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

Zanoni, Paolo; Panteloglou, Grigorios; Othman, Alaa; Haas, Joel T; Meier, Roger; Rimbert, Antoine; Futema, Marta; Abou-Khalil, Yara; Nørrelykke, Simon Flyvbjerg; Rzepiela, Andrzej

Published in:
Circulation research

DOI:
10.1161/CIRCRESAHA.120.318141

Publication date:
2022

Circulation Research

ORIGINAL RESEARCH

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

Paolo Zanoni, Grigorios Panteloglou, Alaa Othman, Joel T. Haas, Roger Meier, Antoine Rimbert, Marta Futema, Yara Abou Khalil, Simon F. Norrelykke, Andrzej J. Rzepiela, Szymon Stoma, Michael Stebler, Freerk van Dijk, Melinde Wijers, Justina C. Wolters, Nawar Dalila, Nicolette C. A. Huijkman, Marieke Smit, Antonio Gallo, Valérie Carreau, Anne Philippi, Jean-Pierre Rabès, Catherine Boileau, Michele Visentin, Luisa Vonghia, Jonas Weyler, Sven Francque, An Verrijken, Ann Verhaegen, Luc Van Gaal, Adriaan van der Graaf, Belle V. van Rosmalen, Jerome Robert, Srividya Velagapudi, Mustafa Yalcinkaya, Michaela Keel, Silvija Radosavljevic, Andreas Geier, Anne Tybjaerg-Hansen, Mathilde Varret, Lucia Rohrer, Steve E. Humphries, Bart Staels, Bart van de Sluis, Jan Albert Kuivenhoven, Arnold von Eckardstein

BACKGROUND: The LDLR (low-density lipoprotein receptor) 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: We performed a genome-wide RNA interference screen for genes limiting the uptake of fluorescent LDL into Huh-7 hepatocarcinoma cells. Top hit genes were validated by in vitro experiments as well as analyses of data sets on gene expression and variants in human populations.

RESULTS: The knockdown of 54 genes significantly inhibited LDL uptake. Fifteen of them encode for 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 translated LDLR fragment lacks 88% of the full length LDLR and is detectable neither in nontransfected cells nor in human plasma. The hepatic expression of the intron 3 retention transcript is increased in nonalcoholic fatty liver disease as well as after bariatric surgery. Its expression in blood cells correlates with LDL-cholesterol and age. Single nucleotide polymorphisms and 3 rare variants of one spliceosome gene, *RBM25*, are associated with LDL-cholesterol in the population and familial hypercholesterolemia, respectively. Compared with overexpression of wild-type *RBM25*, overexpression of the 3 rare *RBM25* mutants in Huh-7 cells led to lower LDL uptake.

CONCLUSIONS: We identified a novel mechanism of posttranscriptional regulation of LDLR activity in humans and associations of genetic variants of *RBM25* with LDL-cholesterol levels.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: hepatocytes ■ cardiovascular diseases ■ endocytosis ■ hypercholesterolemia ■ spliceosomes

Meet the First Author, see p 4

Hypercholesterolemia is a causal and treatable risk factor of atherosclerotic cardiovascular diseases.[1] The most important determinant of LDL-C (low-density lipoprotein cholesterol) levels in plasma is the hepatic removal of circulating LDL by binding to the LDLR (LDL receptor) 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 atherosclerotic cardiovascular diseases.[1]

Correspondence to: Arnold von Eckardstein, MD, Institute for Clinical Chemistry, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland. Email arnold.voneckardstein@usz.ch
*P. Zanoni and G. Panteloglou contributed equally.
Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCRESAHA.120.318141.
For Sources of Funding and Disclosures, see page 94.

Circulation Research is available at www.ahajournals.org/journal/res

Novelty and Significance

What Is Known?

- The LDLR (LDL [low-density lipoprotein] receptor) regulates LDL-cholesterol levels in blood by mediating the uptake of LDL into hepatocytes.
- The transcriptional and posttranslational regulation of LDLR activity is targeted by cholesterol lowering drugs.

What New Information Does This Article Contribute?

- Loss of subunits or interactors of the U2 spliceosome decreases the uptake of LDL into Huh7 hepatocarcinoma cells. Loss of subunits or interactors of the U2 spliceosome also causes intron 3 retention of the *LDLR* mRNA and, thereby, loss of LDR function.
- Intron 3 retention of *LDLR* in human liver and peripheral blood cells is increased by nonalcoholic fatty liver disease and aging, respectively.
- Single nucleotide polymorphisms of the spliceosome gene *RBM25* are associated with higher *RBM25* expression in tissues and lower LDL-cholesterol.
- Expression of rare structural variants of *RBM25* that are associated with familial hypercholesterolemia (FH) decrease LDL uptake into Huh7 cells.

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

Nonstandard Abbreviations and Acronyms

AQR	aquarius intron-binding spliceosomal factor
FH	familial hypercholesterolemia
HDL	high-density lipoprotein
LDL-C	low-density lipoprotein cholesterol
LDLR	LDL receptor
NAFLD	nonalcoholic fatty liver disease
NASH	nonalcoholic steatophepatitis
qRT-PCR	quantitative real-time polymerase chain reaction
RBM25	RNA binding motif protein 25
RNAi	RNA interference

To identify novel regulators of LDL uptake into the liver, we performed an image-based genome-wide RNA interference (RNAi) 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.[3] 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, determining 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 Supplemental Material 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 Supplemental Material.

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 (high-density lipoprotein), Huh-7 human hepatocarcinoma cells were reverse-transfected using 3 different siRNA oligonucleotides against each of the 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 72 hours posttransfection to 33 µg/mL each of Atto594-labelled LDL and Atto655-HDL for 4 hours. As background controls, wells with cells transfected with a nontargeting siRNA were incubated in the absence of fluorescent lipoproteins. After washing,

fixation, and staining of the nuclei with Hoechst 33258, the plates were imaged at 4× and 20× with 2 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 (Figure S1A). Analysis and validation of HDL image data will be subject of a separate report.

For the uptake of fluorescent LDL, the 5 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 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, −023 to 0.20]) and positive control (median, −0.56 [interquartile range, −0.99 to −0.20]) clustered mostly around the 0-line, indicating a suboptimal but analytically exploitable signal-to-noise ratio (Figure S1B). Dimensionality reduction of main assay features did not significantly alter the outcome (Figure S1C and S1D). At a redundant siRNA activity *P* value cutoff of $P<10^{-3}$, interference with 54 and 37 genes decreased and increased LDL uptake, respectively (Table, Table S1). By contrast to the findings of a previous genome-wide CRISPR-based screening in Huh7 cells,[4]

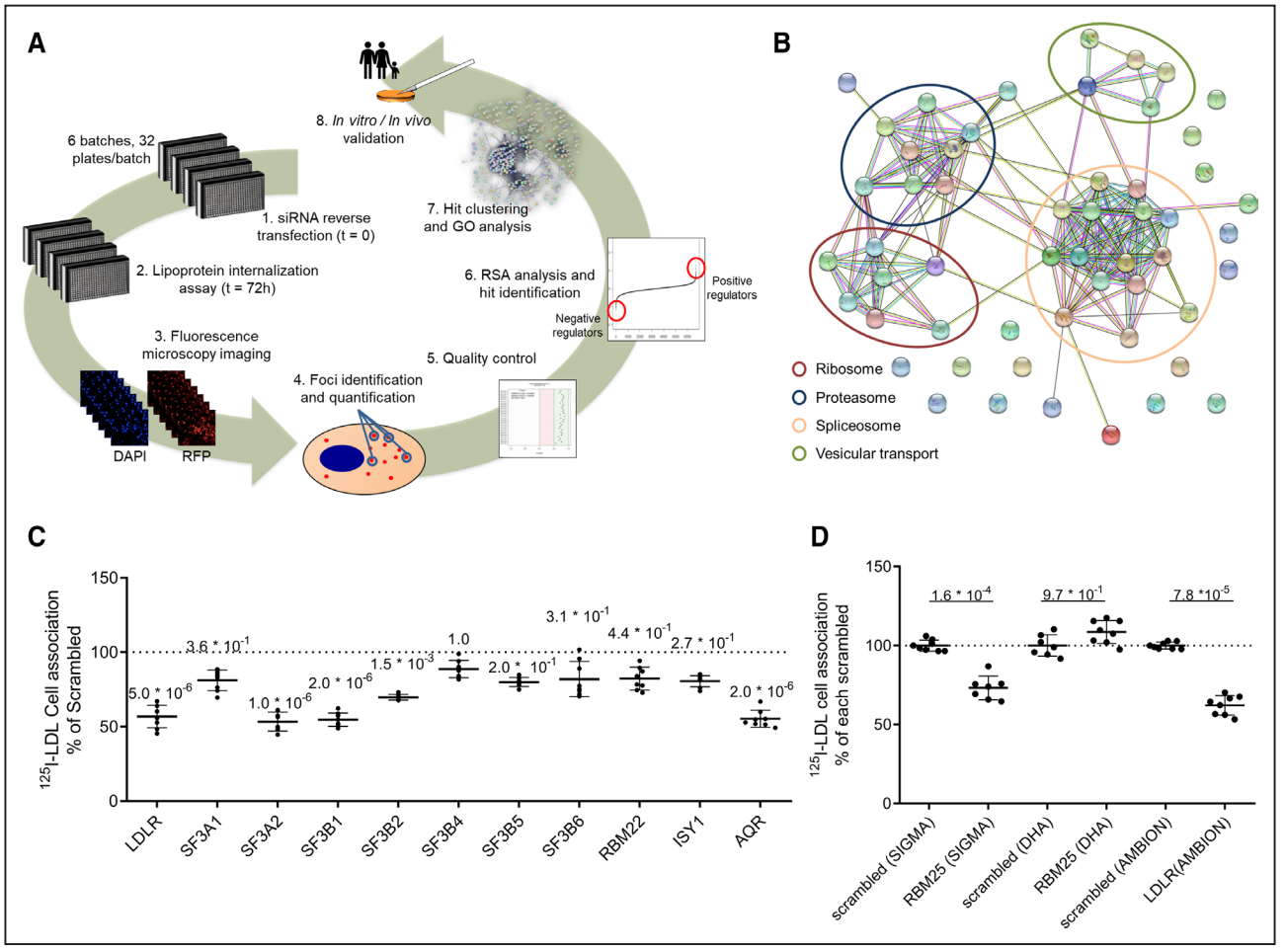

Figure 1. Identification and validation of U2-spliceosome genes as limiting factors for the uptake of LDL (low-density lipoprotein) by Huh-7 cells.

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 $P<1.0\times10^{-3}$ 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 on cell association of ^{125}I-LDL in Huh-7 cells. Seventy-two h after transfection with siRNAs from Ambion (*LDLR*), Sigma (*RBM25*), or Dharmacon (all other genes), cells were incubated for 2 h at 37 °C in the presence of 33.3 µg/mL of ^{125}I-LDL in the presence or absence of 40× excess unlabelled LDL. Specific cell association was calculated as the difference between the 2 conditions. The data are expressed as means±SD of 2 quadruplicate experiments. Statistical analysis was performed using Kruskal-Wallis test with Dunn multiple comparisons test between the nontargeting (scrambled) and each targeting siRNA (**C**) or Mann-Whitney test (1-sided) between each vendor's targeting and nontargeting (scrambled) siRNAs (**D**). The respective *P* values are shown above each condition. GO indicates Gene Ontology; and LDLR, LDL receptor.

Table. 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* avg†	Assay score* SEM‡	RSA, *P* value§	Gene	Assay score* avg†	Assay score* SEM‡	RSA, *P* value§
AP2M1	−3.103179681	0.346648222	3.36×10^{-8}	*PROX1*	6.53057396	0.631260417	3.19×10^{-9}
CHMP2A	−3.130900347	0.359445533	2.51×10^{-7}	*ITGAV*	7.431175355	1.519558432	2.96×10^{-8}
NFKB2	−2.59157417	0.136886566	8.07×10^{-7}	*TGFBR1*	3.464028514	0.397588943	7.31×10^{-6}
AQR‖	−2.484868551	0.199482589	4.57×10^{-6}	*CDC37*	3.747034032	1.072191825	2.35×10^{-5}
PSMD11	−2.557101583	0.239773488	4.77×10^{-6}	*DTNBP1*	57.92944887	57.2617451	4.46×10^{-5}
SF3B2‖	−2.107311389	0.015210399	4.81×10^{-6}	*CYP27C1*	32.06817221	31.61438081	8.92×10^{-5}
RPL35	−2.346954606	0.150946677	5.45×10^{-6}	*PNPLA2*	2.279278207	0.266420784	1.26×10^{-4}
PSMD8	−2.988677308	0.491086915	6.34×10^{-6}	*C22orf39*	7.995494448	8.342242785	1.78×10^{-4}
SON‖	−2.164748153	0.201099955	1.46×10^{-5}	*TMEM133*	3.049034762	1.060165442	1.84×10^{-4}
COPA	−2.307675328	0.213018879	1.61×10^{-5}	*TMEM130*	6.466491664	5.317700355	2.23×10^{-4}
RBM25‖	−1.993998657	0.055265194	1.92×10^{-5}	*PM20D2*	2.155202336	0.176491898	2.29×10^{-4}
RBM22‖	−2.818121291	0.622885617	3.36×10^{-5}	*PET117*	3.001069341	1.652765441	2.68×10^{-4}
PSMD3	−2.21302629	0.224903034	3.98×10^{-5}	*CWF19L2*	3.806757511	4.571696977	3.12×10^{-4}
SF3B5‖	−2.285064158	0.25878823	4.32×10^{-5}	*ENY2*	2.420424347	0.514153659	3.28×10^{-4}
SF3B1‖	−2.267932169	0.253122099	4.55×10^{-5}	*NME4*	2.711491413	0.954425612	3.39×10^{-4}
SALL4	−1.937993523	1.13065979	6.02×10^{-5}	*ZC3H4*	4.545156994	3.551478266	3.57×10^{-4}
RPL5	−2.106905493	0.373542859	7.40×10^{-5}	*WASF2*	2.310874515	0.449202822	3.61×10^{-4}
CCDC180	−1.132459235	1.398333381	9.52×10^{-5}	*HELZ2*	2.546828237	0.984740435	3.87×10^{-4}
SF3B6‖	−2.277000896	0.332074616	9.83×10^{-5}	*RILP*	1.995550072	0.267567916	4.23×10^{-4}
HNRNPU	−1.724036435	0.093847304	1.23×10^{-4}	*MAT2A*	3.705559066	3.611891772	4.91×10^{-4}
RPL17	−2.226845162	0.329575956	1.46×10^{-4}	*NRM*	1.710898743	0.050727817	5.02×10^{-4}
ISY1‖	−2.74487386	0.698388989	1.55×10^{-4}	*CEP295NL*	2.189792071	0.474598108	5.02×10^{-4}
ZNF641	−1.034460324	1.453444444	2.58×10^{-4}	*ACSM2A*	2.207444199	1.531809937	5.32×10^{-4}
COPB1	−1.693933632	0.103029465	2.64×10^{-4}	*RTL9*	3.759306708	3.473297986	5.35×10^{-4}
SF3A1‖	−2.225755015	0.412106586	2.72×10^{-4}	*KIAA1522*	3.362058267	3.27466253	6.25×10^{-4}
SNW1‖	−1.76539067	0.142531611	2.76×10^{-4}	*ZNF84*	2.204388764	0.765657329	6.55×10^{-4}
EIF2S1	−1.486721463	0.790741651	3.45×10^{-4}	*TFAP4*	3.032765033	3.340175625	6.69×10^{-4}
CCDC73	−1.041204586	1.27682775	3.50×10^{-4}	*TMEM182*	3.227517874	1.666669492	7.29×10^{-4}
RPL9	−1.715182797	0.249911985	3.55×10^{-4}	*WDR55*	1.967286849	1.365170916	7.32×10^{-4}
NXNL2	−1.199311468	1.135835784	3.83×10^{-4}	*DYNLL1*	2.268266743	0.467997927	7.72×10^{-4}
WBP11‖	−1.50591484	0.062444555	4.03×10^{-4}	*ADPRHL2*	2.078229013	0.322800093	8.51×10^{-4}
C2CD5	−1.097951788	1.954971449	4.46×10^{-4}	*ELAVL1*	1.945364959	0.968117905	8.70×10^{-4}
RPL21	−1.655773242	0.156797718	4.72×10^{-4}	*CFAP298*	1.883199038	0.378258022	8.87×10^{-4}
EPOP	−1.837314819	0.25795876	4.80×10^{-4}	*PMM1*	2.80926863	3.200260012	8.92×10^{-4}
RMND5B	−1.523957521	0.076773849	5.07×10^{-4}	*CASKIN2*	1.681223061	0.149986926	9.07×10^{-4}
TAPBPL	−1.52965773	0.154207886	5.27×10^{-4}	*CIZ1*	3.454694336	2.803876145	9.37×10^{-4}
STARD10	−1.527795273	0.115135889	5.45×10^{-4}	*BRICD5*	1.962503862	0.408074057	9.41×10^{-4}
PSMD1	−2.207116523	0.551747426	5.63×10^{-4}				
PFDN6	−0.881689024	1.740601376	5.80×10^{-4}				
PSMA1	−1.528976301	0.119805079	5.85×10^{-4}				
RTF2	−1.573924771	0.169765686	6.14×10^{-4}				
LSM2‖	−1.448888015	0.056454941	6.40×10^{-4}				
UBD	−1.171691024	1.530009178	6.69×10^{-4}				
LRRC14	−1.258311764	1.067910962	6.84×10^{-4}				
SUPT6H‖	−1.451332382	0.095214513	7.27×10^{-4}				
COPB2	−2.037140764	0.468882876	7.34×10^{-4}				

(*Continued*)

Table. Continued

Decreased LDL uptake				Increased LDL uptake			
Gene	Assay score* avg†	Assay score* SEM‡	RSA, *P* value§	Gene	Assay score* avg†	Assay score* SEM‡	RSA, *P* value§
SF3A2‖	−1.347147433	0.758462926	7.89×10^{-4}				
ATP6V0C	−1.823918839	0.263639476	7.90×10^{-4}				
EMILIN3	−1.598631472	2.238859705	8.03×10^{-4}				
DMTN	−1.559252376	0.142024687	8.20×10^{-4}				
MRPL19	−0.755460842	1.688052373	8.92×10^{-4}				
MRO	−0.986783025	1.102624895	9.14×10^{-4}				
DDX59	−1.380513222	1.040634076	9.25×10^{-4}				
PSMD12	−1.761325035	0.367766123	9.45×10^{-4}				

LDL indicates low-density lipoprotein.
*Assay score: normalized score for the median cytoplasm intensity assay feature.
†Average.
‡SEM.
§*P* values are not adjusted for multiple testing ($P<3.6\times10^{-6}$ after Bonferroni adjustment for 14 000 genes with expressed transcripts).
‖The 15 hit genes involved in RNA splicing and validated.

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). Gene Ontology enrichment analysis showed significant clustering for genes whose loss of function decreased LDL uptake (Table S2). Functional clustering of these genes with the STRING tool revealed 4 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, 6 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 [aquarius intron-binding spliceosomal factor], ISY1 [ISY1 splicing factor homolog], and RBM25 [RNA binding motif protein 25]) or indirectly (RBM22).[3]

To confirm the role of the U2 spliceosome in LDL endocytosis in vitro, we performed ^{125}I-LDL cell association assays in Huh-7 and HepG2 cells. *SF3B4* was also included in these experiments as it is part of the U2 spliceosome and barely missed the redundant siRNA activity *P* value cutoff ($P=1.4\times10^{-3}$). Knockdown was achieved using 4 pooled siRNA molecules against each hit gene acquired from vendors other than that of the siRNA screening library, namely Dharmacon or Sigma instead of Ambion (see Major Resource Table and Figure S2A). For *RBM25,* we replaced Dharmacon's siRNAs with those from Sigma because of their presumable off-target effects on LDLR protein expression (Figure S3). 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 and 1D and Figure S2B). The association of ^{125}I-LDL was equally decreased by knockdown of *SF3B1* (−45±5%), *SF3A2* (−47±6%), *AQR* (−45±6%), and *LDLR* (−43±8%; Figure 1C). RNAi with *RBM25* reduced the specific cellular association of ^{125}I-LDL and fluorescent LDL by 27±8% and 52±5%, respectively (Figure 1D and Figure S3F). Of note, the specific cell association of ^{125}I-HDL was unaltered or even increased upon knockdown of *AQR* and *SF3A1* in either Huh-7 or HepG2 cells (Figure S2C and S2D).

Loss of U2-Spliceosome Genes and Their Interactors Causes Selective Retention of LDLR Intron 3 (IVS3)

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 U2-spliceosome genes or a nontargeting control siRNA. Sequences can be accessed by codes PRJEB46899 and PRJEB46898 in the data bank of the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/support). 72 hours after transfection, we measured both expression at the gene level and alternative exon usage in polyA-selected transcripts. 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, Figure S4). This effect was confirmed in Huh-7 cells by quantitative real-time polymerase chain reaction (qRT-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[5] (Figure S5A). By contrast to the RNA sequencing (Figure S4), qRT-PCR unravelled increased expression of the *LDLR* IVS3 retention transcript upon knockdown of *RBM25*, albeit not as much as with knockdown of *SF3B1* and *AQR* (Figures S5B and S5C).

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 data set 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, (r=–0.73, P=1.4×10^{-2}, Figure 2C).

To investigate reasons for intron 3 retention in *LDLR*, we transfected HEK293T cells with 2 minigenes containing different portions of the *LDLR* genomic sequence flanked by 2 artificial exons (Figure 3A). The first minigene (MG$_1$) encoding only for exon 3 of *LDLR* and the adjacent intronic regions cloned between 2 artificial exons (SD6 and SA2), displayed very low if any RNA sequencing reads mapping to the first ≈130 bp 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 indicates incomplete splicing of intron 3 when the physiological exon 4 acceptor site and the branch point site were present in the larger minigene MG$_2$ (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/)[6] identified one plausible U2-spliceosome dependent branch point site located 30 bp

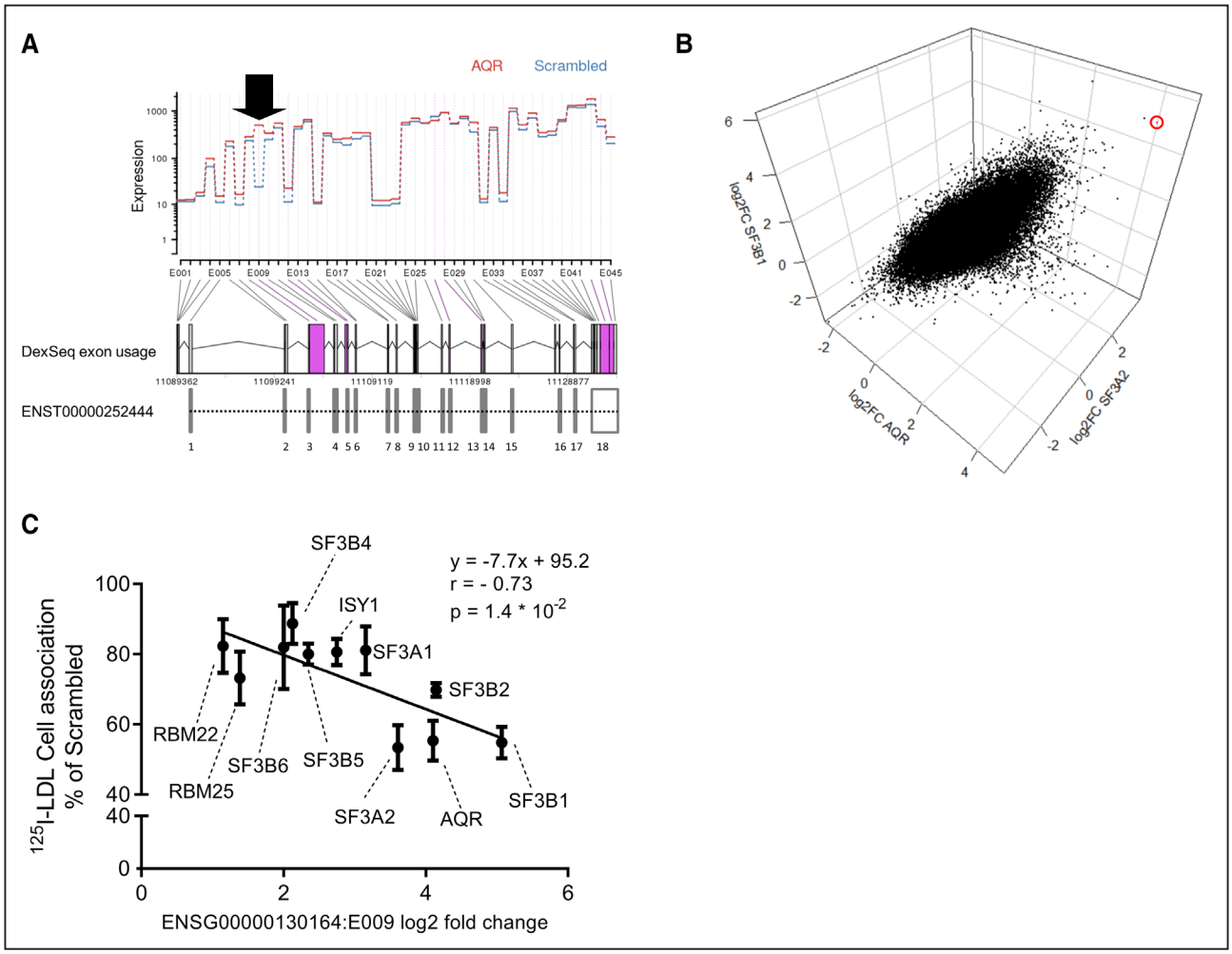

Figure 2. Loss of U2-spliceosome genes causes intron 3 retention in *LDLR*.
A, *LDLR* Exon-level expression upon *AQR* knockdown. Expression of the *LDLR* exons was recorded by RNA sequencing of Huh-7 cells 72 h after knockdown of *AQR*. Segments represent differential exon usage in each sector of the *LDLR* genomic sequence as identified by the DEXSeq algorithm and as summarized in the linear representation below the graph. Canonical exons of the ENST00000252444 full length transcript 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 three independent 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 (low-density lipoprotein receptor) 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 U2-spliceosome hit gene. Cells treated with a nontargeting siRNA were used as reference. Cell association is expressed as mean±SD. *r* and *P* value were calculated according to Spearman.

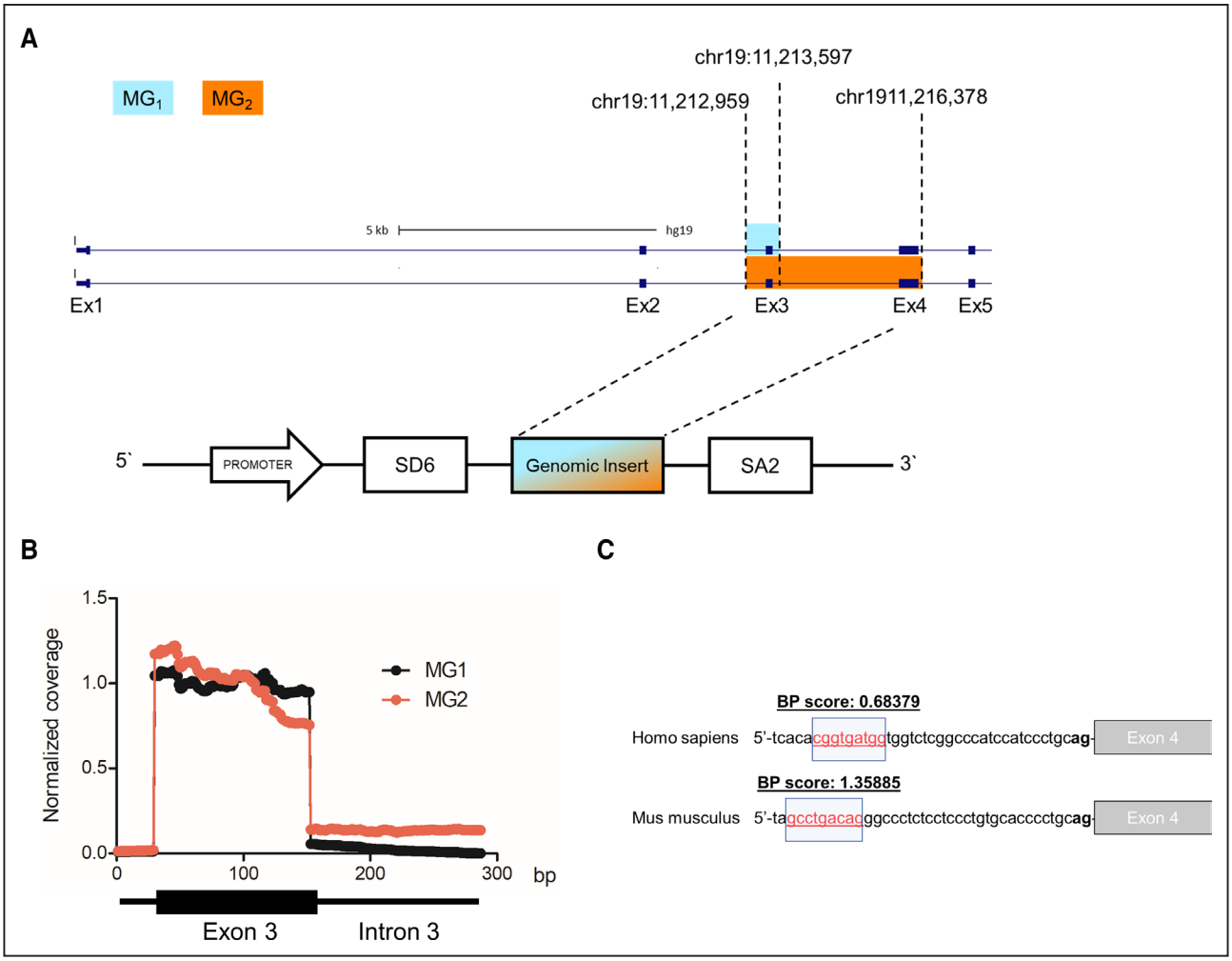

Figure 3. Determination of LDLR (low-density lipoprotein receptor) intron 3 splice patterns.
A, Cloning strategy and structure of the minigenes. The **upper** part shows the genomic location of the 2 segments of the LDLR gene that were cloned in each minigene, while the lower half shows a simplified 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 1 bp shorter at its 5′ end, starting at chr19:11212960. **B**, Characterization of the splice products. The graphs represent the mean RNA sequencing coverage at the Exon 3-Intron 3 junction in 2 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 branch point site (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.

upstream of the acceptor site (Table S3). The *gtgat* pentamer in the center of the *cggtgatgg* branchpoint sequence was associated with very low U2 binding energy and occurs at low frequency in the branchpoint database.[6] We discarded another predicted branchpoint 124 bp 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 bp 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 qRT-PCR in mouse liver (data not shown). Taken together, these data suggest that the branch point site 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 part of the second class A domains (labelled as L1 and L2 in Figure 4A[7]) 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 (Figure 4B and 4C). A similar decrease in

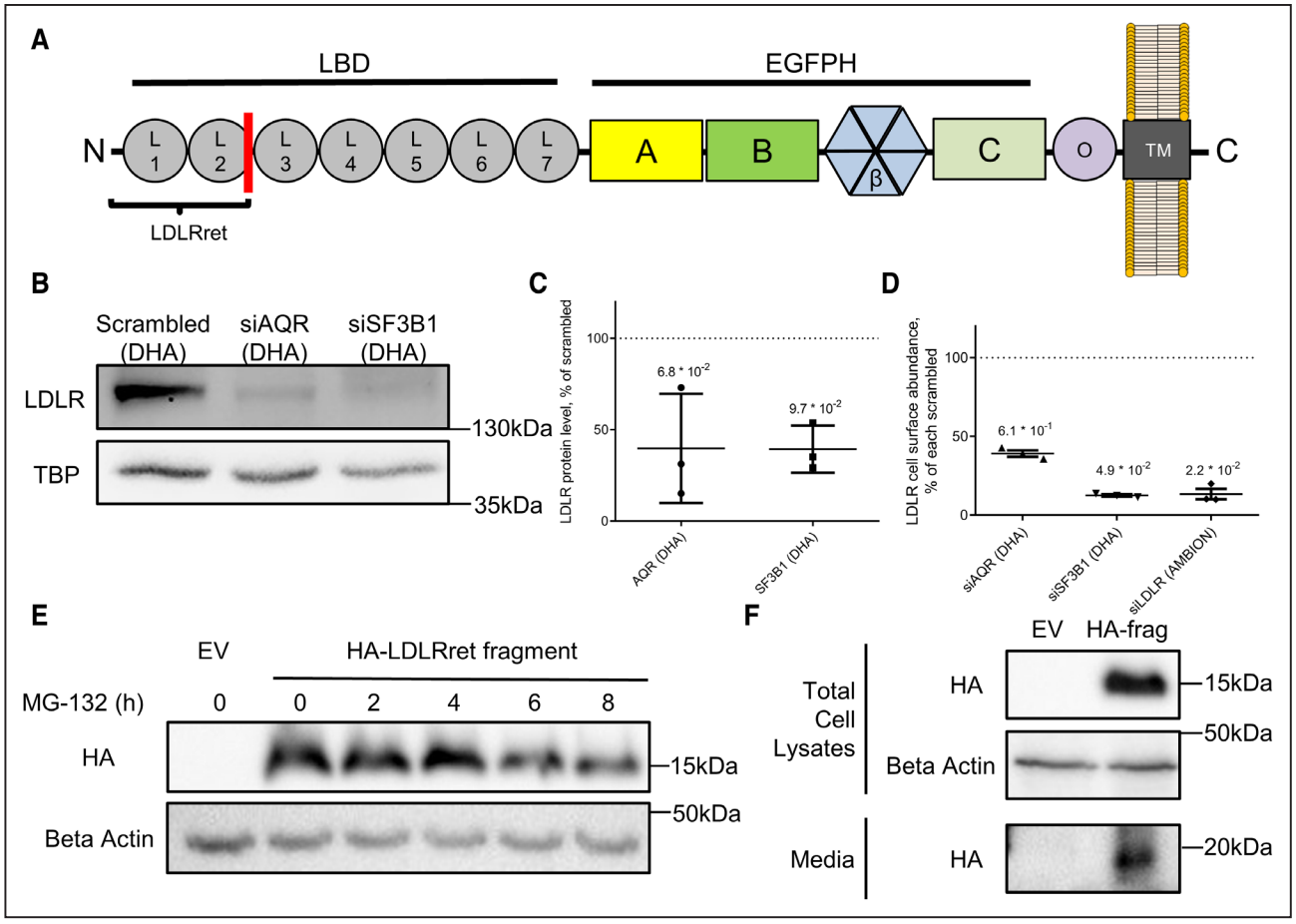

Figure 4. Effect of loss of spliceosome function on LDLR (low-density lipoprotein receptor) protein expression.
A, Schematic structure of the LDLR protein. (modified from[7]). The red line represents the location of the last canonical amino acid found also in the intron 3 retention fragment (LDLRret) 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 h after *SF3B1* or *AQR* knockdown. **B** shows a representative Western blot. **C**, Relative densities of LDLR bands normalized to TBP (TATA-binding-protein, loading control) after knockdown of *AQR* or *SF3B1* relative to the nontargeting control. The data are shown as means±SD of 3 independent experiments. **D**, Effect of *SF3B1* and *AQR* knockdown on LDLR cell surface levels. LDLR cell surface levels in alive Huh-7 cells were measured by flow cytometry 72 h after knockdown of *SF3B1* or *AQR*. siRNAs against *LDLR* were used as positive controls. The data are normalized to a nontargeting control and are shown as means±SD of 3 independent experiments. Numbers in **C** and **D** are *P* values obtained by Kruskal-Wallis test with Dunn multiple comparisons test between the nontargeting (scrambled) and respective targeting siRNA. **E**–**F**, Overexpressed LDLRret fragment is retrieved in cell lysates and cell culture medium. Forty-eight h after transfection in HEK293T cells, the HA-tagged version of the LDLRret fragment was detected by western blot in both total cell lysates (**E** and **F**) and media (**F**). Lysates after 2 and more hours of incubation were obtained after treatment with the proteasome inhibitor MG132 as indicated by the labels in (**E**). A/B/C indicates EGF-type repeat; EGFPH, epidermal growth factor precursor homology domain; EV, pcDNA3.1 empty vector; HA-frag, hemagglutinin-tagged LDLRret fragment; L1-L7, LDLR class A domain; LBD, ligand binding domain; O, O-linked sugar repeat; TM, transmembrane domain; and β, beta propeller.

LDLR protein was seen upon knockdown of *RBM25* with siRNAs from Sigma (−68±10%), whereas the knockdown of *RBM25* with the siRNA of Dharmacon led to an increase in LDLR protein (122±109%), presumably due to off-target effects (Figure S3D and S3E). Flow cytometry experiments on alive 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 knockdown of *RBM25* with siRNAs from Sigma and Dharmacon decreased the cell surface abundance of LDLR by 53±6% and 21±5%, respectively, as compared with nontargeting siRNAs from the respective manufacturers (Figure S3G).

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. Forty-eight 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. 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 (Figure S6) 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 by three different methods, and explored associations with nonalcoholic fatty liver disease (NAFLD), demographic measures, lipid traits, and therapeutic interventions.

qRT-PCR of mRNAs of liver tissue from 17 patients with benign liver tumours and 9 patients with suspected NAFLD, 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, 43% (range, 23%–85%) of the transcripts retained intron 3 (Figure 5A).

The bioinformatics analysis of RNA sequencing data on liver samples of 13 healthy nonobese subjects, 12 obese subjects without NAFLD, 15 patients with NAFLD, and 15 patients with nonalcoholic steatophepatitis (NASH; Gene Expression Omnibus, accession number GSE126848)[8] found 14 different *LDLR* transcripts (Figure S7). Four transcripts showed the largest interindividual variation, namely *LDLR-201* and *LDLR-208*, encoding full length *LDLR*, as well as *LDLR-206*, which corresponds to the retained intron 3 transcript, and the likewise futile *LDR-214* (*LDLR* transcripts are illustrated schematically in Figure S7A). Interestingly, the median concentration of *LDLR-206* was substantially higher in patients with NAFLD or NASH than in normal weight or obese subjects without NAFLD. The median percentages of *LDLR-206* reads relative to total reads from all transcripts of *LDLR* gene increased significantly from 1.8% (range, 0.7%–4.2%) and 1.7% (0.4%–3.7%) in normal weight and obese subjects without NAFLD, respectively, to 5.8% (1.1%–26.7%) and 5.0% (0.9%–29.0%) in patients with NAFLD and NASH, respectively (Figure 5B). Of the 2 most abundant full length encoding *LDLR* transcripts, *LDLR-208* decreased significantly (Figure 5C) while the expression of *LDLR-201* did not change (Figure S7).

We also investigated the expression of *LDLR* transcripts in liver biopsies of 155 obese nondiabetic subjects[9] by using Affymetrix Human Gene 2.0 ST arrays (see Table S4 for clinical and biochemical characteristics). 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 5D). The percent intensities of the IVS3 probe relative to the sum of all *LDLR* probes ranged from 7.5% to 82%. Intron 3 retention correlated significantly only with *SF3B1* (r=0.26, P=1.4×10^{-2}), while no U2-spliceosome gene showed any significant correlation with overall *LDLR* expression (Table S5). Relative intensities of neither the intron 3 probe nor any other of the 24 *LDLR* probes showed significant correlations with plasma levels of total, HDL- or LDL-cholesterol (Figure S8A–S8C and Table S6). Correlations with histological NAFLD stages were inverse by trend but not statistically significant (Figure S8D). Intron 3 relative probe intensity did not correlate with body mass index (Figure S8E). However, in a subgroup of 21 patients who underwent a second liver biopsy after bariatric surgery (median follow-up time, 13 months [interquartile range, 12–15]), the proportion of the intron 3 retention transcript relative to the full length *LDLR* transcript increased significantly after surgery (P=9.8×10^{-3}; Figure S8F; Table S4). This increase was even more pronounced in eleven patients with NASH at baseline but no NASH at follow-up (P=3.6×10^{-2}, Figure S8G).

Finally, we analyzed the RNA sequencing data in whole blood samples from 2462 subjects of the Dutch BIOS-consortium.[10] The *LDLR* ENST00000557958 transcript, predicted to retain intron 3, was detectable in all subjects and represented 21±7% of the total *LDLR* transcripts. The ENST00000557958 transcript levels significantly correlated with age (r=0.25, P=9.2×10^{-36}, Figure 6A) and less strongly with LDL-C (r=0.089, P=3.9×10^{-5}, 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, the only transcript encoding for full length LDLR and expressed in blood cells in all subjects in this data set, was also positively correlated with age (r=0.19, P=8.8×10^{-20}, Figure 6C) but not with LDL-C (r=–0.033, P=4.0×10^{-1}, Figure 6D). Correlation of neither transcript with body mass index was statistically significant.

Single Nucleotide Polymorphisms in RBM25 Are Associated With Lower LDL-Cholesterol

The analysis of whole exome sequencing data of 40 468 UK Biobank subjects[11] did not unravel any significant association between our spliceosome hit genes and LDL-C or any other clinical lipid trait (Table S7). However, constraints data from the gnomAD database indicate a strong intolerance to functional genetic variation for our U2-spliceosome genes, with a probability of intolerance to loss of function[12] of 0.91±0.17 (mean±SD; Table S8). The analysis of SNPs of 11 U2-spliceosome hit genes in 361 194 participants of UK Biobank found 24 SNPs of *RBM25* significantly associated with lower levels of LDL-C (Figure 7A) and apoB (Figure S9A).

In Europeans, 4 SNPs in introns or downstream of the *RBM25* coding sequence including the lead SNP rs17570658 and 2 upstream SNPs are in almost

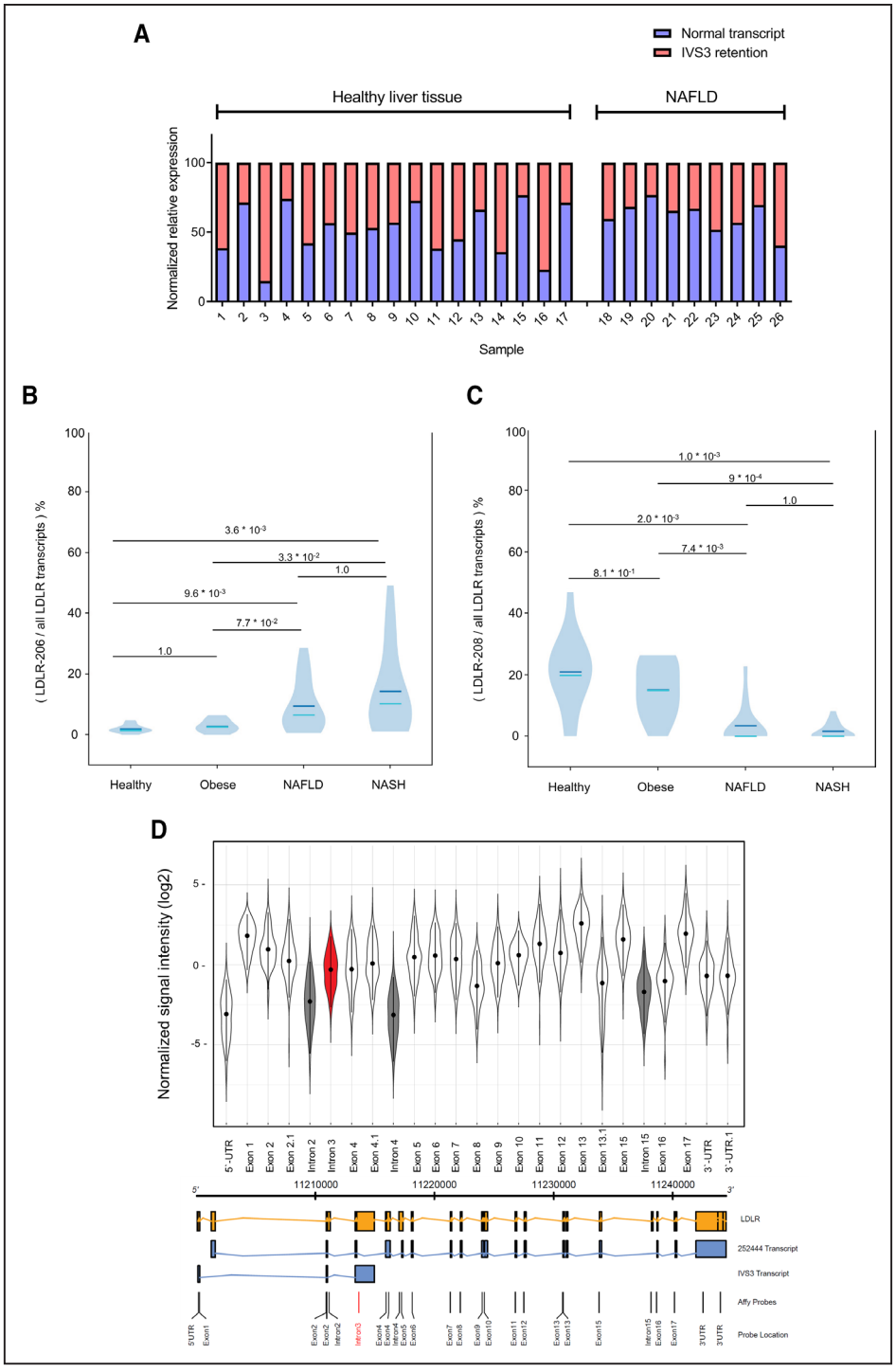

Figure 5. *LDLR* intron 3 retention in human liver.

A, Detection of intron 3 retention in human liver by quantitative real-time polymerase chain reaction (qRT-PCR). Transcripts encoding full-length *LDLR* or the IVS3 retention variant were measured by qRT-PCR and normalized to GAPDH mRNA levels in healthy liver tissue of 17 patients with benign liver tumours and in liver biopsies of 9 patients with suspected nonalcoholic fatty liver disease (NAFLD). Each bar shows the relative expression of the 2 *LDLR* transcripts in one subject. **B** and **C**, Percent expression of the *LDLR* transcript *LDLR-206* with retention of intron 3 (**B**) and a full length *LDLR* transcript *LDLR-208* (**C**) relative to the sum of all 14 *LDLR* transcripts in livers of 13 healthy subjects or 12 obese patients without NAFLD, 15 patients with NAFLD and 15 patients with nonalcoholic steatohepatitis (NASH). Computational analysis of previously published RNA sequencing data (Gene Expression Omnibus, accession number GSE126848).[8] For all transcripts, see Figure S7. The dark and light blue lines within the violin plots represent means and medians, respectively. Numbers indicate *P* values obtained by comparisons of indicated groups using the Kruskal-Wallis test and adjusted for multiple testing using the Bonferroni correction. **D**, Expression of *LDLR* exons and introns in human liver. The violin plots show 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 nondiabetic subjects. Dots indicate median values. Error bars span from the 2.5th to the 97.5th percentile. Intron 3 is highlighted in red while the other introns are shown in grey. The location of each probe is depicted in the diagram below.

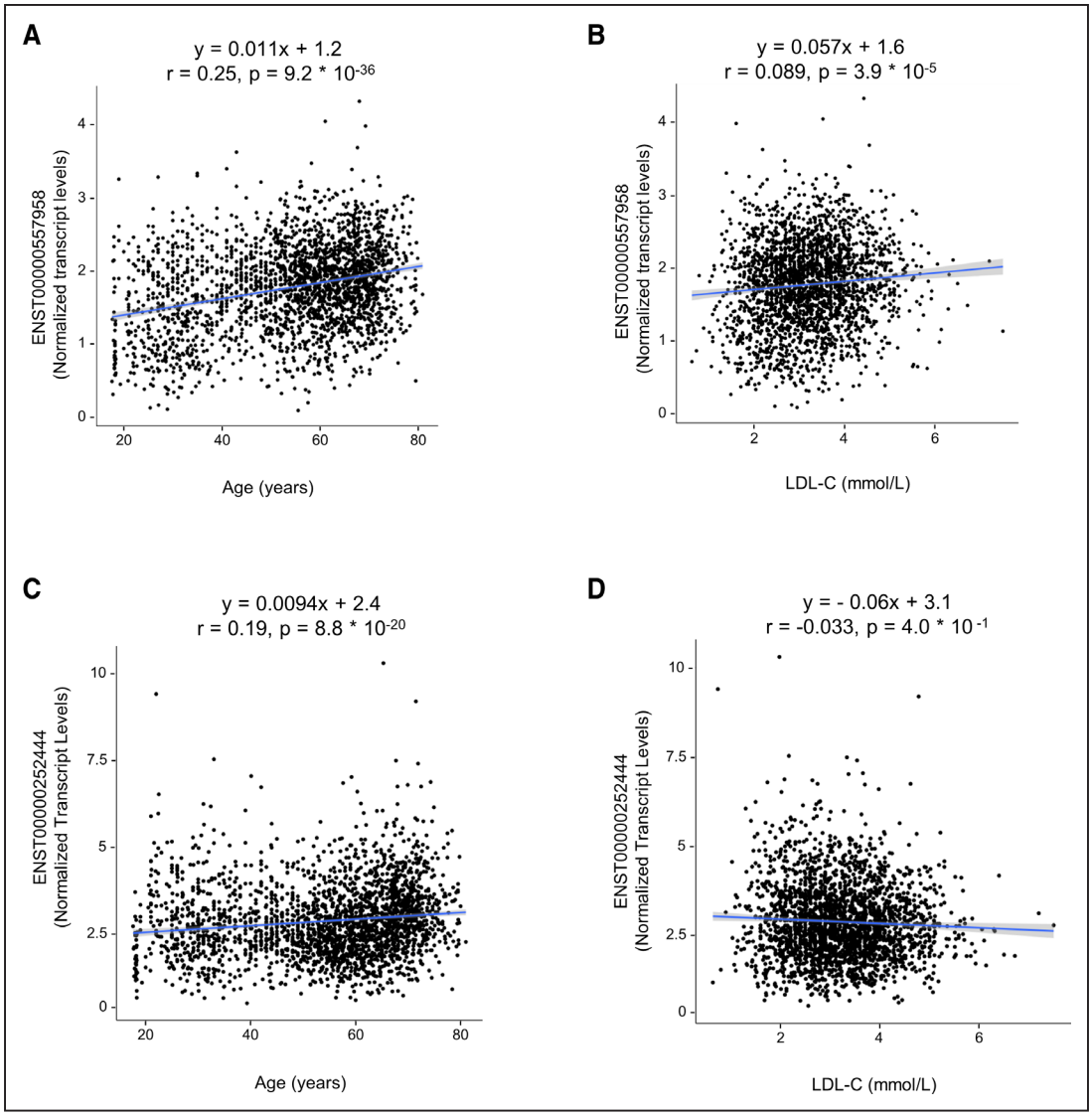

Figure 6. Correlations of the *LDLR* transcript retaining intron 3 and of a full length *LDLR* transcript in whole blood samples with age and LDL-C (low-density lipoprotein cholesterol) levels.

Data is from 2462 subjects of the BIOS population.[10] ENST00000557958 represents the intron 3 retention transcript (**A** and **B**). ENST00000252444 (**C** and **D**) was the only full length *LDLR* transcript detected in all samples analyzed. *r* and *P* values (adjusted for multiple testing using the Bonferroni correction) refer to a Spearman correlation analysis. Linear regression lines and their 95% CIs are shown in blue and gray, respectively.

complete LD (Figure S9B). With R^2 >0.8 no other SNP of RBM25 is in strong LD. A meta-analysis of 8 studies with 455 537 samples (https://cvd.hugeamp.org/variant.html?variant=rs17570658) and data of the Copenhagen City Heart and General Population Studies[13] according to METAL[14] showed the association of rs17570658 with LDL-C (*Z* Score=−4.181, P=2.9×10^{-5}, Table S9).

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 LDL-C levels.

Impaired LDL Uptake by Cells Expressing Rare RBM25 Mutants Found in Patients With Familial Hypercholesterolemia

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 familial hypercholesterolemia (FH)-causing genes.[15] We reanalyzed the burden of variants in the *RBM25* gene, using previously published whole exome sequencing data from 71 FH patients negative for mutations in *LDLR*, *APOB* and

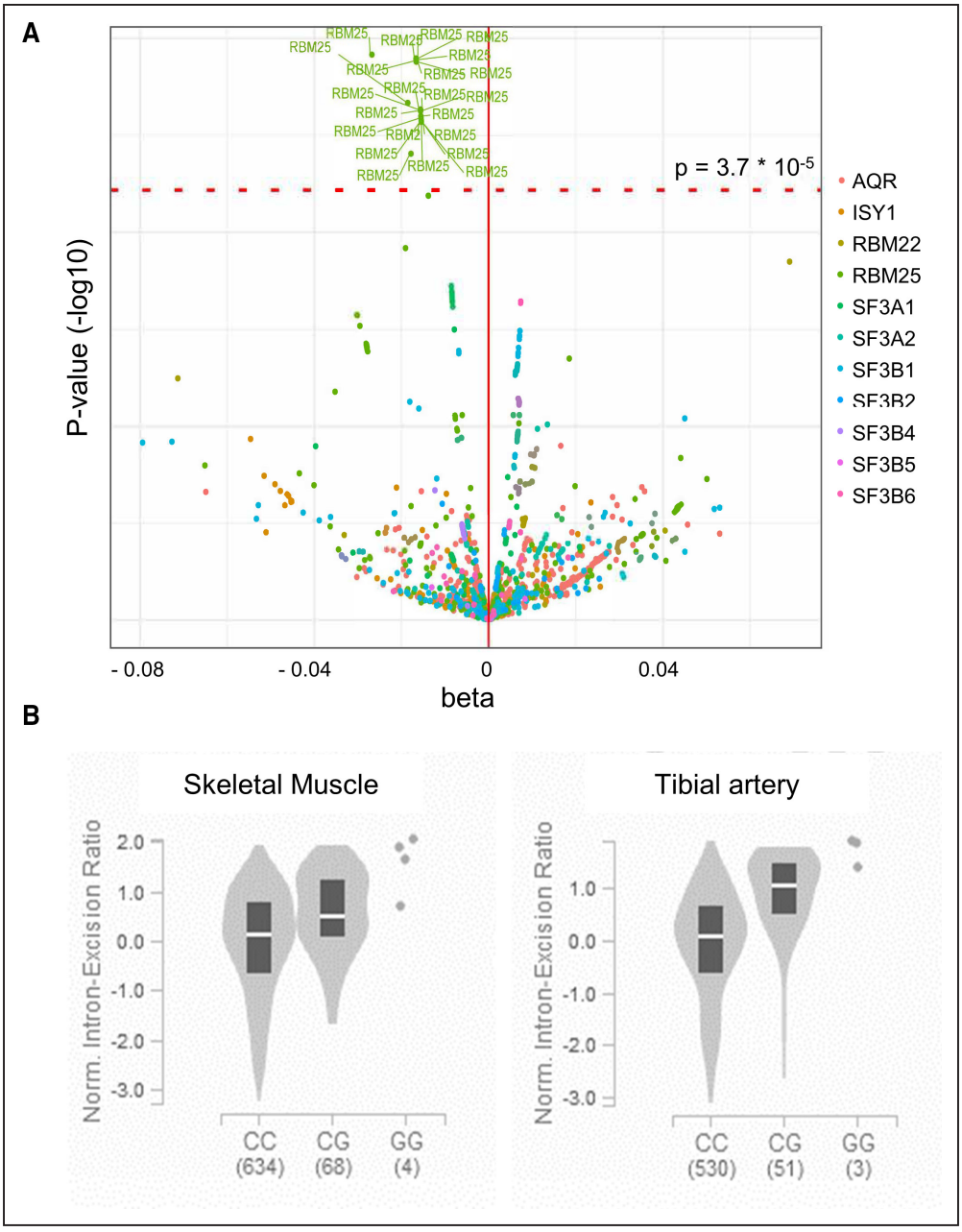

Figure 7. Association between *RBM25* variants and LDL-C (low-density lipoprotein cholesterol) in the UK Biobank data set. **A**, Association of GWAS SNPs from 11 spliceosome genes with LDL-C in the UK Biobank data set. The dashed red horizontal line indicates the threshold for statistical significance after Bonferroni 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 empirical P<1.0×10^{-8} corrected for multiple testing across genes using Storey q value method[26,27]). The horizontal white lines reflect medians; the upper and lower borders of the grey boxes reflect the 75th and 25th percentiles, respectively.

PCSK9, and 56352 European data provided by the gnomAD study.[12] Missense, splice site, frameshift, and stop-gained variants identified by whole exome sequencing in both FH cases and gnomAD were filtered to select those with minor allele frequency (MAF) <1.0×10^{-4}. After filtering, three *RBM25* variants were found in the FH cohort and 163 in the gnomAD Europeans cohort. (Table S10). Two variants, p.I152F (c.454A >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 versus gnomAD using a binomial test demonstrated the enrichment of rare variants in *RBM25* in the FH cohort (P=1.0×10^{-3}). Within the UK10K cohort, no other U2-spliceosome gene was found to carry a rare presumable LOF mutation.

We investigated the functional consequences of overexpressing the 3 FH-associated RBM25 mutants in Huh-7 cells. Overexpression of all *RBM25* constructs was confirmed by qRT-PCR (Figures S10A and S11A) and—for wild-type RBM25—Western blotting (Figure S10B). The overexpression of neither wild-type RBM25 nor any RBM25 mutant in Huh-7 cells caused significant changes in the expression of full length or IVS3 retention transcripts of *LDLR* (Figures S10C, S10D, S11B, and S11C). Compared with empty vector, overexpression of wild-type RBM25 in Huh-7 cells changed neither the cell surface abundance of LDLR nor LDL uptake significantly (Figure S10E and S10F). Comparisons with cells overexpressing wild-type RBM25 revealed minor decreases of LDLR cell surface levels (Figure S11D) but more pronounced or even significant decreases of Atto655-LDL uptake of cells overexpressing the RBM25 mutants p.L17P (−15±16%), I152F (−23±12%), or p.A455D (−28±12%, $P=2.6\times10^{-2}$; Figure S11E).

DISCUSSION

Through genome-wide siRNA screening, 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 nonfunctional receptor. In several cohorts of healthy individuals and patients, we observed considerable interindividual variation of *LDLR's* IVS3 retention in liver as well as in peripheral blood cells. Finally, we obtained initial evidence that rare genetic variants as well as SNPs associated with its expression levels in the U2-spliceosome-associated gene *RBM25* 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-C.

A previous siRNA screen also found U2-spliceosome genes to limit the uptake of LDL into EA.hy926 cells but the authors excluded them from further analysis and validation.[16] 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 Huh-7 cells.[4] As these authors discussed, CRISPR-based screens may overlook genes that are essential or confer a fitness advantage in culture, since guide RNAs targeting those genes will be progressively depleted from the pooled population.[4]

As a preliminary mechanistic explanation, our minigene data as well as our in silico predictions suggest that the branch point site 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 IVS3 retention.[5]

Medina and Krauss[17] 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. Interestingly, PTBP1 works as an inhibitor of the U2AF splice component, and thus inhibits the recognition of 3′ splice sites by the U2-spliceosome.[18] However, the knockdown of *PTBP1* resulted in very limited changes in the expression levels of the different splice forms,[17] especially when compared with the drastic changes observed in our study.

In our in vitro experiments, the knockdown of several U2-spliceosome genes and the resulting IVS3 retention compromised LDLR cell surface expression and LDL uptake as much as the knockdown of *LDLR* itself. 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[19] but may also correspond to secreted alternative splice variants.

The relative expression of *LDLR's* IVS3 transcript in human liver varies strongly due to both analytical and biological reasons, namely between 0.4% and 29% upon RNA sequencing, between 7.5% and 81% upon chip array analysis, and between 23% and 85% upon qRT-PCR. Very likely, RNA sequencing yielded the most realistic data, because this method recorded the different LDLR transcripts most comprehensively. The large interindividual variation of IVS3 expression recorded by each method indicates relevant regulatory mechanisms and consequences. We made seemingly contradictory findings on the association of IVS3 retention with NAFLD. On the one hand, the percentage of IVS3 transcripts was significantly higher in 30 patients with NAFLD or NASH than in 25 normal weight and obese control subjects without NAFLD. On the other hand, the chip array analysis found significant increases of IVS3 transcripts after bariatric surgery, which rather causes regression of NAFLD. However, although causing regression of NASH, bariatric surgery may not necessarily undo all regulatory abnormalities associated with NAFLD. In this regard, it is noteworthy, that neither RNA sequencing nor chip array analysis found any significant effect of NASH on IVS3 retention (Figure 5B and Figure S8D). Larger studies are hence needed to answer the question how NAFLD influences the expression of functional and nonfunctional *LDLR* transcripts.

In peripheral blood cells but not in liver tissue, we found a significant correlation between plasma LDL-C

levels and the IVS3 retention *LDLR* transcript, 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. However, the associations of *RBM25* SNPs with differences in *RBM25* expression and LDL-C levels 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 probability of intolerance values close to 1. Also of note, the analysis of whole exome sequencing data from the UK biobank only retrieved heterozygous mutations in U2-spliceosome genes whereas our knockdown experiments rather mimic homozygous conditions. Opposite effects on upstream regulators of *LDLR* may be another reason why the majority of SNPs and rare exome variants of the spliceosome genes 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 the intron 3 retention. In fact, RBM25 also partakes in other spliceosomal subunits.[20] Of note, RNAi with *RBM25* had the weakest effects on *LDLR* splicing and overexpression of hypercholesterolemia associated *RBM25* mutants in Huh-7 cells resulted in lower LDL uptake without affecting the expression of the *LDLR* IVS3 transcript.

The correlation between ENST00000557958 expression in blood cells with age makes us hypothesize that age-related changes in the activity of the U2-spliceosome contributes to the increase in LDL-C that parallels ageing[21] but is not mechanistically understood. The functionality of the splicing process changes with ageing.[22] Somatic mutations or decreased expression of splice factor genes, notably *SF3B1* and *RBM25* have been implicated in age-related processes, including cancer.[22,23] The total number of alternatively spliced genes also increases with age.[24] Until recently, *SIRT1* is the only known gene involved in cholesterol metabolism and atherosclerosis[25] whose alternative splicing may be disrupted with age.[22] 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 *SREBP2*. A general reason is the not optimal signal-to-noise ratio of our screening. A specific reason for missing *MYLIP* or *PCSK9* is the optimization of our 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 IVS3 retention within an *LDLR* transcript which is translated into a truncated and nonfunctional receptor protein. In both human liver and peripheral blood cells, we demonstrate that this process happens at considerable quantity and interindividual variability, possibly influenced by aging and NAFLD. 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 knockdown of *RBM25* and overexpression of loss of function mutants associated with hypercholesterolemia exerted in Huh-7 cells stronger and more consistent effects on LDL uptake than on IVS3 retention in *LDLR*.

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

ARTICLE INFORMATION

Received September 3, 2020; revision received November 18, 2021; accepted November 19, 2021.

Affiliations

Institute for Clinical Chemistry, University and University Hospital Zurich, Switzerland (P.Z., G.P., J.R., S.V., M.Y., M.K., S.R., L.R., A.v.E.); Now with Institute of Medical Genetics, University of Zurich, Switzerland (P.Z.); Division of Molecular Medicine, Department of Medicine, Columbia University, New York, NY (M.Y.); Center for Molecular Cardiology, University of Zurich, Switzerland (S.V.). Center for Integrative Human Physiology, University of Zurich, Switzerland (P.Z., G.P., S.V., M.Y., M.K., S.R., L.R., A.v.E.). Institute of Molecular Systems Biology, ETH Zurich, Switzerland (A.O.). University of Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1011-EGID, France (J.T.H., B.S.). Scientific center for optical and electron microscopy (ScopeM), ETH Zurich, Switzerland (R.M., S.F.N., A.J.R., S.S., M. Stebler). Department of Pediatrics, Section Molecular Genetics, University of Groningen, University Medical Center Groningen, the Netherlands (A.R., M.W., J.C.W., N.C.A.H., M. Smit, B.v.d.S., J.A.K.); Now with Inserm UMR 1087/CNRS UMR 6291 IRS-UN, Nantes, France (A.R.). Cardiology Research Centre, Molecular and Clinical Sciences Research Institute, St George's, University of London, United Kingdom (M.F.). LVTS-INSERM UMRS 1148 and University of Paris, CHU Xavier Bichat, Paris, France (Y.A.K., J.-P.R., C.B., M. Varret). Laboratory of Biochemistry and Molecular Therapeutics (LBTM), Faculty of Pharmacy and Pôle technologie Santé (PTS), Saint-Joseph University, Beirut, Lebanon (Y.A.K.). Department of Genetics, University of Groningen, University Medical Center Groningen, the Netherlands (F.v.D., A.v.d.G.). Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark (N.D., A.T.-H.). AP-HP, Endocrinology and Metabolism Department, Human Research Nutrition Center, Pitié-Salpêtrière Hospital, Paris, France (A. Gallo, V.C.). Université de Paris, Faculté de Médecine Paris-Diderot, UMR-S958 Paris, France; Now with Université de Paris, Institut Cochin, INSERM U1016, CNRS UMR-8104, Paris, France (A.P.). AP-HP, Université Paris-Saclay, Paris, France (J.-P.R.). UFR Simone Veil des Sciences de la Santé, UVSQ, Montigny-Le-Bretonneux, France (J.-P.R.). AP-HP, Genetics Department, CHU Xavier Bichat, Université de Paris, France (C.B.). Department of Clinical Pharmacology and Toxicology, University Hospital Zurich, Switzerland (M. Visentin). Department

of Gastroenterology and Hepatology, Antwerp University Hospital, Edegem, Belgium (L.V., J.W., S.F.). Laboratory of Experimental Medicine and Paediatrics, Faculty of Medicine, University of Antwerp, Belgium (L.V., J.W., S.F., A. Verrijken, A. Verhaegen, L.V.G.). Department of Endocrinology, Diabetology and Metabolism, Antwerp University Hospital, Edegem, Belgium (A. Verrijken, A. Verhaegen, L.V.G.). Department of Surgery, Academic Medical Center, University of Amsterdam, the Netherlands (B.V.v.R.). Division of Hepatology, Department of Medicine II, University Hospital Würzburg, Germany (A. Geier). Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, United Kingdom (S.E.H.).

Sources of Funding

This work was conducted as part of the TransCard project of the seventh Framework Program (FP7) granted by the European Commission, to J. Albert Kuivenhoven, A. Tybjaerg-Hansen, and A. von Eckardstein (number 603091) as well as partially the FP7 RESOLVE project (to J.T. Haas, B. Staels, A. Verhaegen, S. Francque, L. Van Gaal, and A. von Eckardstein) and the European Genomic Institute for Diabetes (EGID, ANR-10-LABX-46 to B. Staels). Additional work by A. von Eckardstein's team was funded by the Swiss National Science Foundation (31003A-160126, 310030-185109) and the Swiss Systems X program (2014/267 [Medical Research and Development (MRD)] HDL-X). P. Zanoni received funding awards from the Swiss Atherosclerosis Society (Arbeitsgruppe Lipide und Atherosklerose [AGLA] and the DACH Society for Prevention of Cardiovascular Diseases). G. Panteloglou received funding from the University of Zurich (Forschungskredit, grant no. FK-20-037). J. Albert Kuivenhoven is an Established Investigator from the Dutch Heart Foundation (2015T068). J. Albert Kuivenhoven was also supported by GeniusII (CVON2017-2020). The Laboratory for Vascular Translational Science (L.V.T.S.) team is supported by Fondation Maladies Rares, Programme Hospitalier de Recherche Clinique (PHRC) (AOM06024), and the national project CHOPIN (CHolesterol Personalized Innovation), granted by the Agence Nationale de la Recherche (ANR-16-RHUS-0007). Y. Abou Khalil is supported by a grant from Ministère de l'Education Nationale et de la Technologie (France). J.T. Haas was supported by an EMBO Long Term Fellowship (ALTF277-2014). B. Staels is a recipient of an ERC Advanced Grant (no. 694717). Both are also supported by PreciNASH (ANR 16-RHUS-0006). Research at the Antwerp University Hospital was supported by the European Union: FP6 (HEPADIP Contract LSHM-CT-2005-018734). S. Francque has a senior clinical research fellowship from the Fund for Scientific Research (FWO) Flanders (1802154 N). S.E. Humphries received grants RG3008 and PG008/08 from the British Heart Foundation, and the support of the UCLH NIHR BRC. S.E. Humphries directs the UK Children's FH Register which has been supported by a grant from Pfizer (24052829) given by the International Atherosclerosis Society.

Acknowledgments

We acknowledge 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 RNAi screening was performed at the ETH ScopeM facility (R. Meier, M. Stebler, S.F. Norrelykke, A.J. Rzepiela, S. Stoma). As by contract, one-third of the service costs had to be paid by grants of A. von Eckardstein to cover part of the costs for personnel, infrastructure, and maintenance. S.E. Humphries received fees for Advisory Boards of Novartis and Amryt and is the Medical Director of a UCL spin-out company StoreGene that offers to clinicians genetic testing for patients with FH. The other authors report no conflicts.

Supplemental Material

Materials & Methods
Figures S1–S11
Tables S1–S12
Major Resource Table
Original Western Blots

REFERENCES

1. Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, Hegele RA, Krauss RM, Raal FJ, Schunkert H, 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. *Eur Heart J*. 2017;38:2459–2472. doi: 10.1093/eurheartj/ehx144
2. Zanoni P, Velagapudi S, Yalcinkaya M, Rohrer L, von Eckardstein A. Endocytosis of lipoproteins. *Atherosclerosis*. 2018;275:273–295. doi: 10.1016/j.atherosclerosis.2018.06.881
3. Wilkinson ME, Charenton C, Nagai K. RNA splicing by the spliceosome. *Annu Rev Biochem*. 2020;89:359–388.
4. Emmer BT, Sherman EJ, Lascuna PJ, Graham SE, Willer CJ, Ginsburg D. Genome-scale CRISPR screening for modifiers of cellular LDL uptake. *PLoS Genet*. 2021;17:e1009285. doi: 10.1371/journal.pgen.1009285
5. 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. *Clin Chim Acta*. 2009;403:131–135. doi: 10.1016/j.cca.2009.02.001
6. Corvelo A, Hallegger M, Smith CW, Eyras E. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol*. 2010;6:e1001016. doi: 10.1371/journal.pcbi.1001016
7. Surdo PL, Bottomley MJ, Calzetta A, Settembre EC, Cirillo A, Pandit S, Ni YG, Hubbard B, Sitlani A, Carfí A. Mechanistic implications for LDL receptor degradation from the PCSK9/LDLR structure at neutral pH. *EMBO Rep*. 2011;12:1300–1305. doi: 10.1038/embor.2011.205
8. Suppli MP, Rigbolt KTG, Veidal SS, Heeboll S, Eriksen PL, Demant M, Bagger JI, Nielsen JC, Oró D, Thrane SW, et al. Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am J Physiol Gastrointest Liver Physiol*. 2019;316:G462–G472. doi: 10.1152/ajpgi.00358.2018
9. Lefebvre P, Lalloyer F, Baugé E, Pawlak M, Gheeraert C, Dehondt H, Vanhoutte J, Woitrain E, Hennuyer N, Mazuy C, et al. Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPARα-regulated dermatopontin. *JCI Insight*. 2017;2:92264. doi: 10.1172/jci.insight.92264
10. Zhernakova DV, Le TH, Kurilshikov A, Atanasovska B, Bonder MJ, Sanna S, Claringbould A, Võsa U, Deelen P, Franke L, et al; LifeLines cohort study; BIOS consortium. Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nat Genet*. 2018;50:1524–1532. doi: 10.1038/s41588-018-0224-7
11. Cirulli ET, White S, Read RW, Elhanan G, Metcalf WJ, Tanudjaja F, Fath DM, Sandoval E, Isaksson M, Schlauch KA, et al. Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun*. 2020;11:542. doi: 10.1038/s41467-020-14288-y
12. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291. doi: 10.1038/nature19057
13. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, et al; Charge Diabetes Working Group; EPIC-InterAct Consortium; EPIC-CVD Consortium; GOLD Consortium; VA Million Veteran Program. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet*. 2017;49:1758–1766. doi: 10.1038/ng.3977
14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340
15. Futema M, Plagnol V, Li K, Whittall RA, Neil HAW, Seed M, Bertolini S, Calandra S, Descamps OS, Graham CA, et al. Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR / APOB / PCSK9 mutations. *J Med Genet*. 2014;51:537–544. doi: 10.1136/jmedgenet-2014-102405
16. Kraehling JR, Chidlow JH, Rajagopal C, Sugiyama MG, Fowler JW, Lee MY, Zhang X, Ramírez CM, Park EJ, Tao B, et al. Genome-wide RNAi screen reveals ALK1 mediates LDL uptake and transcytosis in endothelial cells. *Nat Commun*. 2016;7:13516. doi: 10.1038/ncomms13516
17. Medina MW, Krauss RM. Alternative splicing in the regulation of cholesterol homeostasis. *Curr Opin Lipidol*. 2013;24:147–152. doi: 10.1097/MOL.0b013e32835cf284
18. Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, Zhou J, Qiu J, Jiang L, Li H, et al. Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol*. 2014;21:997–1005. doi: 10.1038/nsmb.2906
19. Mayne J, Ooi TC, Tepliakova L, Seebun D, Walker K, Mohottalage D, Ning Z, Abujrad H, Mbikay M, Wassef H, et al. Associations between soluble LDLR and lipoproteins in a white cohort and the effect of PCSK9 loss-of-function. *J Clin Endocrinol Metab*. 2018;103:3486–3495. doi: 10.1210/jc.2018-00777
20. 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. *J Biol Chem*. 2017;292:13381–13390. doi: 10.1074/jbc.M117.784371
21. 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. *J Clin Lipidol*. 2017;11:1055–1064.e6. doi: 10.1016/j.jacl.2017.05.007
22. Deschênes M, Chabot B. The emerging role of alternative splicing in senescence and aging. *Aging Cell*. 2017;16:918–933. doi: 10.1111/acel.12646
23. Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Buonamici S, Yu L; Cancer Genome Atlas Research Network. Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Rep*. 2018;23:282–296.e4. doi: 10.1016/j.celrep.2018.01.088
24. 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. doi: 10.1111/acel.12433
25. Miranda MX, van Tits LJ, Lohmann C, Arsiwala T, Winnik S, Tailleux A, Stein S, Gomes AP, Suri V, Ellis JL, et al. The Sirt1 activator SRT3025 provides atheroprotection in Apoe-/- mice by reducing hepatic Pcsk9 secretion and enhancing Ldlr expression. *Eur Heart J*. 2015;36:51–59. doi: 10.1093/eurheartj/ehu095
26. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park YS, Parsana P, Segrè AV, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204–213.
27. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 2003;100:9440–9445. doi: 10.1073/pnas.1530509100
28. Havel RJ, Eder HA, Bragdon JH. The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. *J Clin Invest*. 1955;34:1345–1353. doi: 10.1172/JCI103182
29. Velagapudi S, Yalcinkaya M, Piemontese A, Meier R, Nørrelykke SF, Perisa D, Rzepiela A, Stebler M, Stoma S, Zanoni P, et al. VEGF-A regulates cellular localization of SR-BI as well as transendothelial transport of HDL but not LDL. *Arterioscler Thromb Vasc Biol*. 2017;37:794–803. doi: 10.1161/ATVBAHA.117.309284
30. McFARLANE AS. Efficient trace-labelling of proteins with iodine. *Nature*. 1958;182:53. doi: 10.1038/182053a0
31. Kamentsky L, Jones TR, Fraser A, Bray MA, Logan DJ, Madden KL, Ljosa V, Rueden C, Eliceiri KW, Carpenter AE. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics*. 2011;27:1179–1180. doi: 10.1093/bioinformatics/btr095
32. 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. doi: 10.1371/journal.pone.0008471
33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–2120. doi: 10.1093/bioinformatics/btu170
34. 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. doi: 10.1093/bioinformatics/bts635
35. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013;41:e108. doi: 10.1093/nar/gkt214
36. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 2012;22:2008–2017. doi: 10.1101/gr.133744.111
37. Fedoseienko A, Wijers M, Wolters JC, Dekker D, Smit M, Huijkman N, Kloosterhuis N, Klug H, Schepers A, van Dijk KW, et al. The COMMD family regulates plasma LDL levels and attenuates atherosclerosis through stabilizing the CCC complex in endosomal LDLR traffcking. *Circ Res*. 2018;122:1648–1660. doi: 10.1161/CIRCRESAHA.117.312004
38. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010;26:966–968. doi: 10.1093/bioinformatics/btq054
39. Nisson PE, Watkins PC, Krizman DB. Isolation of exons from cloned DNA by exon trapping. *Curr Protoc Hum Genet*. 2001;Chapter 6:Unit 6.1. doi: 10.1002/0471142905.hg0601s03
40. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–527. doi: 10.1038/nbt.3519
41. Cirulli ET, Washington NL. Helix Research. UK Biobank Exome rare variant analysis v1.3 [Internet]. *Helix Blog*.
42. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR, Funke B, Cook SA, MacArthur D, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med*. 2017;19:1151–1158. doi: 10.1038/gim.2017.26
43. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, et al. Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods*. 2009;6:569–575. doi: 10.1038/nmeth.1351
44. König R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, et al. A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods*. 2007;4:847–849. doi: 10.1038/nmeth1089
45. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000;290:2323–2326. doi: 10.1126/science.290.5500.2323