SUPPLEMENTAL DATA

RBM20 orchestrates pre-mRNA processing by targeting distinct intronic splicing silencers in the heart

Henrike Maatz[1], Marvin Jens[2, 12], Martin Liss[3, 12], 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 JR Barton[6, 7], Stuart A Cook[6, 7, 8, 9], Nikolaus Rajewsky[2], Michael Gotthardt[3, 10], Markus Landthaler[11] & Norbert Hubner[1, 10]

[1] Max-Delbrück-Center for Molecular Medicine (MDC), 13125 Berlin, Germany; [2] Systems Biology of Gene Regulatory Elements, Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany, [3] Neuromuscular and Cardiovascular Cell Biology, Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany; [4] Department of Computational Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany; [5] Laboratory of Cell Signaling and Mass Spectrometry, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany; [6] National Heart and Lung Institute, Cardiovascular Genetics and Genomics, Sydney Street, London SW3 6NP, UK; [7] Royal Brompton NIHR Cardiovascular Biomedical Research Unit, London, SW3 6NP, UK; [8] Duke-NUS Graduate Mediacl School, 8 College Road, Singapore 169857; [9] National Heart Center Singapore, 5 Hospital Drive Singapore, 169609; [10] DZHK (German Centre for Cardiovascular Research), partner site Berlin, Berlin, Germany; [11] RNA Biology and Posttranscriptional Regulation, Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany; [12] These authors contributed equally to this work

Correspondence to: Norbert Hubner, e-mail: nhuebner@mdc-berlin.de, telephone number: +49 30 9406 2530

SUPPLEMENTAL MATERIAL AND METHODS

PAR-CLIP and HITS-CLIP cluster generation

Reads were collapsed into distinct sequences (counting each sequence only once) and aligned to the reference genome assembly (hg19 for human, rn4 for rat) allowing for up to one mismatch, insertion or deletion using BWA (human) or BOWTIE2 (rat). Only uniquely mapping reads were retained. We identified clusters of aligned CLIP-seq reads that continuously covered regions of genomic sequence. To define a PAR-CLIP consensus, we discarded clusters that were not supported by reads from at least 2/3 biological replicates (2x 4SU, 21x 6SG, see Supplemental Table 1). For PAR-CLIP, the number of T to C or G to A mismatches (4SU, 6SG characteristic transitions) served as a crosslink score (Hafner et al. 2010), for HITS-CLIP we treated deletions as crosslink score (Zhang, Darnell 2011).

As the reads should originate from RBM20-bound transcripts we regarded clusters aligning antisense to the annotated direction of transcription as false positives, and indeed the quality score distributions are typically different for "sense" and "antisense" clusters (not shown). The algorithm then tries to select cutoffs on each score such as to keep the estimated false-positive rate (FDR) below a defined limit. Among all scoring functions, the one that retains the largest number of "sense" clusters while limiting the FDR is selected as optimal. For our PAR-CLIP consensus set, the cluster length was selected as optimal, for HITS-CLIP the cumulative uniqueness (difference between best and second-best alignment scores reported by BOWTIE2, summed over all reads that contribute to a cluster) was selected. After filtering by these cutoffs we expect each remaining cluster to harbor at least one RBM20 binding site with high confidence (FDR < 5% for PAR-CLIP,

FDR < 25% for HITS-CLIP). Assuming statistical independence of individual read clusters the FDR for a gene to be targeted by RBM20 is < 5% for a single PAR-CLIP cluster and < 6.25% (< 1.6%) for two (three) or more HITS-CLIP clusters.

We annotated the consensus HITS-CLIP clusters against the ENSEMBL/RGCS3.4 annotation (1). Hits were aggregated by gene name and conflicting annotations, e.g. alternative exons, were resolved by splitting the cluster count and assigning a fraction to each category in proportion to the number of transcripts supporting it.

PAR-CLIP and HITS-CLIP read distribution

To determine the read distribution of the PAR-CLIP and HITS-CLIP libraries the ratio of CLIP-seq versus RNA-seq percent coverage of reads mapping to 5□ UTRs, CDS, introns, and 3□ UTRs was calculated as described before (2).

Analysis of natural selection in RBM20 binding sites

To detect a signature of negative selection on RBM20 binding sites we performed an analysis similar to (3). We obtained SNP data from the 1000 genomes project (4). The variant call file was downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.wgs.phase1_release_v3.20101123.snps_indels_sv.sites.vcf.gz. It contains information about the allele frequencies summarized for populations grouped by continent: Africans (AFR), Americans (AMR), Asians (ASN) and Europeans (EUR). Moreover it also contains information about the ancestral allele derived from phylogenetic analysis implemented in the Ensembl comparative genomics pipeline. We used gene annotation from Ensembl version 66 and identified all SNPs that were located in genes. On the one hand we classified SNPs as

exonic if they overlapped with any annotated exon or intronic otherwise and cases where an intron on one strand overlapped with an exon on the other strand were called ambiguous. On the other hand we classified SNPs by their overlap with motif hits (UCUU) or CLIP-clusters detected in HEK293 cells. Then we obtained the distribution of derived allele frequencies for each of the subgroups. We compared the DAF distributions between groups using the Wilcoxon-Mann-Whitney test with the alternative hypothesis that the DAF is smaller in the motif or cluster group.

Global splicing analyses

We tested for differential exon usage in pair wise comparisons for all three groups of animals (RBM20 -/-, -/+, and +/+) to detect exonic regions regulated by RBM20. DEXSeq scores all disjunct exonic parts that are derived from the Ensembl database release 66 using a negative binomial test based on count data. We defined an exon to be differentially spliced if it overlapped with exonic parts with DEXSeq FDR < 0.05. We considered only exons of expressed genes and removed all duplicated exons with the exact same start and end coordinates to minimize effects of double counting.

Association between RBM20 binding and splicing dysregulation

Differential exon usage was quantified using the DEXSeq package (see above). In order to quantify the direction of the splicing effects we used the difference of PSI values (ΔPSI). Genes that showed differential exon usage and carried a cluster were considered as direct RBM20-bound target transcripts. Within direct RBM20 targets, we calculated the number of RBM20 cluster detected within regulated exons or 400bp up- or

downstream of the exon itself. We tested for independence of differential exon usage and RBM20 binding using Fisher's exact test.

Adjustment of gene expression levels in the heart failure cohort

Heart tissue is mainly composed of cardiomyocytes and fibroblast cells. In order to avoid confounding by cell type heterogeneity in the heart tissue samples we defined a fibroblast gene expression signature by analyzing RNA-seq data from cultured rat cardiomyocytes and heart fibroblasts isolated from SD rats using a modification of the Kasten's technique as described (5). We selected genes with high expression levels in fibroblasts, at least ten-fold higher expression compared to cardiomyocytes and FDR < 0.05. From this list, we selected genes that had human homologues and computed a fibroblast score for each human sample by summing up the log10 transformed expression levels.

We adjusted gene expression levels for available covariates using a linear model. We modeled the expression data after scaling, adding a pseudo count and log10 transformation as response and the following covariates as predictors: barcode id, RNA quality, library preparation date, library concentration and fibroblast score. Finally, for each gene we used its mean expression level plus the residuals of the model.

RRE sequence logos

The sequence logos were computed with MEME 4.9 (6) from the 1000 binding sites (CCRs or clusters) with the highest number of characteristic nucleotide conversions, using default parameters and the '-DNA' switch for nucleic acids.

FLAG/HA-tagged RBM20 immunoprecipitation

For immunoprecipitation of FLAG/HA-tagged RBM20, RBM20-expressing HEK293 cells were lysed on ice in NP40 lysis buffer (50 mM HEPES-K pH 7.5, 150 mM KCl, 2 mM EDTA, 0.5% (v/v) NP-40, 0.5 mM DTT, protease inhibitor cocktail (Roche)). FLAGHA-tagged RBM20 was immunoprecipitated for 4h at 4 ºC using anti-FLAG antibody (Sigma, F3165) coupled to Protein G magnetic beads (Invitrogen). Beads were washed with NP40 buffer and protein was eluted by incubating with 3x FLAG peptide (150ng/µl in NP40 buffer) for 30 min at 4ºC.

GST-tagged rRBM20 RRM domain expression & purification

Bacterially- expressed N-terminal GST-tagged rRBM20 RRM domain (a.a. 508-605) or GST alone were purified under native conditions using GST affinity chromatography following the manufacturer's instructions (Sigma-Aldrich). Protein preparations were stored in buffer D (20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol). Buffer was exchanged using Zeba™ Spin Desalting Columns 7K MWCO (Thermo Scientific).

RT-PCR primers

Name	**Sequence (5' to 3')**
Lmo7_FW	GTT GCG AAA GTG ACA ACG ATT
Lmo7_FW_qRT	GAA GAA AGA CGA CAT GTT GAC
Lmo7_RV	CCG TTA TCA CCA TAA ATC TTT TG
Lmo7_RV_2	CGA AGA CAA GAT ATC ATC AGA G
Mlip_FW	CCA GGG GCA ACA GGC AAC ATA
Mlip_RV	CTC CAC TGG TAG GTC TGA CAC
Mlip_RV_2	GAT CAG CGA CAT TTG AGC CAC AT
Pdlim3_FW	GGC AGA AAC TCG CTT ATG GTC CC
Pdlim3_RV	GTT CAC CAG CTC CTG CAG GAC

Pdlim3_RV_qRT	GCT GCT ACG AAA GGC TGG GCC C
Ryr2_FW	GTT GTC ACG ATG AAG AAG ACG ATG
Ryr2_FW_qRT	CAC AGG ATC CCA ACG CAG CAA
Ryr2_RV	CTT TGC TGG CAC TGA AGG TCT G

Plasmid Construction

Constructs for expression in the mammalian system were generated in pcDNA3.1 (Invitrogen). Constructs for protein expression in bacteria were generated in pGEX-6P-1 (Promega). RBM20 cDNA was amplified from total RNA, extracted from rat left ventricle (LV) using Trizol reagent (Invitrogen) and reverse transcribed by RevertAid™ reverse transcriptase (Fermentas) using oligo (dT)20 as the anchor primer. Full length RBM20 or truncated RBM20 (AA 508-605) were amplified using Phusion™ high fidelity DNA polymerase (Finnzyme) with cloning primers listed in the table "Cloning Primers". The splice reporter constructs contain a) the RBM20-dependent rat titin PEVK exons 4 through 13 (rPEVK4-13wt) (PEVK exon 12 is not present in rat) and b) the RBM20 independent rat titin M-band exons 4 through 6 (rMEx4-6wt). Firefly and renilla luciferase coding regions were PCR amplified from pGL3 and pGL4.7 (Promega) and inserted in frame into a) exon PEVK 8 and 13, and b) exon MEx 5 and 6, respectively. The rat titin MEx reporter was assembled using the InFusion HD cloning kit (Clontech Laboratories, Inc.) following the manufacturer's instructions. The rat titin minigene PEVK exons 4 through 8 + PEVK intron 11-13 + PEVK exon 13 was generated by PCR-based gene assembly using the primers listed in table "Cloning Primers". Splice reporter mutations were introduced by site-directed mutagenesis in a two-step cycle PCR approach using the primer pairs listed in the table "Splice Reporter Mutation Primers".

Splice reporter assay

We established a dual luciferase splice reporter assay based on the RBM20-dependent titin PEVK exons 4 through 13 (see *Plasmid Construction*). HEK 293 cells (25000 cells/well on 96 well plates) were transfected 24h after seeding using Polyethylenimine 40 kDa (Polysciences Europe GmbH). The reporter-plasmid was 30-fold in excess of the RBM20 expression construct or control plasmid (pcDNA3.1). 72 h after transfection cells were lysed and assayed using the Dual-Luciferase® Reporter Assay System (Promega) with 8 wells per condition following the manufacturer's instructions and analyzed using an infinite M200 Pro plate reader (TECAN). Ratios of firefly to renilla luciferase activity were normalized to the pcDNA3.1 negative control co-transfected with the reporter and empty pcDNA3.1 vector alone. Results were confirmed in three independent experiments.

Cloning Primers

Name	Sequence (5' to 3')	PCR Product (bp)
RBM20 fNotI	gagcggccgcgaatgGTGCTGGCTGCAGC	3646
RBM20 rNotI	tcgcggccgcgCATAGCTTCTTCCTTTCCAA	
rRBM20_0508fC	ggatccTCCTCTGGGACAAATTTTGCACAGAGG	303
rRBM20_0605rE	ctcgagctaAGGTTTCTTCAGCTGCAATTCCTTGTATCTCG	
luc for	ccctgcagGAAGACGCCAAAAACATAAAGAA	1647
luc rew	ccgcggccgcCACGGCGATCTTTCCG	
Rluc for	ccctcgaGCTTCCAAGGTGTACGACC	933
Rluc rew	ccgggcccttaCTGCTCGTTCTTCAGCAC	
PEVK Ex 4 for	ccggatccaccatgGAGGAGATCAAGGTGGAAGC	1960
PEVK Ex 8 rew	ccctgcagccCGGTGTGAGAGGCAAAGAAC	
PEVK Ex 8 for	ccgcggccgcACCAGCTGTGCACACAAAGA	3620
PEVK Ex 13 rew	ccctcgagTCTTCTTTGCCACAGGAACG	
MEx4-5FUSf	TTGGTACCGAGCTCGGATCCACCATGATTTCCATTTCTTCAAATATC AGTGTAAGTCGC	856

MEx4-5FUSr	GTTTTTGGCGTCTTCCATCATGCTAGCGCTCCCG	
MEx5FLucFUSf	AGCGCTAGCATGATGGAAGACGCCAAAAACATAAAGAAAGGC	1676
MEx5FLucFUSr	AAACTTCATTTCAGTCACGGCGATCTTTCCGCCC	
MEx5-6FUSf	GGAAAGATCGCCGTGaCTGAAATGAAGTTTGCGAGCATGTCTG	809
MEx5-6FUSr	ACCTTGGAAGCTCGAgaCATAGAGCGGATATTAATATTCACAGTGGC	
PEVK11-13f	GTATCCAGAACACCTGGTGG	1559
PEVK Ex 13 rew	ccctcgagTCTTCTTTGCCACAGGAACG	
PEVK Ex 8 for	ccgcggccgcACCAGCTGTGCACACAAAGA	114
PEVK8r	CACCAGGTGTTCTGGATACCTGTGACAGACACCTCCTCCTCG	

* restriction sites are indicated in red, start and stop codons introduced in blue, and RNA degradation sequence in green. Grey indicates overlapping sequence for PCR based gene assembly. Regions that are not homologous to the original sequence are provided in small caps.

Splice Reporter Mutation Primers

Name	Sequence (5' to 3') *	Position (bp)
PEVKIn11-13mTC1f	GTTTCACCCTTATGTGTGTGTCcctccCATTAAGACATTGTTGTGTG	3:59547789-59547835
PEVKIn11-13mTC1r	CACACAACAATGTCTTAATGggaggGACACACACATAAGGGTGAAAC	3:59547789-59547835
PEVKIn11-13mTC2f	GATTTTCCCTGCCTGTCTTGGAAcctccTAAAGCTACTATAAACATTTGG	3:59547640-59547689
PEVKIn11-13mTC2r	CCAAATGTTTATAGTAGCTTTAggaggTTCCAAGACAGGCAGGGAAAATC	3:59547640-59547689
PEVKIn11-13mTCf	GTGTATTGTAAAGGACAGAAccctccACAGAAAAACACCTGTGCATAC	3:59547590-59547637
PEVKIn11-13mTCr	GTATGCACAGGTGTTTTTCTGTggagggTTCTGTCCTTTACAATACAC	3:59547590-59547637
PEVKIn11-13mTC4f	CCTACACACTCTACCCACAcctccCATTAGCAGTATACTAATGG	3:59547299-59547342
PEVKIn11-13mTC4r	CCATTAGTATACTGCTAATGggaggTGTGGGTAGAGTGTGTAGG	3:59547299-59547342
PEVKIn11-13mTC-1f	CAGCTGTGCTTTGCACAGctccCATTTGTACTGTGTTGTCC	3:59548060-59548100
PEVKIn11-13mTC-1r	GGACAACACAGTACAAATGggagCTGTGCAAAGCACAGCTG	3:59548060-59548100
PEVKIn11-13mTC-2-3f	GTACTGTGTTGTCCTTGTCctccATCctccCAGCAAGTTACCCATGTTTAAAGC	3:59548020-59548073

PEVKIn11-13mTC-2-3r	GCTTTAAACATGGGTAACTTGCTGggagGATggagGACAAGGACAACACAGTAC	3:59548020-59548073
PEVKIn11-13mTC-4f	GGGAAGGATTTGTGGAACActccTTACAGACATGTTTATACACC	3:59547847-59547890
PEVKIn11-13mTC-4r	GGTGTATAAACATGTCTGTAAggagTGTTCCACAAATCCTTCCC	3:59547847-59547890

* Regions that are not homologous to the original sequence are provided in small caps.

Primers for EMSA probe generation

Name	Sequence (5' to 3') *	Position (bp)
T7-PEVK 11-13_Jf	TAATACGACTCACTATAGGGTATCCAGAACACCTGGTGGAG	3:59548122-59548143
PEVK11-13_Jr	GGATGATATTAAACATTTAAGGTTTGACATAATTACTTC	3:59547893-59547931
T7-PEVK 11-13_Kf	TAATACGACTCACTATAGGGTCAAACCTTAAATGTTTAATATCATCCTTGGG	3:59547890-59547920
PEVK11-13_Kr	CAAGACAGGCAGGGAAAATCAATG	3:59547670-59547693
T7_PEVK11-13_Pf	TAATACGACTCACTATAGGGACAGATAACTTTACCATGACTTTAAAGTGG	3:59546782-59546812
PEVK11-13_Pr	CTTAAGGTTTTAAAATAGCAGGTCAGTATAGG	3:59546654-59546685

SILAC samples

Cell culture and transfection

Human embryonic kidney HEK293T cells were cultivated for 2 weeks in SILAC DMEM medium (Dulbecco's Modified Eagle Medium, 4.5g/L glucose without L-lysine and L-arginine; (Gibco)) with 10% dialyzed fetal bovine serum (Gibco) and 2 mM Glutamine, containing 49mg/ml and 28 mg/ml L-lysine (Lys0) and L-arginine (Arg0) (light), 4,4,5,5,-D4-L-lysine (Lys4) and 13C614N4-L-arginine (Arg6) (medium) or 13C6N2-L-lysine (Lys8) and 13C615N4-L-arginine (Arg10) (heavy) at 37°C and 5% CO2.

SILAC labeled cells were transiently transfected with empty vector control and FLAG-tagged RBM20 wildtype or mutant constructs using polyethylenimine (linear PEI 'Max', nominally Mw 40,000, Polysciences) with 15 μg plasmid DNA and 30 μg PEI per $2x10^7$

cells. After 24 hours cells were washed in ice-cold PBS, harvested and lysed in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM MgCl2, 1 mM EGTA, 10% glycerol, 1 % Triton X-100, complete protease inhibitor (Roche) and phosphatase inhibitor cocktail (1:100, Sigma). After incubation on ice for 30 min samples were centrifuged at 14,000g for 15 min at 4°C to remove insoluble material. The supernatant was transferred to a fresh tube for affinity purification experiments.

Affinity purification and protein digest

Cleared SILAC extracts were incubated separately with 50 µl anti-FLAG M2 magnetic beads (Sigma) for 60 min at 4°C under constant rotation. The beads were washed two times with lysis buffer and once with PBS. Corresponding samples of triple SILAC experiments were combined and proteins bound to the beads were eluted together by applying two times 100 µl elution buffer (6M urea/2M thiourea). Protein eluates were ethanol precipitated, followed by in-solution digestion as previously decribed (7). Briefly, protein eluates were reduced for 30 min in 10 mM dithiothreitol solution followed by alkylation with 55 mM iodacetamide for 20 min. The endoproteinase LysC (Wako, Japan) was added (protein:enzyme ratio 50:1) for 4h at room temperature. After 4:1 dilution in digestion buffer (50mM ammonium bi-carbonate, pH 8.0) sequence grade modified trypsin (Promega) was added (protein:enzyme ratio 50:1) and incubated over night. Trypsin and LysC activity was quenched by acidification with 10% trifluoroacetic acid. Peptides were extracted and desalted using StageTips (8).

Cardiomyoctes

48 h past isolation, neonatal rat ventricular myocytes were washed in ice-cold PBS, harvested and lysed and insoluble material was removed as described for the SILAC samples. Cleared tissue extracts were incubated with anti-RBM20 antibody (9) covalently coupled to Protein G Dynabeads (Invitrogen) for 60 min at 4°C under constant rotation. The beads were washed two times with lysis buffer and once with PBS. Protein eluates were ethanol precipitated, followed by in-solution digestion as decribed above. For the MS/MS analysis peptide samples were processed in triplicates.

LC-MS/MS analysis

Peptide mixtures were separated by reversed phase chromatography using the EASY-nLC system (Thermo Scientific) on in-house manufactured 20 cm fritless silica microcolumns with an inner diameter of 75 μm. Columns were packed with ReproSil-Pur C18-AQ 3 μm resin (Dr. Maisch GmbH). Peptides were eluted on a 8-60% acetonitrile gradient (214 min) with 0.5% formic acid at a nanoflow rate of 200 nl/min. Eluting peptides were directly ionized by electrospray ionization and transferred into a Q Exactive mass spectrometer (Thermo Scientific). Mass spectrometry was performed in the data dependent positive mode with one full scan (m/z range = 300-1,700; R = 70,000; target value: 3 x 10^6; maximum injection time = 120 ms). The 10 most intense ions with a charge state greater than one were selected (R = 35,000, target value = 5 x 10^5; isolation window = 4 m/z; maximum injection time = 120 ms). Dynamic exclusion for selected precursor ions was set to 30s. MS/MS data were analyzed by MaxQuant software v1.2.2.5 as described (10). The internal Andromeda search engine was used to search MS/MS spectra against a decoy *Homo sapiens* UniProt database (HUMAN.2012-06) or

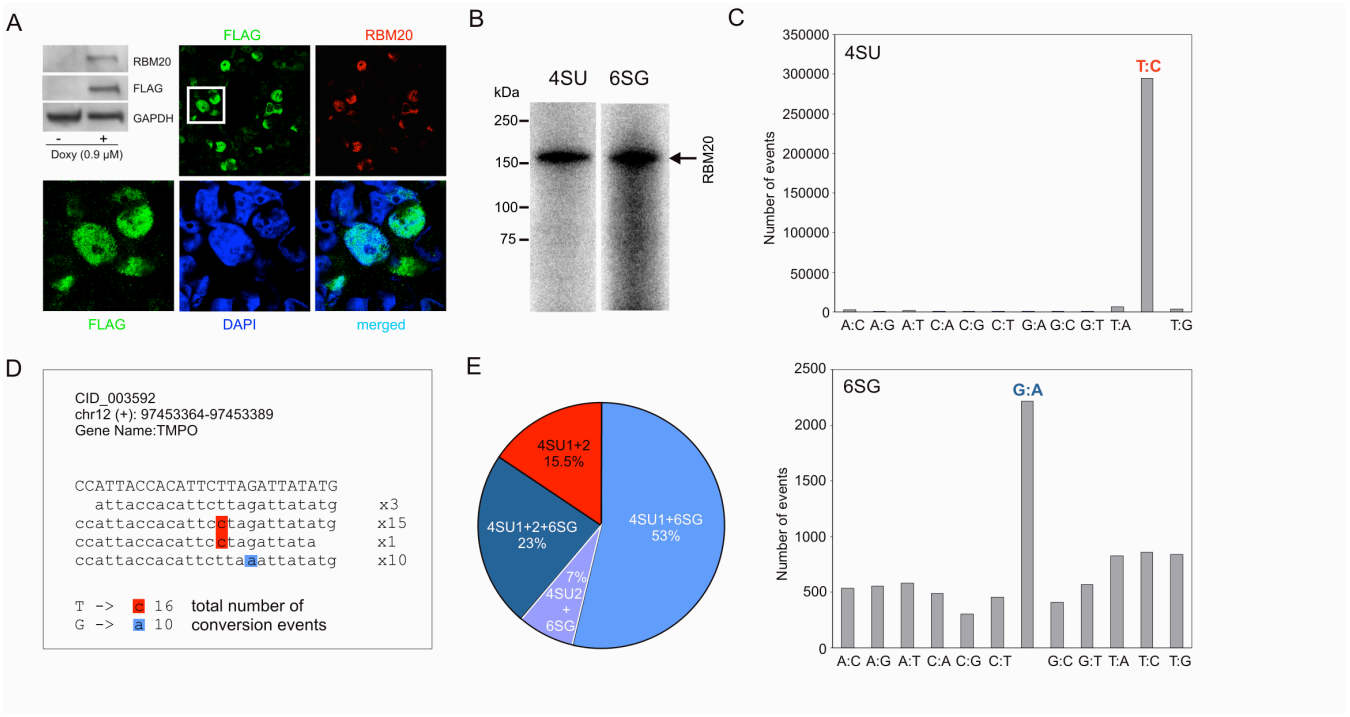
Rattus norgvegicus UNIProt database (RAT.2012-06) containing forward and reverse sequences including common contaminants. The search included variable modifications of methionine oxidation and N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. Minimal peptide length was set to six amino acids and a maximum of two missed cleavages was allowed. The false discovery rate (FDR) was set to 0.01 for peptide and protein identifications. If the identified peptide sequence set of one protein was equal to or contained another protein's peptide set, these two proteins were grouped together and the proteins were not counted as independent hits. For SILAC, the protein quantification was based on unique peptides and non-unique peptides assigned to the protein group with highest number of peptides (razor peptides). At least two peptide SILAC ratio counts were required for quantification. Data analysis, gene ontology annotation and statistical analyses were performed in the MaxQuant environment (Perseus). Label swap experiments were used to filter out contaminants. Biological replicates (n=4) for each condition (control, wildtype, mutant) were assigned as a group and potential interaction partners were selected based on the mean of their log2 fold changes (cut off: 0.4). Significance was calculated by one sample t-test. Label-free quantitation (LFQ) for cardiomyocyte samples was performed in MaxQuant as described (11). Unique and razor peptides were considered for quantification with a minimum ratio count of 1. Retention times were recalibrated based on the built-in nonlinear time-rescaling algorithm. MS/MS identifications were transferred between LC-MS/MS runs with the “Match between runs” option in which the maximal retention time window was set to 2 min. RBM20 pull-down samples and plain-beads control samples were selected as individual groups of 3 technical replicates each; significantly enriched

proteins were determined by combining standard two-sample t-test p-values with ratio information. Significance corresponding to an FDR of 1% was determined by a permutation-based method (12).

Confirmation of q-AP-MS hits

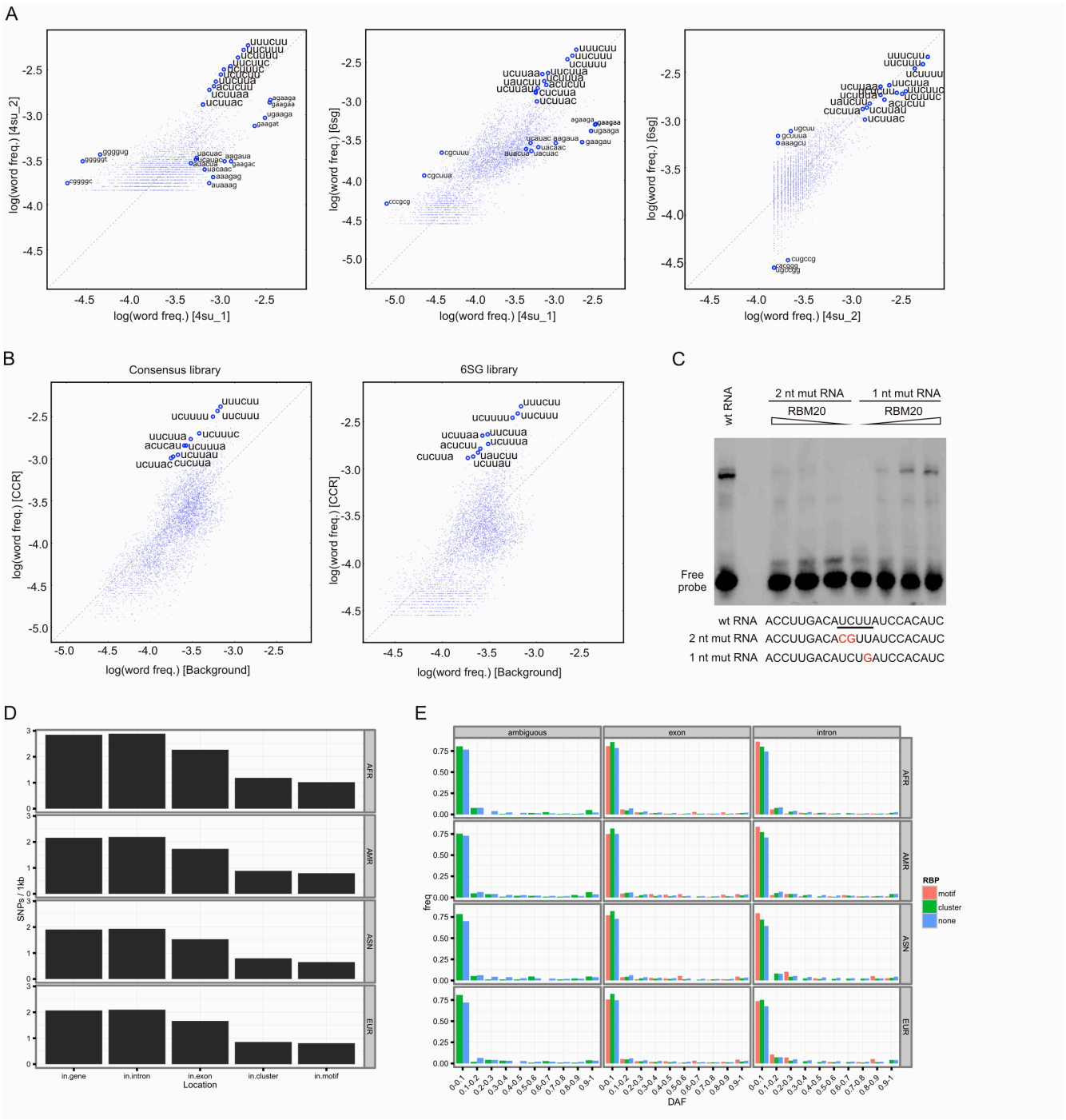
HEK293 cells were transiently transfected with empty vector control or FLAG-tagged RBM20 wildtype construct, harvested and incubated with anti-FLAG M2 magnetic beads (Sigma) as described above (SILAC samples, cell culture and transfection). Proteins bound to the beads were eluted with SDS-loading buffer (Bio-Rad) and subjected to Western blotting. Western blotting was done using standard protocols. The primary antibodies used were anti-SF3B1, anti-FUS (Santa Cruz Biotechnology), anti-SF3A1 (Abcam), and anti-FLAG (Sigma).

For the confirmation of hnRNPL interaction with RBM20, the hnRNPL ORF was amplified from total HEK293 cell RNA using the forward primer ATGCCTAAAAAG AGACAAGCA and the reverse primer GGAGGCGTGCTGAGCAGTGG and cloned into pcDNA3.1 (-) myc-his expression vector. HEK293 cells were co-transfected with either empty vector control and hnRNPL or with RBM20 and hnRNPL. Co-IP with RBM20 was detected using anti-myc (Milipore).



Supplemental Figure 1: PAR-CLIP in FLAG/HA-RBM20 HEK293 cells

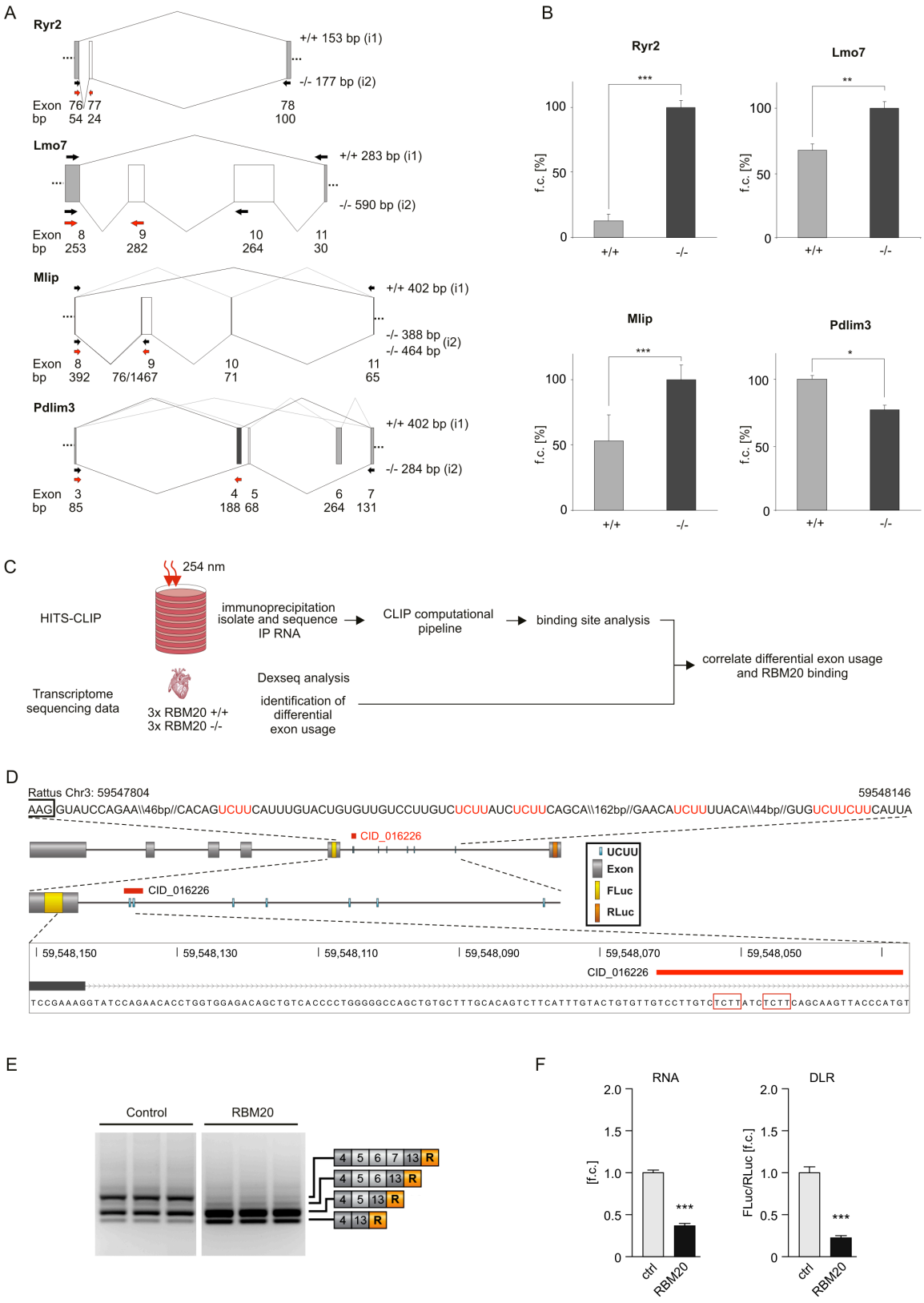
(**A**) FLAG/HA-RBM20 expression in HEK293 cells upon doxycyclin induction and IHC following induced expression of N-terminally FLAG tagged RBM20 in HEK293 cells. The lower panels depict higher magnification images of the detail indicated by the white box in the left upper panel. The lower right panel shows merged lower left FLAG and DAPI images. RBM20 localized predominantly to the nucleus, showing that the tag did not interfere with its physiological localization. (**B**) Phosphoimages of SDS gels resolving 5'-32P-labled RNA-FLAG/HA-RBM20 immunoprecipitates. IPs were prepared from lysates of UV365 crosslinked cells cultured with 4SU or 6SG in the media, respectively. The lanes were run on the same gel but were noncontiguous. (**C**) Enrichment of T to C or G to A transitions for the 4SU or 6SG libraries. (**D**) Representative example of a RBM20 PAR-CLIP consensus cluster. Exemplary alignments of PAR-CLIP cDNA sequence reads to the corresponding regions of the *TMPO* mRNA shown in uppercase letters at the top. Red-letter nucleotides indicate T to C and blue-letter nucleotides G to A sequence changes. Counts for each of the individual reads are indicated. (**E**) Read composition of the consensus library. The pie chart shows the proportions of clusters with support from the individual 4SU and 6SG libraries.



Supplemental Figure 2: RBM20 RNA recognition element analysis

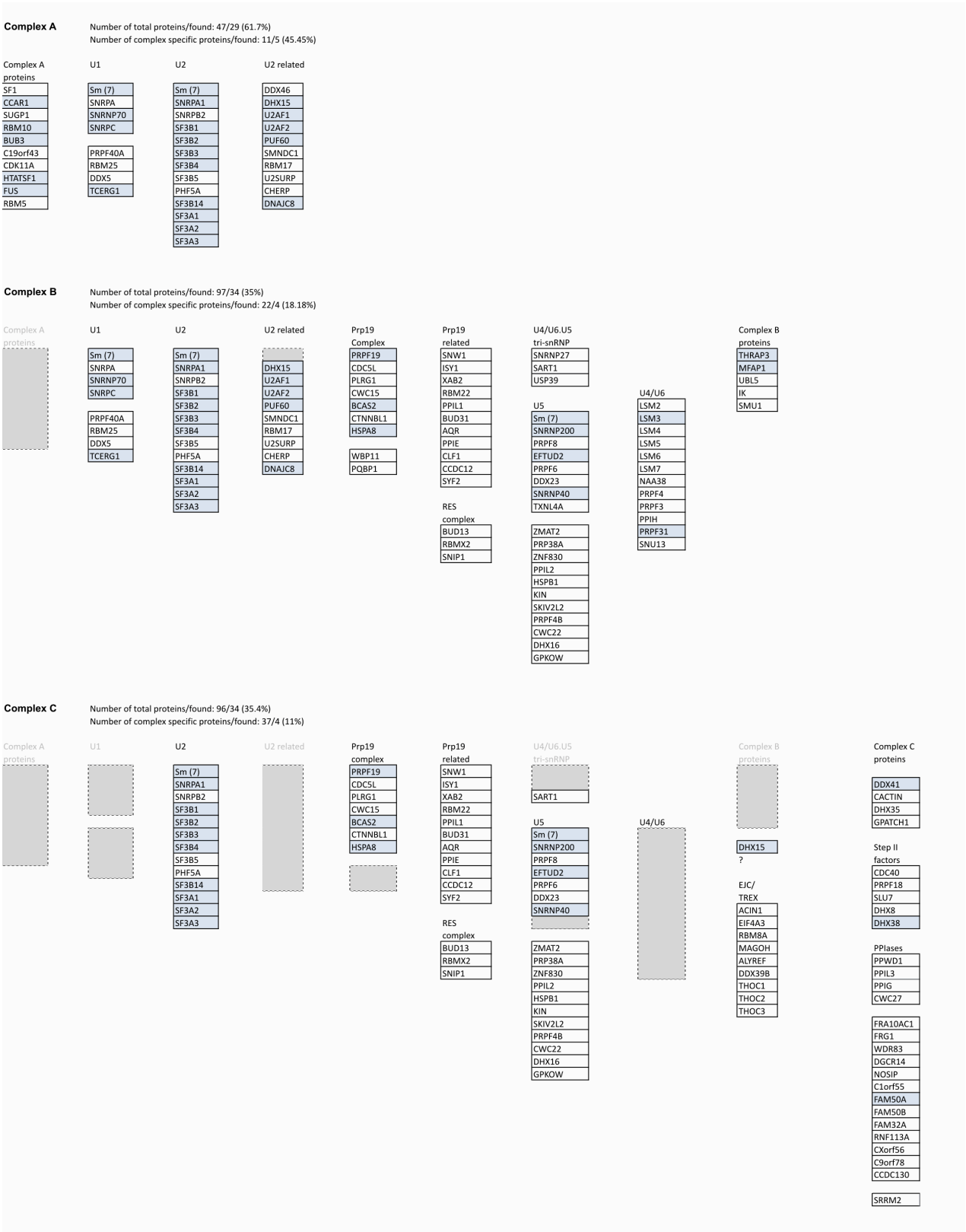
(**A**) Correlations of log10 frequencies of 6-mers in cluster-centered regions in transcript targets of individual PAR-CLIP libraries, 4SU1 vs. 4SU2 (left), 4SU1 vs. 6SG (middle), 4SU2 vs. 6SG (right). (**B**) Binding element enrichment in the RBM20 PAR-CLIP consensus (left) and 6SG library (right). Log10 frequencies of 6-mers in cluster-centered regions of the libraries are correlated with frequencies of 6-mers in control sequence. (**C**) Phosphorimage of native PAGE resolving complexes of immunoprecipitated full-length RBM20 protein with *Ryr2* target RNA oligonucleotides with one or two mutations within

the core RBM20 RRE compared to wild-type. (**D**) Density of SNPs in genes, introns, exons, RBM20 CLIP-clusters in HEK293 cells and RRE (UCUU) within clusters. (**E**) Distribution of derived allele frequencies. The figure shows derived allele frequency distributions for SNPs in RBM20 CLIP-clusters in HEK293 cells (green) and RREs (UCUU) within clusters (red) compared to background SNPs (blue) in each of the four populations from the 1000 genomes project: Africans (AFR), Americans (AMR), Asians (ASN) and Europeans (EUR). SNPs are further classified by their location in introns, exons or ambiguous if an intron on one strand overlaps an exon on the other strand.



Supplemental Figure 3: Integration of RBM20 binding sites and transcriptome data in the heart

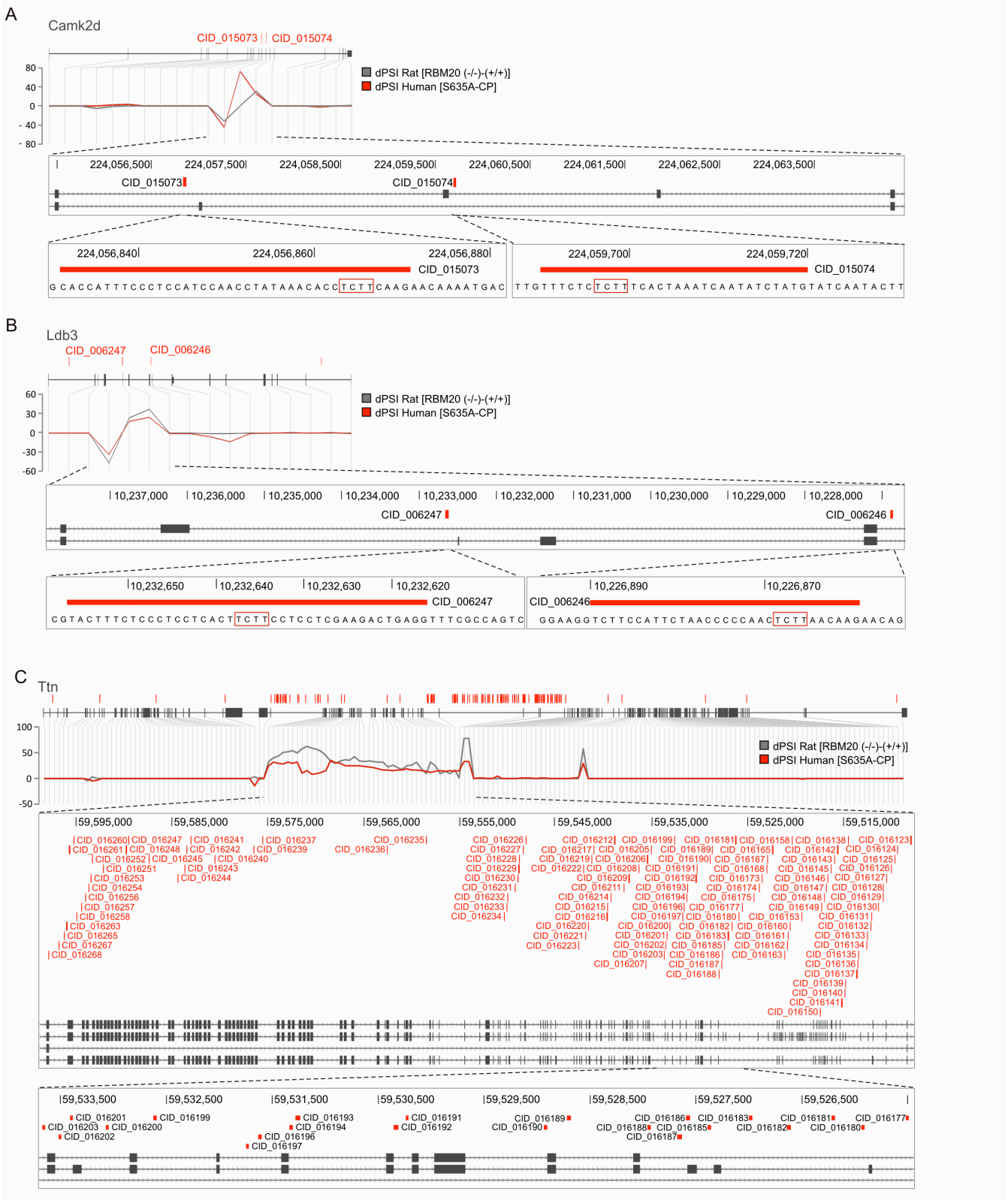
(**A**) Exon structure of alternatively spliced regions in *Ryr2*, *Mlip*, *Lmo7*, and *Pdlim3*. Primer positions used for PCR-based splice analysis are indicated as black arrows and for qPCR as red arrows. Exons expressed primarily in RBM20 deficient animals are indicated in white, those primarily expressed in wild type animals are depicted in dark grey. Exon sizes are indicated below the corresponding exon numbers. Length of PCR fragments produced at higher levels in wild-type than in mutant rats (i1) and those expressed at higher levels in mutant than in wild-type rats (i2) are indicated on the right. The exon compositions of i1 and i2 are indicated by black exon connecting lines while grey lines indicate isoforms equally observed in +/+ and -/-. (**B**) Confirmation of relative exon expression as determined by qPCR in the rat. Exons analyzed for Ryr2, Lmo7, and Mlip were included in transcripts of -/- at higher levels. The alternatively spliced exon tested for Pdlim3 was increasingly included in +/+. Data are represented as mean +/- SEM. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ (n=3 per group). (**C**) HITS-CLIP of RBM20 was performed in primary cardiomyocytes. Genome-wide splicing regulation on the exonic level by RBM20 was investigated by analyzing existing heart transcriptome sequencing data of three wild-type and three *Rbm20* -/- rats using DEXseq. To identify direct RBM20-regulated targets in the heart the information on RBM20 binding sites and RBM20-dependent splicing was correlated. (**D**) Schematic of the *Ttn* splice-reporter construct and RBM20 binding site. Exons that encode the most N-terminal TTN PEVK region (5□ PEVK exons) and firefly and *Renilla* luciferase reporters (Fluc and Rluc, respectively) were used to monitor RBM20 splicing activity. The zoom-in view shows location and sequence of the mapped RBM20 binding site downstream of the differential spliced exon with introduced Fluc reporter as well as positions of random UCUU elements that were mutated to CUCC as controls. (**E**) PCR based analysis of reporter-isoforms in reporter transfected HEK293. cells (ctrl) vs. cells cotransfected with RBM20. Co-transfection with RBM20 favors the exclusion of exons 6 to 8. The predominant isoforms exclude all alternative exons (238 bp) or retain only exon 5 (298 bp). The lanes were run on the same gel but were noncontiguous. (**F**) Quantification of PEVK exon 8 exclusion on RNA level. Values are presented as fold change and normalized to the reporter + vector control (ctrl). (**G**) Quantification of RBM20 dependent reporter activity. Values are presented as FLuc/RLuc and normalized to the reporter + vector control. N=3, *** $p<0.001$.



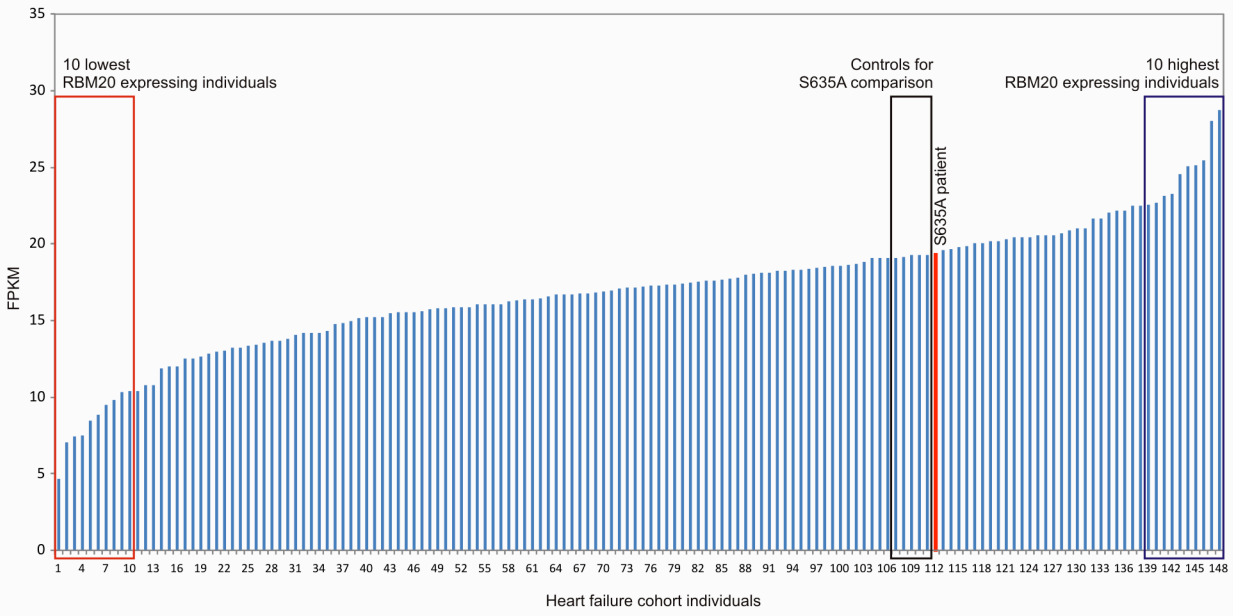
Supplemental Figure 4: Spliceosomal complex A, B, and C proteins identified to interact with RBM20

Protein composition of spliceosomal complex A, B, and C are shown (13). Proteins are highlighted in blue when identified by mass spectrometry following RBM20

immunoprecipitation.


A
Camk2d
CID_015073 | CID_015074
dPSI Rat [RBM20 (-/-)-(+/+)]
dPSI Human [S635A-CP]
224,056,500 224,057,500 224,058,500 224,059,500 224,060,500 224,061,500 224,062,500 224,063,500
CID_015073
CID_015074
224,056,840 224,056,860 224,056,880
CID_015073
224,059,700 224,059,720
CID_015074
B
Ldb3
CID_006247 CID_006246
dPSI Rat [RBM20 (-/-)-(+/+)]
dPSI Human [S635A-CP]
10,237,000 10,236,000 10,235,000 10,234,000 10,233,000 10,232,000 10,231,000 10,230,000 10,229,000 10,228,000
CID_006247
CID_006246
10,232,650 10,232,640 10,232,630 10,232,620
CID_006247
10,226,890 10,226,870
CID_006246
C
Ttn
dPSI Rat [RBM20 (-/-)-(+/+)]
dPSI Human [S635A-CP]
59,595,000 59,585,000 59,575,000 59,565,000 59,555,000 59,545,000 59,535,000 59,525,000 59,515,000
59,533,500 59,532,500 59,531,500 59,530,500 59,529,500 59,528,500 59,527,500 59,526,500

Supplemental Figure 5: RBM20 binding sites in *Camk2d*, *Ldb3*, and *Ttn*
Shown are the Alignment of orthologous rat and human exons of (**A**) *Camk2d*, (**B**) *Ldb3*, and (**C**) *Ttn* and the distribution of RBM20 clusters in these genes. The comparison of dPSI values of.*Rbm20* deficient vs. wildtype rats and a patient with a heterozygote S635A mutation in RBM20 vs. 5 control individuals with DCM are shown below. The zoom-in views show the exon composition of the differential spliced regions of pre-mRNA transcribed from these genes and underlying cluster sequences. Numbers indicate genomic positions in the rat.



Supplemental Figure 6: Endogenous RBM20 expression in human heart tissue
FPKM values for the expression of endogenous RBM20 in heart tissue of 148 end stage heart failure patients. The red bar indicates RBM20 expression in the S635A patient and the five controls with similar expression for comparison are marked by the red box. The 10 highest and the 10 lowest identified RBM20 expressing individuals are marked by the blue and the red box, respectively

Supplemental Table 1: PAR-CLIP and HITS-CLIP statistics

CLIP library	after adapter removal	unique sequences	median/upper quartile length	unique mappings	filtered cluster
PAR-CLIP libraries					
4SU_1	23.1 M	5.5 M (23%)	26/31	839.3 K	87436
4SU_NH_HM_03	26.1 M	2.2 M (8%)	24/29	192.6 K	4066
6SG_NH_HM_04	23.8 M	4.7 M (19%)	25/30	447.1 K	11964
Consensus PAR-CLIP (Clusters require reads from 2/3 libraries)					27623
HITS-CLIP libraries					
hits_NH_HM_001	80.3 M	4.9 M (6 %)	29/36	222.9 K	N/A
hits_NH_HM_002	88.5 M	10.7 M (12%)	37/43	252.2 K	N/A
hits_NH_HM_004	89.9 M	5.4 M (5.0%)	35/43	359.6 K.	N/A
hits_NH_HM_005	117.2 M	9.0 M (7 %)	35/43	359.6 K	N/A
Consensus HITS-CLIP (Clusters require reads from 2/4 libraries)					2454

The numbers of reads are given for each replicate of CLIP experiments. Unique mappings show number of reads that were collapsed into distinct sequences (counting each sequence only once) and uniquely mapped to the reference genome allowing for up to one mismatch, insertion or deletion. Clusters were called on the collapsed CLIP data of pooled libraries (consensus libraries). For PAR-CLIP experiments, clusters were also called for individual libraries.

Supplemental Table 3: Derived allele frequency (DAF) distribution

population	location of SNPs	reference location	p-value	number of SNPs
AFR	intron.motif	intron.none	0.0035	101
AFR	exon.motif	exon.none	0.3188	203
AFR	intron.cluster	intron.none	7.54E+05	1116
AFR	exon.cluster	exon.none	6.48E-42	4633
AMR	intron.motif	intron.none	0.0029	73
AMR	exon.motif	exon.none	0.3395	152
AMR	intron.cluster	intron.none	0.0102	792
AMR	exon.cluster	exon.none	2.07E-16	3210
ASN	intron.motif	intron.none	0.1536	39
ASN	exon.motif	exon.none	0.0552	131
ASN	intron.cluster	intron.none	1.29E+07	595
ASN	exon.cluster	exon.none	6.24E-23	2891
EUR	intron.motif	intron.none	0.0087	57
EUR	exon.motif	exon.none	0.2966	152
EUR	intron.cluster	intron.none	1.56E+09	721
EUR	exon.cluster	exon.none	1.47E-24	3154

Gene annotation from Ensembl version 66 was used to identify all SNPs located in genes. SNPs were classified by their overlap with RBM20 RREs (UCUU) or CLIP-clusters detected in HEK293 cells. The distribution of derived allele frequencies for each of the subgroups were compared between groups using the Wilcoxon-Mann-Whitney test. AFR: Africans, AMR: Americans, ASN: Asians, EUR: Europeans.

Supplemental Table 8: MeSH Analysis – Disease Relevance

MeSH Term	# associated genes	Corrected P-Value
Cardiomyopathy, Dilated	**10**	**1.33E-13**
Muscle Proteins	13	3.50E-12
Sarcomeres	8	2.47E-11
Cardiomyopathy, Hypertrophic	**7**	**1.80E-09**
Actinin	8	2.65E-09
Muscle Fibers, Skeletal	8	5.57E-08
Muscles	9	7.58E-07
Myocardial Contraction	6	1.54E-06
Microscopy, Immunoelectron	9	2.14E-06
Myosins	7	2.54E-06
Myocytes, Cardiac	8	3.53E-06
Cardiomyopathies	**6**	**3.97E-06**
Myofibrils	5	6.33E-06
Calmodulin	7	4.76E-05
Heart Ventricles	6	8.17E-05
Heart	7	8.48E-05
Microscopy, Electron	9	1.08E-04
Heart Failure	**6**	**1.81E-04**
Fetus	9	4.55E-04
Troponin T	4	5.03E-04
Chick Embryo	7	5.11E-04
Stress Fibers	5	7.31E-04
Microfilaments	6	8.71E-04
Axons	6	1.01E-03
Animals, Newborn	7	1.07E-03
Troponin I	4	1.76E-03
Cytoskeleton	8	2.61E-03
Biomechanics	4	4.71E-03
Myosin Subfragments	3	5.41E-03
Circular Dichroism	7	5.48E-03
Ventricular Function, Left	4	6.69E-03
Cardiomyopathy, Hypertrophic, Familial	**3**	**9.67E-03**

Significant enrichments of associated Medical Subject Headings (MESH) for RBM20 splice substrates were evaluated with the hypergeometric test and significant P-values shown were corrected for multiple testing (Bonferroni). Disease related MeSH – Terms are highlighted.

Supplemental Table 9: RNA-seq statistics

Sample	Total Read Count	Mapped	Paired
DCM control patients			
20GD01815	173521076	163768023	133305978
20IS01302	183107417	172532315	137629504
20JD01438	214339071	199125659	156965502
20MR01285	179462664	169583938	135775636
20RP01386	201289173	190159011	153746580
Low RBM20 expressing individuals			
20RC01659	156518118	145767697	111687700
20AA01165	163290286	155968589	130339020
20RV01241	203255517	194784350	158676452
20LM01196	172026830	160849730	128991202
20PC1455	174366830	161000846	124512706
20BP01200	172155638	150958113	116907632
20MD01251	173656889	129540240	101966664
20IP01292	192684183	170412553	135915466
20SH01483	207060339	196122248	159588768
20SH01318	168031905	158717714	132605568
High RBM20 expressing individuals			
20TH01341	185519154	169514270	137933734
20RP01833	184602053	176382452	143660706
20EL01751	279434873	245984537	188373502
20JM01263	191848687	179993914	143670140
20JC01851	199592929	186826638	150994494
20BB01870	187104896	174868904	142934686
20PL01827	220524421	211585922	172138918
20VB01876	196441917	118757759	36083728
20JO01157	214472802	203003283	162880302
20MF01454	190300160	177636072	136882210

The table shows the statistics for cardiac RNA-seq experiments from five subjects with DCM that do not carry mutations in RBM20 and the 10 lowest and highest RBM20 expressing individuals from the human heart failure cohort .

SUPPLEMENTAL REFERENCES

1. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., et al. 2002. The Ensembl genome database project. *Nucleic Acids Res* 30:38-41.
2. Wang, E.T., Cody, N.A., Jog, S., Biancolella, M., Wang, T.T., Treacy, D.J., Luo, S., Schroth, G.P., Housman, D.E., Reddy, S., et al. 2012. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* 150:710-724.
3. Chen, K., and Rajewsky, N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 38:1452-1456.
4. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56-65.
5. Takahashi, N., Calderone, A., Izzo, N.J., Jr., Maki, T.M., Marsh, J.D., and Colucci, W.S. 1994. Hypertrophic stimuli induce transforming growth factor-beta 1 expression in rat ventricular myocytes. *J Clin Invest* 94:1470-1476.
6. Bailey, T.L., and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28-36.
7. Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127:635-648.

8. Rappsilber, J., Ishihama, Y., and Mann, M. 2003. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75:663-670.

9. Guo, W., Schafer, S., Greaser, M.L., Radke, M.H., Liss, M., Govindarajan, T., Maatz, H., Schulz, H., Li, S., Parrish, A.M., et al. 2012. RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med* 18:766-773.

10. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10:1794-1805.

11. Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. 2010. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J Cell Biol* 189:739-754.

12. Tusher, V.G., Tibshirani, R., and Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98:5116-5121.

13. Wahl, M.C., Will, C.L., and Luhrmann, R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136:701-718.