



**HAL**  
open science

## Posttranscriptional Regulation of the Human LDL Receptor by the U2-Spliceosome

Paolo Zanoni, Grigorios Panteloglou, Alaa Othman, Joël T. Haas, Roger Meier, Antoine Rimbert, Marta Futema, Yara Abou Khalil, Simon Flyvbjerg, Norrelykke, Andrzej Jerzy Rzepiela, et al.

► **To cite this version:**

Paolo Zanoni, Grigorios Panteloglou, Alaa Othman, Joël T. Haas, Roger Meier, et al.. Posttranscriptional Regulation of the Human LDL Receptor by the U2-Spliceosome. *Circulation Research*, 2022, 130 (1), pp.80-95. 10.1161/CIRCRESAHA.120.318141 . hal-03554168

**HAL Id: hal-03554168**

**<https://hal.science/hal-03554168v1>**

Submitted on 22 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Posttranscriptional regulation of the human LDL Receptor by the U2-spliceosome

Paolo Zanoni<sup>1,2,\*</sup>, Grigorios Panteloglou<sup>1,2,\*</sup>, Alaa Othman<sup>3</sup>, Joel T. Haas<sup>4</sup>, Roger Meier<sup>5</sup>, Antoine Rimbert<sup>6</sup>, Marta Futema<sup>7</sup>, Yara Abou Khalil<sup>8,9</sup>, Simon F. Norrelykke<sup>5</sup>, Andrzej J. Rzepiela<sup>5</sup>, Szymon Stoma<sup>5</sup>, Michael Stebler<sup>5</sup>, Freerk van Dijk<sup>10</sup>, Melinde Wijers<sup>6</sup>, Justina C. Wolters<sup>6</sup>, Nawar Dalila<sup>11</sup>, Nicolette C. A. Huijckman<sup>6</sup>, Marieke Smit<sup>6</sup>, Antonio Gallo<sup>12</sup>, Valérie Carreau<sup>12</sup>, Anne Philippi<sup>13</sup>, Jean-Pierre Rabès<sup>8,14,15</sup>, Catherine Boileau<sup>8,16</sup>, Michele Visentin<sup>17</sup>, Luisa Vonghia<sup>18,19</sup>, Jonas Weyler<sup>18,19</sup>, Sven Francque<sup>18,19</sup>, An Verrijken<sup>19,20</sup>, Ann Verhaegen<sup>19,20</sup>, Luc Van Gaal<sup>19,20</sup>, Adriaan van der Graaf<sup>10</sup>, Belle V. van Rosmalen<sup>21</sup>, Jerome Robert<sup>1</sup>, Srividya Velagapudi<sup>1,2</sup>, Mustafa Yalcinkaya<sup>1,2</sup>, Michaela Keel<sup>1,2</sup>, Silvija Radosavljevic<sup>1,2</sup>, Andreas Geier<sup>22</sup>, Anne Tybjaerg-Hansen<sup>11</sup>, Mathilde Varret<sup>8</sup>, Lucia Rohrer<sup>1,2</sup>, Steve E. Humphries<sup>23</sup>, Bart Staels<sup>4</sup>, Bart van de Sluis<sup>6</sup>, Jan Albert Kuivenhoven<sup>6</sup>, and Arnold von Eckardstein<sup>#,1,2</sup>

\* These authors contributed equally to this work

### Author affiliations:

1. Institute for Clinical Chemistry, University and University Hospital Zurich, Zurich, 8091 Switzerland.
2. Center for Integrative Human Physiology, University of Zurich, Zurich, 8057 Switzerland.
3. Institute of Molecular Systems Biology, ETH Zurich, 8093, Zurich, Switzerland.
4. University of Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1011- EGID, 59000 Lille, France.
5. Scientific center for optical and electron microscopy (ScopeM), ETH Zurich, 8093, Zurich, Switzerland.
6. Department of Pediatrics, Section Molecular Genetics, University of Groningen, University Medical Center Groningen, 9713 AV Groningen, the Netherlands.
7. Cardiology Research Centre, Molecular and Clinical Sciences Research Institute, St George's, University of London, London, UK.
8. LVTS-INSERM UMRS 1148 and University of Paris, CHU Xavier Bichat, 75018, Paris, France.
9. Laboratory of Biochemistry and Molecular Therapeutics (LBTM), Faculty of Pharmacy and Pôle technologie Santé (PTS), Saint-Joseph University, 1104 2020, Beirut, Lebanon.
10. University of Groningen, University Medical Center Groningen, Department of Genetics, 9713AV, Groningen, The Netherlands
11. Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Faculty of Health and Medical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark.
12. AP-HP, Endocrinology and Metabolism Department, Human Research Nutrition Center, Pitié-Salpêtrière Hospital, 75013 Paris, France.
13. Université de Paris, Faculté de Médecine Paris-Diderot, UMR-S958, Paris, France. Current address: Université de Paris, Institut Cochin, INSERM U1016, CNRS UMR-8104, Paris, France.
14. AP-HP, Université Paris-Saclay, 75004, Paris, France.
15. UFR Simone Veil des Sciences de la Santé, UVSQ, 78180 Montigny-Le-Bretonneux, France.
16. AP-HP, Genetics Department, CHU Xavier Bichat, Université de Paris, 75018 Paris, France.
17. Department of Clinical Pharmacology and Toxicology, University Hospital Zurich, 8091, Zürich, Switzerland
18. Department of Gastroenterology and Hepatology, Antwerp University Hospital, 2650 Antwerp, Belgium.

19. Laboratory of Experimental Medicine and Paediatrics, Faculty of Medicine, University of Antwerp, 2610 Antwerp, Belgium.
20. Department of Endocrinology, Diabetology and Metabolism, Antwerp University Hospital, 2650 Edegem, Belgium.
21. Department of Surgery, Academic Medical Center, University of Amsterdam, 1105 AZ, Amsterdam, The Netherlands
22. Division of Hepatology, Department of Medicine II, University Hospital Würzburg, 97080 Würzburg, Germany.
23. Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, WC1E 6JJ London, United Kingdom.

**Actual affiliations at the moment of submission**

- P.Z. is affiliated to the Institute of Medical Genetics, University of Zürich, 8952 Schlieren, Switzerland
- A.R. is affiliated to Inserm UMR 1087 / CNRS UMR 6291 IRS-UN, 44007, Nantes, France
- M.Y. is affiliated to the Division of Molecular Medicine, Department of Medicine, Columbia University, New York, 10032, NY, USA
- S.V. is affiliated to the Center for Molecular Cardiology, University of Zurich, 8952 Schlieren, Switzerland

**Short title:** U2 spliceosome regulates LDL receptor

**Contact information:**

Arnold von Eckardstein; Institute for Clinical Chemistry, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland; +41 44 255 22 60; [arnold.voneckardstein@usz.ch](mailto:arnold.voneckardstein@usz.ch)

**Word count:** 7888 excluding title page and abstract

*Keywords:* LDL-cholesterol, LDL receptor, spliceosome, intron retention, RNA interference, genome wide screening

**1 Abstract**

2 *Background:* The low-density lipoprotein receptor (LDLR) in the liver is the major determinant of  
3 LDL-cholesterol levels in human plasma. The discovery of genes that regulate the activity of LDLR  
4 helps to identify pathomechanisms of hypercholesterolemia and novel therapeutic targets against  
5 atherosclerotic cardiovascular disease.

6 *Methods* We performed a genome-wide RNA interference screen for genes limiting the uptake of  
7 fluorescent LDL into Huh-7 hepatocarcinoma cells. Top hit genes were validated by in vitro  
8 experiments as well as analyses of datasets on gene expression and variants in human populations.

9 *Results:* The knockdown of 54 genes significantly inhibited LDL uptake. Fifteen of them encode for  
10 components or interactors of the U2-spliceosome. Knocking down any one of 11 out of 15 genes  
11 resulted in the selective retention of intron 3 of *LDLR*. The translated LDLR fragment lacks 88% of  
12 the full length LDLR and is detectable neither in non-transfected cells nor in human plasma. The  
13 hepatic expression of the intron 3 retention transcript is increased in non-alcoholic fatty liver disease  
14 as well as after bariatric surgery. Its expression in blood cells correlates with LDL-cholesterol and age.  
15 Single nucleotide polymorphisms and three rare variants of one spliceosome gene, *RBM25*, are  
16 associated with LDL-cholesterol in the population and familial hypercholesterolemia, respectively.  
17 Compared to overexpression of wild type *RBM25*, overexpression of the three rare *RBM25* mutants in  
18 Huh-7 cells led to lower LDL uptake.

19 *Conclusions:* We identified a novel mechanism of post-transcriptional regulation of LDLR activity in  
20 humans and associations of genetic variants of *RBM25* with LDL-cholesterol levels.

21

## 1 **Introduction**

2           Hypercholesterolemia is a causal and treatable risk factor of atherosclerotic cardiovascular  
3 diseases (ASCVD)<sup>1</sup>. The most important determinant of LDL-cholesterol (LDL-C) levels in plasma is  
4 the hepatic removal of circulating LDL by binding to the LDL receptor (LDLR) for subsequent  
5 endocytosis and degradation<sup>2</sup>. The expression of LDLR is tightly regulated by transcription factors,  
6 proteasomal and lysosomal degradation, endosomal recycling, and cleavage at the cell surface<sup>1,2</sup>. The  
7 unravelling of this complex regulation led to the development of drugs that effectively lower plasma  
8 levels of cholesterol and, as the consequence, risk of ASCVD<sup>1</sup>.

9           To identify novel regulators of LDL uptake into the liver, we performed an image-based genome-wide  
10 RNA interference (RNAi) screen in Huh-7 human hepatocarcinoma cells. Fifteen out of 54 genes  
11 significantly reducing LDL uptake upon knockdown encode for proteins involved in pre-mRNA  
12 splicing. The majority of them are either core components or interactors of the U2-spliceosome<sup>3</sup>. By  
13 functionally validating this finding *in vitro* as well as in human tissues, we provide evidence that a  
14 functional U2 spliceosome is needed for the expression of full length LDLR and, hence, determining  
15 LDLR activity in humans.

16

## 1 Methods

### 2 Data Availability

3 The authors declare that all data and methods supporting the findings of this study are  
4 available in the Data Supplement or from the corresponding authors on reasonable request.

5 A detailed description of materials and methods is provided in the text and Major Resources  
6 Table of the Online Supplement

## 8 Results

### 9 The U2-spliceosome and its interactors are rate-limiting for LDL endocytosis

10 For the genome wide RNAi screen of genes limiting uptake of LDL or HDL, Huh-7 human  
11 hepatocarcinoma cells were reverse-transfected using three different siRNA oligonucleotides against  
12 each of the 21,584 different human genes. To control efficacy and specificity of transfection, each  
13 plate contained wells with cells transfected with siRNAs against *PLK1* whose knockdown results in  
14 cell death, and *LDLR*, respectively. Based on results of time and dose finding experiments, the cells  
15 were exposed 72 hours post transfection to 33 µg/mL each of Atto594-labelled LDL and Atto655-  
16 HDL for 4 hours. As background controls, wells with cells transfected with a non-targeting siRNA  
17 were incubated in the absence of fluorescent lipoproteins. After washing, fixation, and staining of the  
18 nuclei with Hoechst 33258, the plates were imaged at 4x and 20x with two twin wide-field automated  
19 microscopes. Nuclei, the relative cytoplasm, and fluorescent LDL-containing vesicles were identified  
20 through automated image analysis (Figure 1A). Transfection efficiency was very high (Figure S1a).  
21 Analysis and validation of HDL image data will be subject of a separate report.

22 For the uptake of fluorescent LDL, the five best performing assay features (foci count per cell, foci  
23 mean intensity, cytoplasm granularity 1 and 2, cytoplasm median intensity) showed a high degree of  
24 correlation. Therefore and because of the widest dynamic range based on  $Z'$ -factor values from control  
25 wells, we identified gene hits by the Redundant siRNA Activity (RSA) analysis of data from the  
26 median cytoplasm intensity feature.  $Z'$ -factor values for median cytoplasm intensity in each assay  
27 plate for both the background (median 0.00, interquartile range [IQR] -0.23 to 0.20) and positive  
28 control (median -0.56, IQR -0.99 to -0.20) clustered mostly around the 0-line, indicating a suboptimal  
29 but analytically exploitable signal-to-noise ratio (Figure S1b). Dimensionality reduction of main assay  
30 features did not significantly alter the outcome (Figures S1c and S1d). At an RSA p value cut-off of  $p$   
31  $< 10^{-3}$ , interference with 54 and 37 genes decreased and increased LDL uptake, respectively (Table 1,  
32 Table S1). By contrast to the findings of a previous genome wide CRISPR-based screening in Huh7  
33 cells<sup>4</sup>, our list does not include *LDLR* or its modulators such as *SCAP*, *MBTPS1*, or *IDOL/MYLIP*  
34 except *AP2MI*, which is an essential contributor to clathrin-mediated endocytosis (Table 1). Gene  
35 Ontology (GO) enrichment analysis showed significant clustering for genes whose loss of function  
36 decreased LDL uptake (Table S2). Functional clustering of these genes with the STRING tool  
37 revealed four major groups: the ribosome (N = 7), the proteasome (N = 8), the spliceosome (N = 15),  
38 and vesicular transport (N = 5) (Figure 1B). Out of the 15 spliceosome genes, six encode for core  
39 components of the U2 spliceosome, namely *SF3A1*, *SF3A2*, *SF3B1*, *SF3B2*, *SF3B5* and *SF3B6*. Other  
40 proteins, interact with the U2-spliceosome either directly (*AQR*, *ISY1* and *RBM25*) or indirectly  
41 (*RBM22*)<sup>3</sup>.

42 To confirm the role of the U2 spliceosome in LDL endocytosis *in vitro*, we performed <sup>125</sup>I-LDL cell  
43 association assays in Huh-7 and HepG2 cells. *SF3B4* was also included in these experiments as it is  
44 part of the U2 spliceosome and barely missed the RSA p value cut-off ( $p = 1.4 \cdot 10^{-3}$ ). Knockdown was  
45 achieved using 4 pooled siRNA molecules against each hit gene acquired from vendors other than that  
46 of the siRNA screening library, namely Dharmacon or Sigma instead of Ambion. (see Major Resource  
47 Table and figure S2a). For *RBM25* we replaced Dharmacon's siRNAs with those from Sigma because  
48 of their presumable off-target effects on *LDLR* protein expression (Figure S3). Knockdown of each of  
49 these genes significantly decreased the specific cell association of <sup>125</sup>I-LDL with both Huh-7 and  
50 HepG2 cells (Figures 1C, 1D and Figure S2b). The association of <sup>125</sup>I-LDL was equally decreased by  
51 knockdown of *SF3B1* (-45±5%), *SF3A2* (-47±6%), *AQR* (-45±6%), and *LDLR* (-43±8%) (Figure 1C).

1 RNA interference with *RBM25* reduced the specific cellular association of <sup>125</sup>I-LDL and fluorescent  
 2 LDL by 27% ±8% and 52%±5%, respectively (Figure 1D and Figure S3f). Of note, the specific cell  
 3 association of <sup>125</sup>I-HDL was unaltered or even increased upon knockdown of *AQR* and *SF3A1* in either  
 4 Huh-7 or HepG2 cells (Figures S2c and S2d).

### 6 **Loss of U2-spliceosome genes and their interactors causes selective retention of *LDLR* intron 3** 7 **(IVS3)**

8 To unravel the mechanism through which the U2-spliceosome and its interactors regulate LDL  
 9 endocytosis, we applied RNA sequencing to Huh-7 cells, which were transfected with either siRNAs  
 10 against eleven U2-spliceosome genes or a non-targeting control siRNA. Sequences can be accessed by  
 11 codes PRJEB46899 and PRJEB46898 in the data bank of the European Nucleotide Archive  
 12 (<https://www.ebi.ac.uk/ena/browser/support>) 72 hours after transfection, we measured both expression  
 13 at the gene level and alternative exon usage in polyA-selected transcripts. Knockdown of all eleven  
 14 genes except *RBM25* induced a marked increase in the retention of intron 3 of *LDLR* in mature  
 15 transcripts without altering the expression of the *LDLR* full length transcript (Figure 2A, Figure S4).  
 16 This effect was confirmed in Huh-7 cells by qRT-PCR upon knockdown of *AQR*, *SF3B1*, or *RBM25*  
 17 by employing a primer set that was previously used to study the effects of the rare *LDLR* c.313+1,  
 18 G>A intronic variant, which leads to *LDLR* loss of function by constitutively promoting intron 3  
 19 retention<sup>5</sup> (Figure S5a). By contrast to the RNA sequencing (Figure S4), qRT-PCR unravelled  
 20 increased expression of the *LDLR* IVS3 retention transcript upon knockdown of *RBM25*, albeit not as  
 21 much as with knockdown of *SF3B1* and *AQR* (Figures S5b and S5c).

22 Among all intronic or exonic sequences in the transcriptome, the expression of the intron 3  
 23 retaining *LDLR* transcript was altered most strongly. Upon knockdown of *SF3B1*, *AQR*, or *SF3A2*, the  
 24 retained intronic sequence of *LDLR* ranked at the top of each respective dataset when the exon-level  
 25 expression data was plotted against each other (Figure 2B). The degree of intron 3 retention upon  
 26 knocking down U2-spliceosome genes was significantly correlated with the decrease in <sup>125</sup>I-LDL cell  
 27 association, ( $r = -0.73$ ,  $p = 1.4 \times 10^{-2}$ , Figure 2C).

28 To investigate reasons for intron 3 retention in *LDLR*, we transfected HEK293T cells with two  
 29 minigenes containing different portions of the *LDLR* genomic sequence flanked by two artificial exons  
 30 (Figure 3A). The first minigene (*MG*<sub>1</sub>) encoding only for exon 3 of *LDLR* and the adjacent intronic  
 31 regions cloned between two artificial exons (*SD*<sub>6</sub> and *SA*<sub>2</sub>), displayed very low if any RNA  
 32 sequencing reads mapping to the first ~130bp of intron 3. On the contrary, upon expression of the  
 33 whole genomic sequence between the 3'-end of intron 2 and the 5'-end of intron 4 of *LDLR* (*MG*<sub>2</sub>) an  
 34 increased number of reads mapped to the first section of intron 3. This indicates incomplete splicing of  
 35 intron 3 when the physiological exon 4 acceptor site and the branch point site (BPS) were present in  
 36 the larger minigene *MG*<sub>2</sub> (Figure 3B). The acceptor splice site of exon 4 of *LDLR* hence appears to be  
 37 poorly defined. The bioinformatic analysis of the portion of intron 3 neighbouring exon 4 by the U2  
 38 branchpoint prediction algorithm SVM-BP-finder  
 39 ([http://regulatorygenomics.upf.edu/Software/SVM\\_BP/](http://regulatorygenomics.upf.edu/Software/SVM_BP/))<sup>6</sup> identified one plausible U2-spliceosome  
 40 dependent BPS located 30 bp upstream of the acceptor site (Table S3). The *gtgat* pentamer in the  
 41 centre of the *cggtgatgg* branchpoint sequence was associated with very low U2 binding energy and  
 42 occurs at low frequency in the branchpoint database<sup>6</sup>. We discarded another predicted branchpoint 124  
 43 bp upstream of the acceptor site as the subsequent AG-exclusion zone does not reach up to the  
 44 acceptor. Contrary to exon 4 of human *LDLR*, exon 4 of murine *Ldlr* contains a strong and frequently  
 45 recurring branchpoint 33 bp upstream of the acceptor site (Figure 3C). This finding is in accordance  
 46 with intron 3 of *Ldlr* being barely detectable at the RNA level by qRT-PCR in mouse liver (data not  
 47 shown). Taken together, these data suggest that the BPS of intron 3 in human *LDLR* is poorly defined  
 48 and therefore very sensitive to alternative splicing.

### 50 **Selective intron 3 retention limits *LDLR* cell surface abundance**

51 The transcript with intron 3 retention encodes for a prematurely truncated proteoform of  
 52 *LDLR* because the 5'-end of intron 3 encodes for 12 novel amino acids followed by a stop codon.

1 Including the signal peptide, this theoretical 116 amino acid residues long and 12.7 kDa large  
 2 'LDLRret fragment' encompasses the complete first and large part of the second class A domains  
 3 (labelled as L1 and L2 in Figure 4A<sup>7</sup>) but lacks all other domains, including the transmembrane  
 4 portion of LDLR. Western blots probed with an antibody against the C-terminus of LDLR revealed  
 5 60±30% and 61±13% lower LDLR protein levels upon knockdown of *AQR* and *SF3B1*, respectively  
 6 (Figures 4B and 4C). A similar decrease in LDLR protein was seen upon knockdown of *RBM25* with  
 7 siRNAs from Sigma (-68%±10%), whereas the knockdown of *RBM25* with the siRNA of Dharmacon  
 8 led to an increase in LDLR protein (+122%±109%), presumably due to off target effects (Figures S3d  
 9 and S3e). Flow cytometry experiments on alive Huh-7 cells after *SF3B1* and *AQR* knockdown showed  
 10 a -87%±1% and -61%±4%, respectively, lower cell surface abundance of LDLR (Figure 4D). The  
 11 knockdown of *RBM25* with siRNAs from Sigma and Dharmacon decreased the cell surface abundance  
 12 of LDLR by 53%±6% and 21%±5%, respectively, as compared to non-targeting siRNAs from the  
 13 respective manufacturers (Figure S3g).

14 To investigate whether cells produce and secrete the LDLRret fragment, we overexpressed a  
 15 C-terminally HA-tagged version of the LDLRret fragment in HEK293T cells. 48 hours after  
 16 transfection, the HA-tagged LDLRret fragment was detectable in the cell lysates (Figure 4E) as well as  
 17 in undiluted cell culture media (Figure 4F). The proteasomal inhibitor MG-132 decreased cellular  
 18 LDLRret protein levels (Figure 4E) suggesting that the LDLRret fragment is not catabolized through  
 19 the proteasome. We also overexpressed an untagged version of the LDLRret fragment in HEK293T  
 20 cells. Targeted mass spectrometry recorded a peptide, which is present in both the full-length protein  
 21 and in LDLRret, over its basal endogenous level in HEK293T cell lysates (Figure S6) but not in  
 22 human plasma (data not shown).

23

### 24 **A large proportion of *LDLR* transcripts in human liver and blood cells retains intron 3**

25 To investigate its physiological or pathological relevance, we quantified *LDLR* intron 3  
 26 retention in liver biopsies as well as in peripheral blood cells by three different methods, and explored  
 27 associations with non-alcoholic fatty liver disease (NAFLD), demographic measures, lipid traits, and  
 28 therapeutic interventions.

29 qRT-PCR of mRNAs of liver tissue from 17 patients with benign liver tumours and 9 patients  
 30 with suspected NAFLD, found the *LDLR* intron 3 retention transcript expressed at considerable and  
 31 interindividually variable amounts (Figure 5A). Taking the sum of the full length and intron 3  
 32 retention transcripts of *LDLR* as the reference, 43 % (range 23% to 85%) of the transcripts retained  
 33 intron 3 (Figure 5A).

34 The bioinformatics analysis of RNA sequencing data on liver samples of 13 healthy non-obese  
 35 subjects, 12 obese subjects without NAFLD, 15 patients with NAFLD, and 15 patients with non-  
 36 alcoholic steatophepatitis (NASH) (Gene Expression Omnibus, accession number GSE126848)<sup>8</sup> found  
 37 14 different *LDLR* transcripts (Figure S7). Four transcripts showed the largest interindividual  
 38 variation, namely *LDLR-201* and *LDLR-208*, encoding full length *LDLR*, as well as *LDLR-206*, which  
 39 corresponds to the retained intron 3 transcript, and the likewise futile *LDR-214* (*LDLR* transcripts are  
 40 illustrated schematically in Figure S7a). Interestingly, the median concentration of *LDLR-206* was  
 41 substantially higher in patients with NAFLD or NASH than in normal weight or obese subjects  
 42 without NAFLD. The median percentages of *LDLR-206* reads relative to total reads from all  
 43 transcripts of *LDLR* gene increased significantly from 1.8% (range 0.7% to 4.2%) and 1.7% (0.4% to  
 44 3.7%) in normal weight and obese subjects without NAFLD, respectively, to 5.8% (1.1 to 26.7%) and  
 45 5.0% (0.9% to 29.0%) in patients with NAFLD and NASH, respectively (figure 5B). Of the two most  
 46 abundant full length encoding *LDLR* transcripts, *LDLR-208* decreased significantly (Figure 5C) while  
 47 the expression of *LDLR-201* did not change (Figure S7).

48 We also investigated the expression of *LDLR* transcripts in liver biopsies of 155 obese non-  
 49 diabetic subjects<sup>9</sup> by using Affymetrix Human Gene 2.0 ST arrays (see Table S4 for clinical and  
 50 biochemical characteristics). The signal intensities from a probe located in intron 3 of *LDLR* were  
 51 significantly higher than the other intronic *LDLR* probes located in introns 2, 4 and 15 and comparable  
 52 to probes located in coding exons (Figure 5D). The percent intensities of the IVS3 probe relative to the

1 sum of all *LDLR* probes ranged from 7.5% to 82%. Intron 3 retention correlated significantly only  
 2 with *SF3B1* ( $r = 0.26$ ,  $p = 1.4 \times 10^{-2}$ ), while no U2-spliceosome gene showed any significant correlation  
 3 with overall *LDLR* expression (Table S5). Relative intensities of neither the intron 3 probe nor any  
 4 other of the 24 *LDLR* probes showed significant correlations with plasma levels of total, HDL- or  
 5 LDL-cholesterol (Figures S8a, S8b and S8c, Table S6). Correlations with histological NAFLD stages  
 6 were inverse by trend but not statistically significant (Figure S8d). Intron 3 relative probe intensity did  
 7 not correlate with BMI (Figure S8e). However, in a subgroup of 21 patients who underwent a second  
 8 liver biopsy after bariatric surgery (median follow-up time = 13 months, IQR = [12, 15]), the  
 9 proportion of the intron 3 retention transcript relative to the full length *LDLR* transcript increased  
 10 significantly after surgery ( $p = 9.8 \times 10^{-3}$ ) (Figure S8f; Table S4). This increase was even more  
 11 pronounced in eleven patients with NASH at baseline but no NASH at follow-up ( $p = 3.6 \times 10^{-2}$ , Figure  
 12 S8g).

13 Finally, we analyzed the RNA-sequencing data in whole blood samples from 2,462 subjects of  
 14 the Dutch BIOS consortium<sup>10</sup>. The *LDLR* ENST00000557958 transcript, predicted to retain intron 3,  
 15 was detectable in all subjects and represented  $21\% \pm 7\%$  of the total *LDLR* transcripts. The  
 16 ENST00000557958 transcript levels significantly correlated with age ( $r = 0.25$ ,  $p = 9.2 \times 10^{-36}$ , Figure  
 17 6A) and less strongly with LDL-C ( $r = 0.089$ ,  $p = 3.9 \times 10^{-5}$ , Figure 6B). The latter correlation lost its  
 18 statistical significance after adjusting for age, suggesting age itself as the main driver of the  
 19 association between ENST00000557958 levels and LDL-C. ENST00000252444, the only transcript  
 20 encoding for full length *LDLR* and expressed in blood cells in all subjects in this dataset, was also  
 21 positively correlated with age ( $r = 0.19$ ,  $p = 8.8 \times 10^{-20}$ , Figure 6C) but not with LDL-C ( $r = -0.033$ ,  $p =$   
 22  $4.0 \times 10^{-1}$ , Figure 6D). Correlation of neither transcript with BMI was statistically significant.

23

#### 24 **Single Nucleotide Polymorphisms in *RBM25* are associated with lower LDL-Cholesterol**

25 The analysis of whole exome sequencing (WES) data of 40,468 UK Biobank subjects<sup>11</sup> did  
 26 not unravel any significant association between our spliceosome hit genes and LDL-C or any other  
 27 clinical lipid trait (Table S7). However, constraints data from the gnomAD database indicate a strong  
 28 intolerance to functional genetic variation for our U2-spliceosome genes, with a probability of  
 29 intolerance to loss of function (pLI)<sup>12</sup> of  $0.91 \pm 0.17$  (mean  $\pm$  SD) (Table S8). The analysis of SNPs of  
 30 11 U2-spliceosome hit genes in 361,194 participants of UK Biobank found 24 SNPs of *RBM25*  
 31 significantly associated with lower levels of LDL-C (Figure 7A) and apoB (Figure S9a).

32 In Europeans, four SNPs in introns or downstream of the *RBM25* coding sequence including  
 33 the lead SNP rs17570658 and two upstream SNPs are in almost complete LD (Figure S9b). , With  $R^2$   
 34  $> 0.8$  no other SNP of *RBM25* is in strong LD. A meta-analysis of eight studies with 455'537 samples  
 35 (<https://cvd.hugeamp.org/variant.html?variant=rs17570658>) and data of the Copenhagen City Heart  
 36 and General Population Studies<sup>13</sup> according to METAL<sup>14</sup> showed the association of rs17570658 with  
 37 LDL-C (Z Score = -4.181,  $p = 2.9 \times 10^{-5}$ , Table S9).

38 *RBM25* is widely expressed in many tissues, but expression is relatively low in liver (GTEx  
 39 <https://gtexportal.org/home/>, data not shown). rs17570658 shows strong association with *RBM25*  
 40 expression in 15 different tissues including skeletal muscle and arteries (Figure 7B) as well as adipose  
 41 and mammary tissue, lung, oesophagus, kidney, and skin. Carriers of the rare allele have higher mean  
 42 *RBM25* mRNA concentration, which is compatible with higher *LDLR* activity and lower LDL-C  
 43 levels.

44

#### 45 **Impaired LDL uptake by cells expressing rare *RBM25* mutants found in patients with** 46 **familial hypercholesterolemia**

47 In the UK10K study, *RBM25* was also among the genes identified to harbour an excess of rare  
 48 novel variants in 71 patients with familial hypercholesterolemia who are negative for mutations in  
 49 *LDLR*, *APOB* and *PCSK9*, the known FH-causing genes<sup>15</sup>. We re-analyzed the burden of variants in  
 50 the *RBM25* gene, using previously published WES data from 71 FH patients negative for mutations in  
 51 *LDLR*, *APOB* and *PCSK9*, and 56,352 European data provided by the gnomAD study<sup>12</sup>. Missense,

1 splice site, frameshift, and stop-gained variants identified by WES in both FH cases and gnomAD  
 2 were filtered to select those with  $MAF < 1.0 \times 10^{-4}$ . After filtering, three *RBM25* variants were found in  
 3 the FH cohort and 163 in the gnomAD Europeans cohort. (Table S10). Two variants, p.I152F  
 4 (c.454A>T) and p.A455D (c.1364C>A), were not found in any publicly available sequencing database  
 5 and hence appear unique to the FH cohort. The third variant, p.L17P (c.50T>C) (rs1167173761), was  
 6 found in one European individual in the gnomAD cohort ( $MAF = 9 \times 10^{-6}$ , allele count = 1/251402).  
 7 The comparison of variant numbers in FH cases vs. gnomAD using a binomial test demonstrated the  
 8 enrichment of rare variants in *RBM25* in the FH cohort ( $p = 1.0 \times 10^{-3}$ ). Within the UK10K cohort, no  
 9 other U2-spliceosome gene was found to carry a rare presumable LOF mutation.

10 We investigated the functional consequences of overexpressing the three FH-associated *RBM25*  
 11 mutants in Huh-7 cells. Overexpression of all *RBM25* constructs was confirmed by qRT-PCR (Figures  
 12 S10a and S11a) and - for wild type *RBM25* - Western blotting (Figure S10b). The overexpression of  
 13 neither wild type *RBM25* nor any *RBM25* mutant in Huh-7 cells caused significant changes in the  
 14 expression of full length or IVS3 retention transcripts of *LDLR* (Figures S10c, S10d, S11b, and S11c).  
 15 Compared to empty vector, overexpression of wild type *RBM25* in Huh-7 cells changed neither the  
 16 cell surface abundance of *LDLR* nor LDL uptake significantly (Figures S10e and S10f). Comparisons  
 17 with cells overexpressing wild type *RBM25* revealed minor decreases of *LDLR* cell surface levels  
 18 (Figure S11d) but more pronounced or even significant decreases of Atto655-LDL-uptake of cells  
 19 overexpressing the *RBM25* mutants p.L17P ( $-15\% \pm 16\%$ ), I152F ( $-23\% \pm 12\%$ ), or p.A455D ( $-$   
 20  $28\% \pm 12\%$ ,  $p = 2.6 \times 10^{-2}$ ) (Figure S11e).

## 22 Discussion

23 Through genome-wide siRNA screening, we discovered that the U2-spliceosome as well as  
 24 some interacting proteins, control *LDLR* levels and LDL uptake in liver cells by modulating the  
 25 selective retention of intron 3 of *LDLR*. The intron 3 retaining *LDLR* transcript encodes a truncated  
 26 and most probably non-functional receptor. In several cohorts of healthy individuals and patients, we  
 27 observed considerable interindividual variation of *LDLR*'s IVS3 retention in liver as well as in  
 28 peripheral blood cells. Finally, we obtained initial evidence that rare genetic variants as well as SNPs  
 29 associated with its expression levels in the U2-spliceosome-associated gene *RBM25* are related to  
 30 LDL-C levels in humans. Taken together, our findings suggest intron 3 retention of *LDLR* as a novel  
 31 mechanism regulating *LDLR* activity and thereby plasma levels of LDL-C.

32 A previous siRNA screen also found U2-spliceosome genes to limit the uptake of LDL into  
 33 EA.hy926 cells but the authors excluded them from further analysis and validation<sup>16</sup>. Basic cellular  
 34 functionality of spliceosome genes may be the reason why U2- spliceosome genes were not found by a  
 35 previous CRISPR-based screen as limiting factors for LDL uptake into Huh-7 cells<sup>4</sup>. As these authors  
 36 discussed, CRISPR-based screens may overlook genes that are essential or confer a fitness advantage  
 37 in culture, since guide RNAs targeting those genes will be progressively depleted from the pooled  
 38 population<sup>4</sup>.

39 As a preliminary mechanistic explanation, our minigene data as well as our *in silico*  
 40 predictions suggest that the BPS in intron 3 of human *LDLR* is poorly defined and thereby highly  
 41 sensitive to alterations in the activity of U2 splice factors. In this regard it is noteworthy that the rare  
 42 c.313+1, G>A intronic variant leads to loss of *LDLR* function by constitutively promoting IVS3  
 43 retention<sup>5</sup>.

44 Medina and colleagues previously found alternative splicing of *HMGCR*, *HMGCS1*, *MVK*, *PCSK9*,  
 45 and *LDLR* to be mediated by the splice protein PTBP1 and regulated by cellular cholesterol levels<sup>17</sup>.  
 46 Interestingly, PTBP1 works as an inhibitor of the U2AF splice component, and thus inhibits the  
 47 recognition of 3' splice sites by the U2-spliceosome<sup>18</sup>. However, the knockdown of *PTBP1* resulted in  
 48 very limited changes in the expression levels of the different splice forms<sup>17</sup>, especially when compared  
 49 to the drastic changes observed in our study.

50 In our *in vitro* experiments, the knockdown of several U2-spliceosome genes and the resulting  
 51 IVS3 retention compromised *LDLR* cell surface expression and LDL uptake as much as the  
 52 knockdown of *LDLR* itself. The sensitivity of our mass spectrometric analysis only allowed detection

1 of the tagged fragment after overexpression in the immortalized kidney cell line HEK293T. The  
 2 artificial construct unlike an endogenously produced protein may have escaped nonsense-mediated  
 3 decay. Nevertheless, we cannot rule out that the theoretical 116 amino acid long aminoterminal  
 4 fragment of the differentially spliced *LDLR* is expressed *in vivo* and secreted. In fact, human plasma  
 5 contains LDLR fragments, which are currently assumed to result from shedding of LDLR at the cell  
 6 surface<sup>19</sup> but may also correspond to secreted alternative splice variants.

7 The relative expression of *LDLR*'s IVS3 transcript in human liver varies strongly due to both  
 8 analytical and biological reasons, namely between 0.4% and 29% upon RNA sequencing, between  
 9 7.5% and 81% upon chip array analysis, and between 23% and 85% upon qRT-PCR. Very likely,  
 10 RNA sequencing yielded the most realistic data, because this method recorded the different LDLR  
 11 transcripts most comprehensively. The large interindividual variation of IVS3 expression recorded by  
 12 each method indicates relevant regulatory mechanisms and consequences. We made seemingly  
 13 contradictory findings on the association of IVS3 retention with NAFLD. On the one hand, the  
 14 percentage of IVS3 transcripts was significantly higher in 30 patients with NAFLD or NASH than in  
 15 25 normal weight and obese control subjects without NAFLD. On the other hand, the chip array  
 16 analysis found significant increases of IVS3 transcripts after bariatric surgery, which rather causes  
 17 regression of NAFLD. However, although causing regression of NASH, bariatric surgery may not  
 18 necessarily undo all regulatory abnormalities associated with NAFLD. In this regard it is noteworthy,  
 19 that neither RNA sequencing nor chip array analysis found any significant effect of NASH on IVS3  
 20 retention (figures 5B and S8d). Larger studies are hence needed to answer the question how NAFLD  
 21 influences the expression of functional and non-functional *LDLR* transcripts.

22 In peripheral blood cells but not in liver tissue, we found a significant correlation between  
 23 plasma LDL-C levels and the IVS3 retention *LDLR* transcript, which was stronger than the correlation  
 24 with the full-length *LDLR* transcript. Smaller sample size and narrower range of LDL-C levels but also  
 25 differences between tissues may be the reasons, why no significant correlations of LDL-C with any  
 26 hepatic LDLR transcript expression were found. However, the associations of *RBM25* SNPs with  
 27 differences in *RBM25* expression and LDL-C levels and the higher than expected prevalence of rare  
 28 *RBM25* loss of function variants in FH patients with no mutation in canonical FH genes suggest that the  
 29 regulation of *LDLR* splicing by the U2-spliceosome contributes to the determination of LDL-C levels  
 30 in humans.

31 The lack of association of hypercholesterolemia with rare variants of any other U2- spliceosome  
 32 gene may reflect their intolerance to gross variation as suggested by pLI values close to 1. Also of note,  
 33 the analysis of WES data from the UK biobank only retrieved heterozygous mutations in U2-  
 34 spliceosome genes whereas our knockdown experiments rather mimic homozygous conditions.  
 35 Opposite effects on upstream regulators of *LDLR* may be another reason why the majority of SNPs and  
 36 rare exome variants of the spliceosome genes do not show any association with LDL-C levels. The  
 37 exclusive association of LDL-C with *RBM25* variants may also indicate that *RBM25* regulates LDL-C  
 38 levels by mechanisms unrelated to the U2-spliceosome and the intron 3 retention. In fact, *RBM25* also  
 39 partakes in other spliceosomal subunits<sup>20</sup>. Of note, RNAi with *RBM25* had the weakest effects on *LDLR*  
 40 splicing and overexpression of hypercholesterolemia associated *RBM25* mutants in Huh-7 cells resulted  
 41 in lower LDL uptake without affecting the expression of the *LDLR* IVS3 transcript.

42 The correlation between ENST00000557958 expression in blood cells with age makes us  
 43 hypothesize that age-related changes in the activity of the U2-spliceosome contributes to the increase  
 44 in LDL-C that parallels ageing<sup>21</sup> but is not mechanistically understood. The functionality of the  
 45 splicing process changes with ageing<sup>22</sup>. Somatic mutations or decreased expression of splice factor  
 46 genes, notably *SF3B1* and *RBM25* have been implicated in age-related processes, including cancer<sup>22,23</sup>.  
 47 The total number of alternatively spliced genes also increases with age<sup>24</sup>. Until recently, *SIRT1* is the  
 48 only known gene involved in cholesterol metabolism and atherosclerosis<sup>25</sup> whose alternative splicing  
 49 may be disrupted with age<sup>22</sup>. One may speculate that either the epigenetic dysregulation of the activity  
 50 of splice factor genes or the accumulation of somatic loss of function variants in liver cells may  
 51 promote increases in LDL-C with age.

52 Our study has several strengths and limitations. First, our screening unravelled several novel  
 53 candidate genes that regulate hepatic LDL uptake but missed canonical LDL uptake regulating genes

1 such as *MYLIP*, *MBTPS1*, *PCSK9* or *SREBBP2*. A general reason is the not optimal signal to noise  
2 ratio of our screening. A specific reason for missing *MYLIP* or *PCSK9* is the optimization of our  
3 screening towards the discovery of loss of function effects. Second, our validation studies did not only  
4 confirm the limiting effect of U2-spliceosome genes on LDL uptake but unravelled a novel  
5 mechanism of LDL receptor regulation, namely IVS3 retention within an *LDLR* transcript which is  
6 translated into a truncated and non-functional receptor protein. In both human liver and peripheral  
7 blood cells, we demonstrate that this process happens at considerable quantity and interindividual  
8 variability, possibly influenced by aging and NAFLD. Third, *RBM25* was the only spliceosome gene  
9 affected by mutations associated with differences in LDL-C, perhaps because *RBM25* may tolerate  
10 loss of function better than other U2-spliceosome genes. However, we cannot rule out that *RBM25*  
11 affects LDL metabolism beyond or even independently of *LDLR* splicing because both knockdown of  
12 *RBM25* and overexpression of loss of function mutants associated with hypercholesterolemia exerted  
13 in Huh-7 cells stronger and more consistent effects on LDL uptake than on IVS3 retention in *LDLR*.

14 In conclusion, we identified IVS3 retention of *LDLR* upon loss of U2-spliceosome activity as  
15 a novel mechanism regulating *LDLR* activity in cells. The importance of this mechanism for the  
16 regulation of plasma LDL-C levels and thus determination of cardiovascular risk remains to be  
17 established by further studies.

18

## 1 **Acknowledgements and Sources of Funding**

2 This work was conducted as part of the “TransCard” project of the 7<sup>th</sup> Framework Program (FP7)  
3 granted by the European Commission, to J.A.K., A.T.H. and A.v.E. (number 603091) as well as  
4 partially the FP7 “RESOLVE” project (to J.T.H., B.S., A.V., S.F. L.V.G., and A.v.E.) and the  
5 European Genomic Institute for Diabetes (EGID, ANR-10-LABX-46 to BS).

6 Additional work by A.v.E.’s team was funded by the Swiss National Science Foundation (31003A-  
7 160126, 310030-185109) and the Swiss Systems X program (2014/267 (MRD) HDL-X).

8 P.Z. received funding awards from the Swiss Atherosclerosis Society (AGLA and the DACH Society  
9 for Prevention of Cardiovascular Diseases.)

10 G.P. received funding from the University of Zurich (Forschungskredit, grant no. FK-20-037).

11 J.A.K. is an Established Investigator from the Dutch Heart Foundation (2015T068).

12 J.A.K. was also supported by GeniusII (CVON2017-2020).

13 The L.V.T.S. team is supported by Fondation Maladies Rares, PHRC (AOM06024), and the national  
14 project CHOPIN (CHolesterol Personalized Innovation), granted by the Agence Nationale de la  
15 Recherche (ANR-16-RHUS-0007).

16 Y.A.K. is supported by a grant from Ministère de l’Education Nationale et de la Technologie (France).

17 J.T.H. was supported by an EMBO Long Term Fellowship (ALTF277-2014). B.S. is a recipient of an  
18 ERC Advanced Grant (no. 694717). Both are also supported by PreciNASH (ANR 16-RHUS-0006).

19 Research at the Antwerp University Hospital was supported by the European Union: FP6 (HEPADIP  
20 Contract LSHM-CT-2005-018734)

21 S. F. has a senior clinical research fellowship from the Fund for Scientific Research (FWO) Flanders  
22 (1802154N).

23 S.E.H. received grants RG3008 and PG008/08 from the British Heart Foundation, and the support of  
24 the UCLH NIHR BRC. S.E.H. directs the UK Children’s FH Register which has been supported by a  
25 grant from Pfizer (24052829) given by the International Atherosclerosis Society.

26 We acknowledge the use of data from BIOS-consortium ([http://wiki.bbmri.nl/wiki/BIOS\\_bios](http://wiki.bbmri.nl/wiki/BIOS_bios)) which  
27 is funded by BBMRI-NL (NWO project 184.021.007).

28 Flow cytometry was performed with equipment of the flow cytometry facility, University of Zurich.

29

1 **Disclosures**

2 The RNAi screening was performed at the ETH ScopeM facility (R.M., M.S., S.F.N., A.J.R, S.S.). As  
3 by contract, one third of the service costs had to be paid by grants of A.v.E to cover part of the costs  
4 for personnel, infrastructure, and maintenance. S.E.H. received fees for Advisory Boards of Novartis  
5 and Amryt and is the Medical Director of a UCL spin-out company StoreGene that offers to clinicians  
6 genetic testing for patients with FH.

7

1	<b>Supplemental Materials</b>
2	Materials & Methods
3	Online Figures S1 to S11
4	Online Tables S1 to S12
5	Major Resource Table
6	Original Western Blots
7	
8	

1 **Table 1. Hit genes that induced upon knockdown in Huh-7 cells either a decrease (left column)**  
 2 **or an increase (right column) in LDL uptake.**

Decreased LDL uptake				Increased LDL uptake			
Gene	Assay score <sup>A</sup> avg <sup>B</sup>	Assay score <sup>A</sup> SEM <sup>C</sup>	RSA p-value <sup>D</sup>	Gene	Assay score <sup>A</sup> avg <sup>B</sup>	Assay score <sup>A</sup> SEM <sup>C</sup>	RSA p-value <sup>D</sup>
<i>AP2M1</i>	-3.103179681	0.346648222	3.36*10 <sup>-8</sup>	<i>PROX1</i>	6.53057396	0.631260417	3.19*10 <sup>-9</sup>
<i>CHMP2A</i>	-3.130900347	0.359445533	2.51*10 <sup>-7</sup>	<i>ITGAV</i>	7.431175355	1.519558432	2.96*10 <sup>-8</sup>
<i>NFKB2</i>	-2.59157417	0.136886566	8.07*10 <sup>-7</sup>	<i>TGFBR1</i>	3.464028514	0.397588943	7.31*10 <sup>-6</sup>
<b><u>AQR</u></b> <sup>E</sup>	-2.484868551	0.199482589	4.57*10 <sup>-6</sup>	<i>CDC37</i>	3.747034032	1.072191825	2.35*10 <sup>-5</sup>
<i>PSMD11</i>	-2.557101583	0.239773488	4.77*10 <sup>-6</sup>	<i>DTNBP1</i>	57.92944887	57.2617451	4.46*10 <sup>-5</sup>
<b><u>SF3B2</u></b>	-2.107311389	0.015210399	4.81*10 <sup>-6</sup>	<i>CYP27C1</i>	32.06817221	31.61438081	8.92*10 <sup>-5</sup>
<i>RPL35</i>	-2.346954606	0.150946677	5.45*10 <sup>-6</sup>	<i>PNPLA2</i>	2.279278207	0.266420784	1.26*10 <sup>-4</sup>
<i>PSMD8</i>	-2.988677308	0.491086915	6.34*10 <sup>-6</sup>	<i>C22orf39</i>	7.995494448	8.342242785	1.78*10 <sup>-4</sup>
<b><u>SON</u></b>	-2.164748153	0.201099955	1.46*10 <sup>-5</sup>	<i>TMEM133</i>	3.049034762	1.060165442	1.84*10 <sup>-4</sup>
<i>COPA</i>	-2.307675328	0.213018879	1.61*10 <sup>-5</sup>	<i>TMEM130</i>	6.466491664	5.317700355	2.23*10 <sup>-4</sup>
<b><u>RBM25</u></b>	-1.993998657	0.055265194	1.92*10 <sup>-5</sup>	<i>PM20D2</i>	2.155202336	0.176491898	2.29*10 <sup>-4</sup>
<b><u>RBM22</u></b>	-2.818121291	0.622885617	3.36*10 <sup>-5</sup>	<i>PET117</i>	3.001069341	1.652765441	2.68*10 <sup>-4</sup>
<i>PSMD3</i>	-2.21302629	0.224903034	3.98*10 <sup>-5</sup>	<i>CWF19L2</i>	3.806757511	4.571696977	3.12*10 <sup>-4</sup>
<b><u>SF3B5</u></b>	-2.285064158	0.25878823	4.32*10 <sup>-5</sup>	<i>ENY2</i>	2.420424347	0.514153659	3.28*10 <sup>-4</sup>
<b><u>SF3B1</u></b>	-2.267932169	0.253122099	4.55*10 <sup>-5</sup>	<i>NME4</i>	2.711491413	0.954425612	3.39*10 <sup>-4</sup>
<i>SALL4</i>	-1.937993523	1.13065979	6.02*10 <sup>-5</sup>	<i>ZC3H4</i>	4.545156994	3.551478266	3.57*10 <sup>-4</sup>
<i>RPL5</i>	-2.106905493	0.373542859	7.40*10 <sup>-5</sup>	<i>WASF2</i>	2.310874515	0.449202822	3.61*10 <sup>-4</sup>
<i>CCDC180</i>	-1.132459235	1.398333381	9.52*10 <sup>-5</sup>	<i>HELZ2</i>	2.546828237	0.984740435	3.87*10 <sup>-4</sup>
<b><u>SF3B6</u></b>	-2.277000896	0.332074616	9.83*10 <sup>-5</sup>	<i>RILP</i>	1.995550072	0.267567916	4.23*10 <sup>-4</sup>
<i>HNRNPU</i>	-1.724036435	0.093847304	1.23*10 <sup>-4</sup>	<i>MAT2A</i>	3.705559066	3.611891772	4.91*10 <sup>-4</sup>
<i>RPL17</i>	-2.226845162	0.329575956	1.46*10 <sup>-4</sup>	<i>NRM</i>	1.710898743	0.050727817	5.02*10 <sup>-4</sup>
<b><u>ISY1</u></b>	-2.74487386	0.698388989	1.55*10 <sup>-4</sup>	<i>CEP295NL</i>	2.189792071	0.474598108	5.02*10 <sup>-4</sup>
<i>ZNF641</i>	-1.034460324	1.453444444	2.58*10 <sup>-4</sup>	<i>ACSM2A</i>	2.207444199	1.531809937	5.32*10 <sup>-4</sup>
<i>COPB1</i>	-1.693933632	0.103029465	2.64*10 <sup>-4</sup>	<i>RTL9</i>	3.759306708	3.473297986	5.35*10 <sup>-4</sup>
<b><u>SF3A1</u></b>	-2.225755015	0.412106586	2.72*10 <sup>-4</sup>	<i>KIAA1522</i>	3.362058267	3.27466253	6.25*10 <sup>-4</sup>
<b><u>SNW1</u></b>	-1.76539067	0.142531611	2.76*10 <sup>-4</sup>	<i>ZNF84</i>	2.204388764	0.765657329	6.55*10 <sup>-4</sup>
<i>EIF2S1</i>	-1.486721463	0.790741651	3.45*10 <sup>-4</sup>	<i>TFAP4</i>	3.032765033	3.340175625	6.69*10 <sup>-4</sup>
<i>CCDC73</i>	-1.041204586	1.27682775	3.50*10 <sup>-4</sup>	<i>TMEM182</i>	3.227517874	1.666669492	7.29*10 <sup>-4</sup>
<i>RPL9</i>	-1.715182797	0.249911985	3.55*10 <sup>-4</sup>	<i>WDR55</i>	1.967286849	1.365170916	7.32*10 <sup>-4</sup>
<i>NXNL2</i>	-1.199311468	1.135835784	3.83*10 <sup>-4</sup>	<i>DYNLL1</i>	2.268266743	0.467997927	7.72*10 <sup>-4</sup>
<b><u>WBPI1</u></b>	-1.50591484	0.062444555	4.03*10 <sup>-4</sup>	<i>ADPRHL2</i>	2.078229013	0.322800093	8.51*10 <sup>-4</sup>
<i>C2CD5</i>	-1.097951788	1.954971449	4.46*10 <sup>-4</sup>	<i>ELAVL1</i>	1.945364959	0.968117905	8.70*10 <sup>-4</sup>
<i>RPL21</i>	-1.655773242	0.156797718	4.72*10 <sup>-4</sup>	<i>CFAP298</i>	1.883199038	0.378258022	8.87*10 <sup>-4</sup>
<i>EPOP</i>	-1.837314819	0.25795876	4.80*10 <sup>-4</sup>	<i>PMM1</i>	2.80926863	3.200260012	8.92*10 <sup>-4</sup>
<i>RMND5B</i>	-1.523957521	0.076773849	5.07*10 <sup>-4</sup>	<i>CASKIN2</i>	1.681223061	0.149986926	9.07*10 <sup>-4</sup>
<i>TAPBPL</i>	-1.52965773	0.154207886	5.27*10 <sup>-4</sup>	<i>CIZ1</i>	3.454694336	2.803876145	9.37*10 <sup>-4</sup>
<i>STARD10</i>	-1.527795273	0.115135889	5.45*10 <sup>-4</sup>	<i>BRICD5</i>	1.962503862	0.408074057	9.41*10 <sup>-4</sup>
<i>PSMD1</i>	-2.207116523	0.551747426	5.63*10 <sup>-4</sup>				
<i>PFDN6</i>	-0.881689024	1.740601376	5.80*10 <sup>-4</sup>				

<i>PSMA1</i>	-1.528976301	0.119805079	5.85*10 <sup>-4</sup>				
<i>RTF2</i>	-1.573924771	0.169765686	6.14*10 <sup>-4</sup>				
<b><u>LSM2</u></b>	-1.448888015	0.056454941	6.40*10 <sup>-4</sup>				
<i>UBD</i>	-1.171691024	1.530009178	6.69*10 <sup>-4</sup>				
<i>LRRC14</i>	-1.258311764	1.067910962	6.84*10 <sup>-4</sup>				
<b><u>SUPT6H</u></b>	-1.451332382	0.095214513	7.27*10 <sup>-4</sup>				
<i>COPB2</i>	-2.037140764	0.468882876	7.34*10 <sup>-4</sup>				
<b><u>SF3A2</u></b>	-1.347147433	0.758462926	7.89*10 <sup>-4</sup>				
<i>ATP6V0C</i>	-1.823918839	0.263639476	7.90*10 <sup>-4</sup>				
<i>EMILIN3</i>	-1.598631472	2.238859705	8.03*10 <sup>-4</sup>				
<i>DMTN</i>	-1.559252376	0.142024687	8.20*10 <sup>-4</sup>				
<i>MRPL19</i>	-0.755460842	1.688052373	8.92*10 <sup>-4</sup>				
<i>MRO</i>	-0.986783025	1.102624895	9.14*10 <sup>-4</sup>				
<i>DDX59</i>	-1.380513222	1.040634076	9.25*10 <sup>-4</sup>				
<i>PSMD12</i>	-1.761325035	0.367766123	9.45*10 <sup>-4</sup>				

1 <sup>A</sup>Assay score: normalized score for the median cytoplasm intensity assay feature. <sup>B</sup>Avg = average.

2 <sup>C</sup>SEM: standard error of the mean.

3 <sup>D</sup> p-values are not adjusted for multiple testing ( $p < 3.6 \times 10^{-6}$  after Bonferroni adjustment for 14'000  
4 genes with expressed transcripts)

5 <sup>E</sup> The 15 hit genes involved in RNA splicing and validated in this study are highlighted in bold and  
6 underlined.

7

8

9

10

11

1 **Figure Legends**

2 **Figure 1. Identification and validation of U2-spliceosome genes as limiting factors for the uptake**  
 3 **of LDL by Huh-7 cells. A. Schematic representation** of the genome-wide image-based siRNA  
 4 screening and data analysis process. **B. Functional association networks** for genes decreasing LDL  
 5 uptake upon siRNA-mediated knockdown. Genes with  $P < 1.0 \times 10^{-3}$  for median cytoplasm intensity were  
 6 selected as top hits. Spheres represent single genes. Edges represent known and predicted gene-gene  
 7 relationships such as protein-protein interactions, co-expression and homology. The graph was produced  
 8 using the STRING online tool (<http://string-db.org/>). The superimposed coloured circles are used to  
 9 highlight the main functional clusters. **C and D. Effects of RNA-interference with U2-spliceosome**  
 10 **genes on cell association of  $^{125}$ I-LDL in Huh-7 cells.** 72 hours after transfection with siRNAs from  
 11 Ambion (*LDLR*), Sigma (*RBM25*), or Dharmacon (all other genes), cells were incubated for 2 hours at  
 12 37°C in the presence of 33.3 µg/mL of  $^{125}$ I-LDL in the presence or absence of 40x excess unlabelled  
 13 LDL. Specific cell association was calculated as the difference between the two conditions. The data are  
 14 expressed as means  $\pm$ SD of 2 quadruplicate experiments. Statistical analysis was performed using  
 15 Kruskal-Wallis test with Dunn's multiple comparisons test between the non-targeting (scrambled) and  
 16 each targeting siRNA (C) or Mann-Whitney test (one-sided) between each vendor's targeting and non-  
 17 targeting (scrambled) siRNAs (D). The respective p values are shown above each condition.

18  
 19 **Figure 2. Loss of U2-spliceosome genes causes intron 3 retention in LDLR. A. LDLR Exon level**  
 20 **expression upon AQR knockdown.** Expression of the *LDLR* exons was recorded by RNA sequencing  
 21 of Huh-7 cells 72 hours after knockdown of *AQR*. Segments represent differential exon usage in each  
 22 sector of the *LDLR* genomic sequence as identified by the DEXSeq algorithm and as summarized in the  
 23 linear representation below the graph. Canonical exons of the ENST00000252444 full length transcript  
 24 are shown below the graph. Normalized read counts are reported on the y axis. The black arrow indicates  
 25 the location of ENSG00000130164:E009, corresponding to the first half of intron 3. Data represent the  
 26 average of three independent experiments. **B. ENSG00000130164:E009 is most strongly upregulated**  
 27 **upon RNA interference with spliceosome genes.** Log<sub>2</sub> fold change in gene expression at the exon  
 28 level for the whole transcriptome after knockdown of *AQR* (x axis), and *SF3B1* (y axis) and *SF3A2* (z  
 29 axis) in Huh-7 cells. The red circle highlights the position of ENSG00000130164:E009 corresponding  
 30 to the first half of intron 3. **C. Correlation between LDLR intron 3 retention and LDL cell**  
 31 **association.** Correlation between the log<sub>2</sub> fold change in ENSG00000130164:E009 expression level  
 32 and the decrease in  $^{125}$ I-LDL cell association (same data as in Figure 1C) upon knockdown of each U2-  
 33 spliceosome hit gene. Cells treated with a non-targeting siRNA were used as reference. Cell association  
 34 is expressed as mean $\pm$ SD. r and p-value were calculated according to Spearman.

35  
 36 **Figure 3. Determination of LDLR intron 3 splice patterns. A. Cloning strategy and structure of**  
 37 **the minigenes.** The upper part of this panel shows the genomic location of the two segments of the  
 38 *LDLR* gene that were cloned in each minigene, while the lower half shows a simplified structure of the  
 39 pSPL3 minigene used to express them. Genomic coordinates refer to the hg19 assembly. Note that, due  
 40 to primer design, MG<sub>1</sub> is 1 bp shorter at its 5' end, starting at chr19:11,212,960. **B. Characterization**  
 41 **of the splice products.** The graphs represent the mean RNA sequencing coverage at the Exon 3-Intron  
 42 3 junction in two replicate samples for each condition. Coverage data were normalized to the average  
 43 coverage for exon 3. MG<sub>1</sub>/MG<sub>2</sub> = short / long minigene. **C. In silico BPS predictions for the acceptor**  
 44 **site of LDLR exon 4.** BP score: final score (svm\_score) according to the SVM-BP-finder algorithm for  
 45 the putative BPS sequence highlighted in red. A BPS is considered valid when located close to the AG  
 46 exclusion zone, with BP-score > 0 and with svm\_score > 0.

47  
 48 **Figure 4. Effect of loss of spliceosome function on LDLR protein expression A. Schematic**  
 49 **structure of the LDLR protein.** (modified from<sup>7</sup>). LDLRret: intron 3 retention fragment, LBD: Ligand  
 50 binding domain; L1-L7: LDLR class A domain; EGFPH: Epidermal growth factor precursor homology  
 51 domain;  $\beta$ : beta propeller; O: O-linked sugar repeat; A/B/C: EGF-type repeat; TM: transmembrane  
 52 domain. The red line represents the location of the last canonical amino acid found also in the LDLRret  
 53 fragment, followed by 12 novel amino acids and by a stop codon. **B and C. Effect of SF3B1 and AQR**  
 54 **knockdown on LDLR protein levels.** LDLR protein levels in Huh-7 cells 72 hours after *SF3B1* or *AQR*

1 knockdown. **B** shows a representative Western blot. **C**, shows the relative densities of LDLR bands  
 2 normalized to TATA-binding-protein (TBP, loading control) after knockdown of *AQR* or *SF3B1* relative  
 3 to the non-targeting control. The data are shown as means  $\pm$ SD of 3 independent experiments. **D. Effect**  
 4 **of *SF3B1* and *AQR* knockdown on LDLR cell surface levels.** LDLR cell surface levels in alive Huh-  
 5 7 cells were measured by flow cytometry 72 hours after knockdown of *SF3B1* or *AQR*. siRNAs against  
 6 *LDLR* were used as positive controls. The data are normalized to a non-targeting control and are shown  
 7 as means $\pm$ SD of 3 independent experiments. Numbers in **C** and **D** are p-values obtained by Kruskal-  
 8 Wallis test with Dunn's multiple comparisons test between the non-targeting (scrambled) and respective  
 9 targeting siRNA. **E-F. Overexpressed LDLRret fragment is retrieved in cell lysates and cell culture**  
 10 **medium.** 48 hours after transfection in HEK293T cells, the HA-tagged version of the LDLRret fragment  
 11 was detected-by western blot in both total cell lysates (**E**, **F**) and media (**F**). Lysates after 2 and more  
 12 hours of incubation were obtained after treatment with the proteasome inhibitor MG132 as indicated by  
 13 the labels in (**E**). EV = pcDNA3.1 empty vector. HA-frag = hemagglutinin-tagged LDLRret fragment.  
 14

15 **Figure 5. LDLR intron 3 retention in human liver. A. Detection of intron 3 retention in human**  
 16 **liver by qRT-PCR.** Transcripts encoding full-length *LDLR* or the IVS3 retention variant were  
 17 measured by qRT-PCR and normalized to GAPDH mRNA levels in healthy liver tissue of 17 patients  
 18 with benign liver tumours and in liver biopsies of 9 patients with suspected NAFLD. Each bar shows  
 19 the relative expression of the two *LDLR* transcripts in one subject. **B and C. Percent expression of**  
 20 **the *LDLR* transcript *LDLR-206* with retention of intron 3 (B) and a full length *LDLR* transcript**  
 21 ***LDLR-208* (C) relative to the sum of all 14 *LDLR* transcripts in livers of 13 healthy subjects or 12**  
 22 **obese patients without non-alcoholic fatty liver disease (NAFLD), 15 patients with NAFLD and**  
 23 **15 patients with non-alcoholic steatohepatitis (NASH).** Computational analysis of previously  
 24 published RNA sequencing data (Gene Expression Omnibus, accession number GSE126848)<sup>8</sup>. For all  
 25 transcripts, see figure S7. The dark and light blue lines within the violin plots represent means and  
 26 medians, respectively. Numbers indicate p values obtained by comparisons of indicated groups using  
 27 the Kruskal-Wallis test and adjusted for multiple testing using the Bonferroni correction. **D.**  
 28 **Expression of *LDLR* exons and introns in human liver.** The violin plots show the normalized signal  
 29 intensities for probes mapping to the 5'-UTR, 3'-UTR, the exons and some introns of the *LDLR* gene  
 30 in 155 obese non-diabetic subjects. Dots indicate median values. Error bars span from the 2.5<sup>th</sup> to the  
 31 97.5<sup>th</sup> percentile. Intron 3 is highlighted in red while the other introns are shown in grey. The location  
 32 of each probe is depicted in the diagram below.  
 33

34 **Figure 6. Correlations of the *LDLR* ENST00000557958 (A, B) and ENST00000252444 transcripts**  
 35 **(C, D) in whole blood samples with age (A, C) and LDL-C levels (B, D).** Data is from 2,462 subjects  
 36 of the BIOS population<sup>10</sup>. ENST00000557958 represents the intron 3 retention transcript (**A**, **B**).  
 37 ENST00000252444 (**C**, **D**) was the only full length *LDLR* transcript detected in all samples analysed. r  
 38 values and p values (adjusted for multiple testing using the Bonferroni correction) refer to a Spearman  
 39 correlation analysis. Linear regression lines and their 95% confidence intervals are shown in blue and  
 40 gray, respectively.  
 41

42 **Figure 7. Association between *RBM25* variants and LDL-C in the UK Biobank dataset. A.**  
 43 **Association of GWAS SNPs from 11 spliceosome genes with LDL-C in the UK Biobank dataset.**  
 44 The dashed red horizontal line indicates the threshold for statistical significance after Bonferroni  
 45 correction for multiple testing of 1,360 variants within the genes of interest ( $p=3.7*10^{-5}$ ). Effect size and  
 46 directionality are reported on the x axis as beta value. **B. Association between the rs17570658**  
 47 **genotype and *RBM25* expression in different tissues.** Data shown for skeletal muscle and tibial artery  
 48 (both empirical  $p < 1.0*10^{-8}$  corrected for multiple testing across genes using Storey's q value  
 49 method<sup>26,27</sup>). The horizontal white lines reflect medians; the upper and lower borders of the grey boxes  
 50 reflect the 75<sup>th</sup> and 25<sup>th</sup> percentiles, respectively.  
 51  
 52

## 1 References

- 2 1. Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, Hegele RA, Krauss  
3 RM, Raal FJ, Schunkert H, et al. Low-density lipoproteins cause atherosclerotic cardiovascular  
4 disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement  
5 from the European Atherosclerosis Society Consensus Panel. *Eur. Heart J.* 2017;38:2459–2472.
- 6 2. Zanon P, Velagapudi S, Yalcinkaya M, Rohrer L, von Eckardstein A. Endocytosis of  
7 lipoproteins. *Atherosclerosis.* 2018;275:273–295.
- 8 3. Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome [Internet]. *Annu.*  
9 *Rev. Biochem.* 2020;89:359–388.
- 10 4. Emmer BT, Sherman EJ, Lascuna PJ, Graham SE, Willer CJ, Ginsburg D. Genome-scale  
11 CRISPR screening for modifiers of cellular LDL uptake. *PLoS Genet.* 2021;17:e1009285.
- 12 5. Cameron J, Holla ØL, Kulseth MA, Leren TP, Berge KE. Splice-site mutation c.313+1, G>A  
13 in intron 3 of the LDL receptor gene results in transcripts with skipping of exon 3 and inclusion  
14 of intron 3. *Clin. Chim. Acta.* 2009;403:131–135.
- 15 6. Corvelo A, Hallegger M, Smith CWJ, Eyraas E. Genome-wide association between branch point  
16 properties and alternative splicing. *PLoS Comput. Biol.* 2010;6:e1001016.
- 17 7. Surdo P Lo, Bottomley MJ, Calzetta A, Settembre EC, Cirillo A, Pandit S, Ni YG, Hubbard B,  
18 Sitlani A, Carfi A. Mechanistic implications for LDL receptor degradation from the  
19 PCSK9/LDLR structure at neutral pH. *EMBO Rep.* 2011;12:1300–1305.
- 20 8. Suppli MP, Rigbolt KT, Veidal SS, Heebøll S, Eriksen PL, Demant M, Bagger JI, Nielsen  
21 JC, Oró D, Thrane SW, et al. Hepatic transcriptome signatures in patients with varying degrees  
22 of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am. J.*  
23 *Physiol. - Gastrointest. Liver Physiol.* 2019;316:G462–G472.
- 24 9. Lefebvre P, Lalloyer F, Baugé E, Pawlak M, Gheeraert C, Dehondt H, Vanhoutte J, Woitrain  
25 E, Hennuyer N, Mazuy C, et al. Interspecies NASH disease activity whole-genome profiling  
26 identifies a fibrogenic role of PPAR $\alpha$ -regulated dermatopontin. *J. Clin. Invest.* 2017;2:1–17.
- 27 10. Zhernakova D V., Le TH, Kurilshikov A, Atanasovska B, Bonder MJ, Sanna S, Claringbould  
28 A, Vösa U, Deelen P, Franke L, et al. Individual variations in cardiovascular-disease-related  
29 protein levels are driven by genetics and gut microbiome. *Nat. Genet.* 2018;50:1524–1532.
- 30 11. Cirulli ET, White S, Read RW, Elhanan G, Metcalf WJ, Tanudjaja F, Fath DM, Sandoval E,  
31 Isaksson M, Schlauch KA, et al. Genome-wide rare variant analysis for thousands of  
32 phenotypes in over 70,000 exomes from two cohorts. *Nat. Commun.* 2020;11:542.
- 33 12. Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH,  
34 Ware JS, Hill AJ, Cummings BB, et al. Analysis of protein-coding genetic variation in 60,706  
35 humans. *Nature.* 2016;536:285–291.
- 36 13. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam  
37 D, Alves AC, et al. Exome-wide association study of plasma lipids in >300,000 individuals.  
38 *Nat. Genet.* 2017;49:1758–1766.
- 39 14. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide  
40 association scans. *Bioinformatics.* 2010;26:2190–2191.
- 41 15. Futema M, Plagnol V, Li K, Whittall RA, Neil HAW, Seed M, Bertolini S, Calandra S,  
42 Descamps OS, Graham CA, et al. Whole exome sequencing of familial hypercholesterolaemia  
43 patients negative for LDLR / APOB / PCSK9 mutations. *J. Med. Genet.* 2014;51:537–544.
- 44 16. Kraehling JR, Chidlow JH, Rajagopal C, Sugiyama MG, Fowler JW, Lee MY, Zhang X,  
45 Ramírez CM, Park EJ, Tao B, et al. Genome-wide RNAi screen reveals ALK1 mediates LDL  
46 uptake and transcytosis in endothelial cells. *Nat. Commun.* 2016;7:13516.
- 47 17. Medina MW, Krauss RM. Alternative splicing in the regulation of cholesterol homeostasis.  
48 *Curr. Opin. Lipidol.* 2013;24:147–52.
- 49 18. Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, Zhou J, Qiu J, Jiang L, Li H, et al.  
50 Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human  
51 genome. *Nat. Struct. Mol. Biol.* 2014;21:997–1005.
- 52 19. Mayne J, Ooi TC, Tepliakova L, Seebun D, Walker K, Mohottalage D, Ning Z, Abujrad H,  
53 Mbikay M, Wassef H, et al. Associations between Soluble LDLR and Lipoproteins in a White  
54 Cohort and the Effect of PCSK9 Loss-of-Function. *J. Clin. Endocrinol. Metab.*  
55 2018;103:3486–3495.

- 1 20. Carlson SM, Soulette CM, Yang Z, Elias JE, Brooks AN, Gozani O. RBM25 is a global  
2 splicing factor promoting inclusion of alternatively spliced exons and is itself regulated by  
3 lysine mono-methylation. *J. Biol. Chem.* 2017;292:13381–13390.
- 4 21. Balder JW, de Vries JK, Nolte IM, Lansberg PJ, Kuivenhoven JA, Kamphuisen PW. Lipid and  
5 lipoprotein reference values from 133,450 Dutch Lifelines participants: Age- and gender-  
6 specific baseline lipid values and percentiles. *J. Clin. Lipidol.* 2017;11:1055-1064.e6.
- 7 22. Deschênes M, Chabot B. The emerging role of alternative splicing in senescence and aging  
8 [Internet]. *Aging Cell.* 2017;16:918–933.
- 9 23. Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Buonamici S, Yu L,  
10 Caesar-Johnson SJ, et al. Somatic Mutational Landscape of Splicing Factor Genes and Their  
11 Functional Consequences across 33 Cancer Types. *Cell Rep.* 2018;23:282-296.e4.
- 12 24. Rodríguez SA, Grochová D, Mckenna T, Borate B, Trivedi NS, Erdos MR, Eriksson M. Global  
13 genome splicing analysis reveals an increased number of alternatively spliced genes with  
14 aging. *Aging Cell.* 2016;15:267–278.
- 15 25. Miranda MX, Van Tits LJ, Lohmann C, Arsiwala T, Winnik S, Tailleux A, Stein S, Gomes AP,  
16 Suri V, Ellis JL, et al. The Sirt1 activator SRT3025 provides atheroprotection in Apoe<sup>-/-</sup> mice  
17 by reducing hepatic Pcsk9 secretion and enhancing Ldlr expression. *Eur. Heart J.* 2015;36:51–  
18 59.
- 19 26. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park YS, Parsana P,  
20 Segrè A V., et al. Genetic effects on gene expression across human tissues. *Nature.*  
21 2017;550:204–213.
- 22 27. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.*  
23 *U. S. A.* 2003;100:9440–9445.
- 24 28. Havel RJ, Eder HA, Bragdon JH. The distribution and chemical composition of  
25 ultracentrifugally separated lipoproteins in human serum. *J. Clin. Invest.* 1955;34:1345–1353.
- 26 29. Velagapudi S, Yalcinkaya M, Piemontese A, Meier R, Nørrelykke SF, Perisa D, Rzepiela A,  
27 Stebler M, Stoma S, Zaroni P, et al. VEGF-A regulates cellular localization of SR-BI as well  
28 as transendothelial transport of HDL but Not LDL. *Arterioscler. Thromb. Vasc. Biol.*  
29 2017;37:794–803.
- 30 30. McFarlane AS. Efficient trace-labelling of proteins with iodine. *Nature.* 1958;182:53.
- 31 31. Kamentsky L, Jones TR, Fraser A, Bray MA, Logan DJ, Madden KL, Ljosa V, Rueden C,  
32 Eliceiri KW, Carpenter AE. Improved structure, function and compatibility for cellprofiler:  
33 Modular high-throughput image analysis software. *Bioinformatics.* 2011;27:1179–1180.
- 34 32. Parsons BD, Schindler A, Evans DH, Foley E. A Direct Phenotypic Comparison of siRNA  
35 Pools and Multiple Individual Duplexes in a Functional Assay. *PLoS One.* 2009;4:e8471.
- 36 33. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data.  
37 *Bioinformatics.* 2014;30:2114–2120.
- 38 34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,  
39 Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
- 40 35. Liao Y, Smyth GK, Shi W. The Subread aligner: Fast, accurate and scalable read mapping by  
41 seed-and-vote. *Nucleic Acids Res.* 2013;41:e108.
- 42 36. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data.  
43 *Genome Res.* 2012;22:2008–2017.
- 44 37. Fedoseienko A, Wijers M, Wolters JC, Dekker D, Smit M, Huijkman N, Kloosterhuis N, Klug  
45 H, Schepers A, van Dijk KW, et al. The COMMD family regulates plasma LDL levels and  
46 attenuates atherosclerosis through stabilizing the CCC complex in endosomal LDLR trafficking.  
47 *Circ. Res.* 2018;122:1648–1660.
- 48 38. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb  
49 DL, Liebler DC, MacCoss MJ. Skyline: An open source document editor for creating and  
50 analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26:966–968.
- 51 39. Nisson PE, Watkins PC, Krizman DB. Isolation of exons from cloned DNA by exon trapping.  
52 *Curr. Protoc. Hum. Genet.* 2001;Chapter 6:Unit 6.1.
- 53 40. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq  
54 quantification. *Nat. Biotechnol.* 2016;34:525–527.
- 55 41. Cirulli, E.T. and Washington NL. Helix Research. UK Biobank Exome rare variant analysis

- 1 v1.3 [Internet]. Helix Blog.  
2 42. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR,  
3 Funke B, Cook SA, Macarthur D, et al. Using high-resolution variant frequencies to empower  
4 clinical genome interpretation. *Genet. Med.* 2017;19:1151–1158.  
5 43. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J,  
6 Dunican DJ, Long A, Kelleher D, et al. Statistical methods for analysis of high-throughput  
7 RNA interference screens. *Nat. Methods.* 2009;6:569–75.  
8 44. König R, Chiang C, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U,  
9 Zhou Y, et al. A probability-based approach for the analysis of large-scale RNAi screens. *Nat.*  
10 *Methods.* 2007;4:847–849.  
11 45. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*  
12 (80-. ). 2000;290:2323–2326.  
13  
14

## 1 Novelty and Significance

2

### 3 What Is Known?

- 4 • The low density lipoprotein (LDL) receptor (LDLR) regulates LDL-cholesterol levels in
- 5 blood by mediating the uptake of LDL into hepatocytes
- 6 • The transcriptional and posttranslational regulation of LDLR activity is targeted by
- 7 cholesterol lowering drugs

### 8 What New Information Does This Article Contribute?

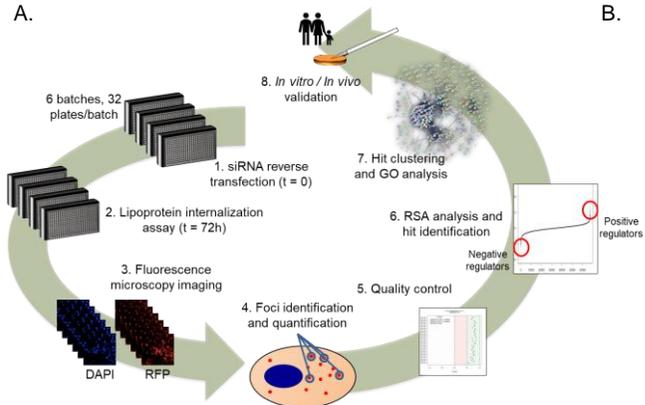
- 9 • Loss of subunits or interactors of the U2 spliceosome decreases the uptake of LDL into
- 10 Huh7 hepatocarcinoma cells
- 11 • Loss of subunits or interactors of the U2 spliceosome cause intron 3 retention of the *LDLR*
- 12 mRNA and, thereby, loss of LDR function
- 13 • Intron 3 retention of *LDLR* in human liver and peripheral blood cells is increased by non-
- 14 alcoholic fatty liver disease and aging, respectively.
- 15 • Single nucleotide polymorphisms of the spliceosome gene *RBM25* are associated with
- 16 higher *RBM25* expression in tissues and lower LDL-cholesterol
- 17 • Expression of rare structural variants of *RBM25* associated with familial
- 18 hypercholesterolemia decrease LDL-uptake into Huh7 cells

19 Low density lipoprotein (LDL) cholesterol is a causal and treatable risk factor of atherosclerotic  
20 cardiovascular diseases whose plasma level is most strongly determined by hepatic removal  
21 through the LDL receptor (LDLR). LDLR activity is known to be regulated both by transcription  
22 of the *LDLR* gene and degradation of the LDLR protein. By genome wide RNA interference we  
23 identified 15 genes encoding subunits and interactors of the U2 spliceosome to limit the uptake of  
24 LDL into Huh7 hepatocarcinoma cells. We identified intron 3 retention of the *LDLR* mRNA as the  
25 underlying mechanism. The mRNA expression analysis of human liver samples and peripheral  
26 blood cells showed the high interindividual variation of this newly identified post-transcriptional  
27 regulation of LDLR. Intron 3 retention increases in non-alcoholic fatty liver disease as well as  
28 with ageing. Moreover, genetic variation in the U2 spliceosome gene *RBM25* is associated with  
29 differences in LDL-cholesterol. Overall, we identified a novel mechanism of LDLR regulation  
30 which might help to better understand the etiology and pathophysiology of LDL-  
31 hypercholesterolemia.

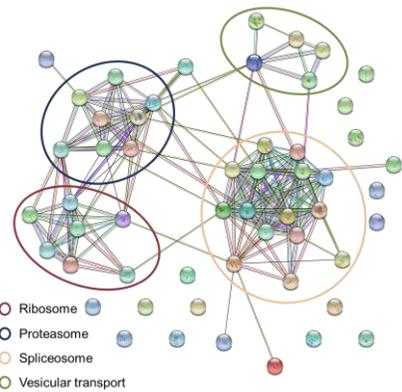
32

# Figure 1

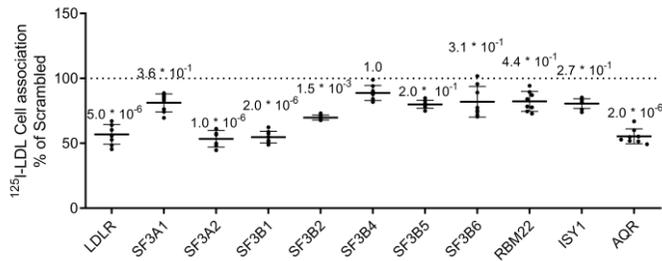
A.



B.



C.



D.

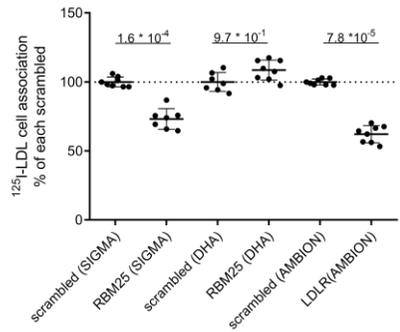
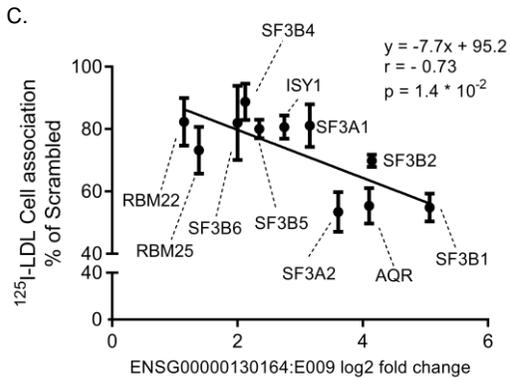
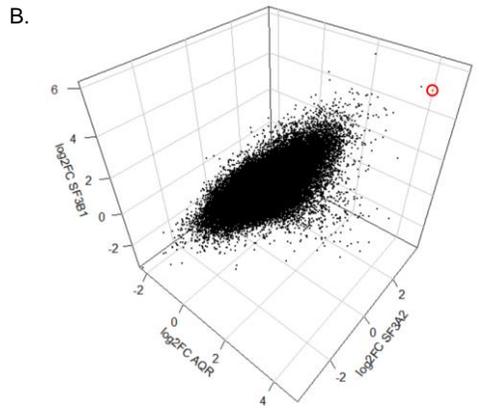
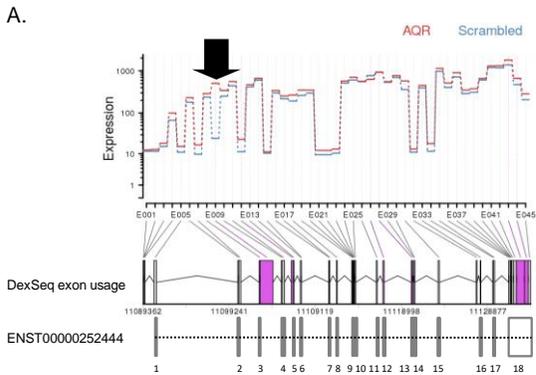
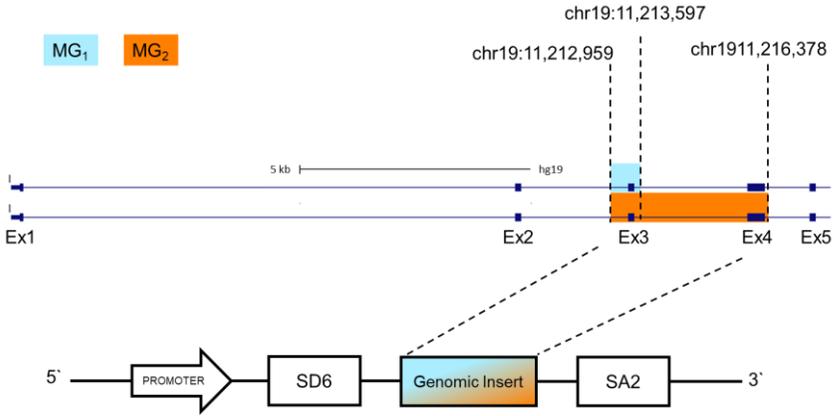


Figure 2

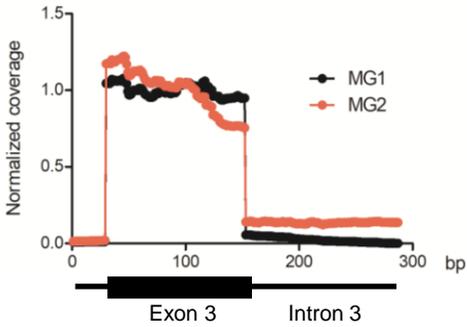


# Figure 3

A.



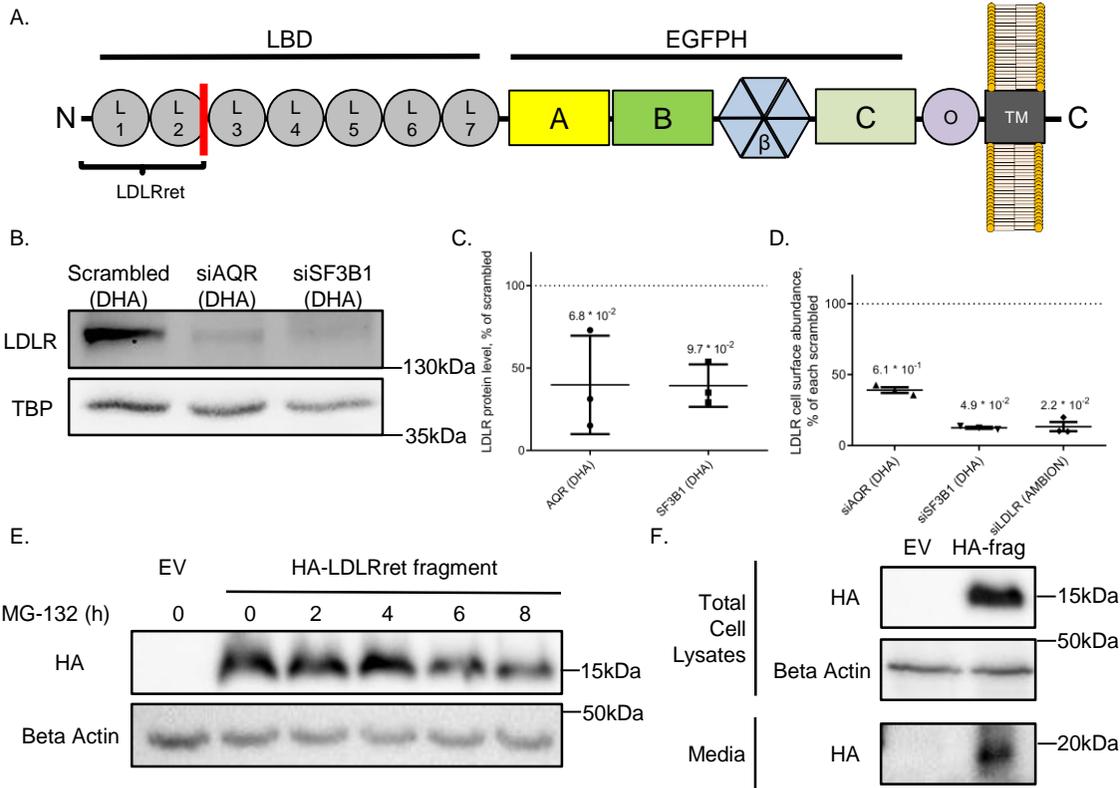
B.



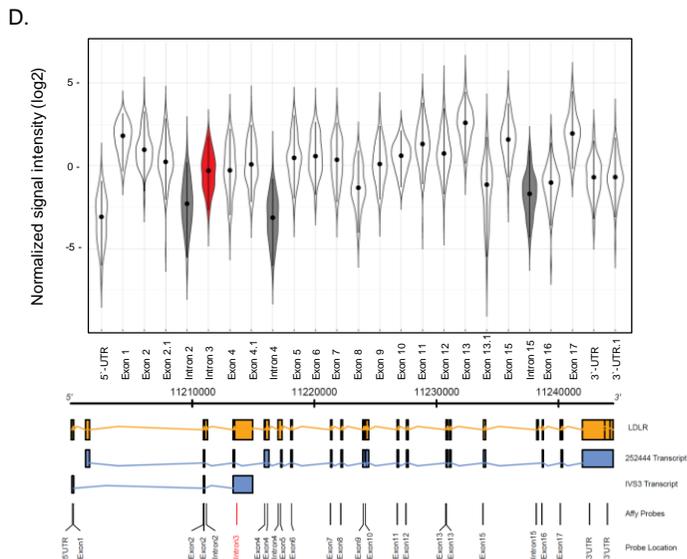
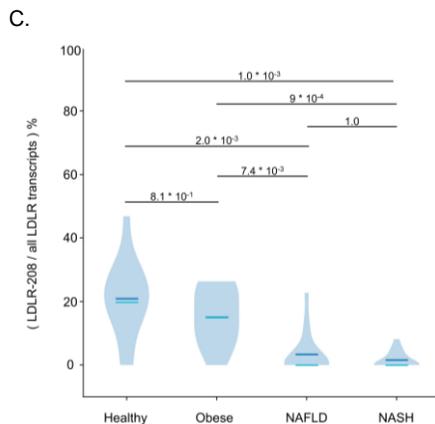
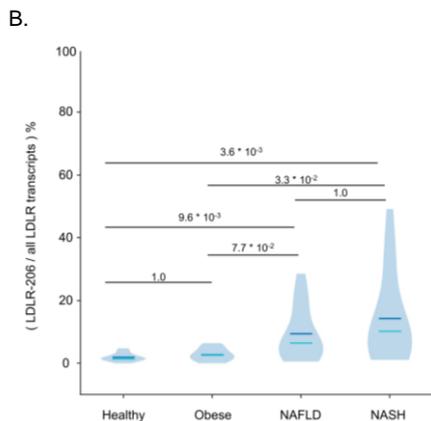
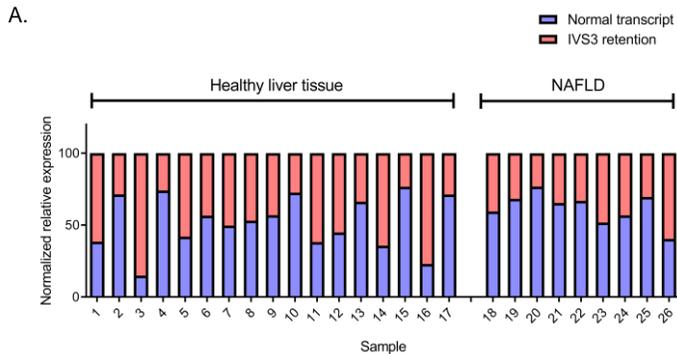
C.



# Figure 4

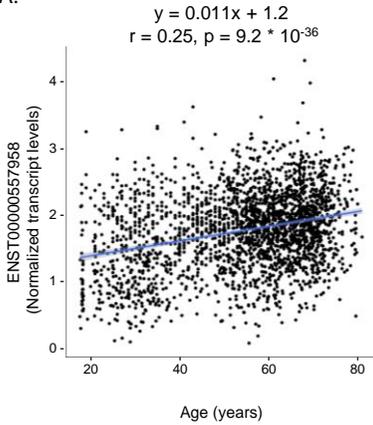


# Figure 5

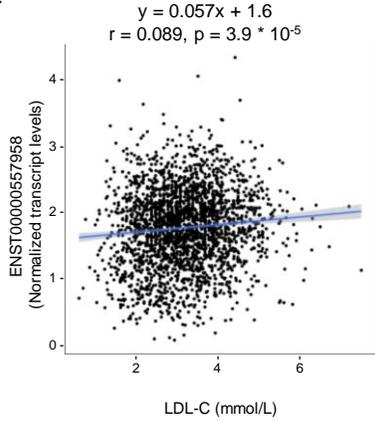


# Figure 6

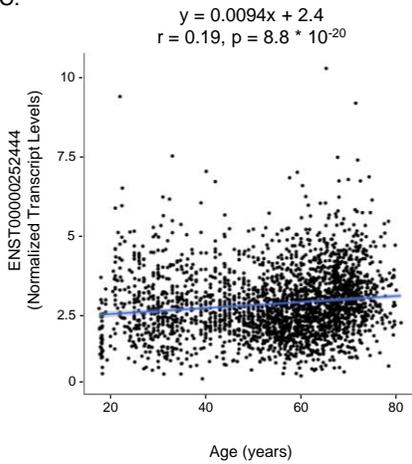
A.



B.



C.



D.

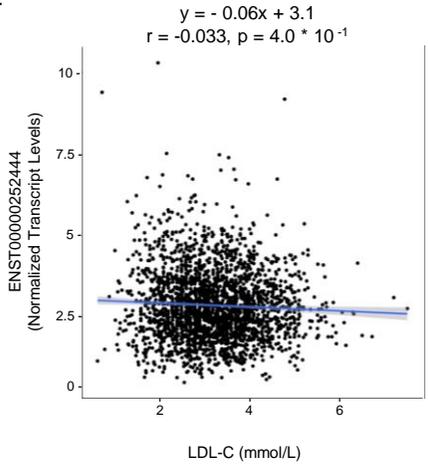
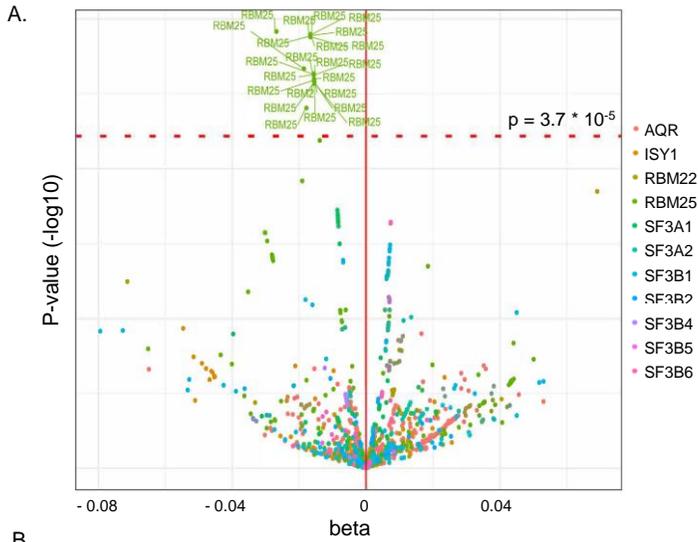
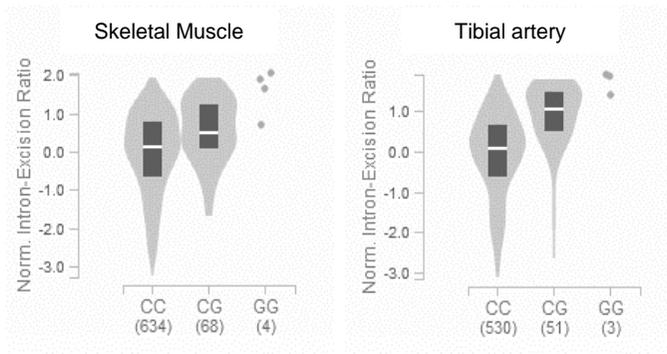


Figure 7



B.



**SUPPLEMENTAL MATERIAL to**

**Posttranscriptional regulation of the human LDL receptor**

**by the U2-spliceosome**

Paolo Zanoni<sup>1,2,\*</sup>, Grigorios Panteloglou<sup>1,2,\*</sup>, Alaa Othman<sup>3</sup>, Joel T. Haas<sup>4</sup>, Roger Meier<sup>5</sup>,  
Antoine Rimbart<sup>6</sup>, Marta Futema<sup>7</sup>, Yara Abou Khalil<sup>8,9</sup>, Simon F. Norrelykke<sup>5</sup>,  
Andrzej J. Rzepiela<sup>5</sup>, Szymon Stoma<sup>5</sup>, Michael Stebler<sup>5</sup>, Freerk van Dijk<sup>10</sup>, Melinde Wijers<sup>6</sup>,  
Justina C. Wolters<sup>6</sup>, Nawar Dalila<sup>11</sup>, Nicolette C. A. Huijkman<sup>6</sup>, Marieke Smit<sup>6</sup>,  
Antonio Gallo<sup>12</sup>, Valérie Carreau<sup>12</sup>, Anne Philippi<sup>13</sup>, Jean-Pierre Rabès<sup>8,14,15</sup>,  
Catherine Boileau<sup>8,16</sup>, Michele Visentin<sup>17</sup>, Luisa Vonghia<sup>18,19</sup>, Jonas Weyler<sup>18,19</sup>,  
Sven Francque<sup>18,19</sup>, An Verrijken<sup>19,20</sup>, Ann Verhaegen<sup>19,20</sup>, Luc Van Gaal<sup>19,20</sup>,  
Adriaan van der Graaf<sup>10</sup>, Belle V. van Rosmalen<sup>21</sup>, Jerome Robert<sup>1</sup>, Srividya Velagapudi<sup>1,2</sup>,  
Mustafa Yalcinkaya<sup>1,2</sup>, Michaela Keel<sup>1,2</sup>, Silvija Radosavljevic<sup>1,2</sup>, Andreas Geier<sup>22</sup>,  
Anne Tybjaerg-Hansen<sup>11</sup>, Mathilde Varret<sup>8</sup>, Lucia Rohrer<sup>1,2</sup>, Steve E. Humphries<sup>23</sup>,  
Bart Staels<sup>4</sup>, Bart van de Sluis<sup>6</sup>, Jan Albert Kuivenhoven<sup>6</sup>, and Arnold von Eckardstein<sup>#,1,2</sup>

\* These authors contributed equally to this work

Materials & Methods	pages 2 - 19
Online Figures S1 to S11	pages 20 - 31
Online Tables S1 to S12	pages 32 - 47
Major Resource Table	pages 48 – 54

## 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 Fisher Scientific, Reinach, Switzerland) and penicillin/streptomycin (100 U/mL each, cat. 15140122, Thermo Fisher Scientific).

### Quantitative Reverse Transcription Real-Time PCR (qRT-PCR)

Total RNA was extracted using Tri reagent<sup>®</sup> (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 Fisher Scientific). Quantitative real time PCR reactions were performed on a Roche Light Cycler 480-II (cat. 05015243001, Roche, Switzerland) using the LightCycler<sup>®</sup> 480 SYBR Green I Master (cat. 04887352001, Roche, Switzerland). For each run, the following thermocycle conditions were used: pre-incubation at 95°C for 5 minutes and 45 cycles of denaturation at 95°C for 10 seconds, primer annealing at 60°C for 10 seconds and extension at 72°C for 20 seconds. For each experiment, 3 technical replicates were used for each condition. At the end of each PCR run, the specificity of the PCR products was confirmed by melting temperature ( $T_m$ ) analysis. The data were analyzed by performing relative quantification based on crossing point ( $C_p$ ) values for the reference gene (*GAPDH*) and the gene of interest using the relative standard curve method. For each condition, the ratio of the

signal for each gene of interest / signal *GAPDH* was calculated and the data were normalized to the respective control condition, as described for each experiment.

### **Isolation and labeling of plasma lipoproteins**

LDL ( $1.019 < d < 1.063$  kg/L) and HDL ( $1.063 < d < 1.21$  kg/L) were isolated from frozen human normolipidemic plasma of blood donors by sequential ultracentrifugation<sup>28</sup> using the Optima™ L-90K Ultracentrifuge and the Type 70 Ti fixed angle rotor both from Beckman Coulter (Brea, USA) and centrifuging at approximately 257,000 g, 15°C for ~16 hours . Density was adjusted by using a KRUESS DS7000 densitometer (Hamburg, Germany). To avoid the oxidation of lipoproteins, EDTA was added to a final concentration of 30 mM. For the genome-wide siRNA screening, LDL and HDL were labeled with Atto594 and Atto655, respectively (cat. AD594-35 and AD655-35, respectively, Atto-Tec, Siegen, Germany) -as published before<sup>29</sup>. For validation experiments, we labeled both LDL and HDL with <sup>125</sup>I according to the McFarlane method<sup>29,30</sup>.

### **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 (cat. 4397926, Thermo Fisher Scientific) 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 Fisher Scientific). 5µL of an 80 nM 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 Fisher 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 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 Fisher Scientific) was diluted in DMEM (0.075 µL of RNAiMax for each 10 µL of DMEM) and 10 µL of diluted RNAiMax were added to each well. The plates were incubated under these conditions for 1 hour at room temperature. 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 26,154 cells/mL (corresponding to a final concentration of 1,700 cells/well) in DMEM 12.3% FBS in 51 Double Sidearm CellStar® Spinner Flasks (Wheaton, Millville, NJ, USA) in a water bath set at 37°C. After 5-10 minutes of gentle stirring, 65µl 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<sub>2</sub>. The cells were maintained under these conditions for 72 hours. Final siRNA concentration was 5 nmol/L. 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, 60 µL of medium were aspirated from each well, followed by the addition of 20 µL of a solution with 66 µg each of Atto594-LDL and Atto655-HDL protein/mL in DMEM, for a final concentration of 33 µg/mL. After incubation for 4 hours at 37°C, 5% CO<sub>2</sub>, the cells of each well were washed six times with PBS, followed by the addition of 50 µL of an isotonic 2% paraformaldehyde solution containing 20 µg/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 remove 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<sup>2</sup> 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, the whole surface of each well was captured at 4x magnification in the DAPI channel for cell counting. Second, 9 tiled sites in each well were then acquired with a 20x objective in the DAPI, RFP and CY5 channels 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)<sup>31</sup>. The data were initially recorded 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 *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-well-plates, using 4 technical replicates for total cellular association (for explanation please see below) and 2 technical replicates for the unspecific cellular association, per each condition respectively, in each experiment. Four pooled siRNA oligonucleotides (Major Resource Table), against each of the top hits were reverse transfected (see screening methods above) using Lipofectamine RNAiMax (cat. 13778150, Thermo Fisher 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<sup>32</sup>. The final siRNA concentration after the addition of the cell suspension was 10 nM. As 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 qPCR using *GAPDH* expression levels as the reference (primers sequences in Table S11). Cell association of <sup>125</sup>I-LDL or <sup>125</sup>I-HDL (33.3 µg/mL, 2 hours incubation at 37°C) was recorded 72 hours after transfection as described previously<sup>29</sup>. The cells were incubated with the respective iodinated lipoprotein without (total) or with 40 times excess of the respective non-labeled lipoprotein (unspecific). The cells were then washed and lysed using 0.1 M NaOH for 1 hour and the <sup>125</sup>I-signal was measured using a Wizard<sup>2</sup>10-Detector Gamma counter (cat. 2470-0100, Perkin Elmer). The protein content of each sample was determined using the

micro-BCA Protein Assay Kit (cat. 23235, Thermo Fisher Scientific) and the ratio of the counts / protein content for each technical replicate of each sample was determined. Specific cellular association was calculated by subtracting the unspecific (mean value of the 2 technical replicates of each condition)) from the total (individual values of each technical replicate-total- of each condition).

### **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. 72 hours after transfection, the cells were washed once with PBS. Total 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  $3.13 \times 10^7$  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<sup>33</sup>. 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)<sup>34</sup>. 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<sup>35</sup>. To detect candidates for differential exon usage we used DEXSeq<sup>36</sup>. Sequences can be accessed by codes

PRJEB46899 and PRJEB46898, respectively, in the data bank of the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/support>)

### **Measurement of intron 3 retention by qRT-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 qRT-PCR using previously published<sup>5</sup> primer couples annealing in exon 3, intron 3 and exon 4 (see Table S11 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 Table S11 and subcloned in pCR-Blunt (Thermo Fisher 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 GeneArt™ Strings™ DNA Fragments (Thermo Fisher 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, Thermo Fisher 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,  $1 \times 10^6$  cells/well in complete medium. 16 hours 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 (7 µL of Lipofectamine 3000 and 5 µL of P3000 reagent per well). The same amount of an empty pcDNA3.1 vector was used as control. For proteasomal inhibition studies, the cells were exposed to the MG-132 (cat. SML1135, Merck) proteasomal

inhibitor at a concentration of 5  $\mu\text{mol/L}$  in DMSO as described previously<sup>37</sup>. Protein lysates and media were collected 48 h after transfection. Media were spun at 1500g for 5 minutes and the supernatant was used to detect the fragment.

### **Molecular cloning of *RBM25* and its mutants**

The coding sequences of the wild type (wt) and the *RBM25* mutants 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 the pcDNA3.1 vector, by digesting the original pLVX-puro constructs with BstBI (cat. R0519S, NEB), Klenow (M0210S, NEB) and XbaI (cat. FD0685, Thermo Fisher scientific) and the pcDNA3.1 empty vector with EcoRV (cat. FD0303, Thermo Fisher scientific) and XbaI, respectively, according to the manufacturer's instructions. After gel purification with Zymoclean Gel DNA Recovery Kit (cat. D4001, Zymo Research), the digested material was ligated using T4 DNA ligase (cat. M0202S, NEB). The correct sequences of the generated plasmids were confirmed by both restriction enzyme digestion and Sanger Sequencing.

For the overexpression experiments,  $4 \times 10^5$  Huh-7 cells were seeded onto 6-well-plates. After 24 hours, the cells were transfected with 4  $\mu\text{g}$  from each plasmid using Lipofectamine 2000 (cat. 11668019, Invitrogen) according to the manufacturer's instructions (6  $\mu\text{L}$  lipofectamine 2,000/well). The same amount of empty pcDNA3.1 vector was used as the control. 24 hours after transfection the cells were treated with complete medium (DMEM containing 10% FBS) supplemented with 750  $\mu\text{g/mL}$  G418 (cat. 10131-027, Gibco) for an additional 48 hours before being harvested.

**Western blotting**

Protein levels of LDLR and RBM25 in the whole cell lysate were determined by Western blotting 72 hours after transfection. 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 (cat. 34577, Thermo Fisher 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) (Figures 4C,4D) or ImageJ (<https://imagej.nih.gov/ij/>; figures S3c, S3e).

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, cat. H6533, Sigma-Aldrich). Beta actin (cat. SC-1616, Santa Cruz) was used as the loading control.

**Flow cytometry based analysis of LDLR cell surface expression**

LDLR cell surface levels were determined by flow cytometry on alive Huh-7 cells. 72 hours after transfection, the medium was aspirated and the cells were washed twice with PBS and detached using Accutase® (cat. A6964, Sigma-Aldrich) for 5 minutes 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 minutes on ice. After blocking, the cells were incubated with a primary anti-LDLR antibody

(Santa Cruz Biotechnology (figure 4D) or Progen Biotechnik (figures S3g, S10e and S11d); see Major Resources Table) diluted to a final concentration of 2  $\mu\text{g}/\text{mL}$  in FACS buffer (PBS containing 0.5% BSA and 0.05%  $\text{NaN}_3$ ) 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 Fisher Scientific) diluted to a final concentration of 4  $\mu\text{g}/\text{mL}$  in FACS buffer for 1 hour on ice in the dark (figure 4D) or anti-Mouse IgG (H+L) Cross-Adsorbed AlexaFluor 647-conjugated Secondary Antibody (cat. A-21236, Thermo Fisher Scientific) (figures S3g, S10e and S11d). Finally, the cells were washed twice with FACS buffer and re-suspended in ice cold FACS buffer for subsequent acquisition. In all cases, cells incubated with the secondary antibody were used as negative controls, while cells not incubated with any antibody were additionally used for the data shown in figure 4D. Prior acquisition, and in order to exclude signal originating from dead cells (figures S3g, S10e and S11d), the cells were incubated with Propidium Iodide (PI) (cat. 81845, Fluka) and FITC-Annexin V (cat. 640945, Biolegend, London, UK) at final concentrations of 1  $\mu\text{g}/\text{mL}$  and 2  $\mu\text{g}/\text{mL}$ , respectively. Sample acquisition was carried out on a BD LSR II Fortessa (BD-Biosciences, Allschwil, Switzerland) and using BD FACSDIVA<sup>TM</sup> software. During acquisition, signal from cell debris (and signal from dead cells for Figure 4D) was excluded by applying the forward size scatter versus the side size scatter gating (FSC-A / SSC-A). Thereafter, the doublets were also excluded through gating (FSC-A / FSC-H). Finally, for data shown in figures S3g, S10e and S11d, to exclude signal originating from the dead cells, the double negative cells for both FITC-Annexin V and PI were selected by gating (FITC-A / PI-A). To account for the spectral overlap of the different fluorophores used, unstained and single stained controls were acquired and analyzed using the compensation function of the FACSDIVA<sup>TM</sup> software. In all experiments, approximately  $10^4$  events per condition recorded at the final gate containing the population of alive cells were used for analysis. Data analysis was carried out using FlowJo version 10 (FlowJO LLC, Ashland, OR,

USA). The Median Fluorescence Intensity (MFI) of each population was used for comparisons between the different conditions.

### **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 at 37°C. After 2 hours, the cells were washed with PBS and detached using Accutase® solution (cat. A6964, Sigma Aldrich, Switzerland). Prior to acquisition, and in order to exclude signal originating from dead cells, the cells were incubated 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. During acquisition, the signals originating from cell debris and from doublets were excluded as described above. Finally, to exclude signal originating from the dead cells, the double negative cells for both FITC-Annexin V and PI were selected by gating (FITC-A /PI-A). To account for the spectral overlap of the different fluorophores used, unstained and single stained controls were acquired and analyzed using the compensation function of the FACSDIVA™ software. In all experiments, approximately 10<sup>4</sup> events per condition recorded at the final gate containing the population of alive cells were used for analysis. 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 sample.

### **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<sup>37</sup>. 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 <sup>13</sup>C<sup>15</sup>N-labeled arginine, PEPotec grade 2 Thermo Scientific; NVVALDTEVASNR containing <sup>13</sup>C<sup>15</sup>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<sup>37</sup>. The MS traces were manually curated using the Skyline software<sup>38</sup> 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 Table S11, cloned first in pCR-Blunt and then subcloned in the pSPL3 exon trapping vector<sup>39</sup> for expression. Reaction conditions for PCR cloning of the short construct: Step 1: 95°C, 2 minutes; Step 2: 95°C, 30 seconds; Step 3: 66°C, 30

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

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 *LDLR*ret fragment overexpression studies. 48 hours after transfection, the cells were lysed in Trizol Reagent (cat. 15596026, Thermo Fisher Scientific) and the lysates were used for total RNA extraction as described above. 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**

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 qRT-PCR analysis. Nine additional liver biopsies were obtained in the Division of Hepatology, University Hospital Würzburg, Germany from patients suspected to have NAFLD and were included into the qRT-PCR analysis as well.

Individual *LDLR* transcript expression was analysed in RNA Sequencing data of human liver samples obtained from 14 healthy individuals, 12 obese individuals without non-alcoholic

fatty liver disease (NAFLD), 15 patients with NAFLD, and 16 patients with non-alcoholic steatohepatitis (NASH) (Gene Expression Omnibus, accession number GSE126848)<sup>8</sup>. Upon a detailed examination of the quality of each sample, 1 sample of a healthy individual (SRR8601557) and 1 sample of a NASH patient (SRR8601555) were excluded from the generated plots due to lower quality of the reads. Although the sample SRR8601570 had 2-3x lower number of mapping reads compared to the rest of the samples, it was retained for analysis. Transcripts per million (TPM) and expected counts were computed using the RNA sequencing quantification software Salmon in quasi-mapping-based mode. The reference transcriptome used was ref. Ensembl 97, GRCh38.p12 (human).

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<sup>9</sup>. 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<sub>2</sub> ratio of each probe to the total *LDLR* transcript expression in each patient.

### **LDLR transcript analysis in peripheral blood cells**

Whole blood samples of 4,000 unrelated individuals were obtained from the Dutch BIOS consortium<sup>10</sup>. 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<sup>40</sup> (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 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. A Spearman correlation test was performed between the TPM values normalized with Kallisto (see above) and the phenotypes of interest using the R `cor.test` function.

### **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<sup>2</sup>, inferred\_sex, age \* inferred\_sex, age<sup>2</sup> \* 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 Table S12. P values for associations 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<sup>11,41</sup>. 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).

**Associations of the *RBM25* SNP rs17570658**

The association of the lead SNP in *RBM25* rs17570658 with LDL cholesterol was meta-analysed according to METAL<sup>14</sup> using data of eight studies with 435,884 subjects summarized by the CardioVascular Disease Knowledge Portal (<https://cvd.hugeamp.org/variant.html?variant=rs17570658>) plus data of 19,653 participants of the Copenhagen City Heart Study and the Copenhagen General Population Studies genotyped with exome chip arrays<sup>13</sup>. The association of rs17570658 with *RBM25* gene expression was analysed by using data of nearly 1000 individuals collected by the Genotype-Tissue Expression (GTEx) project (GTEx <https://gtexportal.org/home/>). Nominal p values for each variant–gene pair were estimated using a two-tailed t-test. The significance of the most highly associated variant per gene was determined from empirical P values, extrapolated from a Beta distribution fitted to adaptive permutations with the setting –permute 1000 10000. These empirical p-values were subsequently corrected for multiple testing across genes using Storey’s q value method<sup>26,27</sup>.

**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<sup>15</sup>. A previous analysis suggested an enrichment of rare (minor allele frequency [MAF] <  $5.0 \times 10^{-3}$ ) *RBM25* gene variants in the FH cohort in comparison to 1926 controls from the UK10K study<sup>15</sup>. The gene burden test was re-analyzed by applying binomial test to a much larger control cohort of 56,352 European data provided by the gnomAD study<sup>12</sup>. Variants identified by WES in both FH cases and gnomAD, were filtered to select those with  $MAF < 1.0 \times 10^{-4}$  that were annotated as missense or loss of function. The MAF cut-off of  $1.0 \times 10^{-4}$  was based on the analysis of another genetically heterogeneous dominant disease, hypertrophic cardiomyopathy, as previously demonstrated by Whiffin et al<sup>42</sup>.

## 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<sup>43</sup> 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<sup>44</sup>, 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<sup>43</sup>. Dimensionality reduction across the five main assay features mentioned above was performed using the Locally-Linear-Embedding<sup>45</sup> method on log<sub>2</sub>-transformed data through the *sklearn* Python implementation of the method (<http://scikit-learn.org/>).

Top hits were clustered by function using the online String tool (<https://string-db.org/>), version 11.0 according to the following settings: the network edges indicate the type of interaction evidence, all active interaction sources (textmining, experiments, databases, co-expression, neighbourhood, gene fusion, co-occurrence) were selected for the analysis, with the minimum interaction score set to 0.400 (medium confidence) and with no network clustering used. 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>). 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.

Due to low numbers of replicate experiments, non-parametric tests of the GraphPad Prism version 8 were used to analyse the data of the validation experiments, namely Kruskal-Wallis test with either Dunn's multiple comparisons test for the experiments shown in figures 1C,

4C, 4D, and in figures S2b-d and S11 or. For one to one comparisons, the Mann-Whitney test (one-tailed) was used for the experiments shown in figure 1D and in figures S3, S5c and S10 or the Wilcoxon test (two-tailed) for the data shown in figures S8f and S8g. Correlations presented in figures 2C, 6, and S8a-e as well as tables S5 and S6 were calculated according to Spearman (two-tailed). Other statistical tests applied to the analysis of genetic data on RBM25 in data banks have been described before in the respective paragraphs. The p values presented in figures 5B, 5C, 6, 7A, S8, and S9a as well as tables S5 and S6 were adjusted for multiple testing according to the Bonferroni correction. Empirical p values shown in Figure 7B were corrected for multiple testing across genes using Storey's q value method<sup>26,27</sup>.

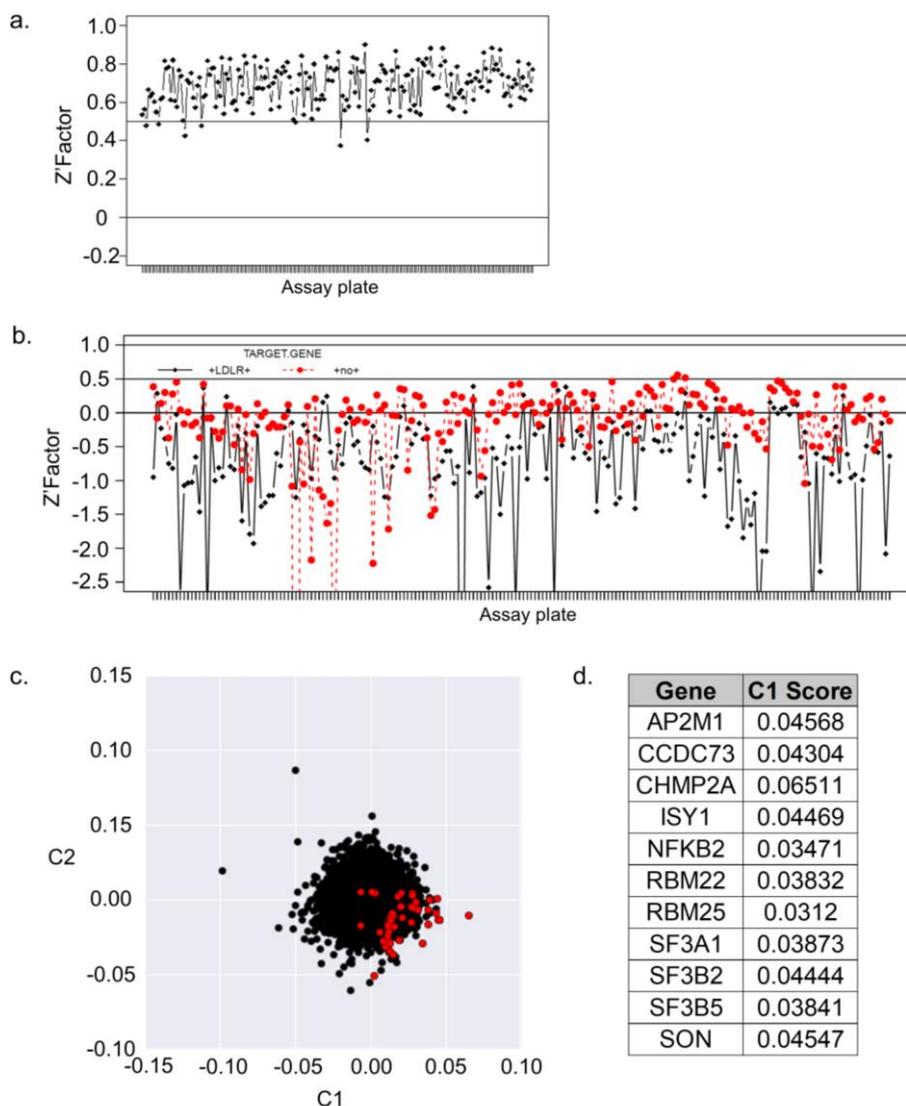
The specific statistical tests and the adjustments for multiple testing used are also described in the legends of the respective figures and tables. P values, coefficients of correlation as well as regression equations are presented within the figures and tables.

### **Study approval**

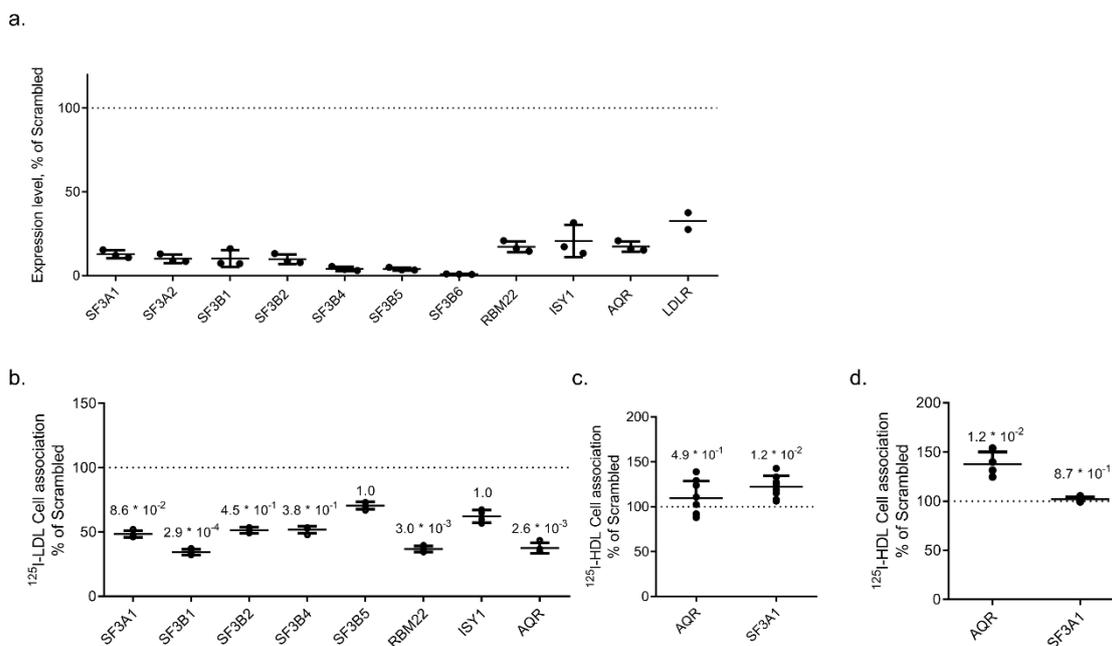
Human studies were approved by institutional boards of each institution involved: the ethics committee of the University Medical Centre Groningen, the Ethical Committee of the Antwerp University Hospital (file 6/25/125) and the Ethics Committee of the University Hospital Würzburg (AZ188/17 and AZ96/12). 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 informed 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.

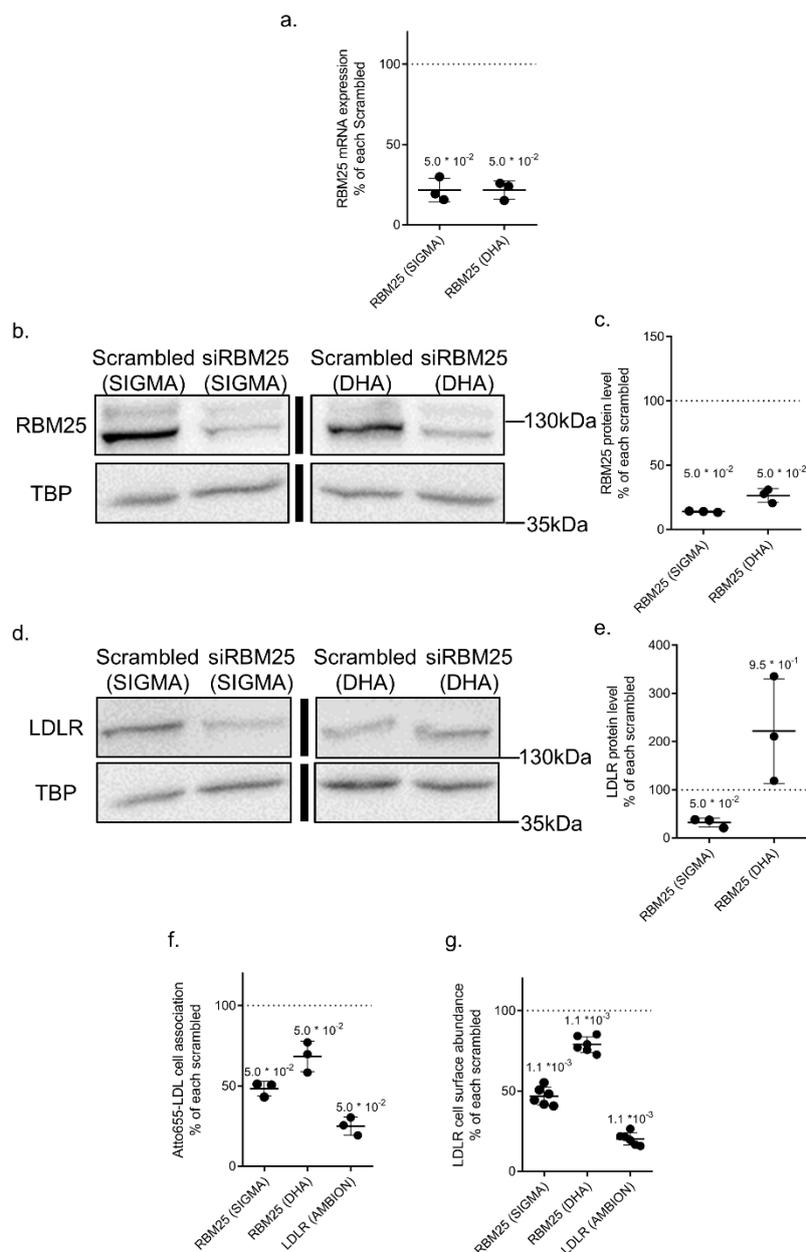
## 2. Supplemental Figures



**Figure S1. 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.

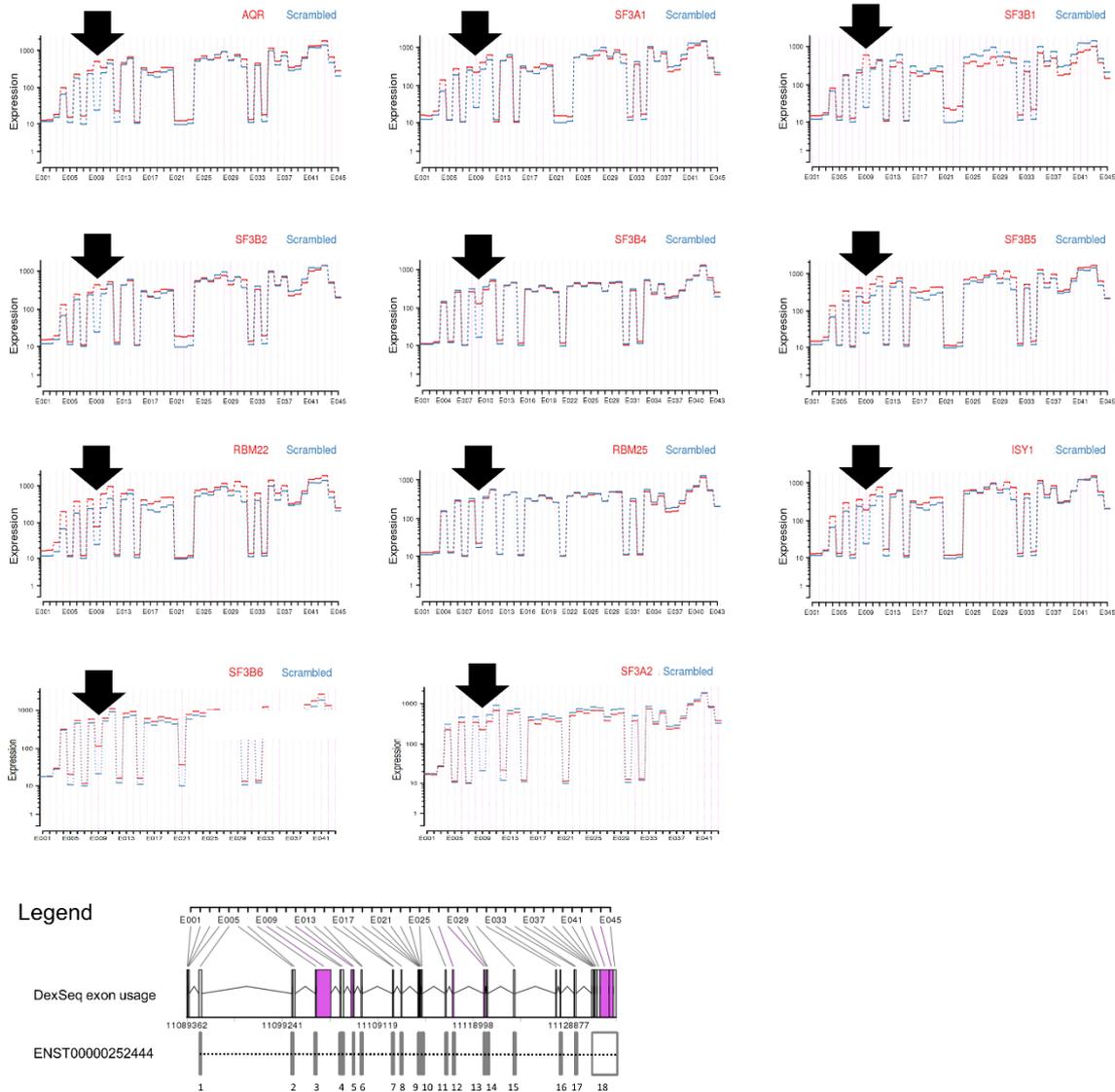


**Figure S2. Additional *in vitro* validation experiments and determination of knockdown efficiency.** **a)** Knockdown efficiency of siRNA pools in Huh-7 cells was measured by real-time PCR 72 hours after siRNA transfection and is expressed as percentage of expression compared to a non-targeting control (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 $\pm$ SD. **b)**  $^{125}\text{I}$ -LDL cell association in HepG2 cells. **c)**  $^{125}\text{I}$ -HDL cell association in Huh-7 cells. **d)**  $^{125}\text{I}$ -HDL cell association in HepG2 cells. Cell association experiments were performed 72 hours after transfection with pooled siRNAs by incubating the cells for 2 hours at 37°C in the presence of 33.3  $\mu\text{g}/\text{ml}$  of radioiodinated lipoproteins or 40 fold excess of the respective unlabeled lipoprotein. Data are means $\pm$ SD of 3 triplicate experiments (**a**), 1 quadruplicate experiment (**b,d**), or 2 quadruplicate experiments (**c**). Statistical analysis (**b-d**) was performed using Kruskal-Wallis test with Dunn's multiple comparisons test between the non-targeting and each targeting siRNA; the respective p values are shown above each condition.

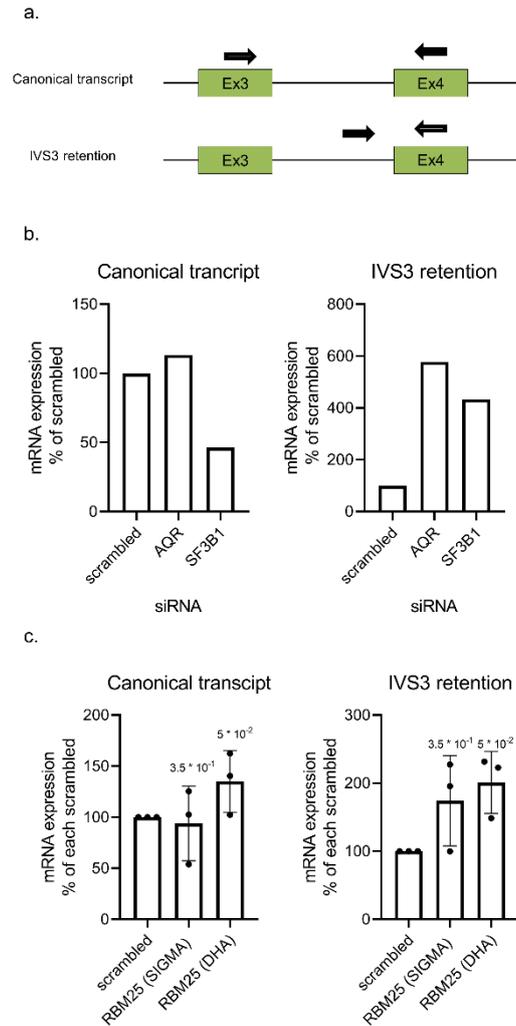


**Figure S3. Validation of RBM25.** **a).** Knockdown efficiency of siRNAs obtained from different vendors (Sigma or Dharmacon) in Huh-7 cells was measured by real-time PCR 72 hours after siRNA transfection and is expressed as percentage of expression compared to the respective vendor's non-targeting control (scrambled). Data are shown as means  $\pm$ SD of 3 independent 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 non-targeting (scrambled) control. TATA-binding-protein (TBP) was used as the loading control. Data are shown as means  $\pm$ SD of 3 independent 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

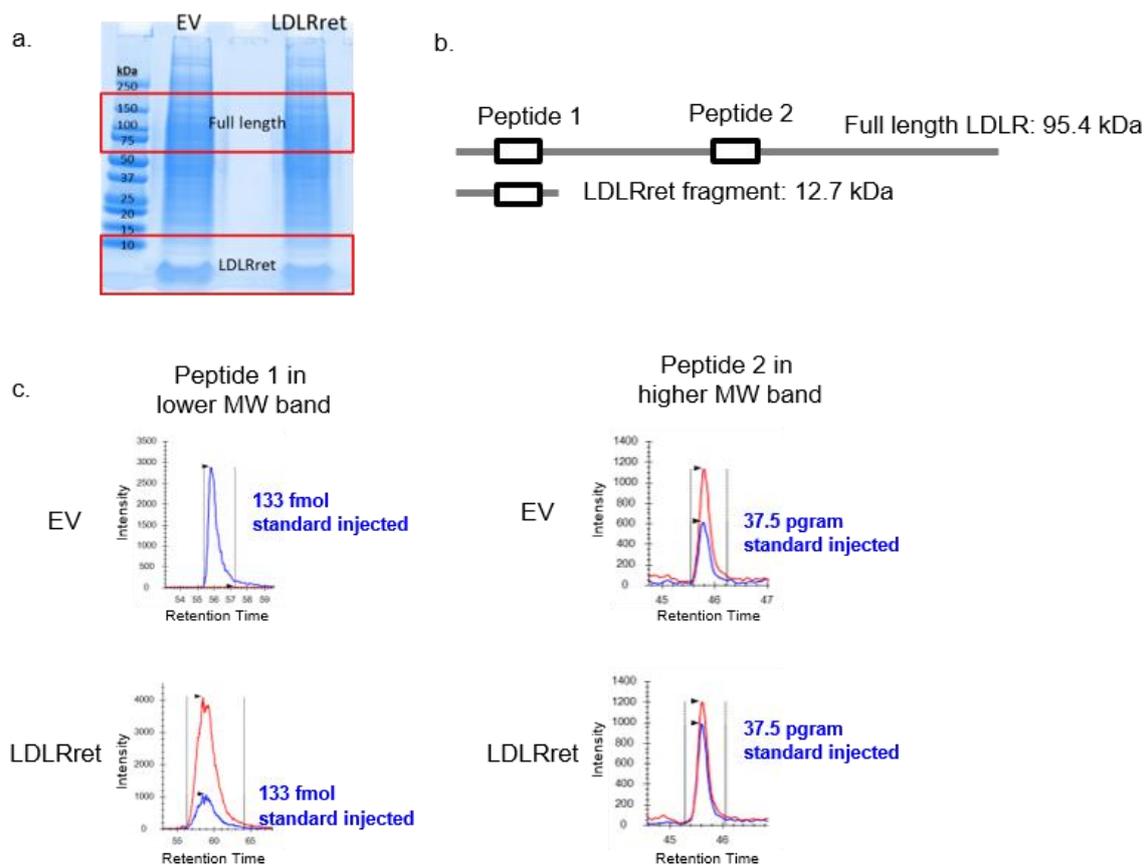
knockdown of *RBM25* with the indicated siRNAs, relative to the respective non-targeting (scrambled) control. TATA-binding-protein (TBP) was used as the loading control. Data are shown as means  $\pm$ SD of 3 independent experiments. **f) Effect of *RBM25* knockdown on Atto655-LDL uptake.** Atto655-LDL 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  $\mu$ g/ml Atto655-LDL in the presence or absence of 100-fold excess of unlabeled LDL. Data are shown as means  $\pm$ SD of 3 independent experiments. **g) Effect of *RBM25* knockdown on LDLR cell surface levels.** LDLR cell surface levels in alive 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-targeting control. Data are shown as means  $\pm$ SD of 6 independent experiments. In all cases, statistical analysis was performed using Mann-Whitney test (one-tailed) between each targeting siRNA and the respective non-targeting control (scrambled) of each vendor. The respective p values are shown above each condition.



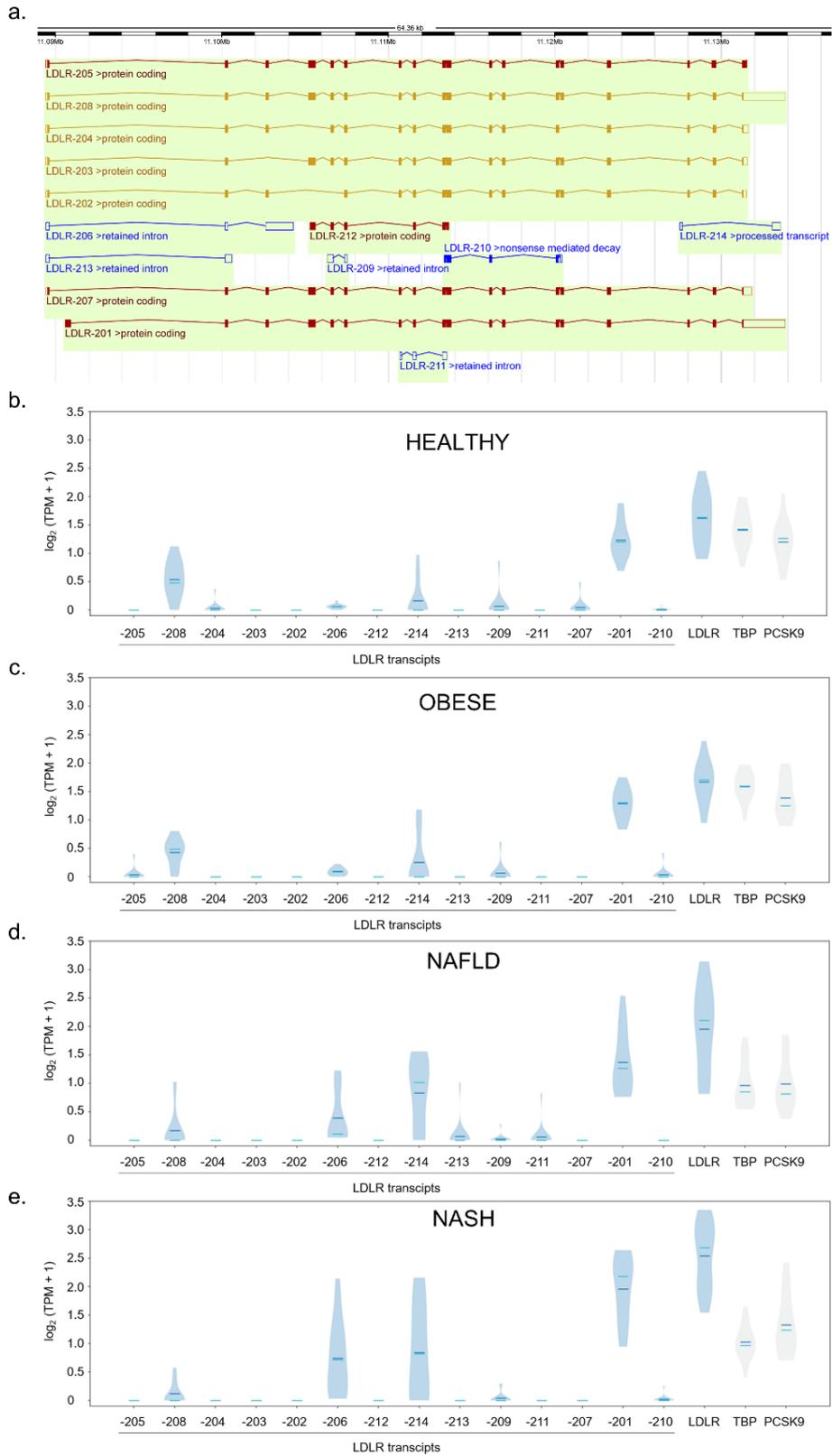
**Figure S4. *LDLR* exon level expression after knockdown of U2-spliceosome genes.** Expression of the *LDLR* exons was recorded by RNA sequencing of Huh-7 cells 72 hours after knockdown of each U2-spliceosome hit gene. Segments represent differential exon usage in each sector of the *LDLR* genomic sequence as identified by the DEXSeq algorithm and as summarized in the linear representation below the graph. Canonical exons belonging to the ENST00000252444 full-length transcript 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 independent experiments.



**Figure S5. Confirmation of *LDLR* intron 3 retention by qRT-PCR. a). qRT-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.,<sup>5</sup>. 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 (scrambled) and are expressed as means of 1 triplicate experiment (b) or means $\pm$ SD of 3 triplicate experiments (c). Statistical analysis in (c) was performed using Mann-Whitney test (one-tailed) by comparing each targeting siRNA with the respective non-targeting control (scrambled) of each vendor. The respective p values are shown above each condition.



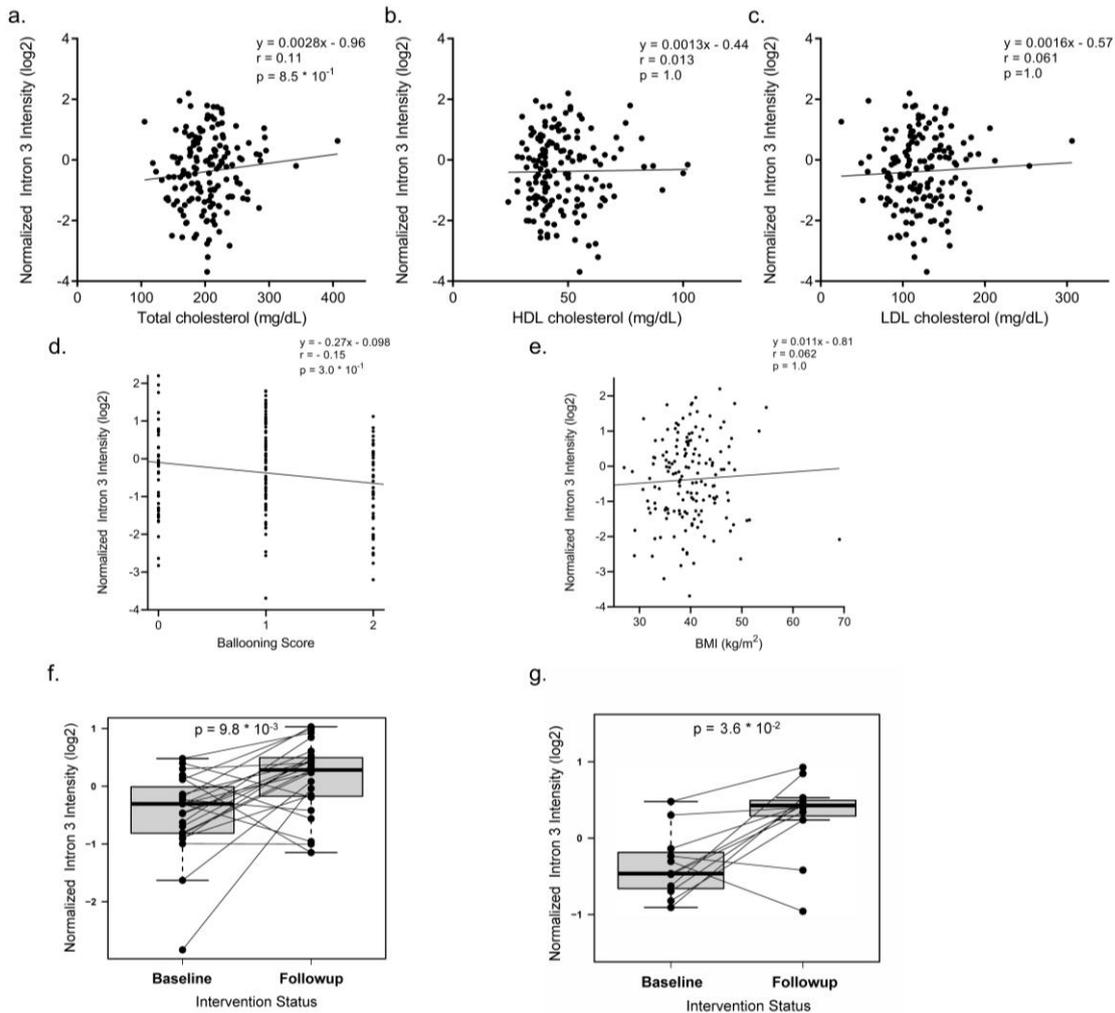
**Figure S6. Detection of the LDLRret fragment by liquid-chromatography coupled mass spectrometry (LC-MS).** The full length LDLR and fragment LDLR were separated on a SDS-PAGE (**a**) and the bands containing the full size protein and the LDLR fragment were excised and trypsinized with in-gel digestion. **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).



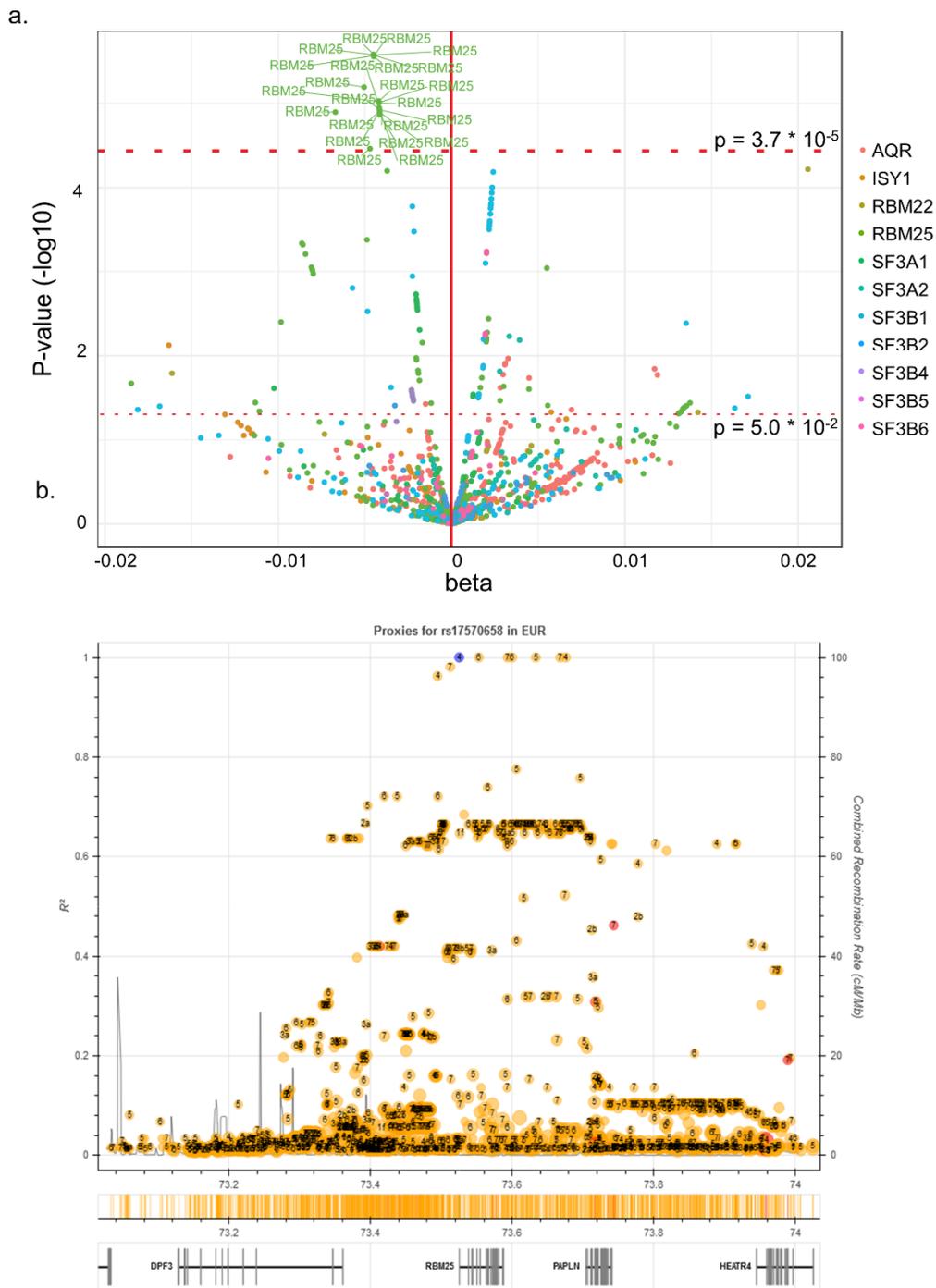
**Figure S7. Quantification of different *LDLR* transcripts expressed in livers of individuals without or with non-alcoholic fatty liver disease (NAFLD) by RNA sequencing. a). Illustration of the different *LDLR* transcripts.** The different *LDLR* transcripts and the exons/introns included for each transcript were retrieved and adapted from ensemble's webpage:

[https://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000130164;r=19:11089462-11133820](https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000130164;r=19:11089462-11133820).

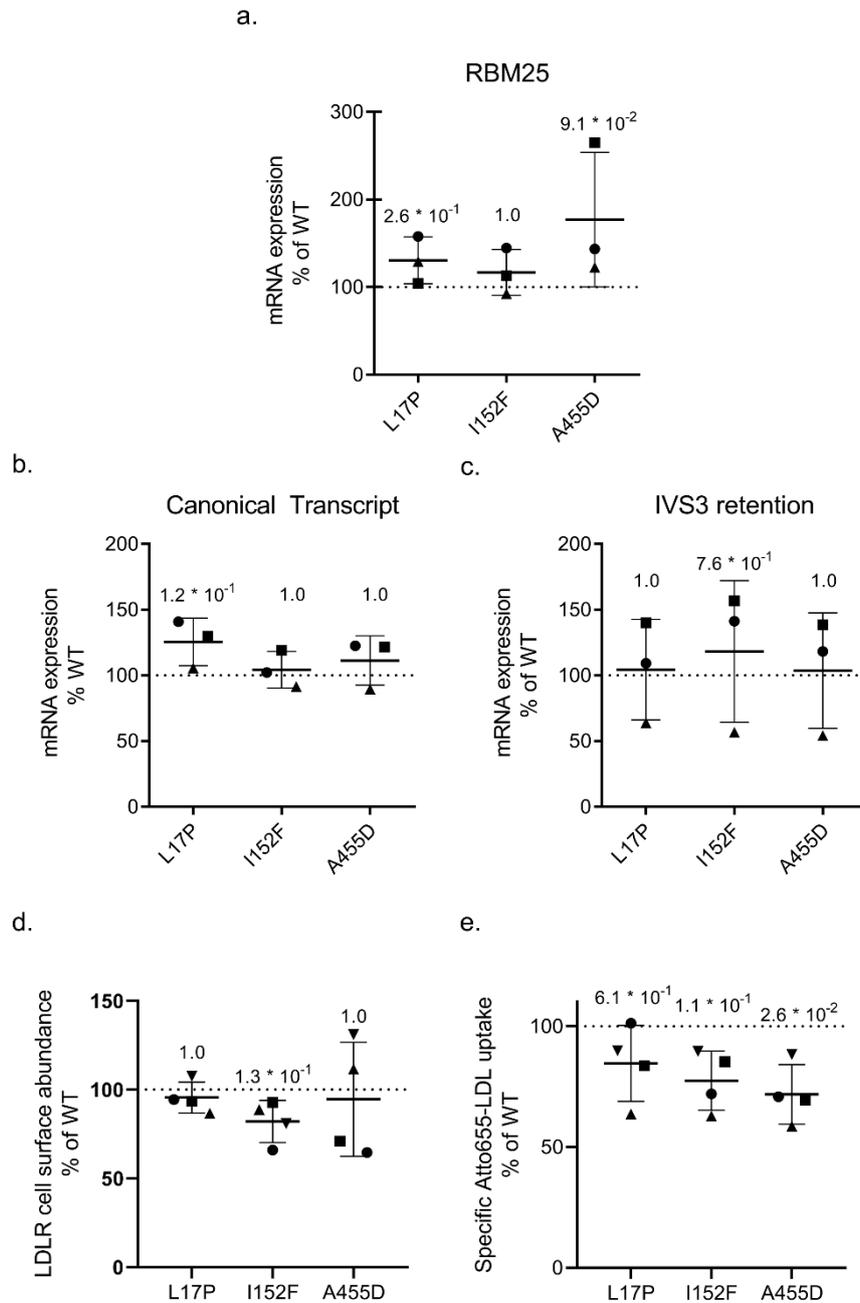
**b-e). Interindividual and disease-dependent variation of *LDLR* transcript expression in human liver.** Computational analysis of previously published RNA sequencing data of liver samples from 13 non-obese (**b**) and 12 obese individuals without NAFLD (**c**) as well as 15 patients with NAFLD, and 15 patients with non-alcoholic steatohepatitis (NASH, **d**). (Gene Expression Omnibus, accession number GSE126848)<sup>8</sup>. Graph depicts the log<sub>2</sub> transcript per million (TPM) levels (y-axis) for each of the annotated *LDLR* transcripts, the sum of all 13 *LDLR* transcripts, and the reference genes *TBP* and *PCSK9* (x axis). The dark blue and the light blue lines within each violin correspond to the mean and median values, respectively.



**Figure S8. Hepatic expression of *LDLR* Intron 3 retention does not correlate with plasma lipids but increases upon bariatric surgery in obese subjects<sup>9</sup>.** a-e. Hepatic Intron 3 probe intensity is shown after log<sub>2</sub> transformation. Correlations of intron 3 probe relative intensity values with plasma levels of total (a), HDL- (b) and LDL- cholesterol (c), histological ballooning score (d), or BMI (e) in the entire cohort at baseline (n = 155). Regression equations, Spearman coefficients of correlation (r) as well as p values adjusted for multiple testing according to Bonferroni correction are reported for each graph f) and g). **Effect of bariatric surgery on intron 3 retention.** Intron 3 retention levels for patients from whom hepatic mRNA was available at baseline and after bariatric surgery (median follow-up time 13 months, IQR = [12 months, 15 months]). Data from all 21 paired biopsies (f) and from 11 paired biopsies of responders (g), defined as having non-alcoholic steatohepatitis (NASH) at baseline but no NASH at follow-up. 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. The p values reported were calculated using the Wilcoxon matched-pairs signed rank test and were adjusted for multiple testing according to Bonferroni correction.



**Figure S9. 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 Bonferroni correction for multiple testing of 1360 variants within the genes of interest ( $p=3.7 \times 10^{-5}$ ), 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  $R^2$ . rs17570658 is indicated by the blue circle



**Figure S11. Impaired LDL uptake by Huh-7 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 IVS3 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. In all graphs, data are normalized to wild type RBM25 (WT) and are shown as means  $\pm$ SD, with each dot representing 1 of 3 independent (**a-c**) or 4 independent (**d, e**) experiments. Numbers in the graphs are p values calculated by Kruskal-Wallis test with Dunn's multiple comparisons test by comparing WT with each mutant.

### **3. Supplemental Tables**

**Table S1 (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-labeled 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.

**Table S2. Results of the Panther GO-Slim BP enrichment test.**

<b>PANTHER GO-Slim Biological Process</b>	<b>RL</b>	<b>HG</b>	<b>EV</b>	<b>FE</b>	<b>p</b>	<b>FDR</b>
<b>intra-Golgi vesicle-mediated transport (GO:0006891)</b>	28	3	0.07	40.9	$7.20 \times 10^{-05}$	$7.61 \times 10^{-3}$
<b>mRNA splicing, via spliceosome (GO:0000398)</b>	164	7	0.43	16.29	$3.02 \times 10^{-7}$	$2.71 \times 10^{-4}$
macromolecule metabolic process (GO:0043170)	2419	22	6.34	3.47	$6.73 \times 10^{-8}$	$1.21 \times 10^{-4}$
organic substance metabolic process (GO:0071704)	3338	22	8.74	2.52	$1.62 \times 10^{-5}$	$2.08 \times 10^{-3}$
metabolic process (GO:0008152)	4072	22	10.67	2.06	$4.42 \times 10^{-4}$	$4.18 \times 10^{-2}$
RNA processing (GO:0006396)	233	7	0.61	11.47	$2.92 \times 10^{-6}$	$4.04 \times 10^{-4}$
gene expression (GO:0010467)	1842	14	4.83	2.9	$2.23 \times 10^{-4}$	$2.23 \times 10^{-2}$
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	164	7	0.43	16.29	$3.02 \times 10^{-7}$	$1.81 \times 10^{-4}$
RNA splicing, via transesterification reactions (GO:0000375)	165	7	0.43	16.2	$3.14 \times 10^{-7}$	$1.41 \times 10^{-4}$
RNA splicing (GO:0008380)	176	7	0.46	15.18	$4.77 \times 10^{-7}$	$1.22 \times 10^{-4}$

Table S2 (continued). Results of the Panther GO-Slim BP enrichment test.

<b>PANTHER GO-Slim Biological Process</b>	<b>RL</b>	<b>HG</b>	<b>EV</b>	<b>FE</b>	<b>p</b>	<b>FDR</b>
<b>proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)</b>	142	5	0.37	13.44	4.07*10 <sup>-5</sup>	4.88*10 <sup>-3</sup>
macromolecule catabolic process (GO:0009057)	300	8	0.79	10.18	1.29*10 <sup>-6</sup>	2.31*10 <sup>-4</sup>
proteolysis involved in cellular protein catabolic process (GO:0051603)	253	8	0.66	12.07	3.69*10 <sup>-7</sup>	1.33*10 <sup>-4</sup>
cellular protein catabolic process (GO:0044257)	255	8	0.67	11.98	3.91*10 <sup>-7</sup>	1.17*10 <sup>-4</sup>
protein catabolic process (GO:0030163)	289	8	0.76	10.57	9.79*10 <sup>-7</sup>	2.20*10 <sup>-4</sup>
proteasomal protein catabolic process (GO:0010498)	153	5	0.4	12.48	5.75*10 <sup>-5</sup>	6.45*10 <sup>-3</sup>
<b>protein ubiquitination (GO:0016567)</b>	200	7	0.52	13.36	1.09*10 <sup>-6</sup>	2.18*10 <sup>-4</sup>
protein modification by small protein conjugation (GO:0032446)	217	7	0.57	12.31	1.85*10 <sup>-6</sup>	2.77*10 <sup>-4</sup>
protein modification by small protein conjugation or removal (GO:0070647)	217	7	0.57	12.31	1.85*10 <sup>-6</sup>	3.02*10 <sup>-4</sup>

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

**Table S3. *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	agttaaca	-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
<b><u>human</u></b>	<b><u>43</u></b>	<b><u>30</u></b>	<b><u>cggtgatgg</u></b>	<b><u>1.7613</u></b>	<b><u>0.68</u></b>	<b><u>13</u></b>	<b><u>13</u></b>	<b><u>21</u></b>	<b><u>0.68379</u></b>
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

**Table S3 (continued). *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
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
<b><u>mouse</u></b>	<b><u>29</u></b>	<b><u>33</u></b>	<b><u>gcctgacag</u></b>	<b><u>2.01588</u></b>	<b><u>0.75</u></b>	<b><u>6</u></b>	<b><u>15</u></b>	<b><u>31</u></b>	<b><u>1.35885</u></b>

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.

**Table S4. Clinical characteristics of the obese non-diabetic patients who underwent liver biopsy at the Antwerp University Hospital (N = 155).<sup>9</sup>**

	<b>mean <math>\pm</math> SD</b>
Gender (% male)	35.5
Age (years)	43.2 $\pm$ 12.9
BMI (kg/m <sup>2</sup> )	39.9 $\pm$ 5.6
Plasma Cholesterol (mg/dL)	203.6 $\pm$ 41.7
HDL-C (mg/dL)	47.5 $\pm$ 14
Plasma TG (mg/dL)	164.9 $\pm$ 86
LDL-C (mg/dL, calculated)	123.6 $\pm$ 37
Fasting Glucose (mg/dL)	86.1 $\pm$ 13.8
Fasting Insulin ( $\mu$ U/mL)	20.1 $\pm$ 18.3
HbA1c (%)	5.6 $\pm$ 0.6
HOMA-IR	4.5 $\pm$ 5.4
ASAT (U/L)	25.2 $\pm$ 14.7
ALAT (U/L)	40.4 $\pm$ 29.7
Alkaline Phosphatase (U/L)	78.1 $\pm$ 20.6
$\gamma$ GT (U/L)	50 $\pm$ 38.4

**Table S5. Correlations between the expression of spliceosome genes, intron 3 ( IVS3), or full length *LDLR* in the livers of obese non-diabetic patients (n=155)<sup>9</sup>.**

Gene	IVS3			Full length LDLR		
	Spearman correlation coefficient	Unadjusted p value	Adjusted p value	Spearman correlation coefficient	Unadjusted p value	Adjusted p value
<b>SF3A1</b>	0.01	$9.0 * 10^{-1}$	1.0	0.13	$1.1 * 10^{-1}$	1.0
<b>SF3A2</b>	-0.21	$8.9 * 10^{-3}$	$9.8 * 10^{-2}$	-0.085	$3.0 * 10^{-1}$	1.0
<b>SF3B1</b>	0.26	$1.3 * 10^{-3}$	$1.4 * 10^{-2}$	0.13	$9.9 * 10^{-2}$	1.0
<b>SF3B2</b>	0.066	$4.1 * 10^{-1}$	1.0	0.024	$7.6 * 10^{-1}$	1.0
<b>SF3B4</b>	0.065	$4.2 * 10^{-1}$	1.0	0.029	$7.2 * 10^{-1}$	1.0
<b>SF3B5</b>	0.067	$4.1 * 10^{-1}$	1.0	0.1	$2.0 * 10^{-1}$	1.0
<b>SF3B6</b>	0.058	$4.7 * 10^{-1}$	1.0	0.061	$4.5 * 10^{-1}$	1.0
<b>RBM22</b>	-0.031	$7.1 * 10^{-1}$	1.0	-0.0012	$9.9 * 10^{-1}$	1.0
<b>RBM25</b>	0.14	$8.4 * 10^{-2}$	1.0	-0.018	$8.2 * 10^{-1}$	1.0
<b>ISY1</b>	0.12	$1.3 * 10^{-1}$	1.0	0.086	$2.9 * 10^{-1}$	1.0
<b>AQR</b>	0.13	$1.2 * 10^{-1}$	1.0	-0.044	$5.9 * 10^{-1}$	1.0

Coefficients of correlation were calculated according to Spearman. Adjustment of p values for multiple testing was done according to Bonferroni correction.

**Table S6: 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
<b>5'UTR</b>	16858387	-0.050	0.062	-0.092	-0.079
<b>Exon1</b>	16858388	-0.034	-0.047	-0.047	-0.030
<b>Exon2</b>	16858390	0.15	-0.0060	0.090	0.13
<b>Exon2</b>	16858391	0.094	-0.000032	0.034	0.059
<b>Intron2</b>	16858392	0.034	-0.055	0.022	0.018
<b>Intron3</b>	16858394	0.11	0.013	0.061	0.096
<b>Exon4</b>	16858395	0.023	-0.018	-0.021	0.029
<b>Exon4</b>	16858396	0.10	-0.070	0.065	0.10
<b>Intron4</b>	16858397	0.062	0.056	0.011	0.0082
<b>Exon5</b>	16858398	0.081	-0.042	0.033	0.077
<b>Exon6</b>	16858399	0.11	-0.11	0.12	0.16
<b>Exon7</b>	16858400	0.013	0.018	-0.019	0.00027
<b>Exon8</b>	16858401	-0.038	-0.032	-0.049	-0.035
<b>Exon9</b>	16858402	0.077	-0.062	0.092	0.084
<b>Exon10</b>	16858403	0.057	0.0017	0.019	0.031
<b>Exon11</b>	16858404	0.013	0.030	-0.029	-0.0092
<b>Exon12</b>	16858405	0.072	-0.032	0.026	0.062
<b>Exon13</b>	16858406	0.058	0.075	0.033	0.023
<b>Exon13</b>	16858407	0.16	0.11	0.061	0.083
<b>Exon15</b>	16858409	0.22	0.037	0.13	0.18
<b>Intron15</b>	16858410	0.067	0.063	0.0088	0.038
<b>Exon16</b>	16858411	0.034	0.059	-0.056	-0.010
<b>Exon17</b>	16858413	0.084	-0.0038	0.048	0.063
<b>3'UTR</b>	16858415	0.078	0.034	0.095	0.069
<b>3'UTR</b>	16858416	0.10	0.042	0.067	0.062

Data measured at baseline in the cohort described in Supplementary Table IV (n=155 patients)<sup>9</sup>. log<sub>2</sub> transformed intensities of probes depicted in figure 5D were correlated with plasma lipid levels by Spearman. None of the correlations is statistically significant after adjustment for multiple testing with the Bonferroni correction (adjusted p value threshold 5\* 10<sup>-4</sup>).

**Table S7. Association between variants in spliceosome hit genes and multiple biochemical phenotypes in 40,468 UK Biobank exomes<sup>11</sup>.**

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

**Table S7 (continued). Association between variants in spliceosome hit genes and multiple biochemical phenotypes in 40,468 UK Biobank exomes<sup>11</sup>.**

	<b>Gene</b>	<b>LDL-C</b>	<b>ApoB</b>	<b>TG</b>	<b>HDL-C</b>	<b>ApoA</b>	<b>Chol.</b>	<b>Lp(a)</b>	<b>Glc</b>	<b>HbA1c</b>
Eur	<i>AQR</i>	0.92	0.86	0.77	0.5	0.67	0.97	0.36	0.34	0.97
	<i>ISY1</i>	0.79	0.87	0.3	0.75	0.92	0.9	0.65	0.52	0.15
Ethnicity	<i>RBM25</i>	0.29	0.15	0.63	0.18	0.072	0.5	0.24	0.53	0.32
	<i>SF3A1</i>	0.96	0.96	0.56	0.89	0.67	0.94	1	0.45	0.88
LOF-based	<i>SF3A2</i>	0.77	0.7	0.79	0.061	0.12	0.78	0.94	0.84	0.25
	<i>SF3B2</i>	0.34	0.22	0.27	0.79	0.85	0.52	0.63	0.45	0.23
model	<i>SF3B5</i>	0.4	0.26	0.3	0.16	0.2	0.52	1	0.075	0.11
	<i>SF3B6</i>	0.83	0.76	0.71	0.92	0.84	0.83	0.79	0.27	0.48

No association was significant, neither at the genome-wide level (p value threshold  $3.13 \times 10^{-6}$ ) nor after correcting the p value for limited hypothesis testing (p value threshold 0.004545). Data are expressed as p values. Associations with p value below 0.05 are highlighted in bold. Non-standard abbreviations: Chol.: Total cholesterol; Glc: Glucose

**Table S8. Gene-based constraint information from the gnomAD database for the spliceosomal hit genes.**

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

pLI = probability of being loss-of-function intolerant. pLI is an estimate of deleteriousness based on variant frequencies within that gene used in the GnomAD database. Metrics to measure a transcript's intolerance to variation by comparing the observed number of variants in the gnomAD dataset with the expected number of variants. Transcripts that are significantly depleted of their expected variation are considered constrained, or intolerant, of such variation. The closer pLI is to 1, the more intolerant of protein variants the transcript appears to be.  $pLI \geq 0.9$  is considered as an indication of extreme intolerance. For more details see supplement of reference<sup>12</sup> and <https://gnomad.broadinstitute.org/>).

**Table S9. Meta analysis of associations between the rs 17570658 SNP of *RBM25* with LDL cholesterol**

<b>Dataset</b>	<b>Sample size</b>	<b>Z Score</b>	<b>p-value</b>
UK Biobank Mendelian Trait GWAS	333541	-4.702	$2.6 * 10^{-6}$
Hoorn DCS2018	3414	-2.636	$8.4 * 10^{-3}$
GoDarts Illumina Infinium GWAS	1884	-1.282	$2.0 * 10^{-1}$
GLGC GWAS	84987	1.072	$2.8 * 10^{-1}$
Hoorn DCS2019	1997	-0.974	$3.3 * 10^{-1}$
Extend GWAS	7159	-0.697	$4.9 * 10^{-1}$
GoDarts Illumina Human OMNI Express	2902	-0.1	$9.2 * 10^{-1}$
Copenhagen City Heart and General Population Studies	19653	-0.722	$4.7 * 10^{-1}$
<b>All together</b>	<b>455537</b>	<b>-4.181</b>	<b>P = <math>2.9 * 10^{-5}</math></b>

All data except from Copenhagen City Heart and General Population Studies<sup>13</sup> are from the CardioVascular Disease Knowledge Portal (<https://cvd.hugeamp.org/variant.html?variant=rs17570658>) and were reanalyzed according to METAL<sup>14</sup>.

**Table S10. *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	$9 \times 10^{-6}$	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<sup>15</sup> 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; GT: genotype; het: heterozygote; MAF: minor allele frequency in the GnomAD database; PPH2: PolyPhen 2; SIFT: Sorting Intolerant From Tolerant ;MT: Mutation Taster; PSD: Possibly Damaging; PD: Probably Damaging; D: Damaging; DC: Disease Causing

**Table S11: PCR primersequences for all primer couples used in this study.**

<b>Experiment/primer</b>	<b>5'-3' Sequence</b>
<b>LDLRret fragment cloning</b>	
F	ATGGGGCCCTGGGGCTGG
R	TCAGATAGGCTCAATAGCAAAGGCA
<b>Short pSPL3 minigene cloning</b>	
F	CTAGCGAATTTCGCAAAGACAGGATTGGCAAGG
R	CTAGCGGATCCTGGTATGAGCCCCCAAGAGA
<b>Long pSPL3 minigene cloning</b>	
F	GCTAGCGAATTCTGCAAAGACAGGATTGGCAAG
R	CTAGCGGATCCTGGAAATCCACTTCGGCACC
<b>Determination of siRNA knockdown efficiency in Huh-7</b>	
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	GGCCCCGAAATGCAGAAAAGG
ISY1 – R	GGCCAGAAAGGGTCTTCGTT

**Table S11 (continued): PCR primer sequences for all primer couples used in this study.**

<b>Experiment/primer</b>	<b>5'-3' Sequence</b>
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	CCCATGTTTCGTCATGGGTGT
GAPDH – R	TGGTCATGAGTCCTTCCACGATA
<b>LDLR expression and IVS3 retention in human liver and Huh-7 cells (from<sup>5</sup>)</b>	
LDLR canonical transcript - F	GACAACGGCTCAGACGAGCA
LDLR canonical transcript - R	CCACAGGTGAGCACCGGGCA
IVS3 retention – F	GTGATGGTGGTCTCGGCCCA
IVS3 retention – R	GGACCACAGGTGAGCACCGG

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

**Table S12: gene boundaries for the annotation of the UK Biobank GWAS data.**

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

Coordinates refer to the GRCh37/hg19 build

**4. Major Resources Tables**

**Antibodies**

<b>Target antigen</b>	<b>Vendor or Source</b>	<b>Catalog #</b>	<b>Working concentration (µg/ml)</b>	<b>Lot #</b>	<b>Persistent ID / URL</b>
LDLR	Abcam,	Ab52818	0.67 (WB)	GR3253668-2	<b>AB881213</b>
RBM25	Abcam	Ab72237	0.2 (WB)	GR3364959-1	<b>AB1270216</b>
Polyclonal Goat anti-Rabbit HRP Conjugated Immunoglobulins	Dako, Agilent Pathology Solutions,	P0448	0.05 (WB)	20061231	<b>AB2617138</b>
Polyclonal Rabbit anti-Mouse HRP conjugated Immunoglobulins	Dako, Agilent Pathology Solutions	P0260	0.26 (WB)	20039216	<b>AB2636929</b>
Polyclonal Rabbit Anti-Goat HRP conjugated Immunoglobulins	Dako, Agilent Pathology Solutions	P0449	0.1 (WB)	00062104	<b>AB2617143</b>
HRP-conjugated anti-HA antibody	Sigma Aldrich	H6533	2 (WB)	NA	<b>AB439705</b>
TBP	Abcam,	Ab51841	1 (WB)	GR103882-2	<b>AB945758</b>
Beta actin	Santa Cruz Biotechnology	SC-1616	0.1 ug/ml	NA	<b>AB630836</b>
LDLR	Santa Cruz Biotechnology	sc- 18823	2 (flow cytometry)	NA	<b>AB627881</b>
LDLR	PROGEN Biotechnik GmbH	61087	2 (flow cytometry)	207100-01	<b>AB1542236</b>
Isotype control antibody	Santa Cruz Biotechnology)	sc-3879	2 (flow cytometry)	L0816	<b>AB737262</b>
chicken anti-Mouse IgG (H+L) Cross-Adsorbed AlexaFluor 488-conjugated Secondary Antibody	Thermo Fisher Scientific	A-21200	4 (flow cytometry)	NA	<b>AB2535786</b>
Goat anti-Mouse IgG (H+L) Cross-Adsorbed AlexaFluor 647-conjugated Secondary Antibody	Thermo Fisher Scientific	A-21236	4 (flow cytometry)	1915660	<b>AB2535805</b>

**DNA/cDNA Clones**

<b>Clone Name</b>	<b>Sequence</b>	<b>Source / Repository</b>	<b>Persistent ID / URL</b>
LDLR ret fragment	<u>ENST00000557958.1: nt 87-437</u>	Cloned in house	NA
HA tagged LDLR fragment	<u>N-terminally HA-tagged version of ENST00000557958.1: nt 87-437</u>	Invitrogen GeneArt Strings DNA Fragments (Thermo Fisher 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**

<b>Name</b>	<b>Vendor or Source</b>	<b>Catalog #</b>	<b>Sex (F, M, or unknown)</b>	<b>Persistent ID / URL</b>
Huh-7	Japanese Collection or Research Bioresources Cell Bank, JCRB Cell Bank, Osaka, Japan	JCRB0403	M	<b>CVCL_0336</b>
HepG2	American Type Culture Collection, ATCC, Manassas, VA, USA	HB-8065	<u>M</u>	<b>CVCL_0027</b>
HEK293T	American Type Culture Collection, ATCC, Manassas, VA, USA	CRL-3216	F	<b>CVCL_0063</b>

**Data & Code Availability**

<b>Description</b>	<b>Source / Repository</b>	<b>Persistent ID / URL</b>
Homo Sapiens reference genome	Data Sciences Platform at the Broad Institute/ build GRCh38 and build GRCh37/h19	<a href="https://gatk.broadinstitute.org/hc/en-us/articles/360035890951-Human-genome-reference-builds-GRCh38-or-hg38-b37-hg19">https://gatk.broadinstitute.org/hc/en-us/articles/360035890951-Human-genome-reference-builds-GRCh38-or-hg38-b37-hg19</a> and <a href="https://gatk.broadinstitute.org/hc/en-us/articles/360035890711-GRCh37-hg19-b37-humanG1Kv37-Human-Reference-Discrepancies">https://gatk.broadinstitute.org/hc/en-us/articles/360035890711-GRCh37-hg19-b37-humanG1Kv37-Human-Reference-Discrepancies</a>
UCSC Genome Browser Variant Annotation Integrator	Variant Annotation Integrator	<a href="https://genome.ucsc.edu/cgi-bin/hgVai">https://genome.ucsc.edu/cgi-bin/hgVai</a>
U2 branchpoint prediction algorithm	SVM-BP-finder	<a href="http://regulatorygenomics.upf.edu/Software/SVM_BP/">http://regulatorygenomics.upf.edu/Software/SVM_BP/</a>
Expression of LDLR transcripts in healthy human liver	Gene Expression Omnibus,	GSE126848/ <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126848L">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126848L</a>
SNP data from UK Biobank	UK Biobank	<a href="http://www.nealelab.is/uk-biobank">http://www.nealelab.is/uk-biobank</a>
Meta analysis of rs 17570658	CardioVascular Disease Knowledge Portal	( <a href="https://cvd.hugeamp.org/variant.html?variant=rs17570658">https://cvd.hugeamp.org/variant.html?variant=rs17570658</a> )
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	<a href="https://www.uk10k.org/data_access.html">https://www.uk10k.org/data_access.html</a>
Expression analysis of spliceosome genes in various tissues	Genotype-Tissue Expression (GTEx) project	<a href="https://gtexportal.org/home">https://gtexportal.org/home</a>

**Data & Code Availability (continued)**

<b>Description</b>	<b>Source / Repository</b>	<b>Persistent ID / URL</b>
gnomAD Databank	genome aggregation database	<a href="https://gnomad.broadinstitute.org/">https://gnomad.broadinstitute.org/</a>
RNA sequencing of Huh7 cells treated with siRNAs against spliceosome genes or scrambled siRNA	European Nucleotide Archive	<a href="https://www.ebi.ac.uk/ena/browser/support">https://www.ebi.ac.uk/ena/browser/support</a> : accession codes PRJEB46898 and PRJEB46899

**siRNAs and Transfection reagents**

<b>Description</b>	<b>Source / Repository</b>	<b>Catalog #</b>	<b>Persistent ID / URL</b>
Lipofectamine™ RNAiMAX Transfection Reagent	Thermo Fisher Scientific	13778150	<a href="https://www.thermofisher.com/order/catalog/product/13778150#/13778150">https://www.thermofisher.com/order/catalog/product/13778150#/13778150</a>
Lipofectamine™ 2000 Transfection Reagent	Thermo Fisher Scientific	11668019	<a href="https://www.thermofisher.com/order/catalog/product/11668019#/11668019">https://www.thermofisher.com/order/catalog/product/11668019#/11668019</a>
Lipofectamine™ 3000 Transfection Reagent	Thermo Fisher Scientific	L3000015	<a href="https://www.thermofisher.com/order/catalog/product/L3000015#/L3000015">https://www.thermofisher.com/order/catalog/product/L3000015#/L3000015</a>
The Ambion Silencer Select Human Genome siRNA library V4	Thermo Fisher Scientific	4397926	<a href="https://www.thermofisher.com/order/catalog/product/4397926?SID=srch-hj-4397926#/4397926?SID=srch-hj-4397926">https://www.thermofisher.com/order/catalog/product/4397926?SID=srch-hj-4397926#/4397926?SID=srch-hj-4397926</a>
Ambion Silencer Select anti- <i>PLK1</i>	Thermo Fisher Scientific	assay ID s448, cat. 4390824	<a href="https://www.thermofisher.com/order/genome-database/details/sirna/s448?CID=&amp;ICID=&amp;subtype=">https://www.thermofisher.com/order/genome-database/details/sirna/s448?CID=&amp;ICID=&amp;subtype=</a>
Ambion Silencer Select anti- <i>LDLR</i>	Thermo Fisher Scientific	assay ID s4, cat. No. 4392420	<a href="https://www.thermofisher.com/order/genome-database/details/sirna/s4?CID=&amp;ICID=&amp;subtype=sirna_silencer_select">https://www.thermofisher.com/order/genome-database/details/sirna/s4?CID=&amp;ICID=&amp;subtype=sirna_silencer_select</a>
Silencer™ Select Negative Control No. 1 siRNA	Thermo Fisher Scientific	4390843	<a href="https://www.thermofisher.com/order/catalog/product/4390843#/4390843">https://www.thermofisher.com/order/catalog/product/4390843#/4390843</a>

**siRNAs and Transfection reagents (continued)**

<b>Description</b>	<b>Source / Repository</b>	<b>Catalog #</b>	<b>Persistent ID / URL</b>
<i>SF3A1</i>	Dharmacon	L-016051-00-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10291">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10291</a>
<i>SF3A2</i>	Dharmacon	L-018282-02-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-8175">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-8175</a>
<i>SF3B1</i>	Dharmacon	L-020061-01-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-23451">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-23451</a>
<i>SF3B2</i>	Dharmacon	L-026599-01-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10992">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10992</a>
<i>SF3B4</i>	Dharmacon	L-017190-00-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10262">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-10262</a>
<i>SF3B5</i>	Dharmacon	L-014706-02-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-83443">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-83443</a>
<i>SF3B6</i>	Dharmacon	L-020260-02-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-51639">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-51639</a>
<i>RBM22</i>	Dharmacon	L-021186-01-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-55696">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-55696</a>
<i>RBM25</i>	Dharmacon	L-021976-00-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-58517">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-58517</a>

**siRNAs and Transfection reagents (continued)**

<i>ISY1</i>	Dharmacon	L-013894-01-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-57461">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-57461</a>
<i>AQR</i>	Dharmacon	L-022214-01-0005	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-9716">https://horizondiscovery.com/en/gene-modulation/knockdown/sirna/products/on-targetplus-sirna-reagents?nodeid=entrezgene-9716</a>
Non-targeting control pool	Dharmacon	D-001810-10-05	<a href="https://horizondiscovery.com/en/gene-modulation/knockdown/controls/products/on-targetplus-non-targeting-control-pool?catalognumber=D-001810-10-05">https://horizondiscovery.com/en/gene-modulation/knockdown/controls/products/on-targetplus-non-targeting-control-pool?catalognumber=D-001810-10-05</a>
MISSION® predesigned siRNA against RBM25	Sigma-Aldrich	SASI_Hs02_00354878	<a href="https://www.sigmaaldrich.com/CH/en/semi-configurators/sirna?activeLink=selectAssays">https://www.sigmaaldrich.com/CH/en/semi-configurators/sirna?activeLink=selectAssays</a>
MISSION® siRNA Universal Negative Control #2	Sigma-Aldrich	SIC002	<a href="https://www.sigmaaldrich.com/CH/en/product/sigma/sic002">https://www.sigmaaldrich.com/CH/en/product/sigma/sic002</a>

