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

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

* These authors contributed equally to this work

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. Univ. 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

Word count: 7993 words

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

Abstract

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

Methods and results: In a genome-wide RNAi screen, the knock-down of 54 genes led to a significant inhibition of LDL uptake. Fifteen of these genes encode for proteins involved in splicing, especially components or interactors of the U2-spliceosome. Knocking down any one of 11 out of 15 genes resulted in the selective retention of intron 3 of *LDLR*. The transcript is translated into an LDLR fragment, which lacks 88% of the full length LDLR and is detectable in cells and their medium upon overexpression, but neither in non-transfected cells nor in human plasma. Surprisingly, the intron 3 retention transcript is expressed in considerable amounts in human liver and in blood cells. Its expression correlates with plasma LDL-cholesterol and age and increases after bariatric surgery. Single nucleotide polymorphisms and three rare variants of one spliceosome gene, *RBM25,* are associated with LDL-cholesterol in the population and familial hypercholesterolemia, respectively. Compared to overexpression of wild type RBM25, overexpression of the three rare RBM25 mutants led to lower LDL uptake by Huh7 cells.

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

Introduction

Hypercholesterolemia is a causal and treatable risk factor of atherosclerotic cardiovascular diseases (ASCVD)[1]. The most important determinant of LDL-cholesterol (LDL-C) levels in plasma is the hepatic removal of circulating LDL by binding to LDL receptors (LDLR) for subsequent endocytosis and degradation[2]. The expression of LDLR is tightly regulated by transcription factors, proteasomal and lysosomal degradation, endosomal recycling, and cleavage at the cell surface[1,2]. The unravelling of this complex regulation led to the development of drugs that effectively lower plasma levels of cholesterol and, as the consequence, risk of ASCVD[1]. Nevertheless, there is still a considerable gap of knowledge in the understanding of the LDLR pathway. For example, LDL-C levels increase with aging from a mean level of 1 mmol/L (39 mg/dl) in newborns to a mean level of 3 to 4 mmol/L (116 to 155 mg/dl) in middle-aged adults, by mechanisms that are not understood[3].

To identify novel regulators of LDL uptake into the liver, we performed an image-based genome-wide RNA interference screen in Huh-7 human hepatocarcinoma cells. Fifteen out of 54 genes significantly reducing LDL uptake upon knockdown encode for proteins involved in pre-mRNA splicing. The majority of them are either core components or interactors of the U2-spliceosome[4]. By functionally validating this finding *in vitro* as well as in human tissues, we provide evidence that a functional U2 spliceosome is needed for the expression of full length LDLR and, hence, a determinant of LDLR activity in humans.

Methods

Data Availability

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

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

Results

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

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

For the uptake of fluorescent LDL, the five best performing assay features (foci count per cell, foci mean intensity, cytoplasm granularity 1 and 2, cytoplasm median intensity) showed a high degree of correlation. Therefore and because of the widest dynamic range based on Z'-factor values from control wells, we identified gene hits by the Redundant siRNA Activity (RSA) analysis of data from the median cytoplasm intensity feature. Z'-factor values for median cytoplasm intensity in each assay plate for both the background (median 0.00, interquartile range [IQR] -023 to 0.20) and positive control (median -0.56, IQR -0.99 to -0.20) clustered mostly around the 0-line, indicating a suboptimal although still analytically exploitable signal-to-noise ratio (Supplemental Figure Ib). Dimensionality reduction of all five main assay features did not significantly alter the outcome (Supplemental Figures Ic and Id). At an RSA p-value cutoff of $P < 0.001$, interference with 54 and 37 genes decreased and increased LDL uptake, respectively (Table 1, Supplemental Table I). Surprisingly and by contrast to the findings in a previous reported genome wide CRISPR-based screening in Huh7 cells[5], our list does not include LDLR or its modulators such as SCAP, MBTPS1, or IDOL/MYLIP except AP2M1, which is an essential contributor to clathrin mediated endocytosis (table 1). Gene Ontology (GO) enrichment analysis showed significant clustering only for genes whose loss of function decreased LDL uptake (Supplemental Table II). Functional clustering of these genes with the STRING tool revealed four major groups: the ribosome (N = 7), the proteasome (N = 8), the spliceosome (N = 15), and vesicular transport (N = 5) (Figure 1B). Out of the 15 spliceosome genes, six encode for core components of the U2 spliceosome, namely *SF3A1*, *SF3A2*, *SF3B1*, *SF3B2*, *SF3B5* and *SF3B6*. Other proteins, interact with the U2-spliceosome either directly (AQR[4,6], ISY1[4,6] and RBM25[7]) or indirectly (RBM22)[8].

To confirm the role of the U2 spliceosome in LDL endocytosis *in vitro*, we performed ^{125}I-LDL cell association assays in Huh-7 as well as in HepG2 cells. *SF3B4* was also included in

these experiments as it is part of the U2 spliceosome and barely missed the RSA p-value cut-off of 0.001 (p = 0.0014). Knockdown was achieved using 4 pooled siRNA molecules against each hit gene acquired from vendors other than that of the siRNAscreening library, namely Dharmacon or Sigma instead of Ambion. (see Major Resource Table).. For RBM25 we replaced Dharmacon's siRNAs with those from Sigma because of their off-target effects on LDLR protein expression (Supplemental Figures IIId and IIIe). Knockdown of each of these genes significantly decreased the specific cell association of ^{125}I-LDL with both Huh-7 and HepG2 cells (Figure 1C, Supplemental Figure IIb). The association of ^{125}I-LDL was equally decreased by knockdown of *SF3B1* (-45±5%), *SF3A2* (-47±6%), *AQR* (-47±8%), and *LDLR* (-43±8%) (Figure 1C). RNA interference with RBM25 reduced the specific cellular association of ^{125}I-LDL and the uptake of fluorescent LDL by 26% ±6% and 52%±5%, respectively (Figure 1D and Supplemental Figure IIIf). Unspecific internalization of lipoproteins was ruled out by the finding of normal or even increased specific cell association of ^{125}I-HDL upon knockdowns of *AQR* and *SF3A1* in either Huh-7 or HepG2 cells (Supplemental Figures IIc and IId).

Loss of U2 spliceosome genes and their interactors causes selective retention of *LDLR* intron 3

To unravel the mechanism through which the U2-spliceosome and its interactors regulate LDL endocytosis, we applied RNA sequencing to Huh-7 cells, which were transfected with either siRNAs against eleven spliceosome genes or a non-targeting control siRNA. 72h after transfection, we measured both expression at the gene level and alternative exon usage in polyA-selected transcripts. Strikingly, knockdown of all eleven genes except *RBM25* induced a marked increase in the retention of intron 3 of *LDLR* in mature transcripts without altering the expression of the *LDLR* full length transcript (Figure 2A, Supplemental

Figure IV). This effect was confirmed in Huh-7 cells by RT-PCR upon knockdown of AQR, *SF3B1, or RBM25* by employing a primer set that was previously used to study the effects of the rare LDLR c.313+1, G>A intronic variant, which leads to *LDLR* loss of function by constitutively promoting intron 3 retention[9] (Supplemental Figure Va). By contrast to the RNA sequencing (Supplemental Figure IV), RT-PCR unravelled increased expression of the LDLR IVS3 retention transcript upon knock-down of RBM25, albeit not as much as with knock-down of SF3B1 and AQR (Supplemental Figures Vb and Vc).

Among all intronic or exonic sequences in the transcriptome, the expression of the intron 3 retaining *LDLR* transcript was altered most strongly (. Upon knockdown of *SF3B1, AQR, or SF3A2*, the retained intronic sequence of *LDLR* ranked at the top of each respective dataset when the exon-level expression data was plotted against each other (Figure 2B). The degree of intron 3 retention upon knocking down U2-spliceosome genes was significantly correlated with the decrease in ^{125}I-LDL cell association, suggesting a direct mechanistic link ($R^2 = 0.5984$, $p = 0.0052$, Figure 2C).

To investigate why the *LDLR* gene allows intron 3 retention, we transfected HEK293T cells with two minigenes containing different portions of the *LDLR* genomic sequence flanked by two artificial exons (Figure 3A). The first minigene (MG_1) encoding only for exon 3 of *LDLR* and the adjacent intronic regions cloned between two artificial exons (SD6 and SA2), displayed very low if any RNA sequencing reads mapping to the first ~130bp of intron 3. On the contrary, upon expression of the whole genomic sequence between the 3'-end of intron 2 and the 5'-end of intron 4 of *LDLR* (MG_2) an increased number of reads mapped to the first section of intron 3. This indicatesincomplete splicing of intron 3 when the physiological exon 4 acceptor site and the branch point site (BPS) were present in the larger minigene MG2 (Figure 3B). The acceptor splice site of exon 4 of *LDLR* hence appears to be poorly defined. The bioinformatic analysis of the portion of intron 3

neighbouring exon 4 by the U2 branchpoint prediction algorithm SVM-BP-finder (http://regulatorygenomics.upf.edu/Software/SVM_BP/)[10] identified one plausible U2-spliceosome dependent BPS located 30 nucleotides upstream of the acceptor site (Supplemental Table III). The *gtgat* pentamer in the centre of the *cggtgatgg* branchpoint sequence was associated with very low U2 binding energy and occurs at low frequency in the branchpoint database[10]. We discarded another predicted branchpoint 124bp upstream of the acceptor site as the subsequent AG-exclusion zone does not reach up to the acceptor. Contrary to exon 4 of human *LDLR*, exon 4 of murine *Ldlr* contains a strong and frequently recurring branchpoint 33 base pairs upstream of the acceptor site (Figure 3C). This finding is in accordance with intron 3 of *Ldlr* being barely detectable at the RNA level by RT-PCR in the mouse liver (data not shown). Taken together, this data suggests that the BPS of intron 3 in human *LDLR* is poorly defined and therefore very sensitive to alternative splicing.

Selective intron 3 retention limits LDLR cell surface abundance

The transcript with intron 3 retention encodes for a prematurely truncated proteoform of LDLR because the 5'-end of intron 3 encodes for 12 novel amino acids followed by a stop codon. Including the signal peptide, this theoretical 116 amino acid residues long and 12.7 kDa large 'LDLRret fragment' encompasses the complete first and large parts of the second class A domains (labelled as L1 and L2 in Figure 4A[11]) but lacks all other domains, including the transmembrane portion of LDLR. Western blots probed with an antibody against the C-terminus of LDLR revealed 60±30% and 61±13% lower LDLR protein levels upon knockdown of AQR and *SF3B1*, respectively *(*Figures 4B and 4C). A similar decrease in LDLR protein was seen upon knock-down of RBM25 with siRNAs from Sigma (-68%±10%), whereas the knock-down of RBM25 with the siRNA of Dharmacon led to an increase in LDLR protein (+122%±109%), presumably due to off target effects

(Supplemental Figures IIId and IIIe). Flow cytometry experiments on live Huh-7 cells after *SF3B1* and *AQR* knockdown showed a -87%±1% and –61%±4%, respectively, lower cell surface abundance of LDLR (Figure 4D). The knock-down of RBM25 with siRNAs from Sigma and Dharmacon decreased the cell surface abundance of LDLR by 54%±4% and 21%±5% , respectively, as compared to scrambled siRNAs from the same manufacturers (Supplemental Figure IIIg).

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

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

To investigate its physiological or pathological relevance, we quantified *LDLR* intron 3 retention in liver biopsies as well as in peripheral blood cells of healthy individuals from the general population, and explored associations with demographic measures, lipid traits, and therapeutic interventions.

The bioinformatics analysis of RNA sequencing data on 13 liver samples of healthy and untreated donors (Gene Expression Omnibus, accession number GSE126848)[12] found 14 rather lowly expressed LDLR transcripts (Supplemental Figure VII). Subsequent to the most

abundant transcripts LDLR-208 and LDLR-201, which encode full length LDLR, LDLR-206, which corresponds to the retained intron 3 transcript, is the third most abundant transcript. It reflects 2.3% of all reads mapping compared to total reads from all transcripts of LDLR gene across 13 examined samples. Other transcripts including those encoding full length LDLR (-202, -203, -204, -205, -207, and -208) as well as transcripts with retained introns (-209, -211, and -213) are even less expressed.

By RT-PCR on mRNAs of healthy liver tissue from 17 patients undergoing partial hepatectomy because of benign liver tumours and nine patients with suspected NAFLD, we found the *LDLR* intron 3 retention transcript expressed at considerable and interindividually variable amounts (Figure 5A). Taking the sum of the full length and intron 3 retention transcripts of *LDLR* as the reference, 48±19 % and 38±11 % of the transcripts retained intron 3 in livers of patients with tumours and NAFLD, respectively (Figure 5A).

We next investigated in 155 obese non-diabetic subjects from the Obesity Clinic of the Antwerp University Hospital[13] whether intron 3 retention of *LDLR* in liver biopsies correlates with the expression of spliceosome genes and plasma levels of lipoproteins, and whether it is influenced by therapeutic interventions. Supplemental Table IV describes the clinical and biochemical characteristics of the study population. As analysed with Affymetrix Human Gene 2.0 ST arrays, the signal intensities from a probe located in intron 3 of *LDLR* were significantly higher than the other intronic *LDLR* probes located in introns 2, 4 and 15 and comparable to probes located in coding exons (Figure 5B). Intron 3 retention correlated significantly with the expression of 5 spliceosome genes, most strongly with *SF3B1* ($r=0.278$, $p=4.7*10^{-4}$) and *AQR* ($r=0.219$, $p=6.0*10^{-3}$; Figure 5C), while only *SF3B1* showed significant correlation with overall *LDLR* expression ($r=0.161$, $p=4.53*10^{-2}$, Figure 5D). Relative intensities of neither theIntron 3 probe nor any other of the 24 *LDLR* probes showed significant correlations with plasma levels of total, HDL- or LDL-cholesterol (Supplemental

Figures VIIIa, VIIIb and VIIIc, Supplemental Table V). . In a subgroup of 21 patients who underwent a second liver biopsy after bariatric surgery (median follow-up time = 13 months, IQR = [12, 15]),the proportion of the intron 3 retention transcript relative to the full length LDLR transcript increased significantly after surgery (p = 0.00077) although LDL-C levels slightly decreased (Supplemental Figure VIIId; supplemental table IV). This increase was even more pronounced in eleven patients with non-alcoholic steatohepatitis (NASH) at baseline but no NASH at follow-up (p = 0.010, Supplemental Figure VIIIe). Interestingly, however, intron 3 relative probe intensities did not correlate with BMI. Correlations with histological NAFLD stages were inverse by trend but significant only for ballooning, a marker of apoptosis or degeneration of hepatocytes (R^2 = 0.029; P = 0.035) (Supplemental Figures VIIIf and VIIIg).

Finally, we analysed the RNA-sequencing data in whole blood samples from 2,462 subjects of the Dutch BIOS consortium[14]. The *LDLR* ENST00000557958 transcript, predicted to retain intron 3, was detectable in all subjects and represented 21%±7% of the total *LDLR* transcripts on average. The ENST00000557958 transcript levels significantly correlated with age (Spearman r = 0.25, p = $2.3*10^{-36}$, Figure 6A) and less strongly with LDL-C (r = 0.089, p = $9.8*10^{-6}$, Figure 6B). The latter correlation lost its statistical significance after adjusting for age, suggesting age itself as the main driver of the association between ENST00000557958 levels and LDL-C. ENST00000252444 was the only transcript encoding for full length *LDLR* in this dataset that was expressed in blood in all subjects. This transcript was also positively correlated with age (r = 0.185, p = $2.2*10^{-20}$, Figure 6C) but not with LDL-C (r = -0.033, p = 0.10, Figure 6D). Correlation of neither transcript with BMI was statistically significant.

Human genetic data support a role of *RBM25* in the regulation of plasma LDL-C levels in humans

Publicly available whole exome sequencing (WES) data of 40,468 UK Biobank subjects[15] did not unravel any significant association between our spliceosome hit genes and any lipid or lipoprotein level including LDL-C (Supplemental Table VI). However, constraints data from the gnomAD database indicate a strong intolerance to functional genetic variation for our U2-spliceosomeht genes components, with a probability of intolerance to loss of function (pLI)[16] of 0.91±0.17 (mean± SD) (Supplemental Table VII). Upon analysis of SNPs of 11 spliceosome hit genes in 361,194 participants of UK Biobank, we found 24 SNPs of *RBM25* significantly associated with lower levels of LDL-C (Figure 7A) and apoB (Supplemental Figure IXa) (Bonferroni correction threshold of $p < 3 \times 10^{-5}$ for 1360 SNPs in 11 spliceosome genes). The lead SNP is rs17570658 which exhibits an unusual linkage disequilibrium (LD) pattern (Supplemental Figure IXb). In Europeans there are two upstream SNPs in almost complete LD, and four additional SNPs in complete LD, spanning the whole *RBM25* gene, with the downstream SNPs occurring in introns or downstream of the *RBM25* coding sequence (none being exonic). There are essentially no other SNPs in strong LD (i.e. $R^2 > 0.8$). *RBM25* is widely expressed in many tissues, but expression is relatively low in liver (GTeX https://gtexportal.org/home/, data not shown). rs17570658 shows strong association with *RBM25* expression in 15 different tissues including skeletal muscle and arteries (Figure 7B) as well as adipose and mammary tissue, lung, oesophagus, kidney, and skin. Carriers of the rare allele have higher mean RBM25 mRNA concentration, which is compatible with higher LDLR activity and lower plasma LDL-C.

When compared to 1926 controls in the UK10K study, *RBM25* was also among the genes identified to harbour an excess of rare novel variants in 71 patients with familial hypercholesterolemia who are negative for mutations in *LDLR*, *APOB* and *PCSK9*, the

known FH-causing genes[17]. [17]We re-analysed the burden of variants in the *RBM25* gene, using previously published WES data from 71 FH patients negative for mutations in *LDLR*, *APOB* and *PCSK9*, sequenced as part of the UK10K project[18], and 56,352 European data provided by the gnomAD study[16]. Missense, splice site, frameshift, and stop-gained variants identified by WES in both FH cases and gnomAD were filtered to select those with MAF<0.0001). After filtering, three *RBM25* variants were found in the FH cohort and 163 in the gnomAD Europeans cohort. (Supplemental Table VIII). Two variants, p.I152F (c454A>T) and p.A455D (c.1364C>A), were not found in any publicly available sequencing database and hence appear unique to the FH cohort. The third variant, p.L17P (c.50T>C) (rs1167173761), was found in one European individual in the gnomAD cohort (MAF=$9*10^{-6}$, allele count = 1/251402). The comparison of variant numbers in FH cases vs. gnomAD using a binomial test demonstrated the enrichment of rare variants in *RBM25* in the FH cohort (p = 0.001). Within the UK10K cohort, no other spliceosome gene was found to carry a rare presumable LOF mutation.

Functional characterization of RBM25 mutants

Finally, we investigated the functional consequences of the three RBM mutants associated with FH. RBM25 overexpression was confirmed for all constructs via qPCR (Supplemental Figures Xa and XIa) and - for wild type RBM25 - Western blotting (Supplemental Figure Xb). The overexpression of neither wild type RBM25 nor any RBM25 mutant caused significant changes in the expression of full length or IVS3 retention transcripts of LDLR (Supplemental Figures Xc, Xd, XIb, and XIc). Compared to empty vector, overexpression of wild type RBM25 changed neither cell suface abundance of the LDL-receptor nor LDL uptake (Supplemental Figures Xe and Xf). Comparisons with cells overexpressing wild type RBM25 revealed minor decreases of LDLR cell surface levels but more pronounced or even

significant decreases of Atto594-LDL-uptake by cells overexpressing the p.L17P (-15%+16%), I152F (-23%+12%, P<0.05), or p.A455D (-32%+12%, P<0.05) mutants of RBM25 (Supplemental Figure XIe).

Discussion

Through an unbiased genome-wide siRNA screening strategy, we discovered that the U2 spliceosome as well as some interacting proteins, control LDLR levels and LDL uptake in liver cells by modulating the selective retention of intron 3 of *LDLR*. The intron 3 retaining *LDLR* transcript encodes a truncated and most probably non-functional version of the receptor. In several cohorts of healthy individuals and patients, we observed that an interindividually variable and frequently large proportion of intron 3 of *LDLR* is retained in the liver as well as in peripheral blood cells. Finally, we obtained initial evidence that rare genetic variants in at least one spliceosome-associated gene, namely *RBM25*, as well as SNPs associated with its expression levels are related to LDL-C levels in humans. Taken together, our findings suggest intron 3 retention of *LDLR* as a novel mechanism regulating LDLR activity and thereby plasma levels of LDL cholesterol.

A previous siRNA screen also found U2 spliceosome genes to limit the upake of LDL into EA.hy926 cells but the authors excluded them from further analysis and validation[19]. Basic cellular functionality of spliceosome genes may be the reason why U2 spliceosome genes were not found by a previous CRISPR-based screen as limiting factors for LDL uptake into Huh7 cells[5]. As these authors discuss, CRISPR-based screens may overlook genes that are essential or confer a fitness advantage in culture, since gRNAs targeting those genes will be progressively depleted from the pooled population [5].

As a preliminary mechanistic explanation, our minigene data as well as our *in silico* predictions suggest that the BPS in intron 3 of human LDLR is poorly defined and thereby highly sensitive to alterations in the activity of U2 splice factors. In this regard it is noteworthy that the rare c.313+1, G>A intronic variant leads to loss of *LDLR* function by constitutively promoting intron 3 retention[9].

Interestingly, the knockdown of proteins that assemble on the U2-spliceosome only at later stages such as AQR and ISY1[6] induces the same effect as the knockdown of core-components of the spliceosome. This indicates a more complex mechanism in which not only BPS recognition by small nuclear ribonucleic acids and ribonucleoprotein but possibly also the 3-dimensional structure of intron 3 during subsequent splicing steps (such as after the formation of the B_{act} complex)[4] may represent a challenge for the spliceosome[4]. Medina and colleagues previously found alternative splicing of *HMGCR*, *HMGCS1*, *MVK*, *PCSK9*, and *LDLR* to be mediated by the splice protein PTBP1 and regulated by cellular cholesterol levels[20]. Interestingly, PTBP1 works as an inhibitor of the U2AF splice component, and thus inhibits the recognition of 3' splice sites by the U2 spliceosome[21]. However, the knockdown of *PTBP1* resulted in very limited changes in the expression levels of the different splice forms[22], especially when compared to the drastic changes observed in our study.

In our *in vitro* experiments, the knock-down of several U2-spliceosome genes and the resulting intron 3 retention compromised LDLR cell surface expression and LDL uptake as much as *LDLR* knockdown. As yet the sensitivity of our mass spectrometric analysis only allowed detection of the tagged fragment after overexpression in the immortalized kidney cell line HEK293T . The artificial construct unlike an endogenously produced protein may have escaped nonsense-mediated decay. Nevertheless, we cannot rule out that the theoretical 116 amino acid long aminoterminal fragment of the differentially spliced *LDLR* is expressed *in vivo* and secreted. In fact, human plasma contains LDLR fragments, which are currently

assumed to result from shedding of LDLR at the cell surface[23] but may also correspond to secreted alternative splice variants.

In humans, we found a weak but significant correlation between plasma LDL-C levels and the IVS3 retention *LDLR* transcript in peripheral blood cells, which was stronger than the correlation with the full-length *LDLR* transcript. Smaller sample size and narrower range of LDL-C levels but also differences between tissues may be the reasons, why no significant correlations of LDL-C with any hepatic LDLR transcript expression were found. The decrease of LDL-C but increase of LDLR IVS3 retention after bariatric surgery may question whether LDLR transcript expression in peripheral blood cells is a reporter of hepatic LDLR splicing. However, the associations of *RBM25* SNPs with high *RBM25* expression and lower LDL-C levels in UK Biobank and the higher than expected prevalence of rare *RBM25* loss-of-function variants in FH patients with no mutation in canonical FH genes suggest that the regulation of *LDLR* splicing by the U2 spliceosome contributes to the determination of LDL-C levels in humans. The lack of association of hypercholesterolemia with rare variants of any other U2 spliceosome gene may reflect their intolerance to gross variation as suggested by pLI values close to 1. In this regard, it is important to note that our analysis of WES data of UK biobank only retrieved heterozygous mutations in spliceosome genes whereas our knock-down experiments rather mimic homozygous conditions. Therefore, and because in the liver biopsies maximally 38±11% (obesity cohort) to 48±19% (benign liver tumor cohort) of *LDLR* transcripts were affected by intron 3 retention, a complete loss of a spliceosome gene may be needed to manifest severe hypercholesterolemia. Opposite effects on upstream regulators of *LDLR* may be another reason why the majority of SNPs and rare exome variants of the spliceosome gene do not show any association with LDL-C levels. The exclusive association of LDL-C with *RBM25* variants may also indicate that *RBM25* regulates LDL-C levels by mechanisms unrelated to the U2-spliceosome and intron 3 retention. In fact, *RBM25* also

partakes in other spliceosomal subunits[24]. In this regard, it is also important to note, that RNA interference with *RBM25* had the weakest effects on *LDLR* splicing and that overexpression of hypercholesterolemia associated RBM25 mutants resulted in lower LDL uptake without affecting the expression of the LDLR IVS3 transcript in comparison to overexpression of the wild type RBM25

The correlation between ENST00000557958 expression levels in blood cells with age suggests the intriguing possibility that age-related changes in the activity of the U2-spliceosome provide a mechanism for the increase in LDL-C that parallels ageing [3]. The reason for this increase is not yet understood. In general, the functionality of the splicing process changes with ageing[25] . Somatic mutations or decreased expression of splice factor genes notably SF3B1 and RBM25, respectively have been implicated in age-related processes, including cancer[7,26,27]. The total number of alternatively spliced genes also increases with age[28]. Until recently, *SIRT1* is the only known gene involved in cholesterol metabolism and atherosclerosis[29] whose alternative splicing may be disrupted with age[25]. One may speculate that either the epigenetic dysregulation of the activity of splice factor genes or the accumulation of somatic loss-of-function variants in liver cells may promote increases in LDL-C with age.

Our study has several strengths and limitations. First, our screening unravelled several novel candidate genes that regulate hepatic LDL uptake but missed canonical LDL uptake regulating genes such as MYLIP, MBTPS1, PCSK9 or SREBBP2 which were rediscovered by another previous screen that used CRISPR/Cas9 technology. A general reason is the not optimal signal to noise ratio of our screen. A specific reason for the missing of MYLIP or PCSK9 is the optimization ofour screening towards the discovery of loss of function effects. . Second, our validation studies did not only confirm the limiting effect of U2 spliceosome genes on LDL uptake but unravelled a novel mechanism of LDL receptor regulation, namely

intron 3 retention within an LDLR transcript which is translated into a truncated and non-functional receptor protein. In both liver and peripheral blood cells, we demonstrate that this process happens at considerable quantity in the human organism and is influenced by aging as well as interventions such as bariatric surgery.. Third, RBM25 was the only spliceosome gene affected by mutations associated with differences in LDL-C, perhaps because RBM25 may tolerate loss of function better than other U2 spliceosome genes. However, we cannot rule out that RBM25 affects LDL metabolism beyond or even independently of LDLR splicing because both knock-down of RBM25 and overexpression of loss of function mutants associated with hypercholesterolemia exerted stronger and more consistent effects on LDL uptake than on intron 3 retention in *LDLR.* ..

In conclusion, we identified intron 3 retention of *LDLR* by the U2-spliceosome and its interactors as a novel mechanism regulating LDLR activity in cells. The importance of this mechanism for the regulation of plasma LDL-C levels especially with increasing age and thus determination of cardiovascular risk remains to be f established by further studies.

Acknowledgements and Sources of Funding

This work was conducted as part of the "TransCard" project of the 7th Framework Program (FP7) granted by the European Commission, to J.A.K., A.T.H. and A.v.E. (number 603091). This work was also partially supported by the FP7 "RESOLVE" project (to J.T.H., B.S., A.V., S.F. L.V.G., and A.v.E.) and the European Genomic Institute for Diabetes (EGID, ANR-10-LABX-46 to BS).

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

P.Z. received funding by the AGLA-Award of the Swiss Atherosclerosis Society and the DACH Award from the DACH Society for Prevention of Cardiovascular Diseases

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

The project was also supported by GeniusII (CVON2017-2020).

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

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

J.T.H. was supported by an EMBO Long Term Fellowship (ALTF277-2014). B.S. is a recipient of an ERC Advanced Grant (no. 694717).

Research at the Antwerp University Hospital was supported by the European Union: FP6 (HEPADIP Contract LSHM-CT-2005-018734) and FP7-HEALTH (RESOLVE nr. 305707).

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

We facknowledge the use of data from BIOS-consortium
(http://wiki.bbmri.nl/wiki/BIOS_bios). which
is funded by BBMRI-NL (NWO project 184.021.007).
Flow cytometry was performed with equipment of the flow cytometry facility, University of Zurich.

Disclosures

.

The collaborators at the ETH ScopeM facility (R.M., M.S., S.F.N., A.J.R, S.S.), which hosted the high-content screening reported in this manuscript, report payments from A.v.E.'s as, by contract, 1/3 of the service costs needs to be paid by customers to cover part of the costs for personnel, infrastructure, and maintenance.

Supplemental Materials

Materials & Methods

Online Figures I to XI

Online Tables I to X

Major Resource Table

Original Western Blots

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

Decreased LDL uptake				Increased LDL uptake			
Gene	**Assay score[A] avg[B]**	**Assay score[A] SEM[C]**	**RSA p-value**	**Gene**	**Assay score[A] avg[B]**	**Assay score[A] SEM[C]**	**RSA p-value**
AP2M1	-3.103179681	0.346648222	3.36E-08	*PROX1*	6.53057396	0.631260417	3.19E-09
CHMP2A	-3.130900347	0.359445533	2.51E-07	*ITGAV*	7.431175355	1.519558432	2.96E-08
NFKB2	-2.59157417	0.136886566	8.07E-07	*TGFBR1*	3.464028514	0.397588943	7.31E-06
AQR[D]	-2.484868551	0.199482589	4.57E-06	*CDC37*	3.747034032	1.072191825	2.35E-05
PSMD11	-2.557101583	0.239773488	4.77E-06	*DTNBP1*	57.92944887	57.2617451	4.46E-05
SF3B2	-2.107311389	0.015210399	4.81E-06	*CYP27C1*	32.06817221	31.61438081	8.92E-05
RPL35	-2.346954606	0.150946677	5.45E-06	*PNPLA2*	2.279278207	0.266420784	1.26E-04
PSMD8	-2.988677308	0.491086915	6.34E-06	*C22orf39*	7.995494448	8.342242785	1.78E-04
SON	-2.164748153	0.201099955	1.46E-05	*TMEM133*	3.049034762	1.060165442	1.84E-04
COPA	-2.307675328	0.213018879	1.61E-05	*TMEM130*	6.466491664	5.317700355	2.23E-04
RBM25	-1.993998657	0.055265194	1.92E-05	*PM20D2*	2.155202336	0.176491898	2.29E-04
RBM22	-2.818121291	0.622885617	3.36E-05	*PET117*	3.001069341	1.652765441	2.68E-04
PSMD3	-2.21302629	0.224903034	3.98E-05	*CWF19L2*	3.806757511	4.571696977	3.12E-04
SF3B5	-2.285064158	0.25878823	4.32E-05	*ENY2*	2.420424347	0.514153659	3.28E-04
SF3B1	-2.267932169	0.253122099	4.55E-05	*NME4*	2.711491413	0.954425612	3.39E-04
SALL4	-1.937993523	1.13065979	6.02E-05	*ZC3H4*	4.545156994	3.551478266	3.57E-04
RPL5	-2.106905493	0.373542859	7.40E-05	*WASF2*	2.310874515	0.449202822	3.61E-04
CCDC180	-1.132459235	1.398333381	9.52E-05	*HELZ2*	2.546828237	0.984740435	3.87E-04
SF3B6	-2.277000896	0.332074616	9.83E-05	*RILP*	1.995550072	0.267567916	4.23E-04
HNRNPU	-1.724036435	0.093847304	1.23E-04	*MAT2A*	3.705559066	3.611891772	4.91E-04
RPL17	-2.226845162	0.329575956	1.46E-04	*NRM*	1.710898743	0.050727817	5.02E-04
ISY1	-2.74487386	0.698388989	1.55E-04	*CEP295NL*	2.189792071	0.474598108	5.02E-04
ZNF641	-1.034460324	1.453444444	2.58E-04	*ACSM2A*	2.207444199	1.531809937	5.32E-04
COPB1	-1.693933632	0.103029465	2.64E-04	*RTL9*	3.759306708	3.473297986	5.35E-04
SF3A1	-2.225755015	0.412106586	2.72E-04	*KIAA1522*	3.362058267	3.27466253	6.25E-04
SNW1	-1.76539067	0.142531611	2.76E-04	*ZNF84*	2.204388764	0.765657329	6.55E-04
EIF2S1	-1.486721463	0.790741651	3.45E-04	*TFAP4*	3.032765033	3.340175625	6.69E-04
CCDC73	-1.041204586	1.27682775	3.50E-04	*TMEM182*	3.227517874	1.666669492	7.29E-04
RPL9	-1.715182797	0.249911985	3.55E-04	*WDR55*	1.967286849	1.365170916	7.32E-04
NXNL2	-1.199311468	1.135835784	3.83E-04	*DYNLL1*	2.268266743	0.467997927	7.72E-04
WBP11	-1.50591484	0.062444555	4.03E-04	*ADPRHL2*	2.078229013	0.322800093	8.51E-04
C2CD5	-1.097951788	1.954971449	4.46E-04	*ELAVL1*	1.945364959	0.968117905	8.70E-04
RPL21	-1.655773242	0.156797718	4.72E-04	*CFAP298*	1.883199038	0.378258022	8.87E-04
EPOP	-1.837314819	0.25795876	4.80E-04	*PMM1*	2.80926863	3.200260012	8.92E-04
RMND5B	-1.523957521	0.076773849	5.07E-04	*CASKIN2*	1.681223061	0.149986926	9.07E-04
TAPBPL	-1.52965773	0.154207886	5.27E-04	*CIZ1*	3.454694336	2.803876145	9.37E-04

STARD10	-1.527795273	0.115135889	5.45E-04	*BRICD5*	1.962503862	0.408074057	9.41E-04
PSMD1	-2.207116523	0.551747426	5.63E-04				
PFDN6	-0.881689024	1.740601376	5.80E-04				
PSMA1	-1.528976301	0.119805079	5.85E-04				
RTF2	-1.573924771	0.169765686	6.14E-04				
LSM2	-1.448888015	0.056454941	6.40E-04				
UBD	-1.171691024	1.530009178	6.69E-04				
LRRC14	-1.258311764	1.067910962	6.84E-04				
SUPT6H	-1.451332382	0.095214513	7.27E-04				
COPB2	-2.037140764	0.468882876	7.34E-04				
SF3A2	-1.347147433	0.758462926	7.89E-04				
ATP6V0C	-1.823918839	0.263639476	7.90E-04				
EMILIN3	-1.598631472	2.238859705	8.03E-04				
DMTN	-1.559252376	0.142024687	8.20E-04				
MRPL19	-0.755460842	1.688052373	8.92E-04				
MRO	-0.986783025	1.102624895	9.14E-04				
DDX59	-1.380513222	1.040634076	9.25E-04				
PSMD12	-1.761325035	0.367766123	9.45E-04				

[A]Assay score: normalized score for the median cytoplasm intensity assay feature. [B]Avg = average.

[C]SEM: standard error of the mean.

[D]The 15 hit genes inducing a decrease in LDL uptake that, based on a literature research, are involved in RNA splicing are highlighted in bold and underlined.

Figure Legends

Figure 1. Identification and validation of U2 spliceosome genes as limiting factors for the uptake of LDL by Huh7 hepatocytes. A. Schematic representation of the genome-wide image-based siRNA screening and data analysis process. **B. Functional association networks** for genes decreasing LDL uptake upon siRNA-mediated knockdown. Genes with $P < 0.001$ for median cytoplasm intensity were selected as top hits ,. Spheres represent single genes. Edges represent known and predicted gene-gene relationships such as protein-protein interactions, co-expression and homology. The graph was produced using the STRING online tool (http://string-db.org/). The superimposed coloured circles are used to highlight the main functional clusters. **C and D. Effects of RNA-interference with U2-spliceosome genes with cell association of ^{125}I-LDL with Huh-7 cells.** 72h after transfection with siRNAs from Ambion (LDLR), Sigma (RBM25), or Dharmacon (all other genes)**,** cells were incubated for 2h at 37°C in the presence of 33.3µg/ml of ^{125}I-LDL in the presence or absence of 100x excess unlabelled LDL. Specific cell association was calculated as the difference between the two conditions. A minimum of at least 8 replicates were measured for each condition. Data are expressed as means ±SD; Significance was calculated by one-way ANOVA followed by Dunnett`s post-hoc test (C) or unpaired t-test between each vendor's coding and non-coding (scrambled) siRNAs (D). ns = not significant, * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$.

Figure 2. Loss of U2 spliceosome genes causes intron 3 retention in LDLR. A. ***LDLR*** **Exon level expression upon *AQR* knockdown**.Expression of the *LDLR* exons was recorded by RNA sequencing of Huh-7 cells 72h after knockdown of *AQR*. Segments represent differential exon usage in each sector of the *LDLR* genomic sequence as identified by the DEXSeq algorithm. Canonical exons within the ENSG00000130164 genomic reference are shown below the graph.

Normalized read counts are reported on the y axis. The black arrow indicates the location of ENSG00000130164:E009, corresponding to the first half of intron 3. Data represent the average of 3 replicate experiments. **B. ENSG00000130164:E009 is most strongly upregulated upon RNA interference with spliceosome genes.** Log2 fold change in gene expression at the exon level for the whole transcriptome after knockdown of *AQR* (x axis), and *SF3B1* (y axis) and *SF3A2* (z axis) in Huh-7 cells. The red circle highlights the position of ENSG00000130164:E009 corresponding to the first half of intron 3. **C. Correlation between LDLR intron 3 retention and LDL cell association.** Correlation between the log2 fold change in ENSG00000130164:E009 expression level and the decrease in ^{125}I-LDL cell association (same data as in Figure 1C) upon knockdown of each spliceosome hit gene. Cells that received a non-targeting siRNA were used as the reference. Cell association is expressed as mean±SD. R^2, p-value and the regression line were calculated by the linear regression function of GraphPad Prism, ver. 5, without constraints and according with its default parameters.

Figure 3. Determination of LDLR intron 3 splice patterns. A. Cloning strategy and structure of the minigenes. The upper part of this panel shows the genomic location of the two segments of the LDLR gene that were cloned in each minigene, while the lower half shows a simplified structure of the pSPL3 minigene used to express them. Genomic coordinates refer to the hg19 assembly. Note that, due to primer design, MG_1 is actually 1nt shorter at its 5' end, starting at chr19:11,212,960. **B. Characterization of the splice products.** The graphs represent the RNA sequencing coverage at the Exon3-Intron 3 junction in two replicate samples for each condition. Coverage data were normalized to the average coverage for exon 3. MG_1/MG_2 = short / long minigene. **C. *In silico* BPS predictions for the acceptor site of *LDLR* exon 4.** BP score: final score (svm_score) according to the SVM-BP-finder algorithm for the

putative BPS sequence highlighted in red. A BPS is considered valid when located close to the AG exclusion zone, with BP-score > 0 and with svm_score > 0.

Figure 4. Effect of loss of spliceosome function on LDLR protein expression A. Schematic structure of the LDLR protein. (modified from[11]). LDLRret: intron 3 retention fragment, LBD: Ligand binding domain; L1-L7: LDLR class A domain; EGFPH: Epidermal growth factor precursor homology domain; β: beta propeller; O: O-linked sugar repeat; A/B/C: EGF-type repeat; TM: transmembrane domain. The red line represents the location of the last canonical amino acid found also in the LDLRret fragment, followed by 12 novel amino acids and by a stop codon. **B and C. Effect of *SF3B1* and *AQR* knockdown on LDLR protein levels.** LDLR protein levels in Huh-7 cells 72h after *SF3B1* or *AQR* knockdown. **B** shows a representative Western blot. **C,** shows the relative densities of LDLR bands normalized to TATA-binding-protein (TBP, loading control) after knockdown of AQR or SF3B1 relative to scrambled control.. Bars represent means ±SD of three replicate experiments. **D. Effect of *SF3B1* and *AQR* knockdown on LDLR cell surface levels.** LDLR cell surface levels in live Huh-7 cells measured by flow cytometry 72 hours after knockdown of *SF3B1* or *AQR*. siRNAs against *LDLR* were used as the positive control. The data are normalized to a non-silencing control. Each point represents one of three experiments. Bars represent means±SD. . * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$ **** = $p<0.0001$ (One-way ANOVA followed by Dunnett's post-hoc test). **E-F. Overexpressed LDLRret fragment is retrieved in cell lysates and cell culture medium.** 48 hours after transfection in HEK293T cells, the HA-tagged version of the LDLRret fragment was detected-by western blot in both total cell lysates **(E,F)** and media **(F)**. Lysates after 2 and more hours of incubation were obtained after incubation with the proteasome inhibitor MG132 as indicated by the labels in (**E**) EV = pcDNA3.1 empty vector. HA-frag = hemagglutinin-tagged LDLRret fragment.

Figure 5. *LDLR* intron 3 retention in human liver. A. Detection of intron 3 retention in the human liver by RT-PCR. Transcripts encoding full-length *LDLR* transcript or the IVS3 retention variant were measured by RT-PCR and normalized to GAPDH mRNA levels in the healthy liver tissue of 17 patients with benign liver tumors and in liver biopsies of 9 patients with suspected NAFLD.. Data are expressed as mean relative expression ±SD after normalizing to the sum of both *LDLR* transcripts **B. Exon-level *LDLR* expression in the human liver.** The boxplot shows the normalized signal intensities for probes mapping to the 5'-UTR, 3'-UTR, the exons, and some introns of the *LDLR* gene in 155 obese non-diabetic subjects. The red box indicates the normalized signal levels for intron 3. The other introns are shown in grey. The location of each probe is depicted in the diagram below. **C and D. Correlations between the expression of spliceosome genes and the intron 3 probe (C) or the full length *LDLR* probe (D).** The scatterplot depicts the Pearson's correlation coefficients and their $-\log_{10}$(p-values) for the correlations between the signal intensities of probes for the spliceosome genes and the intron 3 retention probe (IVS3) (**C**) or the *LDLR* full length probe gene (**D**). Negative Pearson's coefficients (-0.092 and -0.145) for correlations of SF3A2 with IVS3 and full length LDLR, respectively, are not displayed for simplicity.

Figure 6. Correlations of the *LDLR* ENST00000557958 (A, B) and ENST00000252444 transcripts (C, D) in whole blood samples with age (A, C) and LDL-C levels (B, D). Data is from 2,462 subjects of the BIOS population[14]. ENST00000557958 represents the intron 3 retention transcript (**A, B**). ENST00000252444 (**C, D**) was the only full-length *LDLR* transcript detected in all samples analysed. Violin plots describe the frequency distribution of the transcripts in quartiles of age or LDL-C quartile. The red horizontal lines reflect medians; the

upper and lower borders of the boxes reflect the 75^{th} and 25^{th} percentiles, respectively. r-values and p-values refer to a Spearman correlation analysis.

Figure 7. Association between *RBM25* variants and LDL-C in the UK Biobank dataset. A. Association of GWAS SNPs from 11 spliceosome genes with LDL-C in the UK Biobank dataset. The dashed red horizontal line indicates the threshold for statistical significance after correction for multiple testing of 1360 variants within the genes of interest ($p=3.7*10^{-5}$). Effect size and directionality are reported on the x axis as beta value. **B. Association between the rs17570658 genotype and *RBM25* expression in different tissues.** Data shown for skeletal muscle and tibial artery (both $p < 10^{-8}$). The horizontal white lines reflect medians; the upper and lower borders of the grey boxes reflect the 75^{th} and 25^{th} percentiles, respectively.

References

1. Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement fromthe European Atherosclerosis Society Consensus Panel. *European Heart Journal*. 2017;38:2459–2472.
2. Zanoni P, Velagapudi S, Yalcinkaya M, Rohrer L, von Eckardstein A. Endocytosis of lipoproteins. *Atherosclerosis*. 2018;275:273–295.
3. Balder JW, de Vries JK, Nolte IM, Lansberg PJ, Kuivenhoven JA, Kamphuisen PW. Lipid and lipoprotein reference values from 133,450 Dutch Lifelines participants: Age- and gender-specific baseline lipid values and percentiles. *Journal of Clinical Lipidology*. 2017;11:1055-1064.e6.
4. Kastner B, Will CL, Stark H, Lührmann R. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harbor Perspectives in Biology*. 2019;a032417.
5. Emmer BT, Sherman EJ, Lascuna PJ, Graham SE, Willer CJ, Ginsburg D. Genome-scale CRISPR screening for modifiers of cellular LDL uptake. *PLoS Genetics*. 2021;17:e1009285.
6. De I, Bessonov S, Hofele R, dos Santos K, Will CL, Urlaub H, Lührmann R, Pena V. The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nature Structural & Molecular Biology*. 2015;22:138–144.
7. Ge Y, Schuster MB, Pundhir S, Rapin N, Bagger FO, Sidiropoulos N, Hashem N, Porse BT. The splicing factor RBM25 controls MYC activity in acute myeloid leukemia. *Nature Communications*. 2019;10.
8. Haselbach D, Komarov I, Agafonov DE, et al. Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. *Cell*. 2018;172:454-464.e11.
9. Cameron J, Holla ØL, Kulseth MA, Leren TP, Berge KE. Splice-site mutation c.313+1, G>A in intron 3 of the LDL receptor gene results in transcripts with skipping of exon 3 and inclusion of intron 3. *Clinica Chimica Acta*. 2009;403:131–135.
10. Corvelo A, Hallegger M, Smith CWJ, Eyras E. Genome-wide association between branch point properties and alternative splicing. *PLoS Computational Biology*. 2010;6:e1001016.
11. Surdo P lo, Bottomley MJ, Calzetta A, et al. Mechanistic implications for LDL receptor degradation from the PCSK9/LDLR structure at neutral pH. *EMBO Reports*. 2011;12:1300–1305.
12. Suppli MP, Rigbolt KTG, Veidal SS, et al. Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *American Journal of Physiology - Gastrointestinal and Liver Physiology*. 2019;316:G462–G472.
13. Lefebvre P, Lalloyer F, Baugé E, et al. Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPARα-regulated dermatopontin. *Journal of clinical investigation*. 2017;2:1–17.
14. Zhernakova D v., Le TH, Kurilshikov A, et al. Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nature Genetics*. 2018;50:1524–1532.

15. Cirulli ET, White S, Read RW, et al. Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nature Communications*. 2020;11:542.
16. Lek M, Karczewski KJ, Minikel E V., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291.
17. Futema M, Plagnol V, Li K, et al. Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR / APOB / PCSK9 mutations. *Journal of Medical Genetics*. 2014;51:537–544.
18. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526:82–89.
19. Kraehling JR, Chidlow JH, Rajagopal C, et al. Genome-wide RNAi screen reveals ALK1 mediates LDL uptake and transcytosis in endothelial cells. *Nature Communications*. 2016;7:13516.
20. Medina MW, Krauss RM. Alternative splicing in the regulation of cholesterol homeostasis. *Current opinion in lipidology*. 2013;24:147–52.
21. Shao C, Yang B, Wu T, et al. Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nature Structural & Molecular Biology*. 2014;21:997–1005.
22. Medina MW, Gao F, Naidoo D, Rudel LL, Temel RE, McDaniel AL, Marshall SM, Krauss RM. Coordinately regulated alternative splicing of genes involved in cholesterol biosynthesis and uptake. *PLoS ONE*. 2011;6:e19420.
23. Mayne J, Ooi TC, Tepliakova L, et al. Associations between Soluble LDLR and Lipoproteins in a White Cohort and the Effect of PCSK9 Loss-of-Function. *Journal of Clinical Endocrinology and Metabolism*. 2018;103:3486–3495.
24. Carlson SM, Soulette CM, Yang Z, Elias JE, Brooks AN, Gozani O. RBM25 is a global splicing factor promoting inclusion of alternatively spliced exons and is itself regulated by lysine mono-methylation. *Journal of Biological Chemistry*. 2017;292:13381–13390.
25. Deschênes M, Chabot B. The emerging role of alternative splicing in senescence and aging. Aging Cell. 2017;16:918–933.
26. Li H, Wang Z, Ma T, Wei G, Ni T. Alternative splicing in aging and age-related diseases. *Translational Medicine of Aging*. 2017;1:32–40.
27. Seiler M, Peng S, Agrawal AA, et al. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Reports*. 2018;23:282-296.e4.
28. Rodríguez SA, Grochová D, Mckenna T, Borate B, Trivedi NS, Erdos MR, Eriksson M. Global genome splicing analysis reveals an increased number of alternatively spliced genes with aging. *Aging Cell*. 2016;15:267–278.
29. Miranda MX, Van Tits LJ, Lohmann C, et al. The Sirt1 activator SRT3025 provides atheroprotection in Apoe-/- mice by reducing hepatic Pcsk9 secretion and enhancing Ldlr expression. *European Heart Journal*. 2015;36:51–59.
30. Velagapudi S, Yalcinkaya M, Piemontese A, et al. VEGF-A regulates cellular localization of SR-BI as well as transendothelial transport of HDL but Not LDL. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2017;37:794–803.
31. McFarlane AS. Efficient trace-labelling of proteins with iodine. *Nature*. 1958;182:53.
32. Kamentsky L, Jones TR, Fraser A, et al. Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software. *Bioinformatics*. 2011;27:1179–1180.

33. Parsons BD, Schindler A, Evans DH, Foley E. A Direct Phenotypic Comparison of siRNA Pools and Multiple Individual Duplexes in a Functional Assay. *PLoS ONE*. 2009;4:e8471.
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–2120.
35. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
36. Liao Y, Smyth GK, Shi W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*. 2013;41:e108.
37. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Research*. 2012;22:2008–2017.
38. Fedoseienko A, Wijers M, Wolters JC, et al. The COMMD family regulates plasma LDL levels and attenuates atherosclerosis through stabilizing the CCC complex in endosomal LDLR traffcking. *Circulation Research*. 2018;122:1648–1660.
39. MacLean B, Tomazela DM, Shulman N, et al. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010;26:966–968.
40. Nisson PE, Watkins PC, Krizman DB. Isolation of exons from cloned DNA by exon trapping. *Current protocols in human genetics / editorial board, Jonathan L Haines . [et al]*. 2001;Chapter 6:Unit 6.1.
41. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 2016;34:525–527.
42. Cirulli, E.T. and Washington NL. Helix Research. UK Biobank Exome rare variant analysis v1.3 [Internet]. Helix Blog. Available from: blog.helix.com/uk-biobank-helix-research
43. Whiffin N, Minikel E, Walsh R, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genetics in Medicine*. 2017;19:1151–1158.
44. Birmingham A, Selfors LM, Forster T, et al. Statistical methods for analysis of high-throughput RNA interference screens. *Nature methods*. 2009;6:569–75.
45. König R, Chiang C, Tu BP, et al. A probability-based approach for the analysis of large-scale RNAi screens. *Nature Methods*. 2007;4:847–849.
46. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000;290:2323–2326.

Figure 1

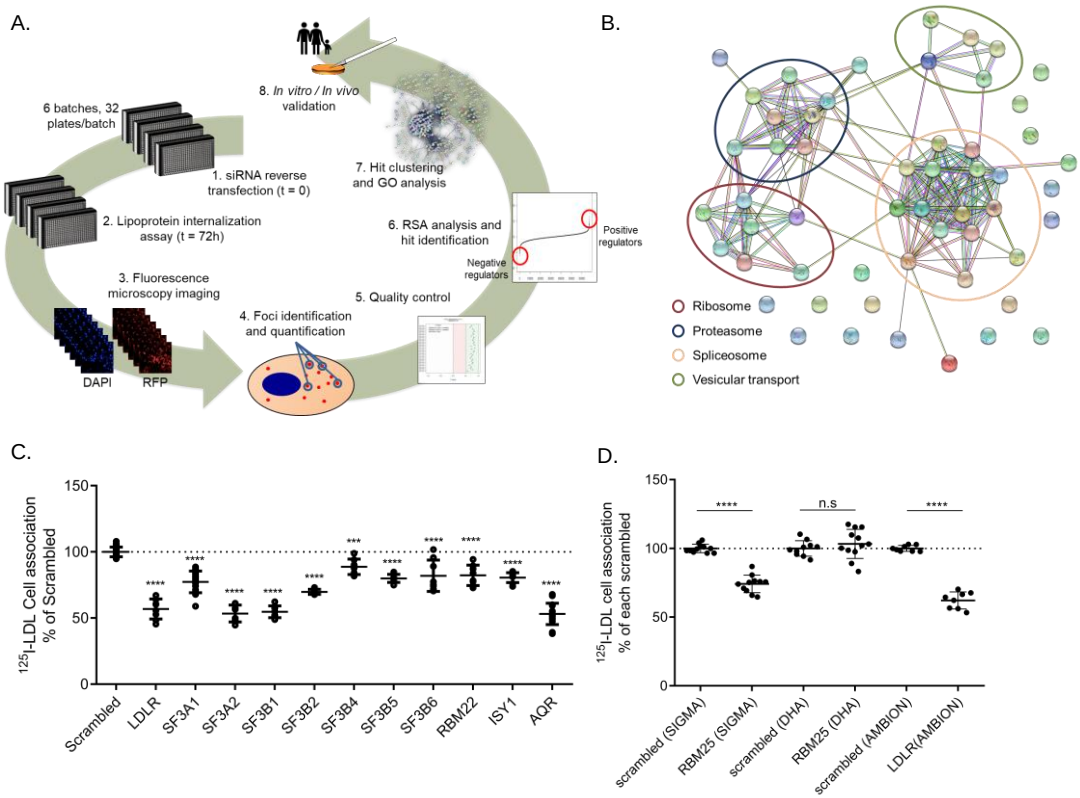

Figure 1. Identification and validation of U2 spliceosome genes as limiting factors for the uptake of LDL by Huh7 hepatocytes. A. Schematic representation of the genome-wide image-based siRNA screening and data analysis process. **B. Functional association networks** for genes decreasing LDL uptake upon siRNA-mediated knockdown. Genes with P < 0.001 for median cytoplasm intensity were selected as top hits ,. Spheres represent single genes. Edges represent known and predicted gene-gene relationships such as protein-protein interactions, co-expression and homology. The graph was produced using the STRING online tool (http://string-db.org/). The superimposed coloured circles are used to highlight the main functional clusters. **C and D. Effects of RNA-interference with U2-spliceosome genes with cell association of ^{125}I-LDL with Huh-7 cells.** 72h after transfection with siRNAs from Ambion (LDLR), Sigma (RBM25), or Dharmacon (all other genes)**,** cells were incubated for 2h at 37°C in the presence of 33.3µg/ml of ^{125}I-LDL in the presence or absence of 100x excess unlabelled LDL. Specific cell association was calculated as the difference between the two conditions. A minimum of at least 8 replicates were measured for each condition. Data are expressed as means ±SD; Significance was calculated by one-way ANOVA followed by Dunnett`s post-hoc test (C) or unpaired t-test between each vendor's coding and non-coding (scrambled) siRNAs (D). ns = not significant, * = p<0.05, ** = p<0.01, *** = p<0.001, **** = p<0.0001.

Figure 2

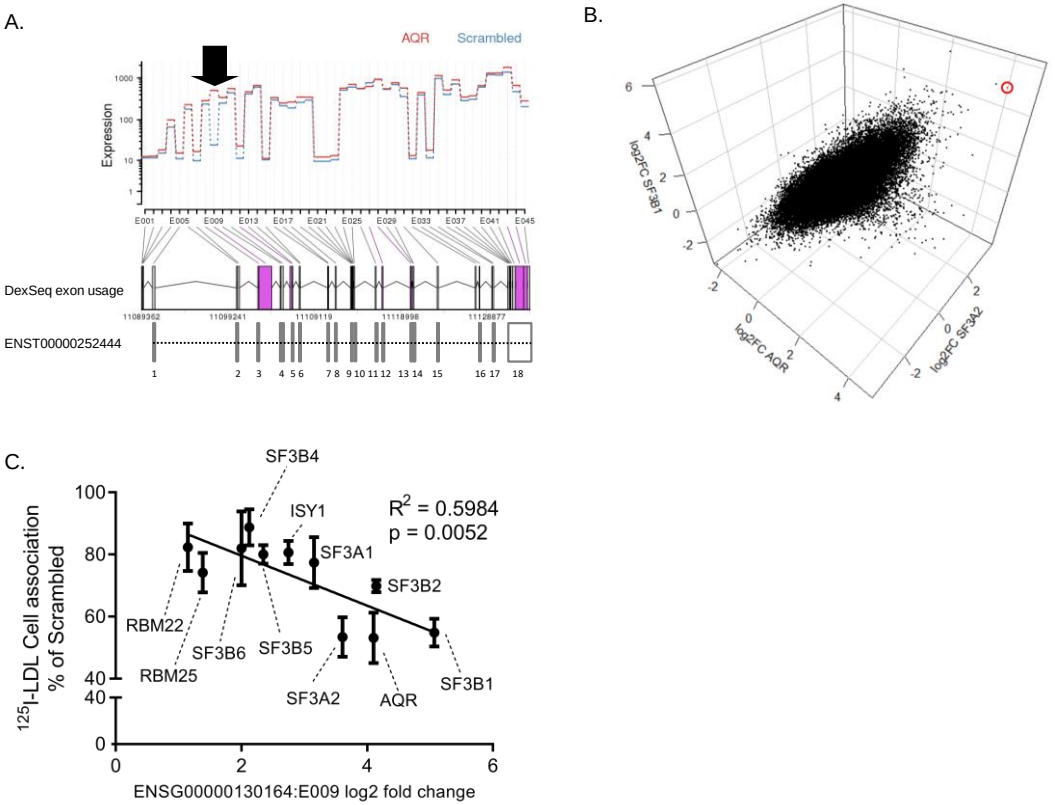

Figure 2. Loss of U2 spliceosome genes causes intron 3 retention in LDLR. A. ***LDLR*** **Exon level expression upon *AQR* knockdown**.Expression of the *LDLR* exons was recorded by RNA sequencing of Huh-7 cells 72h after knockdown of *AQR*. Segments represent differential exon usage in each sector of the *LDLR* genomic sequence as identified by the DEXSeq algorithm. Canonical exons within the ENSG00000130164 genomic reference are shown below the graph. Normalized read counts are reported on the y axis. The black arrow indicates the location of ENSG00000130164:E009, corresponding to the first half of intron 3. Data represent the average of 3 replicate experiments. **B. ENSG00000130164:E009 is most strongly upregulated upon RNA interference with spliceosome genes.** Log2 fold change in gene expression at the exon level for the whole transcriptome after knockdown of *AQR* (x axis), and *SF3B1* (y axis) and *SF3A2* (z axis) in Huh-7 cells. The red circle highlights the position of ENSG00000130164:E009 corresponding to the first half of intron 3. **C. Correlation between LDLR intron 3 retention and LDL cell association.** Correlation between the log2 fold change in ENSG00000130164:E009 expression level and the decrease in ^{125}I-LDL cell association (same data as in Figure 1C) upon knockdown of each spliceosome hit gene. Cells that received a non-targeting siRNA were used as the reference. Cell association is expressed as mean±SD. R^2, p-value and the regression line were calculated by the linear regression function of GraphPad Prism, ver. 5, without constraints and according with its default parameters

Figure 3

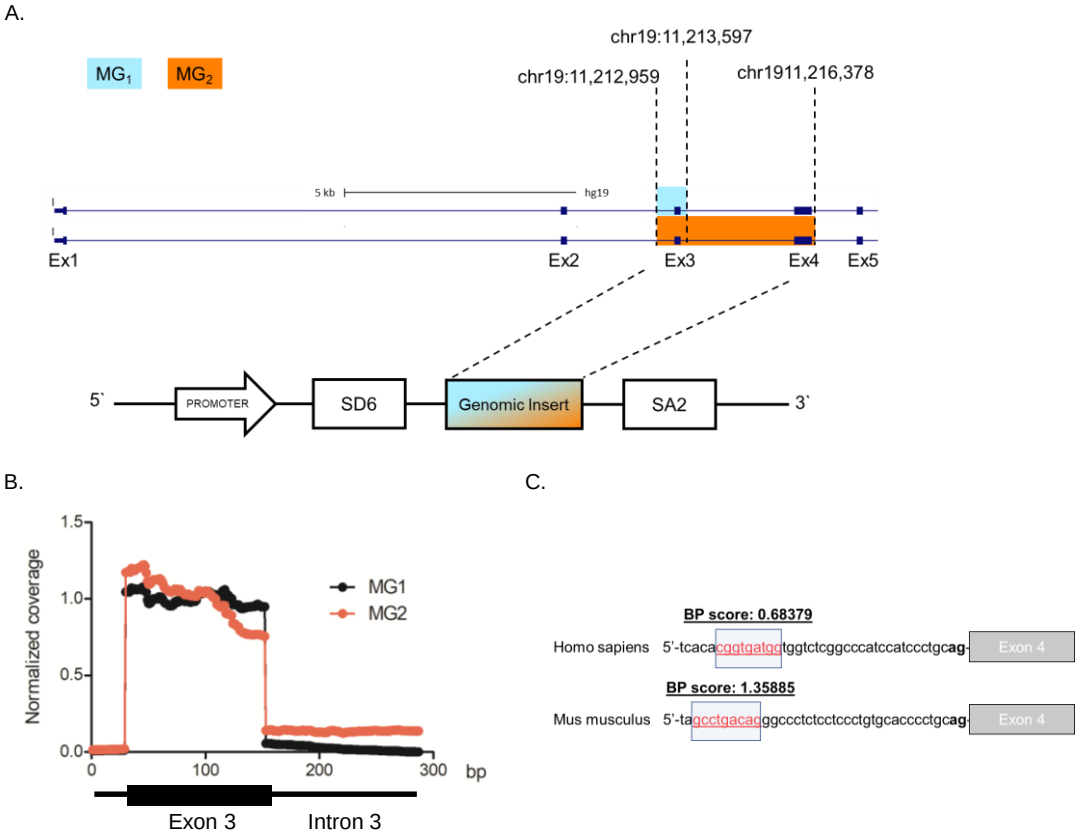

Figure 3. Determination of LDLR intron 3 splice patterns. A. Cloning strategy and structure of the minigenes. The upper part of this panel shows the genomic location of the two segments of the LDLR gene that were cloned in each minigene, while the lower half shows a simplified structure of the pSPL3 minigene used to express them. Genomic coordinates refer to the hg19 assembly. Note that, due to primer design, MG_1 is actually 1nt shorter at its 5' end, starting at chr19:11,212,960. **B. Characterization of the splice products.** The graphs represent the RNA sequencing coverage at the Exon3-Intron 3 junction in two replicate samples for each condition. Coverage data were normalized to the average coverage for exon 3. MG_1/MG_2 = short / long minigene. **C. *In silico* BPS predictions for the acceptor site of *LDLR* exon 4.** BP score: final score (svm_score) according to the SVM-BP-finder algorithm for the putative BPS sequence highlighted in red. A BPS is considered valid when located close to the AG exclusion zone, with BP-score > 0 and with svm_score > 0

Figure 4

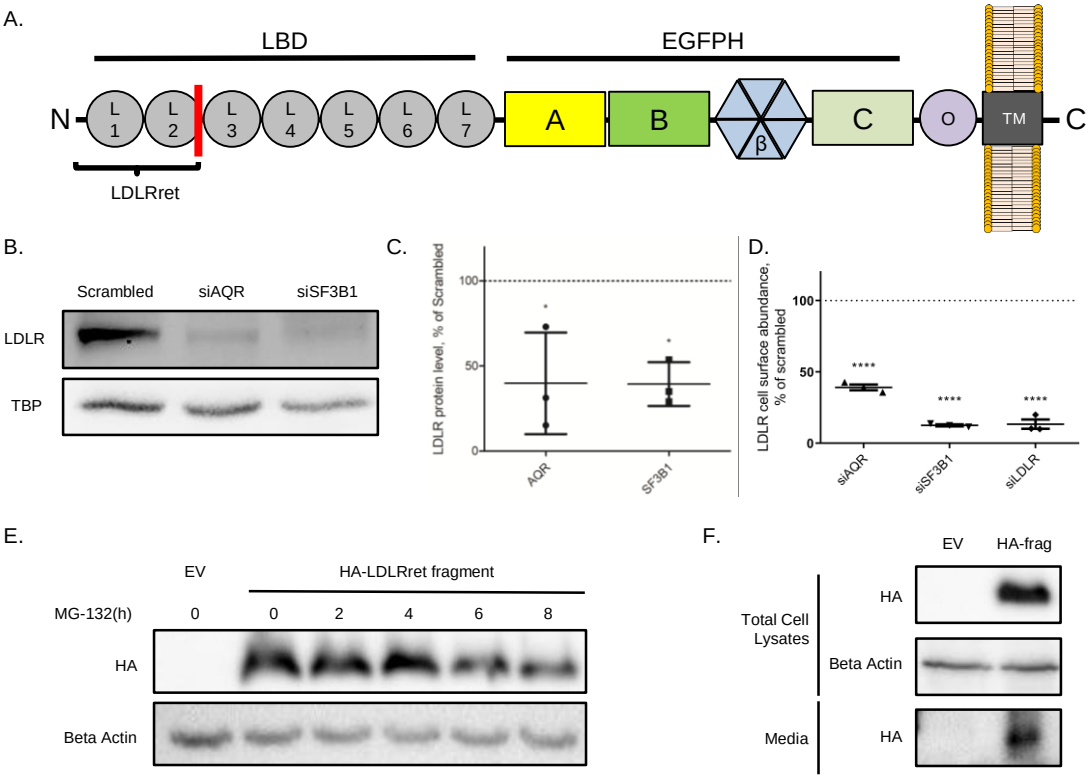

Figure 4. Effect of loss of spliceosome function on LDLR protein expression A. Schematic structure of the LDLR protein. (modified from[32]). LDLRret: intron 3 retention fragment, LBD: Ligand binding domain; L1-L7: LDLR class A domain; EGFPH: Epidermal growth factor precursor homology domain; β: beta propeller; O: O-linked sugar repeat; A/B/C: EGF-type repeat; TM: transmembrane domain. The red line represents the location of the last canonical amino acid found also in the LDLRret fragment, followed by 12 novel amino acids and by a stop codon. **B and C. Effect of *SF3B1* and *AQR* knockdown on LDLR protein levels.** LDLR protein levels in Huh-7 cells 72 hours after *SF3B1* and *AQR* knockdown measured by western blot. A representative blot is shown in **B**. **C** shows relative densities of LDLR bands after knockdown of AQR or SF3B1 relative to scrambled control. TATA-binding-protein (TBP) was used as the loading control. Bars represent means ±SD of three replicate experiments. **D. Effect of *SF3B1* and *AQR* knockdown on LDLR cell surface levels.** LDLR cell surface levels in live Huh-7 cells measured by flow cytometry 72 hours after knockdown of *SF3B1* and *AQR*. A pool of siRNAs against *LDLR* was used as the positive control. The data are normalized to a non-silencing control. Each point represents one of three identical experiments. Bars represent mean±SD. Significance in C and D was calculated by One-way ANOVA followed by Dunnett's post-hoc test. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$. **E-F. Overexpressed LDLRret fragment is retrieved in cell lysates and cell culture medium.** 48 hours after transfection in HEK293T cells, the HA-tagged version of the LDLRret fragment was detected by western blot in both total cell lysates **(E,F)** and media **(F)**. Lysates after 2 and more hours of incubation were obtained after incubation with the proteasome inhibitor MG132 as indicated by the labels in (**E**) EV = pcDNA3.1 empty vector. HA-frag = hemagglutinin-tagged LDLRret fragment.

Figure 5

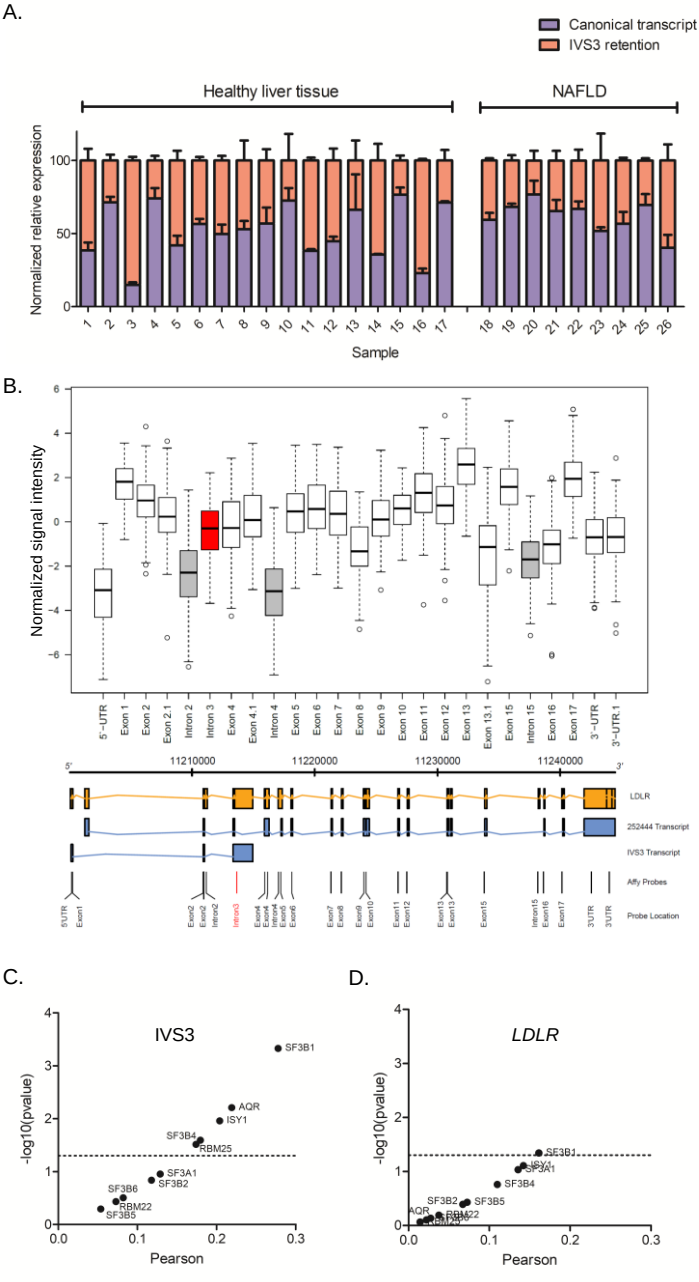

Figure 5. *LDLR* intron 3 retention in human liver. A. Detection of intron 3 retention in the human liver by RT-PCR. Transcripts encoding full-length *LDLR* transcript or the IVS3 retention variant were measured by RT-PCR and normalized to GAPDH mRNA levels in the healthy liver tissue of 17 patients with benign liver tumors and in liver biopsies of 9 patients with suspected NAFLD.. Data are expressed as mean relative expression ±SD after normalizing to the sum of both *LDLR* transcripts **B. Exon-level *LDLR* expression in the human liver.** The boxplot shows the normalized signal intensities for probes mapping to the 5'-UTR, 3'-UTR, the exons, and some introns of the *LDLR* gene in 155 obese non-diabetic subjects. The red box indicates the normalized signal levels for intron 3. The other introns are shown in grey. The location of each probe is depicted in the diagram below. **C and D. Correlations between the expression of spliceosome genes and the intron 3 probe (C) or the full length *LDLR* probe (D).** The scatterplot depicts the Pearson's correlation coefficients and their $-\log_{10}$(p-values) for the correlations between the signal intensities of probes for the spliceosome genes and the intron 3 retention probe (IVS3) (**C**) or the *LDLR* full length probe gene (**D**). Negative Pearson's coefficients (-0.092 and -0.145) for correlations of SF3A2 with IVS3 and full length LDLR, respectively, are not displayed for simplicity.

Figure 6

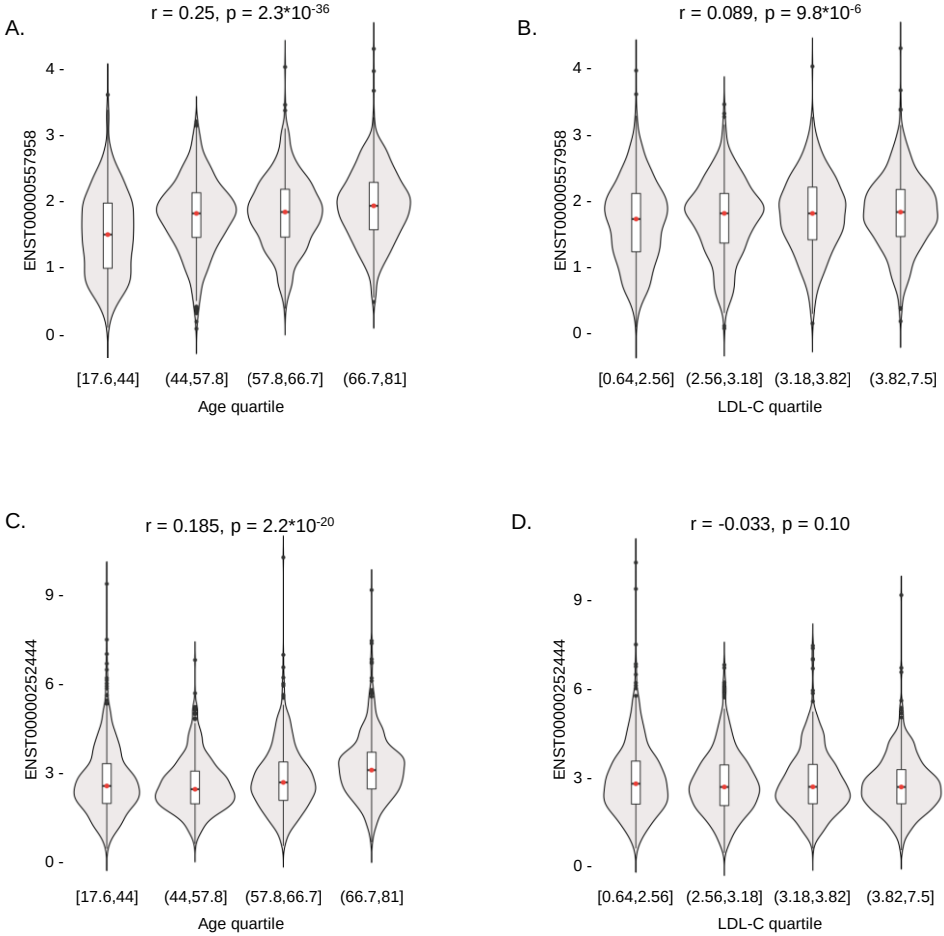

Figure 6. Correlations of the *LDLR* ENST00000557958 (A, B) and ENST00000252444 transcripts (C, D) in whole blood samples with age (A, C) and LDL-C levels (B, D). Data is from 2,462 subjects of the BIOS population[9]. ENST00000557958 represents the intron 3 retention transcript (**A, B**). ENST00000252444 (**C, D**) was the only full-length *LDLR* transcript detected in all samples analysed. Violin plots describe the frequency distribution of the transcripts in quartiles of age or LDL-C quartile. The red horizontal lines reflect medians; the upper and lower borders of the boxes reflect the 75th and 25th percentiles, respectively. r-values and p-values refer to a Spearman correlation analysis.

Figure 7

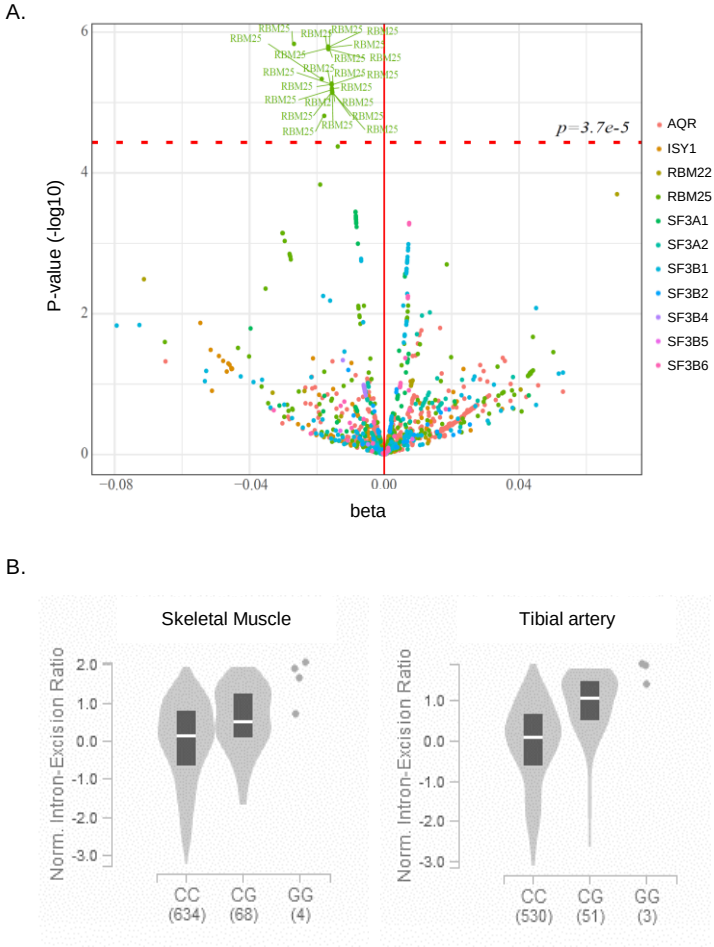

Figure 7. Association between *RBM25* variants and LDL-C in the UK Biobank dataset. A. Association of GWAS SNPs from 11 spliceosome genes with LDL-C in the UK Biobank dataset. The dashed red horizontal line indicates the threshold for statistical significance after correction for multiple testing of 1360 variants within the genes of interest ($p=3.7*10^{-5}$). Effect size and directionality are reported on the x axis as beta value. **B. Association between the rs17570658 genotype and *RBM25* expression in different tissues.** Data shown for skeletal muscle and tibial artery (both $p < 10^{-8}$). The horizontal white lines reflect medians; the upper and lower borders of the grey boxes reflect the 75th and 25th percentiles, respectively