR* spanning from exon 3 (Ex 3) to exon 4 (Ex 4) region is represented by black arrows. Primers are from Cameron et al., [22]. IVS3 = intron 3. **b) and c) Expression levels of the full length and intron 3-retaining *LDLR* transcripts after knockdown of AQR, SF3B1 (b) or RBM25 (c) in Huh-7 cells.**. Data are normalized to the respective non-targeting siRNA condition and are expressed as means (**b**) or means±SD (**c**).Significance (**c**) was calculated by t-test between each targeting siRNA and the respective non-targeting control (scrambled) of each vendor. ns = not significant, * = p<0.05
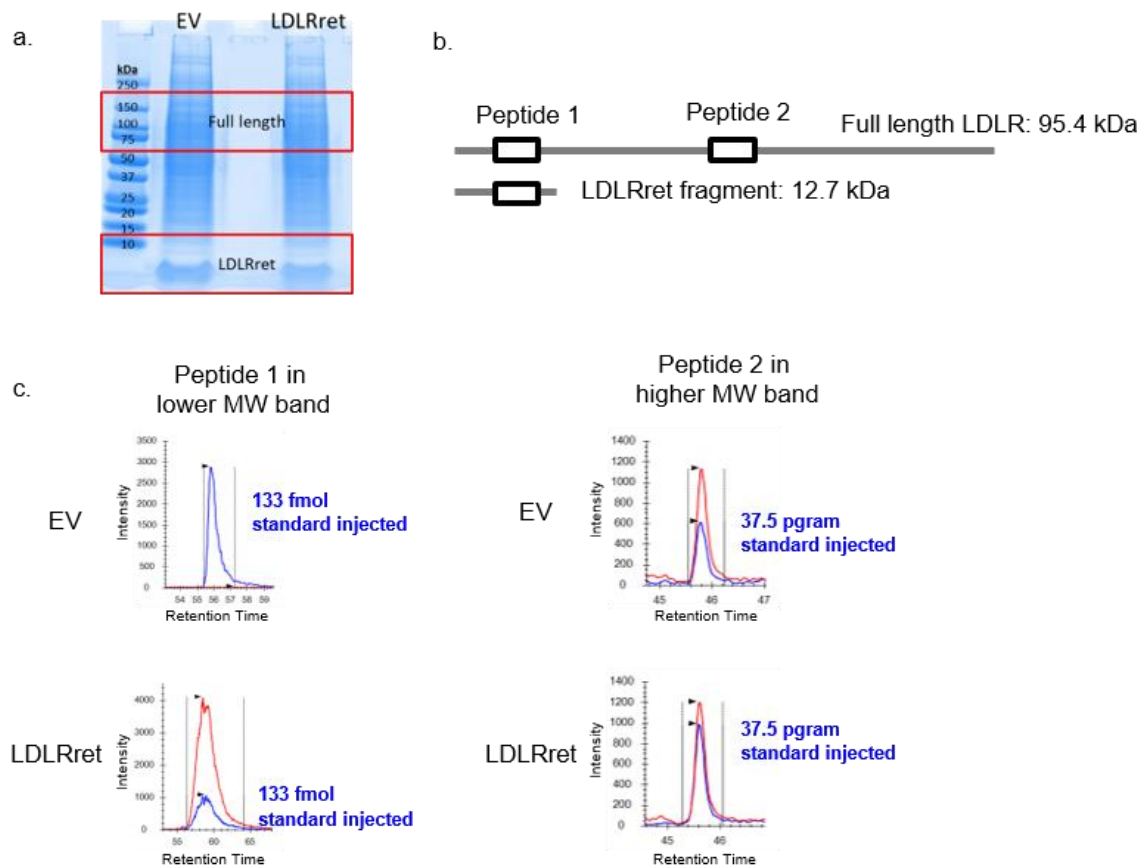
Supplemental Figure VI. Detection of the LDLRret fragment by liquid-chromatography coupled mass spectrometry (LC-MS). The full length LDLR and fragment LDLR were separated on a SDS-PAGE (**a**) and the bands containing the full size protein and the LDLR fragment were excised and trypsinized with in-gel digestion. **b) Localization of the two LDLR peptides recorded by LC-MS:** peptide 1 is located in the overlapping part of the protein (but the fragment and full protein are already separated in the SDS-PAGE gel). Peptide 2 is specific for the full-length protein. Isotopically labeled standard peptides were added to the trypsinized samples for comparison of the different samples. **c) Quantification of endogenous LDLR peptides and isotopic standard peptides by LC-MS.** The peaks of the isotopically labeled standards are shown in blue, with the concentration of each standard depicted next to it. The peaks of the endogenous peptides are shown in red and the peak boundaries are annotated with black dotted lines. These analyses were done for three types of samples: a sample with the LDLR fragment being overexpressed (untagged) annotated as LDLRret, an empty vector control annotated as EV and detection in a plasma pool (not shown).

a.

b.
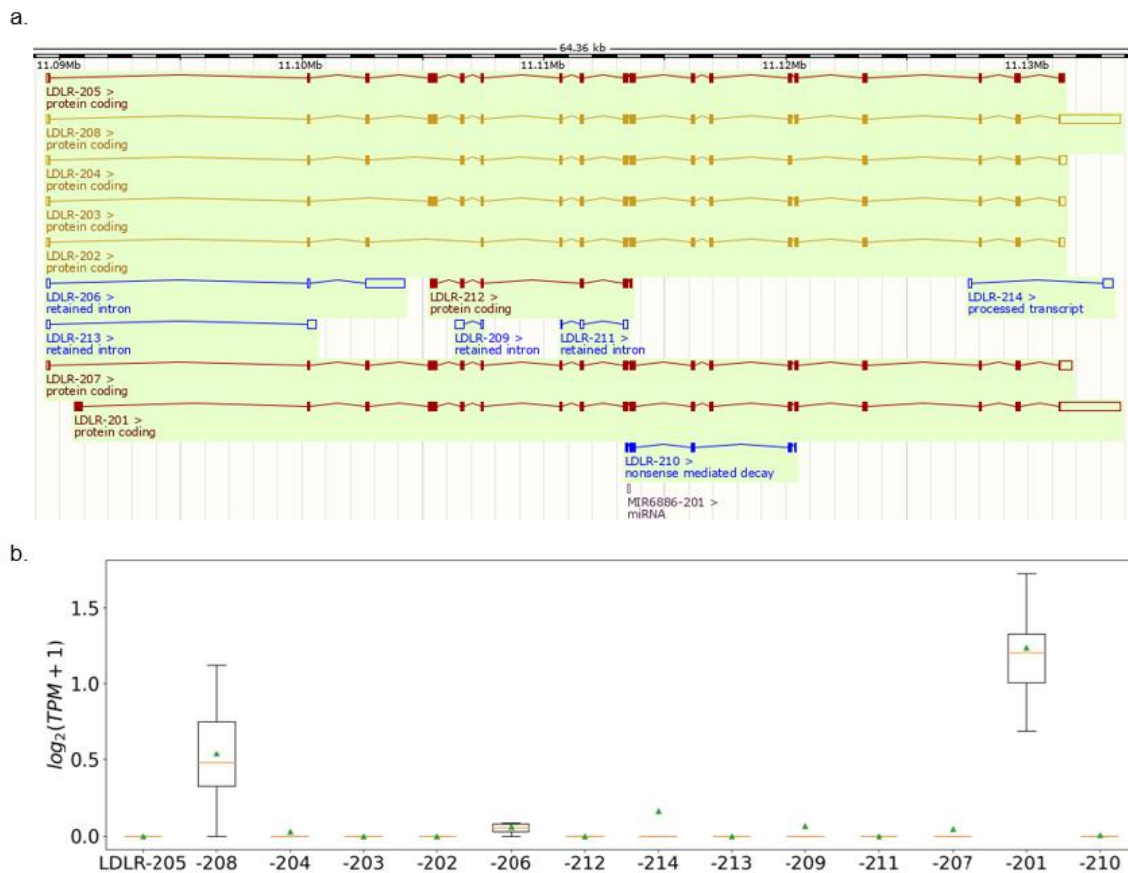
Supplemental Figure VII. Detection of the different LDLR transcripts in the liver of healtjy donors. a). Illustration of the different LDLR transcripts. The different available LDLR transcripts and the exons/introns included for each transcript are shown. Obtained from ensemble's webpage:
https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000130164;r=19:11089462-11133820. **b). In silico analysis of healthy human liver samples confirms the expression of LDLR transcript with retention of intron 3 (206)**. Computational analysis of previously published RNA seq data of livers from healthy subjects (Gene Expression Omnibus, accession number GSE126848)[11]. Graph depicts the log2 transcript per million (TPM) levels for each of the annotated LDLR transcripts (x axis). The orange lines and the green triangles represent median and mean values, respectively. Boxes represent IQR's.
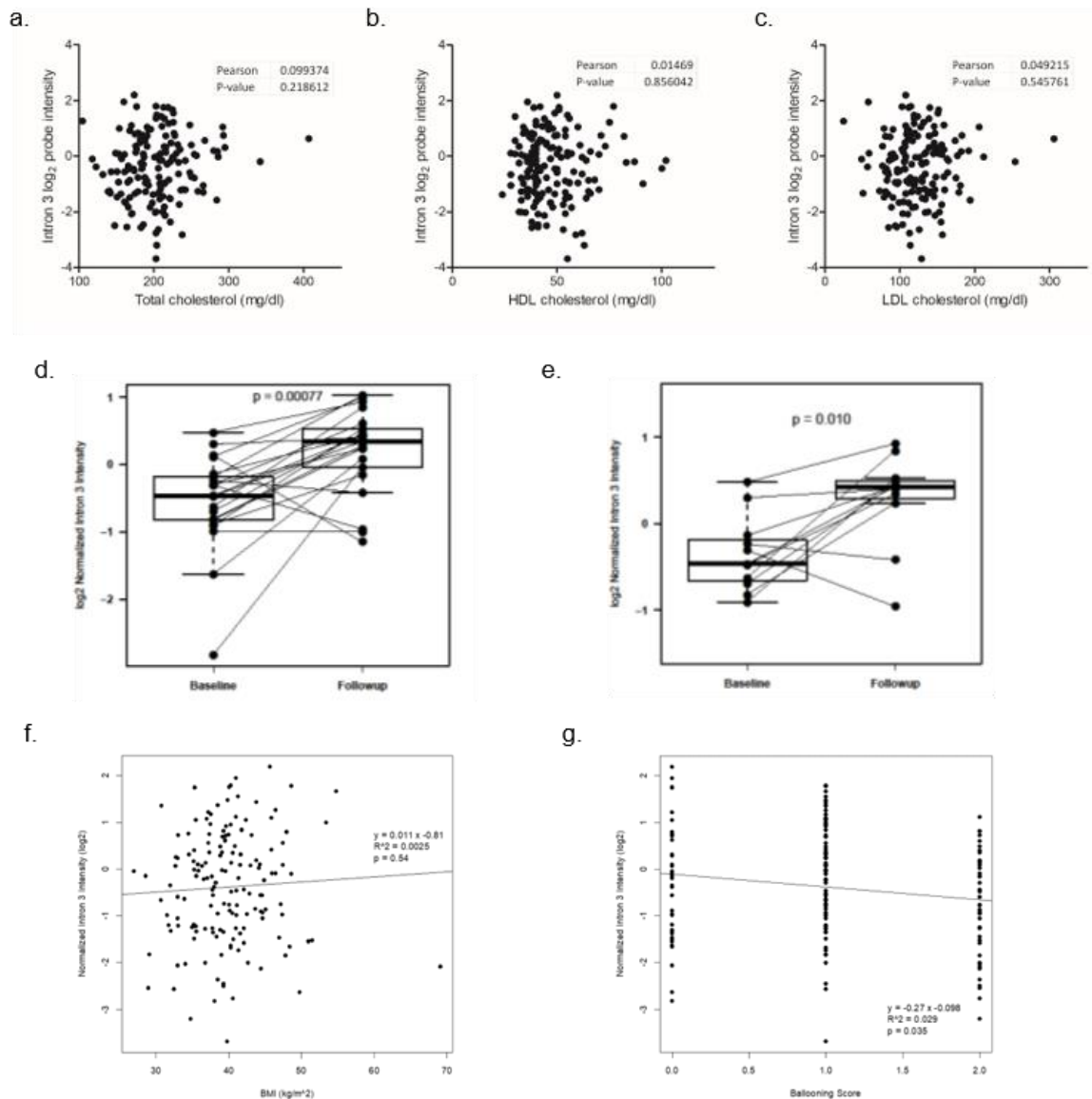
Supplemental Figure VIII. Hepatic expression of *LDLR* Intron 3 retention does not correlate with plasma lipids but is changed upon bariatric surgery in obese subjects. a-c. Correlations of intron 3 probe relative intensity values with plasma levels of total (a), HDL- (b) and LDL cholesterol (c). Hepatic Intron 3 probe intensity in the baseline cohort (n=155 patients) is shown after $\log_2$ transformation. The Pearson coefficient as well as the P-value for correlation are reported in the boxes on the upper right of each panel. **d) and e). Effect of bariatric surgery on intron 3 retention.** Intron 3 retention levels for 21 patients for which hepatic mRNA was available at baseline and after bariatric surgery (median follow-up time 13 months, IQR = [12 months, 15 months]). Normalized intron 3 is calculated as the ratio between Robust Multichip Averaging (RMA)-normalized probe intensity for the intron 3 probe and the RMA-normalized total LDLR transcript level for each sample. The diagonal lines connect the datapoints of each individual. All available data (N=155) are shown in **a)-c)**, while only the data from 21 paired biopsies and 11 responders, defined as having non-alcoholic steatohepatitis (NASH) at baseline but no NASH at follow-up, are shown in **d)** and **e)**, respectively. **f) and g). Correlations of normalized hepatic intron 3 expression with patient BMI (f) or histological ballooning score (g)** in the baseline cohort (n = 155).

9
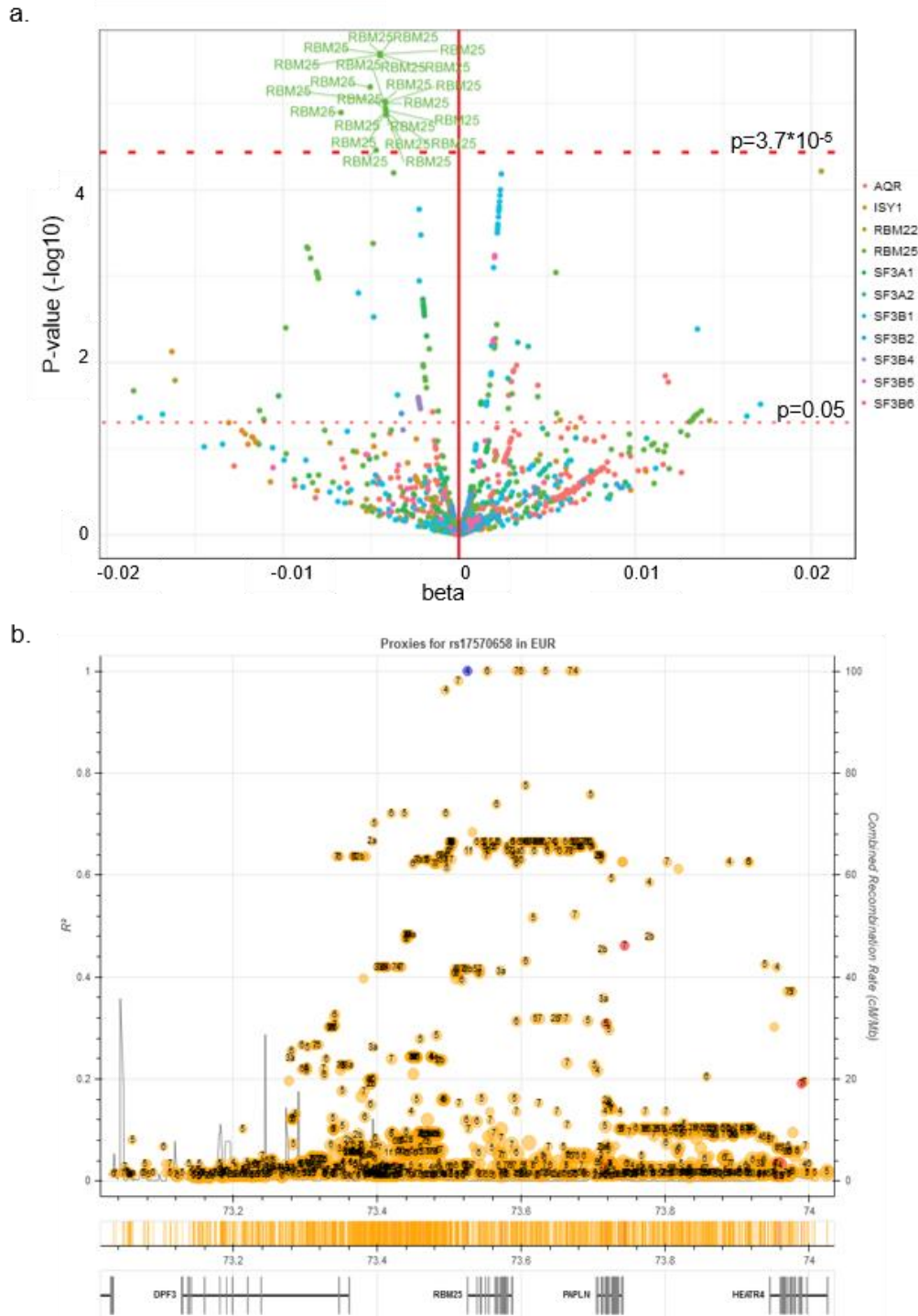
Supplemental Figure IX. a). Association between RBM25 variants and plasma apoB levels in the UK Biobank dataset. The dashed red horizontal lines indicate the p=0.05 threshold as well as the threshold for statistical significance after correction for multiple testing of 1360 variants within the genes of interest (p=3.7*10^{-5}), respectively. Effect size and directionality are reported on the x axis as beta value. **b). Linkage disequilibrium information for the rs17570658 SNP in Europeans.** Data from Phase 3 (Version 5) of the 1000 Genomes Project were plotted using the LDproxy tool at LDlink (https://ldlink.nci.nih.gov/) and are expressed as R2. rs17570658 is indicated by the blue circle
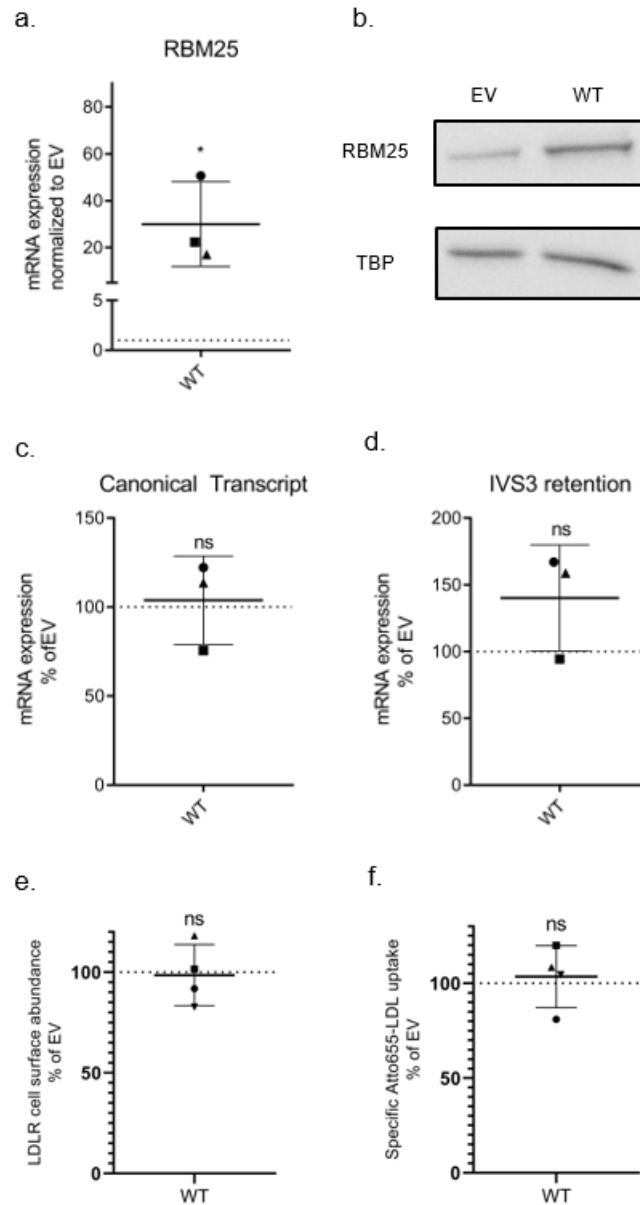
Supplemental Figure X. Overexpression of wild type RBM25 affects neither splicing nor cell surface expression of LDLR nor uptake of LDL. 72 hours after plasmid transfection and 48 hours after introduction of the selection antibiotic G418, the cells were harvested and analyzed. **a) and b).** Overexpression of RBM25 is confirmed by RT-PCR **(a)** and Western blotting **(b). c) and d).** RT-PCR does not reveal any different expression of the canonical **(c)** and the intron 3 retaining LDLR transcripts **(d)** between cells overexpressing the empty vector and the wild type RBM25 protein. **e) and f)** Flow cytometry does not reveal any difference in LDLR cell surface levels **(e)** and LDL uptake **(f)** between cells overexpressing the empty vector and the wild type RBM25 protein. In every graph, each point represents one of three **(a,b,c,d)** or four **(e,f)** experiments. Data are normalized to the empty vector (EV) and are shown as means ±SD.
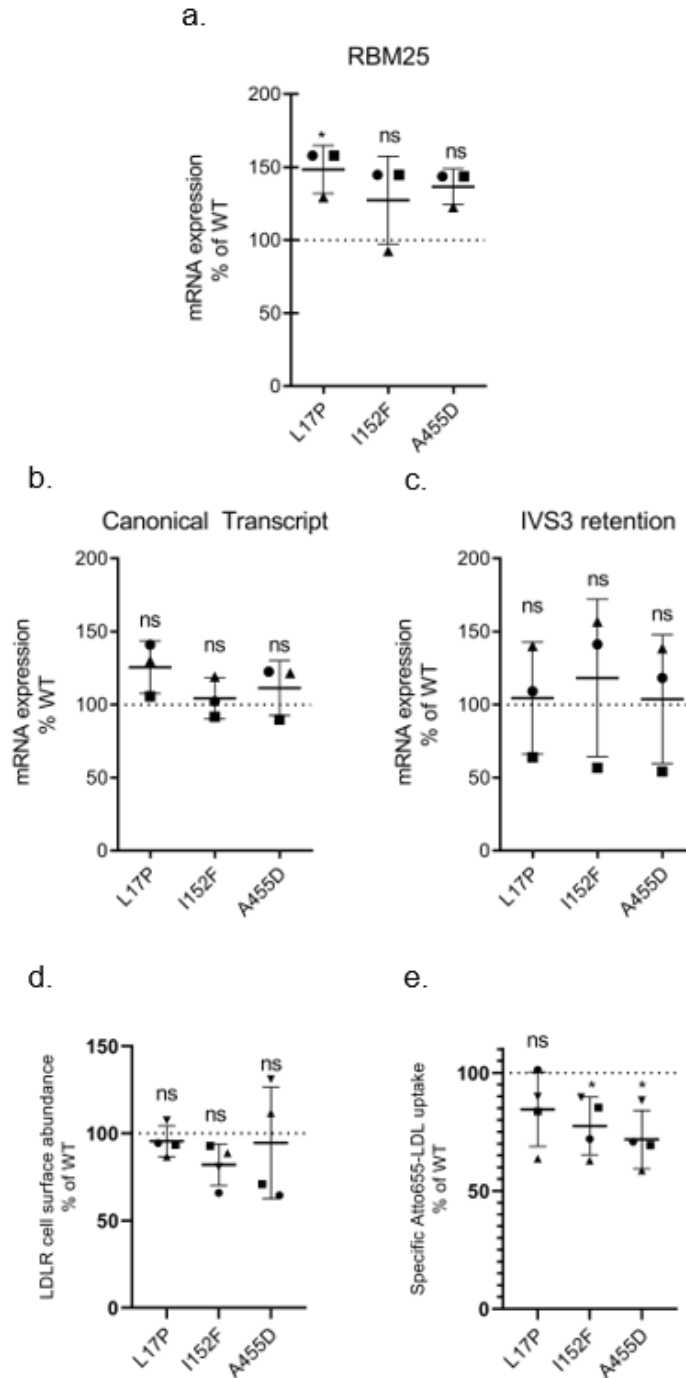
Supplemental Figure XI. Impaired LDL uptake by Huh7 cells overexpressing RBM25 mutants. 72 hours after plasmid transfection and 48 hours after introduction of the selection antibiotic G418, the cells were harvested and analyzed by using RT-PCR (**a, b, c**) and flow cytometry (**d,e**). Overexpression of wild type and mutant RBM25 constructs cause comparable increases in mRNA levels of RBM25 (**a**), LDLR canonical transcript (**b**) and intron 3 retention transcript (**c**). **d)** LDLR cell surface abundance is not altered upon expression of RBM25 mutants compared to wild type RBM25. **e)** Uptake of fluorescently labeled LDL is lower in cells overexpressing RBM25 mutants than in cells overexpressing wild type RBM25. Statistical significance of differences between RBM25 wild type and RBM25 mutants was calculated by one-way ANOVA followed by Dunnett`s post-hoc test. * = $p<0.05$. In every graph, each point represents data of one of three (**a,b,c**) or four (**d,e**) experiments. Data are normalized to RBM25 wild type (WT) and are shown as means ±SD.

3. Supplemental Tables

Supplemental Table I (Provided as Excel file). Complete screening dataset for the five best performing assay features. The assay features values are normalized as reported in the methods section. Non self-explanatory column names are as follows: Gene.ID.20160314: NCBI Gene-ID on 2016.03.14; siRNA.ID: siRNA molecule ID as according to manufacturer; Gene.Symbol.20160314: NCBI Gene Symbol on 2016.03.14; RefSeq.Accession.Number: NCBI RefSeq ID for the targeted transcript; Batch: experimental batch number; FL.LABEL: describes whether the well received fluorescently-labelled LDL; RNASeq_Huh-72..Signal.: normalized gene-level expression as measured by RNA sequencing; RNASeq_Huh-72..Present.: dichotomic gene-level expression label (threshold 7.5). For each assay feature, RSA p-values are given for both directionalities of the RSA analysis (inhibition and enhancement of LDL uptake). No RSA p-values were generated for empty wells, control wells, wells without an associated Gene.ID as well as wells that did not meet the lower ranking threshold for the RSA analysis.

.

Supplemental Table II. Results of the Panther GO-Slim BP enrichment test.

PANTHER GO-Slim Biological Process	RL	HG	EV	FE	p	FDR
intra-Golgi vesicle-mediated transport (GO:0006891)	28	3	0.07	40.9	7.20E-05	7.61E-03
mRNA splicing, via spliceosome (GO:0000398)	164	7	0.43	16.29	3.02E-07	2.71E-04
macromolecule metabolic process (GO:0043170)	2419	22	6.34	3.47	6.73E-08	1.21E-04
organic substance metabolic process (GO:0071704)	3338	22	8.74	2.52	1.62E-05	2.08E-03
metabolic process (GO:0008152)	4072	22	10.67	2.06	4.42E-04	4.18E-02
RNA processing (GO:0006396)	233	7	0.61	11.47	2.92E-06	4.04E-04
gene expression (GO:0010467)	1842	14	4.83	2.9	2.23E-04	2.23E-02
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	164	7	0.43	16.29	3.02E-07	1.81E-04
RNA splicing, via transesterification reactions (GO:0000375)	165	7	0.43	16.2	3.14E-07	1.41E-04
RNA splicing (GO:0008380)	176	7	0.46	15.18	4.77E-07	1.22E-04
proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	142	5	0.37	13.44	4.07E-05	4.88E-03
macromolecule catabolic process (GO:0009057)	300	8	0.79	10.18	1.29E-06	2.31E-04

	RL	HG	EV	FE	P	FDR
proteolysis involved in cellular protein catabolic process (GO:0051603)	253	8	0.66	12.07	3.69E-07	1.33E-04
cellular protein catabolic process (GO:0044257)	255	8	0.67	11.98	3.91E-07	1.17E-04
protein catabolic process (GO:0030163)	289	8	0.76	10.57	9.79E-07	2.20E-04
proteasomal protein catabolic process (GO:0010498)	153	5	0.4	12.48	5.75E-05	6.45E-03
protein ubiquitination (GO:0016567)	200	7	0.52	13.36	1.09E-06	2.18E-04
protein modification by small protein conjugation (GO:0032446)	217	7	0.57	12.31	1.85E-06	2.77E-04
protein modification by small protein conjugation or removal (GO:0070647)	217	7	0.57	12.31	1.85E-06	3.02E-04

Only categories with FDR<0.05 are displayed. The categories are sorted by the Fold Enrichment of the most specific categories (highlighted in bold), with their parent terms (p-value lower than 0.05) indented directly below. RL: Reflist genes, number of genes in the Panther reference list that map to each particular GO-Slim BP category; HG: hit genes, number of screening hit genes that map to each category; EV: expected value, expected number of genes mapping to each category if no enrichment were present; FE: Fold enrichment, FE is calculated as the ratio between the number of hit genes detected by the screen and the expected value. P: raw p-values based on a Fischer's exact test; FDR: False discovery rate

Supplemental Table III. *In silico* BPS predictions in human and mouse.

Species	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
human	43	294	tactcaacc	0.80074	0.50519	35	12	16	-1.18792
human	43	283	aaataagga	-2.45889	0.51079	24	12	16	-1.76613
human	43	266	ctataaatt	-1.48048	0.51724	7	12	16	-0.30488
human	43	240	gtctcagtt	0.00264	0.50638	10	13	20	0.11969
human	43	235	agtttaaca	-3.81564	0.50435	5	13	20	-1.05951
human	43	234	gtttaacag	0.47649	0.50655	4	13	20	0.68507
human	43	226	gctttacac	-1.87418	0.50226	103	22	33	-6.38214
human	43	217	ctattagcg	-2.00591	0.49528	94	22	33	-5.86629
human	43	202	tgctcatag	0.87151	0.49239	79	22	33	-3.7911
human	43	173	agatgagga	-3.29087	0.51786	50	22	33	-3.57699
human	43	164	aactgaggc	0.04083	0.53459	41	22	33	-1.69739
human	43	141	ggttcagag	-1.70368	0.56618	18	22	33	-0.91438
human	43	124	ccctgactg	3.62683	0.57143	1	22	33	2.25053
human	43	87	gcctcactg	1.97921	0.52439	70	13	21	-2.88916
human	43	38	acttcacac	-0.61163	0.60606	21	13	21	-0.7756
human	**43**	**30**	**cggtgatgg**	**1.7613**	**0.68**	**13**	**13**	**21**	**0.68379**
mouse	29	295	gtataattt	-0.89117	0.4931	63	10	13	-4.07236
mouse	29	291	aatttatac	-3.30713	0.48951	59	10	13	-4.70549
mouse	29	283	catttagca	-2.73307	0.48561	51	10	13	-3.8971
mouse	29	274	agataagca	-1.35831	0.49442	42	10	13	-2.70381
mouse	29	253	acctaaagc	0.36538	0.49597	21	10	13	-0.49484
mouse	29	246	gcattaaga	-3.75146	0.49793	14	10	13	-1.55753

mouse	29	245	cattaagat	-0.94489	0.5	13	10	13	-0.40666
mouse	29	234	ggctcaata	-0.62438	0.50218	2	10	13	0.52747
mouse	29	218	cagtgaggg	-1.10555	0.49765	98	11	12	-6.73381
mouse	29	210	ggctcatag	0.16255	0.50244	90	11	12	-5.65641
mouse	29	190	gcttgatag	-0.89592	0.51351	70	11	12	-4.58519
mouse	29	183	agttgacca	1.31323	0.51685	63	11	12	-3.22098
mouse	29	171	ggatgagga	-2.5864	0.5241	51	11	12	-3.83023
mouse	29	132	tcctgagcc	1.0425	0.5748	12	11	12	0.45168
mouse	29	99	tgctaaatg	1.46063	0.56383	6	12	21	1.06719
mouse	29	80	ttttgaagc	-1.06802	0.57333	10	10	16	-0.20359
mouse	29	45	gtgtaagcc	-0.29433	0.675	18	15	31	-0.43572
mouse	**29**	**33**	**gcctgacag**	**2.01588**	**0.75**	**6**	**15**	**31**	**1.35885**

Each entry represents a putative U2-dependent branch point identified by the SVM-BP-finder algorithm. Results for the last 300nt of each intron are displayed. A BPS is considered valid when located close to the AG exclusion zone (AGEZ; the distance to 3'ss is usually approx. within AGEZ + 9nt), with BP-score > 0 and with svm_score > 0. The two entries matching these requirements in man and in mouse are highlighted in bold and underlined. agez: AG dinucleotide Exclusion Zone length; ss_dist: Distance to 3' splice site; bp_seq: BP sequence (nonamer; from -5 to +3 relative to the BP adenine); bp_scr: BP sequence score using a variable order Markov model; y_cont: Pyrimidine content between the BP adenine and the 3' splice site; ppt_off: Polypyrimidine tract offset relative to the BP adenine; ppt_len: Polypyrimidine tract length; ppt_scr: Polypyrimidine tract score; svm_scr: Final BP score using the SVM classifier.

Supplemental Table IV. Clinical characteristics of the obese non-diabetic patients who underwent liver biopsy at the Antwerp University Hospital (N = 155).[12]

	mean ± SD
Gender (% male)	35.5
Age (years)	43.2 ± 12.9
BMI (kg/m2)	39.9 ± 5.6
Plasma Cholesterol (mg/dL)	203.6 ± 41.7
HDL-C (mg/dL)	47.5 ± 14
Plasma TG (mg/dL)	164.9 ± 86
LDL-C (mg/dL, calculated[23])	123.6 ± 37
Fasting Glucose (mg/dL)	86.1 ± 13.8
Fasting Insulin (µU/mL)	20.1 ± 18.3
HbA1c (%)	5.6 ± 0.6
HOMA-IR	4.5 ± 5.4
ASAT (U/L)	25.2 ± 14.7
ALAT (U/L)	40.4 ± 29.7
Alkaline Phosphatase (U/L)	78.1 ± 20.6
γGT (U/L)	50 ± 38.4

Supplemental Table V: Correlations of Total and Lipoprotein Cholesterol levels with the expression of LDLR transcripts in the liver of non-diabetic obese patients

	probeID	total cholesterol	HDL cholesterol	LDL cholesterol	non-HDL cholesterol
5'UTR	16858387	-0.005	0.061	-0.054	-0.025
Exon1	16858388	-0.011	-0.096	-0.011	0.020
Exon2	16858390	0.099	0.020	0.045	0.091
Exon2	16858391	0.076	-0.030	0.039	0.085
Intron2	16858392	0.038	0.003	0.009	0.037
Intron3	16858394	0.099	0.015	0.049	0.093
Exon4	16858395	0.056	-0.046	0.006	0.071
Exon4	16858396	0.112	-0.014	0.055	0.115
Intron4	16858397	0.072	0.070	0.017	0.048
Exon5	16858398	0.103	-0.015	0.056	0.106
Exon6	16858399	0.184	-0.110	0.164	0.218
Exon7	16858400	0.053	0.010	0.015	0.050
Exon8	16858401	0.003	-0.021	-0.024	0.010
Exon9	16858402	0.096	-0.043	0.096	0.109
Exon10	16858403	0.038	0.037	0.009	0.025
Exon11	16858404	0.039	0.050	-0.017	0.022
Exon12	16858405	0.091	-0.031	0.042	0.100
Exon13	16858406	0.120	0.087	0.078	0.090
Exon13	16858407	0.052	0.160	-0.020	-0.001
Exon15	16858409	0.230	0.039	0.133	0.214
Intron15	16858410	0.089	0.029	0.031	0.079
Exon16	16858411	0.104	0.078	0.038	0.077
Exon17	16858413	0.110	0.016	0.074	0.103

Data from the baseline cohort described in Supplementary Table IV (n=155 patients)[12]. $\log_2$ transformed intensities of probes depicted in figure 5B were correlated with plasma lipid levels by Pearson. None of the correlations is statistically significant.

Supplemental Table VI. Association between variants in spliceosome hit genes and

multiple biochemical phenotypes in 40,468 UK Biobank exomes.

	Gene	LDL-C	ApoB	TG	HDL-C	ApoA	Chol.	Lp(a)	Glc	HbA1c
All Ethnicities coding-based model	AQR	0.32	0.45	0.59	0.18	0.35	0.17	0.71	0.29	0.73
	ISY1	0.87	0.67	0.64	0.16	0.17	0.69	0.93	0.42	0.2
	RBM25	0.22	0.31	0.59	0.72	0.81	0.3	0.74	0.39	0.061
	SF3A1	0.69	0.55	**0.007**	0.28	0.97	0.85	0.95	0.62	0.079
	SF3A2	0.66	0.34	0.86	0.1	0.089	0.94	0.18	0.11	**0.028**
	SF3B1	0.092	0.09	0.28	0.68	0.75	**0.043**	0.76	0.81	0.16
	SF3B2	0.66	0.52	0.54	0.49	0.36	0.9	0.7	0.62	0.41
	SF3B4	0.19	0.3	0.41	0.42	0.7	0.29	0.36	0.28	0.99
	SF3B5	0.77	0.51	0.93	0.19	0.089	0.77	0.88	0.25	0.23
	SF3B6	0.8	0.81	0.82	0.66	0.73	0.94	0.89	0.73	0.77
Eur Ethnicity coding-based model	AQR	0.73	0.78	0.69	0.47	0.62	0.6	0.65	0.19	0.88
	ISY1	0.93	0.47	0.43	0.22	0.21	0.94	0.72	0.41	0.29
	RBM25	0.28	0.4	0.81	0.97	0.78	0.35	0.52	0.51	0.12
	SF3A1	0.76	0.92	**0.039**	0.93	0.37	0.27	0.72	0.57	0.12
	SF3A2	0.77	0.6	0.65	**0.044**	0.063	0.74	0.51	0.59	0.081
	SF3B1	0.074	0.1	0.27	0.42	0.41	**0.032**	0.74	0.87	0.1
	SF3B2	0.49	0.38	0.57	0.91	1	0.56	0.76	0.47	0.056
	SF3B4	0.42	0.65	0.62	0.85	0.84	0.6	0.93	0.46	0.3
	SF3B5	0.88	0.8	0.7	0.2	0.23	0.77	0.95	0.07	0.12
	SF3B6	0.98	0.93	0.75	0.9	0.82	0.93	0.77	0.2	0.67
All Ethnicities LOF-based model	AQR	0.76	0.74	0.76	0.54	0.68	0.81	0.59	0.34	0.97
	ISY1	0.78	0.9	0.42	0.8	0.54	0.82	0.55	0.56	0.49
	RBM25	0.36	0.24	0.54	0.19	0.12	0.54	0.44	0.97	0.47
	SF3A1	0.98	0.94	0.5	0.83	0.66	0.88	1	0.44	0.98
	SF3A2	0.76	0.43	0.32	0.09	0.11	0.8	0.48	0.43	**0.037**
	SF3B2	0.47	0.27	0.64	0.94	0.9	0.67	0.37	0.27	0.49
	SF3B4	0.96	0.96	0.11	0.55	0.62	1	0.38	0.28	**0.008**
	SF3B5	0.4	0.16	0.78	0.14	0.065	0.58	0.86	0.24	0.32
	SF3B6	0.68	0.56	0.76	0.75	0.97	0.87	0.93	0.97	0.97

Eur Ethni-city LOF-based model	*AQR*	0.92	0.86	0.77	0.5	0.67	0.97	0.36	0.34	0.97
	ISY1	0.79	0.87	0.3	0.75	0.92	0.9	0.65	0.52	0.15
	RBM25	0.29	0.15	0.63	0.18	0.072	0.5	0.24	0.53	0.32
	SF3A1	0.96	0.96	0.56	0.89	0.67	0.94	1	0.45	0.88
	SF3A2	0.77	0.7	0.79	0.061	0.12	0.78	0.94	0.84	0.25
	SF3B2	0.34	0.22	0.27	0.79	0.85	0.52	0.63	0.45	0.23
	SF3B5	0.4	0.26	0.3	0.16	0.2	0.52	1	0.075	0.11
	SF3B6	0.83	0.76	0.71	0.92	0.84	0.83	0.79	0.27	0.48

1   No association was significant, neither at the genome-wide level (p-value threshold $3.13*10^{-6}$)

2   nor after correcting the p-value for limited hypothesis testing (p-value threshold 0.004545).

3   Data are expressed as p-values. Associations with p-value below 0.05 are highlighted in bold.

4   Non-standard abbreviations: Chol.: Total cholesterol; Glc: Glucose

5

1 **Supplementary Table VII. Gene-based constraint information from the gnomAD**
2 **database for the spliceosomal hit genes**.

Gene	pLI
SF3A1	0.99
SF3A2	0.88
SF3B1	1
SF3B2	1
SF3B4	0.99
SF3B5	0.58
SF3B6	0.59
RBM22	1
RBM25	1
ISY1	1
AQR	1
Average	0.91
SD	0.17

3

4 pLI = probability of being loss-of-function intolerant. pLI is an estimate of deleteriousness
5 based on variant frequencies within that gene used in the GnomAD database. Metrics to
6 measure a transcript's intolerance to variation by comparing the observed number of variants
7 in the gnomAD dataset with the expected number of variants. Transcripts that are
8 significantly depleted of their expected variation are considered constrained, or intolerant, of
9 such variation. The closer pLI is to 1, the more intolerant of protein variants the transcript
10 appears to be. pLI ≥ 0.9 is considered as an indication of extreme intolerance. For more
11 details see supplement of reference[14] and https://gnomad.broadinstitute.org/).

12

Table VIII. *RBM25* gene variants identified in 71 FH patients without any mutation in the three canonical FH genes.

rsID	hg 19 Chr:pos	cDNA	AA	GT	MAF	PPH2	SIFT	MT
NA	14:73554780	c.454A>T	p.I152F	het	0	PSD	D	DC
NA	14:73572776	c.1364C>A	p.A455D	het	0	PD	D	DC
rs1167173761	14:73538399	c.50T>C	p.L17P	het	$9*10^{-6}$	PD	D	DC

The overall p value for a gene burden test of rare variants (MAF<0.0001) in RBM25 identified in 71 FH mutation negative patients in comparison to the gnomAD data of 56,352 individuals was p= 0.001. The three RBM25 variants listed in the table were identified in 3 FH probands and were unique to those individuals. Coordinates refer to the canonical ENST00000261973 transcript. AA: amino acid; nsSNV: nonsynonymous single nucleotide variant; GT: genotype; het: heterozygote; MAF: minor allele frequency in the GnomAD database; PPH2: PolyPhen 2; MT: Mutation Taster; PSD: PoSsibly Damaging; PD: Probably Damaging; D: Damaging; DC: Disease Causing

Supplemental Table IX: PCR primersPrimer sequences for all primer couples used in
this study.

Experiment/primer	5'-3' Sequence
LDLRret fragment cloning	
F	ATGGGGCCCTGGGGCTGG
R	TCAGATAGGCTCAATAGCAAAGGCA
Short pSPL3 minigene cloning	
F	CTAGCGAATTCGCAAAGACAGGATTGGCAAGG
R	CTAGCGGATCCTGGTATGAGCCCCCAAGAGA
Long pSPL3 minigene cloning	
F	GCTAGCGAATTCTGCAAAGACAGGATTGGCAAG
R	CTAGCGGATCCTGGAAATCCACTTCGGCACC
Determination of siRNA knockdown efficiency in Huh-7	
SF3A1 – F	TGTTACCGAGTGGAATGGGC
SF3A1 – R	GATCTGAGCATAGGCCACCC
SF3A2 – F	TCG ACA TCA ACA AGG ACC CG
SF3A2 – R	GCT TCT TCC CCT GCG TAT GT
SF3B1 – F	AGATCGCCAAGACTCACGAAG
SF3B1 – R	ACCTGTAGAATCGAGGCCCA
SF3B2 – F	CTATGGCCCACCCACCAAAT
SF3B2 – R	CGAATGATCTCCCTGCTGCT
SF3B4 – F	TCAGGATGCCACTGTGTACG
SF3B4 – R	TCCTTTGGCATGTGGGTGTT
SF3B5 – F	CCGCTACACCATCCATAGCC
SF3B5 – R	GCCCATGTAGGAGCAGTACG
SF3B6 – F	GGC GAA CAT TCG ACT TCC AC
SF3B6 – R	GGT GTG TTC CCC ACT CTG AT
RBM22 – F	TCGTGACGCAGGATTGTCTT
RBM22 – R	TAGATGTGGCTTTCCCCAGC
RBM25 – F	AGCCAGAATCTACCCTCCGT
RBM25 – R	TCTGGCCTTGCATTCCCATT
ISY1 – F	GGCCCGAAATGCAGAAAAGG
ISY1 – R	GGCCAGAAAGGGTCTTCGTT

AQR – F	GGCGGCATTAGCTGAAACTG
AQR – R	CAGCTGTAAGCCCATCCACA
LDLR – F	AAG GAC ACA GCA CAC AAC CA
LDLR – R	CAT TTC CTC TGC CAG CAA CG
GAPDH – F	CCCATGTTCGTCATGGGTGT
GAPDH – R	TGGTCATGAGTCCTTCCACGATA
LDLR expression and IVS3 retention in human liver and Huh-7 cells (from[11])	
LDLR canonical transcript - F	GACAACGGCTCAGACGAGCA
LDLR canonical transcript - R	CCACAGGTGAGCACCGGGCA
IVS3 retention – F	GTGATGGTGGTCTCGGCCCA
IVS3 retention – R	GGACCACAGGTGAGCACCGG

1   IVS3: intron 3. F: forward primer, R: reverse primer

2

25

1 **Supplemental Table X: gene boundaries for the annotation of the UK Biobank GWAS**
2 **data**.

Gene	chr	Start	End
AQR	15	35148551	35261995
ISY1	3	128846258	128880073
RBM22	5	150070351	150080669
RBM25	14	73525220	73588076
SF3A1	22	30727976	30752936
SF3A2	19	2236815	2248678
SF3B1	2	198256697	198299817
SF3B2	11	65819815	65836382
SF3B4	1	149895208	149900144
SF3B5	6	144416017	144416754
SF3B6	2	24290453	24299314

3 Coordinates refer to the GRCh37/hg19 build

4

Major Resources Table

Antibodies

Target antigen	Vendor or Source	Catalog #	Working concentration (µg/ml)	Lot # (preferred but not required)	Persistent ID / URL
LDLR	Abcam,	Ab52818	0.67 (WB)	GR3253668-2	**AB_881213**
RBM25	Abcam	Ab72237	0.2 (WB)	GR3364959-1	**AB_1270216**
Polyclonal Goat anti-Rabbit HRP Conjugated Immunoglobulins	Dako, Agilent Pathology Solutions,	P0448	0.05 (WB)	20061231	**AB_2617138**
Polyclonal Rabbit anti-Mouse HRP conjugated Immunoglobullins	Dako, Agilent Pathology Solutions	P0260	0.26 (WB)	20039216	**AB_2636929**
Polyclonal Rabbit Anti-Goat HRP conjugated Immunoglobulins	Dako, Agilent Pathology Solutions	P0449	0.1 (WB)	00062104	**AB_2617143**
HRP-conjugated anti-HA antibody	Sigma Aldrich	H6533	2 (WB)	NA	**AB_439705**
TBP	Abcam,	Ab51841	1 (WB)	GR103882-2	**AB_945758**
Beta actin	Santa Cruz Biotechnology	SC-1616	0.1 ug/ml	NA	**AB_630836**
LDLR	Santa Cruz Biotechnology	sc-18823	2 (flow cytometry)	NA	**AB_627881**
LDLR	PROGEN Biotechnik GmbH	61087	2 (flow cytometry)	207100-01	**AB_2892206**
Isotype control antibody	Santa Cruz Biotechnology)	sc-3879	2 (flow cytometry)	L0816	**AB_737262**
chicken anti-Mouse IgG (H+L) Cross-Adsorbed AlexaFluor 488-conjugated Secondary Antibody	Thermo Fischer Scientific	A-21200	4 (flow cytometry)	NA	**AB_2535786**
Goat anti-Mouse IgG (H+L) Cross-Adsorbed AlexaFluor 647-conjugated Secondary Antibody	Thermo Fischer Scientific	A-21236	4 (flow cytometry)	1915660	**AB_2535805**

DNA/cDNA Clones

Clone Name	Sequence	Source / Repository	Persistent ID / URL
LDLR ret fragment	ENST00000557958.1: nt 87-437	Cloned in house	NA
HA tagged LDLR fragment	N-terminally HA-tagged version of ENST00000557958.1: nt 87-437	Invitrogen GeneArt Strings DNA Fragments (Thermo Fischer Scientific)	NA
short LDLR minigene	hg38 chr19:11,102,285-11,102,921	Cloned in house	NA
long LDLR minigene	hg38 chr19:11,102,283-11,105,702	Cloned in house	NA
Wild type RBM25	NM_021239.3:188-2719	proteogenix	NA
RBM25 p.L17P	NM_021239.3:188-2719 (c.50T>C)	proteogenix	NA
RBM25 p.I152F	NM_021239.3:188-2719 (c.454A>T)	proteogenix	NA
RBM25 p.A455D	NM_021239.3:188-2719 (c.1364C>A)	proteogenix	NA

Cultured Cells

Name	Vendor or Source	Catalog #	Sex (F, M, or unknown)	Persistent ID / URL
Huh7	Japanese Collection or Research Bioresources Cell Bank, JCRB Cell Bank, Osaka, Japan	JCRB0403	M	**CVCL_0336**
HepG2	American Type Culture Collection, ATCC, Manassas, VA, USA	HB-8065	M	**CVCL_0027**
HEK293T	American Type Culture Collection, ATCC, Manassas, VA, USA	CRL-3216	F	**CVCL_0063**

Data & Code Availability

Description	Source / Repository	Persistent ID / URL
Homo Sapiens reference genome	Data Sciences Platform at the Broad Institute/ build GRCh38 and build GRCh37/h19	https://gatk.broadinstitute.org/hc/en-us/articles/360035890951-Human-genome-reference-builds-GRCh38-or-hg38-b37-hg19 and https://gatk.broadinstitute.org/hc/en-us/articles/360035890711-GRCh37-hg19-b37-humanG1Kv37-Human-Reference-Discrepancies
UCSC Genome Browser Variant Annotation Integrator	Variant Annotation Integrator	https://genome.ucsc.edu/cgi-bin/hgVai
U2 branchpoint prediction algorithm	SVM-BP-finder	http://regulatorygenomics.upf.edu/Software/SVM_BP/
Expression of LDLR transcripts in healthy human liver	Gene Expression Omnibus,	GSE126848/ https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126848L
SNP data from UK Biobank	UK Biobank	http://www.nealelab.is/uk-biobank
exome variants of UK Biobank	Helix Research and UK Biobank	server s3://helix-researchpublic/ukbb_exome_analysis_results/V1.3
UK10K	Patient specific data not shared. Summary data previously published in DOI: 10.1136/jmedgenet-2014-102405	https://www.uk10k.org/data_access.html
Expression analysis of spliceosome genes in various tissues	Genotype-Tissue Expression (GTEx) project	https://gtexportal.org/home
gnomAD Databank	genome aggregation database	https://gnomad.broadinstitute.org/

siRNAs and Transfection reagents

Description	Source / Repository	Catalog #	Persistent ID / URL
Lipofectamine™ RNAiMAX Transfection Reagent	Thermo Fischer Scientific	13778150	https://www.thermofisher.com/order/catalog/product/13778150#/13778150
Lipofectamine™ 2000 Transfection Reagent	Thermo Fischer Scientific	11668019	https://www.thermofisher.com/order/catalog/product/11668019#/11668019
Lipofectamine™ 3000 Transfection Reagent	Thermo Fischer Scientific	L3000015	https://www.thermofisher.com/order/catalog/product/L3000015#/L3000015
The Ambion Silencer Select Human Genome siRNA library V4	Thermo Fischer Scientific	4397926	https://www.thermofisher.com/order/catalog/product/4397926?SID=srch-hj-4397926#/4397926?SID=srch-hj-4397926
Ambion Silencer Select anti-*PLK1*	Thermo Fischer Scientific	assay ID s448, cat. 4390824	https://www.thermofisher.com/order/genome-database/details/sirna/s448?CID=&ICID=&subtype=
Ambion Silencer Select anti-*LDLR*	Thermo Fischer Scientific	assay ID s4, cat. No. 4392420	https://www.thermofisher.com/order/genome-database/details/sirna/s4?CID=&ICID=&subtype=sirna_silencer_select
Silencer™ Select Negative Control No. 1 siRNA	Thermo Fischer Scientific	4390843	https://www.thermofisher.com/order/catalog/product/4390843#/4390843
SF3A1	Dharmacon	L-016051-00-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10291
SF3A2	Dharmacon	L-018282-02-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-8175
SF3B1	Dharmacon	L-020061-01-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-23451
SF3B2	Dharmacon	L-026599-01-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10992
SF3B4	Dharmacon	L-017190-00-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10262
SF3B5	Dharmacon	L-014706-02-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-83443

siRNAs and Transfection reagents (continued)

Description	Source / Repository	Catalog #	Persistent ID / URL
SF3B6	Dharmacon	L-020260-02-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-51639
RBM22	Dharmacon	L-021186-01-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-55696
RBM25	Dharmacon	L-021976-00-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-58517
ISY1	Dharmacon	L-013894-01-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-57461
AQR	Dharmacon	L-022214-01-0005	https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-9716
Non-targeting control pool	Dharmacon	D-001810-10-05	
MISSION® predesigned siRNA against RBM25	Sigma-Aldrich	SASI_Hs02_00354878	https://www.sigmaaldrich.com/CH/en/semi-configurators/sirna?activeLink=selectAssays
MISSION® siRNA Universal Negative Control #2	Sigma-Aldrich	SIC002	https://www.sigmaaldrich.com/CH/en/product/sigma/sic002