

RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing

Henrike Maatz,¹ Marvin Jens,² Martin Liss,³ Sebastian Schafer,¹ Matthias Heinig,^{1,4} Marieluise Kirchner,⁵ Eleonora Adami,¹ Carola Rintisch,¹ Vita Dauksaite,³ Michael H. Radke,³ Matthias Selbach,⁵ Paul J.R. Barton,^{6,7} Stuart A. Cook,^{6,7,8,9} Nikolaus Rajewsky,² Michael Gotthardt,^{3,10} Markus Landthaler,¹¹ and Norbert Hubner^{1,8,10}

¹Max-Delbrück-Center for Molecular Medicine (MDC), Berlin, Germany. ²Systems Biology of Gene Regulatory Elements and ³Neuromuscular and Cardiovascular Cell Biology, MDC, Berlin, Germany.

⁴Department of Computational Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany. ⁵Laboratory of Cell Signalling and Mass Spectrometry, MDC, Berlin, Germany. ⁶National Heart and Lung Institute, Cardiovascular Genetics and Genomics, London, United Kingdom. ⁷Royal Brompton NIHR Cardiovascular Biomedical Research Unit, London, United Kingdom.

⁸Duke-NUS Graduate Medical School, Singapore. ⁹National Heart Center Singapore, Singapore. ¹⁰DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany.

¹¹RNA Biology and Posttranscriptional Regulation, MDC, Berlin, Germany.

Mutations in the gene encoding the RNA-binding protein RBM20 have been implicated in dilated cardiomyopathy (DCM), a major cause of chronic heart failure, presumably through altering cardiac RNA splicing. Here, we combined transcriptome-wide crosslinking immunoprecipitation (CLIP-seq), RNA-seq, and quantitative proteomics in cell culture and rat and human hearts to examine how RBM20 regulates alternative splicing in the heart. Our analyses revealed the presence of a distinct RBM20 RNA-recognition element that is predominantly found within intronic binding sites and linked to repression of exon splicing with RBM20 binding near 3' and 5' splice sites. Proteomic analysis determined that RBM20 interacts with both U1 and U2 small nuclear ribonucleic particles (snRNPs) and suggested that RBM20-dependent splicing repression occurs through spliceosome stalling at complex A. Direct RBM20 targets included several genes previously shown to be involved in DCM as well as genes not typically associated with this disease. In failing human hearts, reduced expression of RBM20 affected alternative splicing of several direct targets, indicating that differences in RBM20 expression may affect cardiac function. Together, these findings identify RBM20-regulated targets and provide insight into the pathogenesis of human heart failure.

Introduction

Alternative RNA processing is determined by RNA-binding splicing factors to generate different mRNA isoforms in a tissue-specific and developmental-regulated manner (1). Alterations in splicing factors or mutations in their target sequences result in misspliced gene products, which is the underlying cause of several human genetic diseases (2–4).

RBM20 encodes the RNA-binding motif protein 20 with highly conserved functional domains: an RNA-recognition motif 1 (RRM-1) and an arginine/serine-rich (RS-rich) region. These structural features are characteristic of a family of SR and SR-related RNA-binding proteins (RBP) that assemble in the spliceosome and orchestrate the splicing of pre-mRNA. RBM20 is predominantly expressed in striated muscle, with the highest levels in the heart (5). Recently, mutations in *RBM20* have been shown to cause human dilated cardiomyopathy (DCM) (6, 7), a major cause of heart failure and significant source of mortality and morbidity worldwide (8). Given the role of RBM20 in cardiomyopathy, efforts toward the identification of its RNA targets have been employed (5, 9) with the goal that their discovery would shed light upon human heart failure mechanisms and provide options for

molecular therapy. No precise RNA recognition element (RRE) has been defined for RBM20, and with titin (*Ttn*), only a single bona fide mRNA target has been confirmed (5, 7). No protein interaction partners have been identified for RBM20, despite the fact that previously reported RBM20 mutations associated with cardiomyopathy cluster in the RS-rich region (7) that is predicted to mediate protein-protein interactions (7, 10–12).

Here, we reveal the previously elusive RBM20 RRE at nucleotide resolution using photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) (13, 14) and transcriptome-wide RBM20-binding sites in heart-specific transcripts using high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) (15). To identify direct regulatory targets of RBM20, we integrated the crosslinking immunoprecipitation (CLIP-seq) data with the analysis of transcriptomes of heart tissue obtained from WT rats and rats deficient for RBM20 protein as well as human heart failure patients. These data yielded a transcriptome-wide RNA map for RBM20-dependent splicing regulation in cardiac tissue. Using quantitative SILAC-based proteomics (where SILAC indicates stable isotope labeling by amino acids in cell culture) (16), we discovered RBM20 protein-binding partners. We revealed that RBM20 recognizes splicing silencers at defined intronic positions and interacts with components of the spliceosomal complex A but not complex B. Importantly, patients with severe heart failure and low levels of *RBM20* gene expression in heart tissue show a loss of splicing

Authorship note: Marvin Jens and Martin Liss contributed equally to this work.

Conflict of interest: The authors have declared that no conflict of interest exists.

Submitted: December 3, 2013; **Accepted:** May 13, 2014.

Reference information: *J Clin Invest*. 2014;124(8):3419–3430. doi:10.1172/JCI174523.

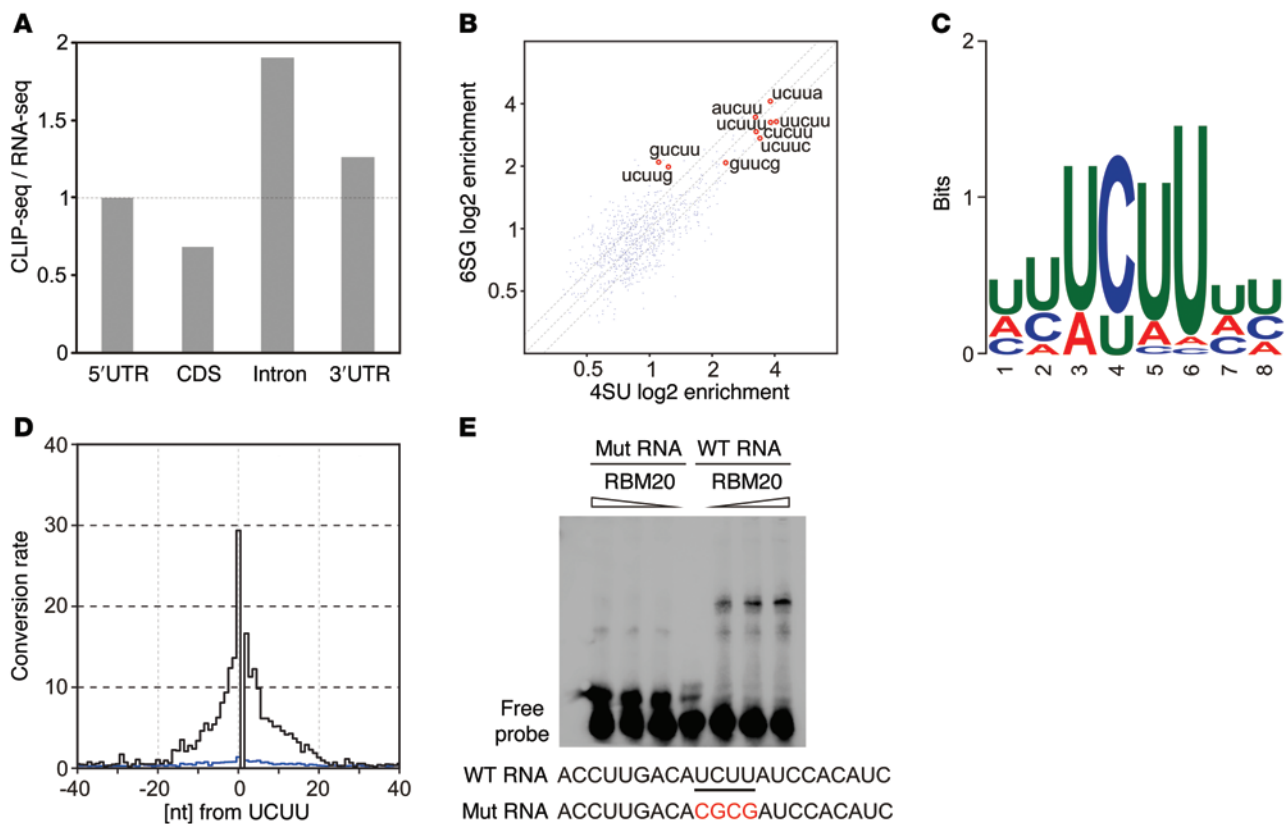


Figure 1.

PAR-CLIP in HEK293 cells identifies RBM20 RRE. (A) Transcriptome-wide distribution of RBM20 PAR-CLIP consensus library clusters. The ratio of PAR-CLIP to control HEK293 RNA-seq percentage coverage in each region is displayed. (B) Binding motif enrichment in RBM20 PAR-CLIP libraries. The log₂ enrichments of 5-mers in cluster-centered regions of the 4SU and 6SG libraries are correlated. UCUU core-containing 5-mers were most abundant. (C) Sequence logo for the RBM20 RRE was computed from the top 1000 intronic binding sites of the consensus library using MEME. (D) Positional transition frequency for intronic PAR-CLIP clusters anchored at the UCUU core of the RBM20 RRE (black line) and control CCRs randomly placed in the same intron (blue line). (E) Phosphorimage of native PAGE resolving complexes of immunoprecipitated full-length RBM20 protein with WT and mutated *Ryr2* target RNA oligonucleotides.

repression for several previously unknown direct RBM20 targets compared with those of high *RBM20*-expressing patients. Endogenous differences of *RBM20* expression may thus modulate cardiac function and may contribute to the molecular events responsible for cardiomyopathy and heart failure.

Results

Transcriptome-wide identification of RBM20-binding sites. To identify *cis*-acting RNA elements recognized by RBM20, we performed PAR-CLIP in human embryonic kidney 293 (HEK293) cells with stable, inducible expression of N-terminally FLAG/HA-tagged human WT RBM20. The epitope-tagged RBM20 localized predominantly to the nucleus, consistent with a role in splicing, with no detectable endogenous RBM20 expression (Supplemental Figure 1A; supplemental material available online with this article; doi:10.1172/JCI74523DS1). PAR-CLIP experiments were carried out independently with 4-thiouridine (4SU) and 6-thioguanosine (6SG) to reduce bias for specific nucleotide compositions (Supplemental Figure 1B). Crosslinked RNA of 2 4SU and 1 6SG experiment was recovered, transcribed into cDNA libraries, and Illumina sequenced. For each library, we generated more than 2 million distinct sequencing reads (Supplemental Table 1) that uniquely

mapped to the human genome with up to 1 mismatch. A hallmark of PAR-CLIP is the occurrence of diagnostic nucleotide transitions of crosslinked thio-nucleosides during reverse transcription. The PAR-CLIP reads aligned with T to C mismatches in 4SU, and G to A in 6SG libraries, respectively, indicating specific mRNA-RBM20 crosslinking (Supplemental Figure 1C). To investigate only those binding sites that were supported by reads from at least 2 out of 3 experiments, we computationally analyzed all 3 libraries jointly (see Methods; ref. 14) with an estimated false discovery rate (FDR) below 5% (Supplemental Figure 1, D and E). Within target transcripts (Supplemental Table 2), the resulting consensus read clusters mapped most frequently to introns (Figure 1A).

Identification of the RBM20 RRE. The diagnostic nucleotide transitions of PAR-CLIP experiments have been shown to occur at crosslinking sites and reflect the immediate positions of cellular RBP-RNA interactions (13). To elucidate the sequence preference of RBM20, we performed 5-mer counting in crosslink-centered regions (CCRs) derived from 4SU and 6SG experiments. The most abundant element was the UCUUA 5-mer, followed by variations of this element, all containing a UCUU core (Figure 1B). These UCUU-containing 5-mers were most strongly enriched not only in the 4SU but also in the 6SG experiment, demonstrat-

Table 1. Direct RBM20-regulated genes

Gene ID	Gene symbol	Name
ENSRNOG00000011589	<i>Camk2d</i>	Calcium/calmodulin-dependent protein kinase II delta
ENSRNOG00000012207	<i>Dst</i>	Dystonin
ENSRNOG00000031934	<i>Enah</i>	Enabled homolog (<i>Drosophila</i>)
ENSRNOG00000009097	<i>Immt</i>	Inner membrane protein, mitochondrial
ENSRNOG00000011343	<i>Ldb3</i>	LIM domain binding 3
ENSRNOG00000009702	<i>Lmo7</i>	LIM domain only protein 7
ENSRNOG00000005934	<i>Mlip</i>	Muscular-enriched A-type laminin-interacting protein
ENSRNOG00000019892	<i>Lrrfip1</i>	Leucine-rich repeat (in FLII) interacting protein 1
ENSRNOG00000016983	<i>Myh7</i>	Myosin heavy chain 7
ENSRNOG00000026700	<i>Myom1</i>	Myomesin 1
ENSRNOG00000012512	<i>Nexn</i>	Nexilin
ENSRNOG00000022790	<i>Obscn</i>	Obscurin
ENSRNOG00000012658	<i>Pdlim3</i>	PDZ and LIM domain 3
ENSRNOG00000004621	<i>Rtn4</i>	Reticulon 4
ENSRNOG00000017060	<i>Ryr2</i>	Ryanodine receptor 2
ENSRNOG00000015658	<i>Sorbs1</i>	Sorbin and SH3 domain containing 1
ENSRNOG00000033734	<i>Tnnt2</i>	Troponin T type 2
ENSRNOG00000022637	<i>Ttn</i>	Titin

The listed direct targets show Rbm20-dependent differential splicing and are directly bound by RBM20.

allele frequencies (DAFs) of SNPs in RREs and clusters to intronic and exonic SNPs and observed an enrichment of small DAFs in motifs and clusters when located in introns (Supplemental Figure 1E). Testing whether the DAF distributions differ among groups confirmed that DAF distributions are shifted significantly toward smaller values (Supplemental Table 3).

HITS-CLIP in rat cardiomyocytes reveals direct heart-specific targets. Cardiomyocytes are terminally differentiated cells with highly specialized functions. Although the majority of the transcriptome from HEK cells is also expressed in cardiomyocytes, expression profiles differ substantially among these cell types, with several transcripts that are specific to the cardiac muscle. To detect heart-specific RBM20-binding sites, we crosslinked endogenous RBM20 protein with pre-mRNA in rat cardiomyocytes and performed high-throughput sequencing of immunoprecipitated RNA (HITS-CLIP) (15). The HITS-CLIP libraries were analyzed essentially with the same computational approaches as for the PAR-CLIP experiments (see Methods). The most prominent RBM20-bound target RNA in cardiomyocytes was *Ttn* (Supplemental Table 4), the only direct RBM20 target previously known (5, 9). Transcriptome-wide, cardiac RBM20 clusters mapped predominantly to introns, as observed in HEK293 cells (Figure 2A). Our motif analysis of the HITS-CLIP data identified the same RRE in cardiomyocytes and mirrored our findings from the PAR-CLIP experiments (Figure 2, B and C).

RBM20 represses cassette and regulates mutually exclusive exon splicing. To detect RBM20-dependent exon usage in the heart, we analyzed existing paired-end sequencing data of RNA from *Rbm20*^{-/-} and WT rat hearts using DEXSeq (18). We quantified changes in alternative splicing by comparison of percentage spliced in (PSI) values, representing the fraction of mRNAs that include an alternative exon (Supplemental Table 5), and validated several candidates

by RT-PCR and quantitative PCR (qPCR) (Figure 2D and Supplemental Figure 3, A and B). To distinguish direct RBM20-associated splicing events from those that may be secondary in nature, we intersected the list of transcripts containing differentially spliced exons with HITS-CLIP-derived RBM20-binding sites (Supplemental Figure 3B). This identified 97 RBM20-regulated exons (Supplemental Table 5), of which 80 (82%) were located in 18 genes that carried RBM20 HITS-CLIP clusters (Table 1). Of the RBM20-regulated exons, 4 times more were repressed than activated, suggesting a role of RBM20 as a splicing repressor in the heart (Figure 2E). We found that mutually exclusive splicing was the most significantly enriched RBM20-regulated splicing event, which we observed approximately 14-fold more often than expected in cardiac-expressed transcripts ($P = 1.55 \times 10^{-11}$) (Figure 2F). Mutually exclusive splicing is frequently involved in regulating highly tissue-specific functions (19) and is in line with the specific expression pattern of RBM20.

Intronic RBM20-binding positions flanking alternative exons correlate significantly with splicing repression. In *Ttn*, the identified RBM20-binding sites were predominantly located in the PEVK- and

the immunoglobulin-rich regions within the I-band and were associated with aberrant splicing of consecutive exons in these regions (Figure 3A). RBM20-binding sites in the direct targets LIM domain binding 3 (*Ldb3*) and calcium/calmodulin-dependent protein kinase II δ (*Camk2d*) were closely located near mutually exclusive spliced exons (Figure 3, B and C). This led us to determine whether the position of RBM20-binding sites on pre-mRNA is indicative for the occurrence of differential exon usage. We found a significant overrepresentation of RBM20-binding sites in a window spanning 400 bp upstream of the 5' splice site to 400 bp downstream of the 3' splice site of differentially spliced exons (Odds ratio = 78.99, $P < 2.2 \times 10^{-16}$). In transcripts with both differential exon usage and RBM20-binding sites, 97% of clusters contained the identified RRE (UCUU), indicating its functional significance. We further differentiated between RBM20-activated and -repressed events and computed the mean CLIP-read density within or near regulated and control exons on a global and high-resolution scale. Very little RBM20 binding was associated with exon activation (Figure 3D). In contrast, RBM20 binding peaked 50 nucleotides upstream and 100 nucleotides downstream of repressed exons. This indicates that RBM20 binding predominantly at these distinct positions is required for exon repression.

We used a reporter assay to determine whether the identified cardiac RBM20-binding positions are necessary for splicing activity. The splice reporter construct was based on the RBM20-dependent N-terminal titin PEVK region with a firefly luciferase in the alternative exon 8 (PEVK8), a cluster containing 2 UC UU RREs approximately 100 bp downstream of exon 8, and a Renilla luciferase in the 3' exon (Figure 3E and Supplemental Figure 3D). Mutation of the UC UU elements within the cluster sequence resulted in the reduced exclusion of PEVK8 (Figure 3E). On the other hand, nucleotide changes in UC UU sequences without evidence

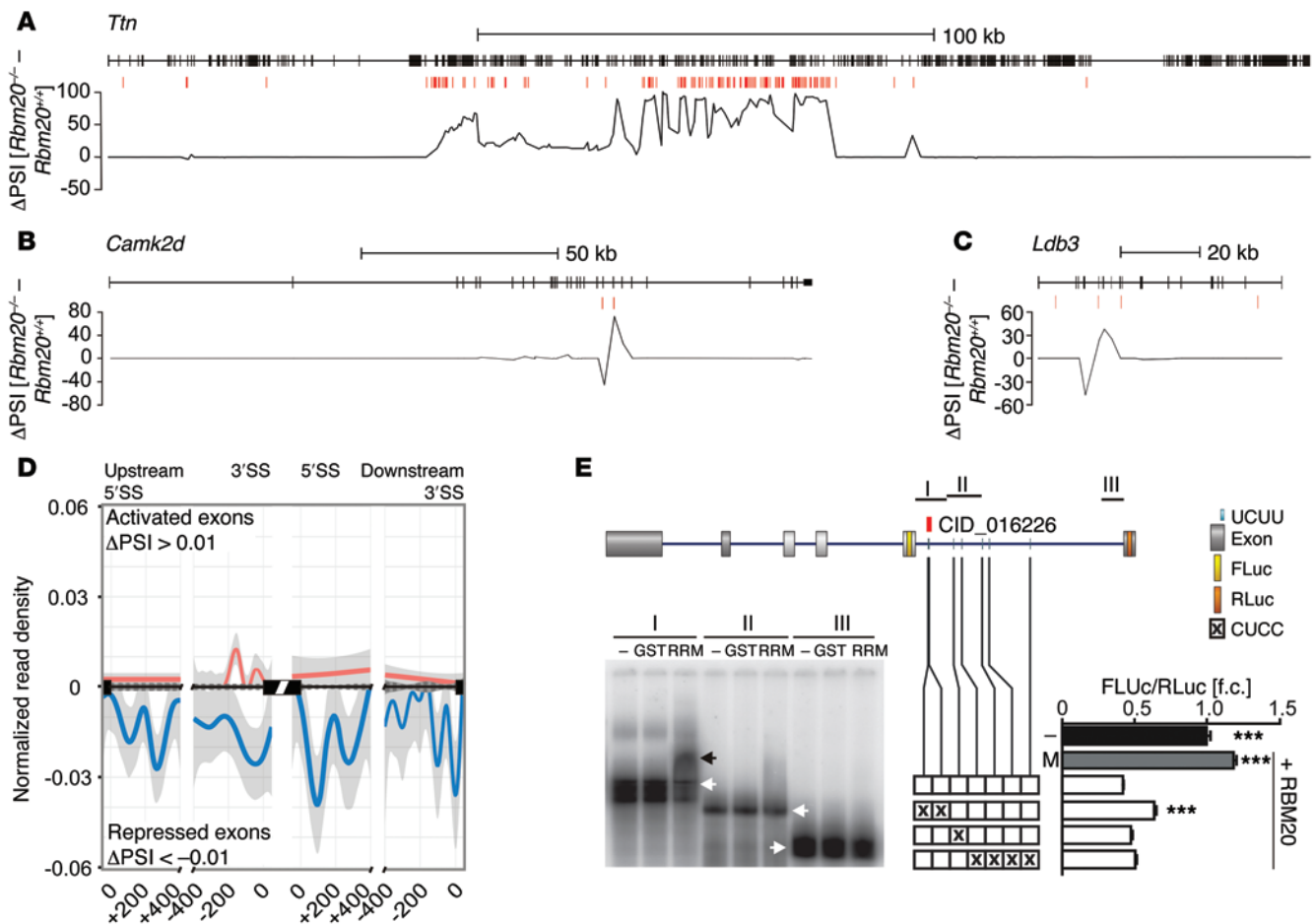


Figure 3. RBM20 binds to intronic splicing silencers upstream and downstream of repressed exons. (A–C) Δ PSIs are shown as black lines (WT compared with homozygous rats). Elevated versus decreased Δ PSIs show inclusion or exclusion of exons, respectively. Red ticks indicate mapped RBM20 cluster positions. (A) *Ttn* is predominantly spliced in the elastic region. Mapped RBM20-binding sites coincide with differential splicing. (B) *Camk2d* exons 14–16 are mutually exclusive and regulated by RBM20. Mapped RBM20-binding sites are located upstream of exon 14 and downstream of exon 15. (C) *Ldb3* undergoes an exon switch affecting exon 4 versus exons 5–7. Mapped binding sites are upstream of exon 5 and downstream of exon 8. (D) Mean CLIP density near RBM20 activated (red) and repressed exons (blue). Dotted line indicates density around exons not regulated by RBM20. Gray shows 90% CIs. (E) Right panel: specific activity of RBM20 on *Ttn* RNA from the PEVK region in a splice-reporter assay. *Rbm20* expression leads to exclusion of the firefly luciferase-containing (FLuc) exon. Ratio of FLuc to Renilla luciferase (RLuc) activity (downstream exon) reflects splice activity. Reporter activity depends on mutation of 2 UCUU elements in cluster CID016226, but not outside the cluster. Mutations are indicated as X. White boxes only show native reporter. Controls: PEVK-construct cotransfected with empty vector (–), *Rbm20* cotransfected with a *Ttn* M-band region construct not affected by RBM20 (M). *** $P < 0.001$ compared with native reporter. Left: EMSA evaluating RBM20 binding to *Ttn* PEVK-derived RNAs. White arrows indicate input RNA; black arrows indicate RNA-protein complex.

for RBM20 binding did not affect reporter activity. This is further supported by mobility shift assays, in which we document that RNA containing the cluster-binding site downstream of PEVK8 interacted with the RRM domain of RBM20 (RNA I), while RNA containing UCUU elements that do not correspond to the cluster (RNA II and III) did not bind to RBM20 (Figure 3E).

Mutation of the RBM20 RS-rich region abrogates protein-protein interactions with other splicing factors. To identify RBM20 interaction partners, we used quantitative affinity purification followed by mass spectrometry (q-AP-MS) in SILAC-labeled HEK293 cells (ref. 16 and Figure 4A). In total, we identified 538 proteins precipitating with RBM20 (2-fold enriched compared with controls) and validated several interactions by Western blotting (Figure 4B and Supplemental Table 6). This set of proteins was significantly enriched for the Gene Ontology (GO) Biological Process “RNA splicing” ($P = 2.7 \times 10^{-38}$)

with 138 proteins being frequently identified in multiple spliceosome preparations (20). In particular, we found all 7 Sm and nearly all U1 and U2 small nuclear ribonucleic particle (snRNP) proteins (except SNRPA, SNRPB2, SF3B5, and PHF5A) to precipitate with RBM20 (Figure 4C). In contrast, only few of the U4/U6 and U5 snRNP proteins were identified in the RBM20 pull-down. In addition, while we identified 45% of complex A-specific proteins, including U2AF65 and U2AF35, we found only 18% and 11% of complex B- or complex C-specific proteins, respectively (Supplemental Figure 4). In particular, only 3 out of 20 proteins of the Prp19 complex and related proteins that are crucial during catalytic activation and join the spliceosome during complex B formation were found. Together, these data suggest that RBM20 interacts with U1 and U2 in early stages of spliceosome assembly, while it is not present in catalytically active complexes. It has been shown that the RS-rich region of

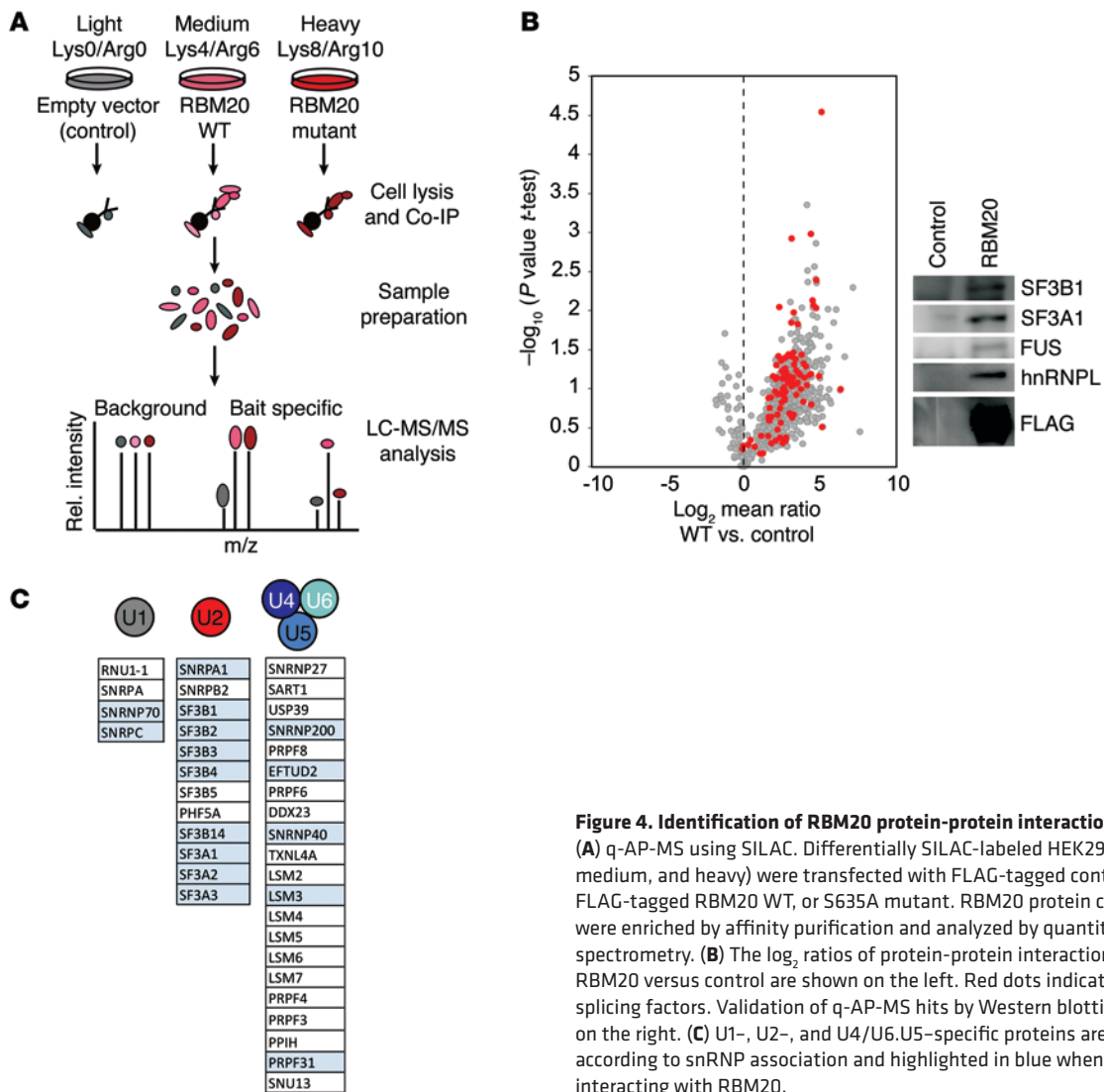


Figure 4. Identification of RBM20 protein-protein interaction partners. (A) q-AP-MS using SILAC. Differentially SILAC-labeled HEK293 cells (light, medium, and heavy) were transfected with FLAG-tagged control vector, FLAG-tagged RBM20 WT, or S635A mutant. RBM20 protein complexes were enriched by affinity purification and analyzed by quantitative mass spectrometry. (B) The \log_2 ratios of protein-protein interactions for WT-RBM20 versus control are shown on the left. Red dots indicate known splicing factors. Validation of q-AP-MS hits by Western blotting is shown on the right. (C) U1-, U2-, and U4/U6.U5-specific proteins are grouped according to snRNP association and highlighted in blue when identified as interacting with RBM20.

RBM20 is a hot spot for mutations that cause DCM. We therefore determined the impact of a previously reported RBM20 S635A mutation located in the RS-rich region (5) on protein binding by comparing proteins precipitating with WT RBM20 versus S635A-RBM20. The mutation did not abrogate interactions with core spliceosomal proteins, but for 38 of the alternative spliceosomal proteins associated with WT RBM20, interactions were significantly reduced when the RS-rich region of RBM20 was mutated (Figure 5A). To distinguish RNase insensitive interactions indicating binders not dependent on RNA as a connector, we repeated the experiment with an RNase digestion step after immunoprecipitation and compared the results with the nondigested data set. This revealed 13 of the WT-specific spliceosomal protein interactions as RNase insensitive (Figure 5B). Using a label-free quantification, 10 of these RNase insensitive WT-specific proteins identified in HEK cells were also found to precipitate with endogenous-expressed RBM20 in cardiomyocytes (Figure 5B). Overall, the set of interaction partners in cardiomyocytes encompassed the majority of splicing factors identified in the SILAC experiment, including U2AF65 and U2AF35, but only 1 protein of the Prp19 complex (Supplemental Table 7), corroborating our findings from HEK cells in cardiomyocytes.

Regulatory impact of RBM20 on cardiomyopathy-associated and novel candidate genes. NCBI Medical Subject Heading (MeSH) analyses of all 18 directly RBM20-regulated targets revealed that the set was significantly enriched for the heading “dilated cardiomyopathy” ($P = 1.33 \times 10^{-13}$) (Supplemental Table 8). Further significantly enriched disease-related MeSH terms were “hypertrophic cardiomyopathy” ($P = 1.8 \times 10^{-09}$) and “heart failure” ($P = 8.48 \times 10^{-05}$). The identification of direct RBM20 target genes and their regulated exons allowed us to specifically test their relevance in a large cohort of heart failure patients, including a single patient that we reported previously with a heterozygous S635A mutation in RBM20 (5).

We compared the rat cardiac-splicing patterns of direct targets to those of the RBM20 S635A individual and 5 control subjects with DCM, but without any nonsynonymous changes in their RBM20-coding sequence (Supplemental Table 9). This analysis revealed conserved differential splicing of orthologous direct targets at the exon level for *Camk2d*, *Ldb3*, LIM domain only protein 7 (*Lmo7*), PDZ and LIM domain 3 (*Pdlim3*), reticulon 4 (*Rtn4*), *Ryr2*, and *Ttn* (Figure 6 and Supplemental Figure 5). As expected from our analysis, the differentially spliced exons were flanked

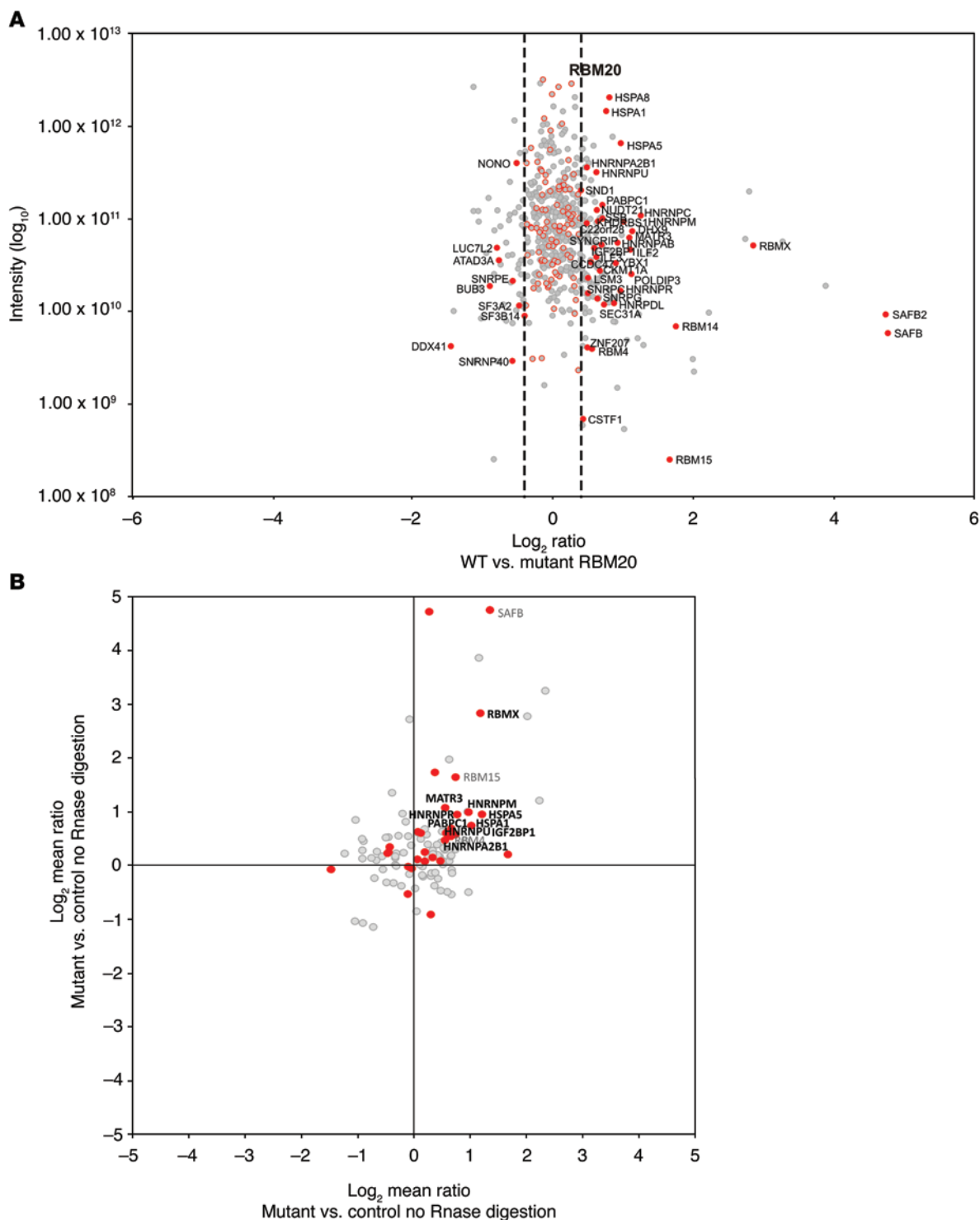


Figure 5. Impact of RBM20 S635A mutation on protein-protein interaction. (A) WT versus S635A comparison. RBM20 was identified as highly and equally abundant in WT- and S635A-RBM20 samples, indicating successful immunoprecipitation. Red circles indicate spliceosomal interaction partners not affected by the mutation. Labeled red dots indicate spliceosomal interaction partners specific to WT- (\log_2 ratio > 0.4) or S635A-RBM20 (\log_2 ratio < -0.4). (B) Comparison of RNase-digested versus nondigested q-AP-MS identifies RNase-insensitive RBM20 interactors. Upper right quadrant contains WT- and lower left quadrant mutant-specific interactors enriched in RNase-digested and nondigested samples. Spliceosomal interaction partners are marked by red dots and labeled when WT- or mutant-specific in both protein-protein interaction screens. Bold labeled proteins were identified as also interacting with RBM20 in cardiomyocytes.

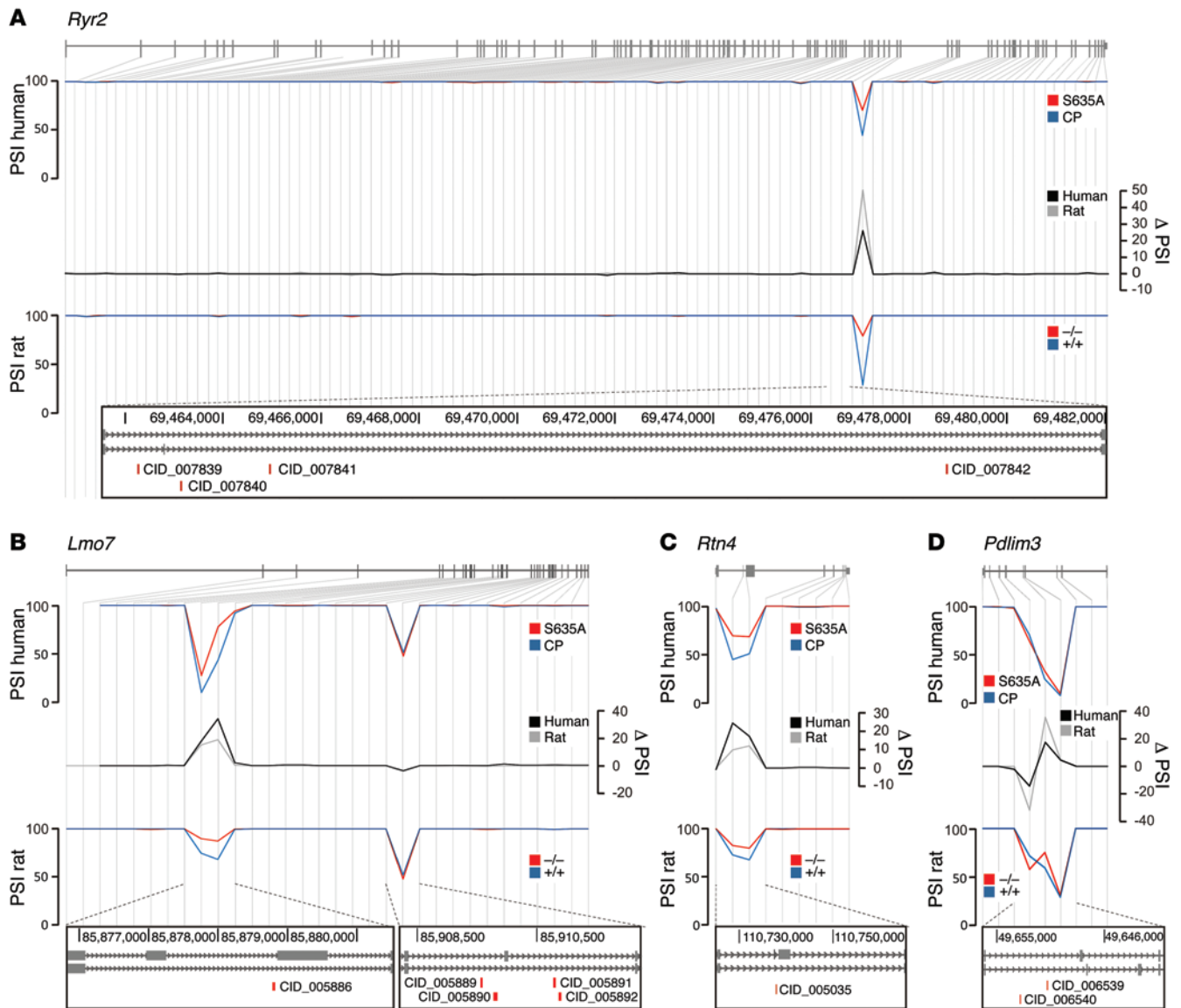


Figure 6. Alignment of orthologous rat and human exons for *Ryr2*, *Lmo7*, *Rtn4*, and *Pdlim3* to compare RBM20-regulated exon usage. PSI scores for the RBM20-deficient individual (S635A) are indicated in red and the average of 5 control subjects from the human heart failure cohort (CP) in blue. The average PSI of 3 rats per genotype is shown below (WT in blue and *Rbm20*^{-/-} in red). Resulting Δ PSIs are shown in black (S635A compared with CP) and in gray (WT compared with *Rbm20*^{-/-} rats). Deflections in Δ PSI values in differentially spliced regions are highly conserved across species. (A) *Ryr2* shows RBM20-dependent splicing of a 24-bp exon included in RBM20-deficient hearts. The magnified view shows RBM20 cluster flanking the differentially spliced rat *Ryr2* exon. (B) In *Lmo7*, RBM20 deficiency causes retention of exons 9 and 10. The zoom-in shows RBM20 cluster flanking exon 10 immediately upstream of its 3' splice site. (C) *Rtn4* exons 3 and 4 are differentially spliced. The magnification shows the intronic RBM20 cluster flanking rat exons 3 and 4. (D) *Pdlim3* exons 4–6 are mutually exclusive and differentially regulated by RBM20. The zoom-in shows the locations of RBM20 cluster upstream of exon 4.

by RBM20 clusters in downstream and/or upstream introns in all cases. In addition to *Ttn*, *Ldb3* and *Camk2d* have previously been suggested as RBM20 regulated. Here, we confirm that the direct interaction between RBM20 and their pre-mRNA leads to the observed differential exon usage in rats and humans (Supplemental Figure 5). Interestingly, *Lmo7*, *Pdlim3*, *Rtn4*, and *Ryr2* have not been implicated in the molecular target framework of RBM20 so far, but are here clearly identified as RBM20 regulated (Figure 6 and Supplemental Figure 3B).

Moreover, we tested the impact of differential expression levels of RBM20 on the splicing of its target exons in cardiac tissue from a human heart-failure cohort. In heart tissue from 148

patients with end-stage heart failure, we observed that cardiac RBM20 expression greatly varies among individuals (Supplemental Figure 6). We compared the splicing patterns of RBM20-regulated exons in 10 individuals, each with the highest versus lowest RBM20 expression (Figure 7 and Supplemental Table 9). The difference in endogenous RBM20 expression (high vs. low) significantly affected the splicing of orthologous exons for *CAMK2d*, *LDB3*, *RYR2*, and *TTN* and mirrors our observations from the S635A mutant patient and *Rbm20*-deficient rats. Given the fundamental role of these targets for proper heart function, this may suggest that endogenous RBM20 expression levels are important for the modulation of cardiac function.

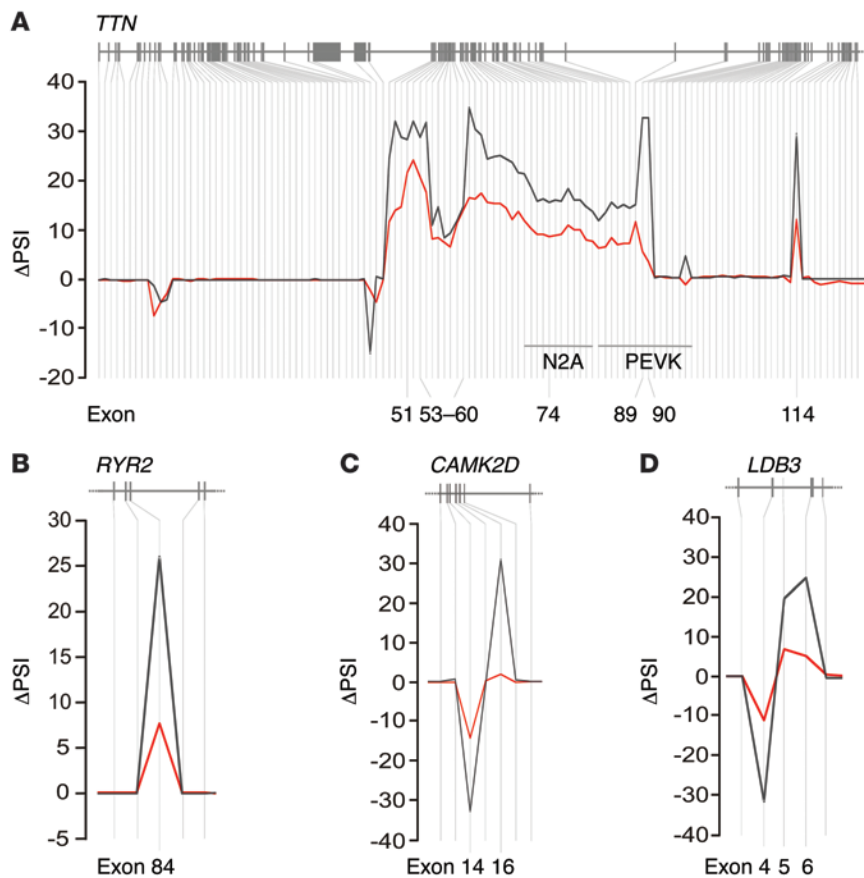


Figure 7. Splicing patterns of RBM20-regulated exons in individuals with differential endogenous RBM20 expression levels. Δ PSI values for RBM20 regulated exons of (A) *TTN*, (B) *RYR2*, (C) *CAMK2D*, and (D) *LDB3* are shown in red for low compared with high RBM20-expressing individuals. Δ PSI values for S635A compared with control subjects shown in gray serve as indicators of RBM20 regulated events.

Discussion

In the current study, we identified direct cardiac targets of RBM20 and revealed its immediate binding sites in these transcripts. We showed that RBM20 binds to a UCUU-containing RNA element in nascent transcripts in rat as well as human cells, indicating the conservation of the binding site and splicing regulation across species. We found evidence for negative selection on RBM20-binding sites in human populations. RREs and CLIP clusters were significantly depleted of SNPs, and SNPs occurring inside motifs or clusters were enriched for small DAFs. These findings underscore the functional significance of RBM20-binding sites in humans, as variants in binding sites are evolutionary selected against throughout the genome.

Importantly, we found that RBM20 predominantly binds to intronic sequence in close proximity to differentially spliced exons. SR and SR-related proteins are usually regarded as splicing activators (10), recognizing exonic splicing enhancers (21). Interestingly, there are sporadic examples of SR protein involvement in splicing repression. For few targets, it has been recently shown that this protein family can repress splicing when bound at intronic positions (22–25), but SR and SR-related proteins have not yet been shown to act transcriptome wide as splicing repressors from intronic positions. Here, we found that RBM20 binds predomi-

nantly to introns near the 3' and 5' splice sites of repressed exons and in close proximity to the binding sites of U1 and U2 snRNPs. The binding of RBM20 to RNA appears to be important, as mutating the RBM20-binding site affects the splicing activity. Furthermore, mice lacking the RBM20 RNA recognition motif-1 domain (inframe deletion of exons 6 and 7) show aberrant splicing of its major target, *Ttn* (26).

We demonstrated the interaction of RBM20 with nearly all U1 and U2 snRNP proteins as well as U2-related proteins (including U2AF65 and U2AF35) by quantitative proteomic approaches. The data suggested a predominant presence of RBM20 in complex A, but not in catalytically active spliceosomes. This was supported by interactions of RBM20 with spliceosomal proteins that are exclusively represented in complex A but missing with complex B proteins such as the U4/U6.U5 tri-snRNP and the Prp19 complex. We speculate that the particular RBM20-binding sites may prevent the recruitment of the U4/U6.U5 tri-snRNP interfering with complex B formation subsequently leading to splicing repression.

Interaction with other proteins is predicted to be mediated by the SR-rich region of RBM20 (7). Here, we showed that a disease-causing mutation underlying DCM within the SR-rich region disrupted the interaction of RBM20 with alternative splicing factors, but did not interfere with core splicing factor binding. This suggests that RBM20 presence in complex A may need to be stabilized by additional alternative splicing factors. Disruption of this process may lead to subsequent displacement from the spliceosome and loss of splicing repression, as observed in *Rbm20* deficiency.

The combination of the splicing analysis and determination of direct RNA-binding sites in the heart revealed several directly regulated RBM20 targets unknown to date. We discovered DCM-related and several other genes implicated in cardiac disorders and validated representatives in human heart failure patients. Interestingly, the identified targets allow further insights and reveal new RBM20-related disease mechanisms: *RYR2* is found primarily in cardiac muscle and is a component of the major Ca^{2+} release channel for excitation-contraction coupling in the sarcoplasmic reticulum membrane. A 24-bp exon in *RYR2* is upregulated in *Rbm20*-deficient rats and human cardiomyopathy patients (Figure 6A). This exon is usually present at low levels in hearts and shifts *RYR2* from the ER to the intranuclear Golgi apparatus, with profound effects on intracellular Ca^{2+} signaling (27). Mutations in N-terminal domains of *RYR2* have been associated with arrhythmic right ventricular tachycardia and sudden death (28). Cardiac arrhythmia and sudden death are both signs of RBM20-related DCM (5), and recent studies suggest that defective *RYR2*

function may be associated not only with cardiac arrhythmias, but also cardiomyopathies (29, 30). Here, we were able to show that the 24-bp exon is also regulated in human heart failure patients with high and low *RBM20* expression.

LMO7 is a transcription factor that previously has not been associated with DCM. In *LMO7*, *RBM20* deficiency resulted in retention of exons 9 and 10 across species (Figure 6B). These exons encode an F-box domain present only in a brain-specific isoform of *LMO7* (31). Its function in regulating expression of genes that are important for skeletal muscle and heart function (32) makes it an interesting modulator in *RBM20*-related DCM.

RTN4 protein, also known as neurite outgrowth inhibitor (NOGO), localizes predominantly in the endoplasmic reticulum. It is well studied in the nervous system, but its function in the heart remains largely unknown. In *RBM20*-mutant rats and human, exons 3 and 4 of *RTN4* are retained (Figure 6C). This isoform is most strongly expressed in the brain and only weakly in the heart (33). It is increased in left ventricular tissue from DCM and ischemic hearts and has been proposed as a potential indicator of heart failure (34–36). Here, we demonstrate that *RBM20* represses the predominantly neuronal isoform in favor of the heart-specific exon usage.

PDLIM3 binds α -actinin 2 in the Z-discs of myofibers. Two different major isoforms of *PDLIM3* are expressed in heart and skeletal muscle tissue that include either exon 4 or exons 5 and 6. The predominant isoform in the heart includes exon 4 (37). In *RBM20*-deficient rats and human, the ratio of these 2 isoforms is shifted toward the skeletal muscle form (Figure 6D). Within exon 4, a Glu106Ala mutation has been linked to hypertrophic cardiomyopathy, indicating the importance of the sequence encoded by this exon for proper heart functioning (37).

Considering the fundamental role of *RBM20* for proper splicing of cardiac transcripts, it was unexpected that its endogenous expression levels in heart failure patients varied drastically. On the other hand, this could also be an indication that differences in *RBM20* expression levels contribute to the modulation of cardiac function under pathophysiological conditions. With the set of direct *RBM20* target genes defined in this study, it was possible to investigate the impact of *RBM20* expression on differential exon usage in its targets in cardiac tissue from end-stage heart failure patients. Despite the heterogeneous nature of the human heart failure samples, we could demonstrate the regulation of differential exon usage in 4 direct *RBM20* target genes in 10 patients with high versus 10 patients with low *RBM20* gene expression levels (*CAMK2D*, *LDB3*, *RYR2*, *TTN*). These 4 genes showed the strongest differences in *RBM20*-regulated exon usage across species. It is therefore conceivable that we have missed other targets with smaller effects in our analysis due to the heterogeneous nature of the patient samples. The validity of our findings, on the other hand, is supported by the fact that the same exons within the 4 genes were also regulated in a patient that carried an *RBM20* nonsynonymous mutation (S635A) in the SR-rich region. Given the suggested impact of *RBM20* expression levels on exon usage of its direct targets, we paid careful attention to comparing the S635A patient to human cardiomyopathy controls that were matched in their expression levels but did not have any nonsynonymous variants in their *RBM20*-coding sequence.

In our study, we describe the function of *RBM20* as a molecular switch that orchestrates isoform transitions of a network of genes with essential cardiac functions. Not only genetic variation, but also the modulation of *RBM20* gene expression levels, is operative in the regulation of this network of direct targets in the human heart. This is of interest because the modulation of *RBM20* levels may be an attractive target for therapeutic intervention. Moreover, the abundance of *RBM20* in heart tissue could play a role in the pathogenesis of more common forms of heart failure irrespective of the *RBM20* genotype. The clinical significance of splicing controlled by *RBM20* expression levels during the pathogenesis of common forms of heart failure will be an interesting avenue to explore in future studies.

In summary, our analyses tie together *RBM20*-regulated direct targets, the underlying mechanism of splicing repression, and molecular consequences in clinical samples. Collectively this provides insights that are extensible into the pathophysiology of human cardiac function.

Methods

PAR-CLIP. HEK293 cells with tetracycline-inducible expression of FLAG/HA-tagged *RBM20* have been described previously (5). For PAR-CLIP, we essentially followed the protocol previously described (13). The libraries were sequenced on an Illumina Genome Analyzer GAII using the 50-bp single-read protocol.

HITS-CLIP. Neonatal rat ventricular myocytes (NRVMs) were isolated from SD rats using a modification of the Kasten's technique, as described (38). At 48 hours after isolation, cardiomyocytes were irradiated 1 time at 400 mJ/cm and an additional 200 mJ/cm with 254 nm UV light and lysed in NP40 lysis buffer. After DNase and RNase T1 treatment, the lysate was immunoprecipitated using *RBM20* antibody (5) covalently coupled to Protein G Dynabeads (Invitrogen). Radiolabeling of immunoprecipitated RNA, protein-RNA complex separation, sequencing library preparation, and sequencing were done following the PAR-CLIP protocol.

Computational analysis of PAR-CLIP and HITS-CLIP data. We used a computational pipeline that performed all steps of the analysis from raw reads to cluster sets and target genes, essentially as previously described (14), with emphasis on stringent filtering and controlling the false-positive rate in the identification of binding sites (Supplemental Methods).

Motif analysis. Word counting was performed on CCRs for PAR-CLIP and complete read clusters for HITS-CLIP. The \log_2 word enrichment was computed for intronic-binding sites as the log ratio of the relative frequency of each 5-mer to the background frequency in RefSeq introns. Words with less than 10 occurrences inside binding sites were discarded from the analysis.

Within PAR-CLIP CCRs, we searched for the UCUU that was closest to the anchor position and selected it as the new center. CCRs without UCUU were discarded. The remaining motif-centered regions (MCRs) were used to build an average profile of nucleotide conversion events, indicative of crosslinking, around the motif.

Human studies. We obtained cardiac tissue from 147 subjects with end-stage heart failure. All subjects were transplanted at the Royal Brompton and Harefield NHS Foundation Trust (London, United Kingdom). On this occasion, cardiac samples were obtained and immediately frozen in liquid nitrogen. The S635A patient was described earlier (5).

RNA-seq and splicing analyses. Total RNA was extracted using TRIzol reagent and RNA-seq libraries prepared with the TruSeq RNA Sample Preparation Kit (Illumina). Barcoded cDNA fragments of poly(A)⁺ RNA were then sequenced on a HiSeq 2000 instrument from Illumina with 2 × 100 bp PE chemistry. Reads were mapped against the reference genome GRCh37 using TopHat version 1.3.1 (39) supplied with the gene annotation of the Ensembl database (40). We used DEXSeq (18) to test all exons for differential usage and to reanalyze previously published RNA-seq data of Rbm20-deficient rats (Sequencing Archive of the European Nucleotide Archive, <http://www.ebi.ac.uk/ena/>; accession code ERP001301), as detailed in Supplemental Methods.

Gene expression analysis in the heart failure cohort. Gene expression in the heart failure cohort individuals was measured using RNA-seq. For each protein coding gene from the NCBI annotation, version 66, the overlapping TopHat aligned reads were counted using the HTSeq package. Counts across samples were normalized using a quantile based scaling method (41). Gene expression levels were adjusted for available covariates as described in Supplemental Methods.

EMSA. For EMSA, FLAG/HA-tagged full-length RBM20 or GST-tagged Rbm20 RRM domain was used (see Supplemental Methods). γ -32P-labeled RNA was incubated with immunoprecipitated RBM20 protein and 100 ng tRNA in 20 μ l binding buffer (10 mM HEPES-KOH, pH 7.9, 50 mM KCl, 3 mM MgCl₂, 1 mM DTT, and 5% glycerol). The resulting complexes were resolved on native 5% acrylamide gels run at 4°C at 200 V for 2 hours in 0.5× TBE and exposed using a PhosphorImager.

Enrichment analyses. To assess functional and disease-related association of RBM20 splice substrates, enrichment analysis was performed with EGAN (42). To estimate the nature of transcript-processing events leading to differential usage of Rbm20-regulated exons, overlapping transcript events as annotated by Ensembl were assigned to exonic regions. The significant enrichments of GO terms and MeSH and those for certain event types of exons in direct Rbm20 targets when compared with all exons expressed in the heart were evaluated with the hypergeometric test, and significant enrichments were adjusted for multiple testing using Bonferroni's method.

Construction of RNA maps. We determined the coverage profiles of HITS-CLIP reads for all differential used exons as defined above. We split the exons into activated (Δ PSI > 0.01) or repressed (Δ PSI < -0.01) exons and compared the coverage profiles of both sets with the read coverage of control sets of the same size that were randomly sampled from nondifferential exons. We only counted the start positions of reads, and duplicated identical reads were only counted once. We investigated regions from 400 bp upstream of the 3' splice site to 50 bp into the exon and from 50 bp upstream of the 5' splice site to 400 bp into the downstream intron of differentially used exons and their adjacent neighbors. We smoothed the coverage profiles using the ggplot2 R package, which applies a generalized additive model with integrated smoothness estimation and plots the smoothed coverage along with its 95% CI.

Splice reporter assay. We quantified RBM20 activity in a dual luciferase splice reporter assay (5) using a mini-gene representing titin PEVK exons 4 through 13 and a RNA destabilization motif for increased signal

amplitude. Several mutations were introduced into the reporter construct, as described in Supplemental Methods. HEK293 cells (25000 cells/well on 96-well plates) were transfected with the reporter-plasmids and CMV-driven RBM20 or control plasmid (pcDNA3.1) using Polyethylenimine 40 kDa (Polysciences Europe GmbH). At 72 hours after transfection, reporter activity was measured using the Dual-Luciferase Reporter Assay System (Promega), following the manufacturer's instructions, and an infinite M200 Pro Plate Reader (TECAN). Ratios of firefly to Renilla luciferase activity were normalized to the RBM20-negative control. Results were confirmed in 3 independent experiments. Based on the half-life of the resulting protein, we found higher sensitivity of the RNA-based splice analysis as compared with the luciferase readout (Supplemental Figure 3, E and F).

SILAC experiments. SILAC labeling of HEK293T cells, immunoprecipitation, and LC-MS/MS analysis were essentially performed as previously described (16) (for details see Supplemental Methods). Label swap experiments were performed to filter out contaminants. Experiments were performed in quadruplicate. Biological replicates were grouped and enriched proteins determined by 1-sample *t* test (ratio cut off: 1.3).

Accession numbers. PAR-CLIP, HITS-CLIP, HEK transcriptome, and human cardiac transcriptome sequencing data are available at the Sequencing Archive of the European Nucleotide Archive (ERP001301 and E-MTAB-2572).

Statistics. Data are presented as mean \pm SD unless otherwise indicated. Differences between experimental groups were analyzed by 2-tailed Student's *t* test unless otherwise specified. *P* < 0.05 was considered statistically significant.

Study approval. All human studies were approved by the institutional review board of the Cardiovascular Biomedical Research Unit, Royal Brompton and Harefield NHS Trust, and individuals gave informed consent. All experiments involving animals were performed in accordance with institutional and NIH guidelines and approved by the Landesamt für Gesundheit und Soziales (LaGeSo), Berlin, Germany.

Acknowledgments

We thank G. Patone, S. Blachut, S. Schmidt, and J. Fröhlich for expert technical assistance and Claudia Langnick and Mirjam Feldkamp for performing the sequencing runs. This work was supported by funding from the European Union EURATRANS award (HEALTH-F4-2010-241504), the Helmholtz Alliance ICEMED, and the Deutsche Forschungsgemeinschaft (Forschergruppe 1054, HU 1522/1-1) (to N. Hubner); the Deutsche Forschungsgemeinschaft (Bonn, Germany) and the European Research Council grant StG282078 (to M. Gotthardt); and by the Fondation Leducq and the NIH Research Cardiovascular Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College, London, United Kingdom (to S.A. Cook).

Address correspondence to: Norbert Hubner, Max-Delbrück-Center for Molecular Medicine (MDC), Robert-Roessler-Strasse 10, 13125 Berlin, Germany. Phone: 49.30.9406.2530; E-mail: nhuebner@mdc-berlin.de.

1. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol.* 2009;10(11):741-754.

2. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet.* 2002;3(4):285-298.

3. Caceres JF, Kornblihtt AR. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* 2002;18(4):186-193.

4. Douglas AG, Wood MJ. RNA splicing: disease and

- therapy. *Brief Funct Genomics*. 2011;10(3):151-164.
5. Guo W, et al. RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med*. 2012;18(5):766-773.
 6. Brauch KM, et al. Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J Am Coll Cardiol*. 2009; 54(10):930-941.
 7. Li D, et al. Identification of novel mutations in RBM20 in patients with dilated cardiomyopathy. *Clin Transl Sci*. 2010;3(3):90-97.
 8. Jefferies JL, Towbin JA. Dilated cardiomyopathy. *Lancet*. 2010;375(9716):752-762.
 9. Li S, Guo W, Dewey CN, Greaser ML. Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Res*. 2013;41(4):2659-2672.
 10. Manley JL, Tacke R. SR proteins and splicing control. *Genes Dev*. 1996;10(13):1569-1579.
 11. Graveley BR. Sorting out the complexity of SR protein functions. *RNA*. 2000;6(9):1197-1211.
 12. Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*. 2009;417(1):15-27.
 13. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010;141(1):129-141.
 14. Lebedeva S, et al. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell*. 2011;43(3):340-352.
 15. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008;456(7221):464-469.
 16. Paul FE, Hosp F, Selbach M. Analyzing protein-protein interactions by quantitative mass spectrometry. *Methods*. 2011;54(4):387-395.
 17. Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet*. 2006;38(12):1452-1456.
 18. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 2012;22(10):2008-2017.
 19. Wang ET, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456(7221):470-476.
 20. Cvitkovic I, Jurica MS. Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res*. 2013;41(Database issue):D132-D141.
 21. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A*. 2011;108(27):11093-11098.
 22. Kanopka A, Muhlemann O, Akusjarvi G. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature*. 1996; 381(6582):535-538.
 23. Wang Y, et al. A complex network of factors with overlapping affinities represses splicing through intronic elements. *Nat Struct Mol Biol*. 2013; 20(1):36-45.
 24. Shen M, Mattox W. Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. *Nucleic Acids Res*. 2012;40(1):428-437.
 25. Ibrahim EC, Schaal TD, Hertel KJ, Reed R, Maniatis T. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc Natl Acad Sci U S A*. 2005; 102(14):5002-5007.
 26. Methawasin M, et al. Experimentally increasing titin compliance in a novel mouse model attenuates the frank-Starling mechanism but has a beneficial effect on diastole. *Circulation*. 2014; 129(19):1924-1936.
 27. George CH, et al. Alternative splicing of ryanodine receptors modulates cardiomyocyte Ca²⁺ signaling and susceptibility to apoptosis. *Circ Res*. 2007;100(6):874-883.
 28. Jiang D, et al. Enhanced store overload-induced Ca²⁺ release and channel sensitivity to luminal Ca²⁺ activation are common defects of RyR2 mutations linked to ventricular tachycardia and sudden death. *Circ Res*. 2005;97(11):1173-1181.
 29. Tang Y, Tian X, Wang R, Fill M, Chen SR. Abnormal termination of Ca²⁺ release is a common defect of RyR2 mutations associated with cardiomyopathies. *Circ Res*. 2012;110(7):968-977.
 30. Tiso N, et al. Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Hum Mol Genet*. 2001;10(3):189-194.
 31. Ooshio T, Irie K, Morimoto K, Fukuhara A, Imai T, Takai Y. Involvement of LMO7 in the association of two cell-cell adhesion molecules, nectin and E-cadherin, through afadin and alpha-actinin in epithelial cells. *J Biol Chem*. 2004; 279(30):31365-31373.
 32. Holaska JM, Rais-Bahrami S, Wilson KL. Lmo7 is an emerin-binding protein that regulates the transcription of emerin and many other muscle-relevant genes. *Hum Mol Genet*. 2006; 15(23):3459-3472.
 33. Huber AB, Weinmann O, Brosamle C, Oertle T, Schwab ME. Patterns of Nogo mRNA and protein expression in the developing and adult rat and after CNS lesions. *J Neurosci*. 2002; 22(9):3553-3567.
 34. Sarkey JP, et al. Nogo-A knockdown inhibits hypoxia/reoxygenation-induced activation of mitochondrial-dependent apoptosis in cardiomyocytes. *J Mol Cell Cardiol*. 2011;50(6):1044-1055.
 35. Bullard TA, Protack TL, Aguilar F, Bagwe S, Massey HT, Blaxall BC. Identification of Nogo as a novel indicator of heart failure. *Physiol Genomics*. 2008;32(2):182-189.
 36. Gramolini AO, et al. Comparative proteomics profiling of a phospholamban mutant mouse model of dilated cardiomyopathy reveals progressive intracellular stress responses. *Mol Cell Proteomics*. 2008;7(3):519-533.
 37. Xia H, Winokur ST, Kuo WL, Altherr MR, Bredt DS. Actinin-associated LIM protein: identification of a domain interaction between PDZ and spectrin-like repeat motifs. *J Cell Biol*. 1997;139(2):507-515.
 38. Takahashi N, et al. Hypertrophic stimuli induce transforming growth factor- β 1 expression in rat ventricular myocytes. *J Clin Invest*. 1994; 94(4):1470-1476.
 39. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105-1111.
 40. Hubbard T, et al. The Ensembl genome database project. *Nucleic Acids Res*. 2002;30(1):38-41.
 41. Schulte JH, et al. Deep sequencing reveals differential expression of microRNAs in favorable versus unfavorable neuroblastoma. *Nucleic Acids Res*. 2010;38(17):5919-5928.
 42. Paquette J, Tokuyasu T. EGAN: exploratory gene association networks. *Bioinformatics*. 2010; 26(2):285-286.