SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure S1; related to Figure 1. (**A**) Correlation heatmap for normalized triplicate splicing array intensities. Analyses excluding the unclustered control experiment (CTRL-1) confirm that our results are unaffected due to this control experiment. (**B-G**) Additional RT-PCR validations for individual hnRNP-regulated events. Bar plots represent densitometric analysis of bands and error bars correspond to standard deviation in triplicate experiments. Absolute separation score as predicted by the splicing array analysis is listed below the gene name. All events shown here are significantly different as measured by a *t*-test ($P < 0.05$). (**H**) Western blot showing successful depletion of hnRNP A2/B1 in primary human fibroblasts using an anti-sense oligonucleotide (ISIS Pharmaceuticals). Quantification of percent knockdown was performed by densitometric analysis of the bands. GAPDH was used as a loading control. (**I**) RT-PCR validation of hnRNP A2/B1 regulated events in primary human fibroblasts ($P < 0.05$, *t*-test). For comparison, see **Figure S3C**.

Supplemental Figure S2; related to Figure 2. (**A**) Additional RT-PCR validations of antagonistic splicing events co-regulated by hnRNP F and U. (**B**) RT-PCR validations of antagonistic splicing events co-regulated by hnRNP A1 and U. (**C**) RT-PCR validations of synergystic splicing events co-regulated by hnRNP A1 and M ($P < 0.05$, *t*-test). Bar plots represent densitometric analysis of bands. Absolute separation score as predicted by the splicing array analysis is listed above the bars. Binding of the regulating hnRNP proteins near the cassette exon is shown below if CLIP derived clusters are detected.

Supplemental Figure S3; related to Figure 3. (**A**) Autoradiograph of hnRNP-RNA complexes digested with increasing amounts of micrococcal nuclease (MNase). The red box is the RNA that was excised and used for CLIP-seq library preparation. (**B**) CLIP derived clusters plotted across composite gene structure displays weak preferences for binding at 3′ or 5′ gene ends. (**C**) Top 4 enriched motifs as determined by the HOMER motif discovery algorithm. (**D**) Fraction of changing AS events with binding (blue), or no binding (gray) in the exon and within 2kb. HnRNP binding is defined by a CLIP-seq cluster for the specific hnRNP whose depletion affected the event. (**E**) For each type of AS event from the top quartile of events (|sep_score| > 1), the percentage of events that was affected in more than one hnRNP depletion (black) and the percentage of events that was affected uniquely in a single hnRNP depletion. For comparison, see **Figure 2A**. (**F**) Comparison of the percent of all splicing events (|sep_score| > 0.05) that are direct targets, and the percent of only the top quartile of events (|sep_score| > 1) that are direct targets. No significant difference is observed. (**G**) Same as **D**, but HnRNP binding is defined by a CLIP-seq cluster for any hnRNP. (**H**) Same as **D**, but HnRNP binding is defined by a CLIP-seq read for any hnRNP.

Supplemental Figure S4; related to Figure 4. Heatmap of the log ratio of hnRNP binding versus random background binding near exons which are alternatively spliced in a tissue-specific manner or show conserved alternative splicing between human and mouse (ACEscan [+] exons). No set of exons is significantly over- or under-represented for hnRNP binding ($P < 0.001$, Fisher exact test).

Supplemental Figure S5; related to Figure 5. Western blots for all 6 hnRNP proteins across all siRNA treated triplicate conditions. GAPDH was used as a loading control.

Supplemental Figure S6; related to Figure 6. Significant splicing changes for 41 cancer-associated events as predicted by the splicing microarray and successfully validated on the Fluidigm micro-fluidic multi-plex qRT-

PCR platform. The name of the affected gene along with the length of the cassette exon in parenthesis, and the p-value determined by Student's *t*-test relative to control ($P < 0.05$) is shown. Green boxes are validations of edges shown in **Figure 6B**.

SUPPLEMENTAL TABLES

Supplemental Table S1. CLIP-seq generates millions of reads and identifies thousands of binding sites for members of the hnRNP protein family. Original raw read count, number of Bowtie mapped reads, number of binding sites (CDCs), and total gene targets containing at least one CDC are listed for each hnRNP.

	A1	A2/B1	F	H1	M	U
Raw reads	7,424,077	21,136,009	41,149,356	(Katz et al., 2010)	10,816,641	27,202,723
Mapped reads	1,778,279 (24%)	16,659,630 (79%)	24,347,839 (59%)	3,856,295	4,727,761 (44%)	11,408,515 (42%)
Reads at unique position in genes	503,416	311,052	586,005	1,674,958	1,540,171	625,481
CDCs	1,956	10,193	18,951	33,209	4,437	17,175
Gene Targets	1,207	4,480	6,479	8,966	1,570	6,057

Supplemental Table S2. Binding for each hnRNP protein (columns) to each set of hnRNP-regulated events (rows). Binding is defined as an event that contains a CDC for a particular hnRNP on or within 2kb of the changing exon. Values represent a ratio of the fraction of changing events with binding compared to the fraction of unchanged events with binding by each hnRNP for each knock down. Larger numbers demonstrate more binding to events that are differentially spliced relative to unchanged events. The gray shading highlights the 3 largest ratios.

		Binding					
		A1	F	M	U	A2/B1	H1
Set of regulated events	**A1**	1.398	1.366	1.344	1.205	1.294	1.143
	F	1.569	1.379	1.101	1.250	1.537	1.250
	M	2.248	1.306	4.364	1.337	1.539	1.182
	U	1.969	1.447	1.822	1.399	1.723	1.427
	A2/B1	2.337	1.160	1.823	1.183	1.585	1.198
	H1	1.795	1.367	1.254	1.401	1.449	1.344

Supplemental Table S3. siRNA and antisense oligonucleotide (ASO) sequences used for hnRNP knock downs in human 293T cells and primary fibroblasts.

Target gene	siRNA sequence 5' to 3'	Source
hnRNP A1	CCACUUAACUGUGAAAAAGAUUU CUUUGGUGGUGGUCGUGGAUU	Venables et al., 2008, Thermo Scientific
hnRNP A2/B1	GUUCAGAGUUCUAGGAGUG GGAGAGUAGUUGAGCCAAA GAACAAUGGGGAAAGCUUA GCAAGACCUCAUUCAAUUG	Zefeng Wang (University of North Carolina), Dharmacon
hnRNP F	GGAAUGUAUGACCACAGAU	Jens Lykke-Anderson (UCSD)
hnRNP H1	CAAACAACGUUGAAAUGGA UGAAAAGGCUCUAAAGAAA GUUCGCAACUCAUGAAGAU	Zefeng Wang (University of North Carolina), Dharmacon

	GAACACAGAUAUGUAGAAC	
hnRNP M	GAAGUCCUAAACAAGCAUAUU AUAUGCCAAUCCAACUAAAUU	Venables et al., 2008, Thermo Scientific
hnRNP U	AGGAUAUUAUUGAAUACCCAA CUGGCCGUGGUAGUUACUCAA	Qiagen, SI02780540 Qiagen, SI02781002
Non-targeting control	UGGUUUACAUGUCGACUAA	Thermo Scientific, D-001810-01-05
ASO	**ASO sequence 5' to 3'**	**Source**
hnRNP A2/B1	AAGAAACTGTCTCAAAGGCA	ISIS Pharmaceuticals
Non-targeting control	CTCAGTAACATTGACACCAC	ISIS Pharmaceuticals

Supplemental Table S4. RT-PCR and qRT-PCR forward and reverse primer sequences (5' to 3') used for validation of hnRNP regulated splicing events. Predicted size (in nucleotides) of the included event (In) and skipping event (Sk) are listed in the final two columns for RT-PCR experiments.

Gene	Forward Primer 5' to 3'	Reverse Primer 5' to 3'	In	Sk
APLP2	AGAATCCTACTGAACCCGGC	ATCAACATCATTGGTTGGCA	269	101
ASPH	TCGAAGATGAAGCAAAAGAACA	CTTCCACGTGGTAACTATGCTC	220	91
C16orf13	GCTCTTCAGAGCAGCAGGAC	CTCCAGGAGGGCTGTGTC	166	98
CAPRIN2	TGATTTTCTTCAAGAGCCGTT	GCTTGAAAGGGTGCTGTGTT	267	162
CASC4	TCCAGTCCTCTTCAGCGTTT	GCAGGATCCATTTGAAGCTC	323	155
CCR6	CCAAAGTCACCAGAGGGAAA	CTGGCTTAGGAATGGGATCA	319	259
DHX32	TTATGGAAACGTGCCTGTCA	GTATCTCTGCCTGGCTCTGG	479	236
DST	ACAAACATGGAACTGCGTGA	TGGCAGCTGCAGTTGTTATC	403	292
EIF4EBP3	GGAGAACTCTCCTGCCCTTT	AGGATACACAGGAGGCATCG	937	178
EML5	ATGCTGTTCACATAGCACCC	GTCGACGGTAGGTGAGCTGT	283	129
FAM126A	TGCAGTAACCAGCATGTCAA	CTAGACGCAGCCCTGGAATA	425	129
G2AD	AGCTGGCCAAACTGCTCTAC	CAGGTCTCGGCACATCTCA	269	144
GAPDH	ATGTTCGTCATGGGTGTGAA	AGTTGTCATGGATGACCTTGG	qRT-PCR	
GAPVD1	ACAGGAACAGACCTTGGTGG	CCGTTTAATGGCATTCCTGT	212	131
HNRNPA1	TGGGGATGGCTATAATGGAT	CGCCATAGCCACCTTGGTTTC	356	200
HNRNPA2/B1 spliced 3'UTR	TCTGCTGCCACAAAGACTGTA	GCAGCAAGACACCTTCCATT	qRT-PCR	
HNRNPDL	ACTTATGGCAAGGCATCTCG	TCTTCAATGTCGTCCTGCAA	231	126
HNRNPH1 intron removal	GAGAAGGCAGACCAAGTGGC	ACTTGAATACTTCAACATATCTGTGTC	qRT-PCR	
HNRNPH1 intron retention	GAGAAGGCAGACCAAGTGGC	TTCAGAAGGGGTGAGACCTCTA	qRT-PCR	
IDS	CATCAGTGTGCCGTATGGTC	GCCGACCTGTGTATCCAAAT	227	97
IVNS1ABP	AGCATCTGGGAGAATGGAGA	CATCATCACTGCCAAACACC	274	133
KIAA0146	TGGAAGTGATTTGTCGGATG	CTGCTCACACATGGCAACTT	646	294
KITLG	TCATTCAAGAGCCCAGAACC	GCCCAGTGTAGGCTGGAGT	272	188
KRIT1	ATCTCGGTGGTCCAACTCAG	TGGCAGTATTCTTTGGACGA	429	281
LAS1L	TTGAACAGTTGGCAGCTTTG	AACTGAGAGAACGGCTTTGG	136	85
MACF1	CCTACTCGTTCCAGCTCCAG	GCAAGGGATGTCCGACTAGA	272	161
MCL1	AGACCTTACGACGGGTTGG	ACATTCCTGATGCCACCTTC	361	133
MEMO1	GCGGATCCTAGTAATCTCTTTGTG	CCAATGGGATGTCTTCCAC	218	141
METTL8	TTCAACCTCCCAGATTCCAG	AGGACTCGCACAGCTGAGTT	294	139
MFF	CGAGCAGTTGGCAGACTAAA	ATGAGGATTAGAAGTGGCGG	291	132
MTF2	GAGCATGTTCTGGAGGCATT	TTGTATATGGGCCAGGTGGA	280	109
MYCBP2	CTCCAAGCCCTTTCTCAGTG	GGCTCATACGTCCATCAGGT	375	195
MYO18A	GCAGGCTGACCTAAAGTTGG	CCTTGGAAGGTCCCTTGTTT	233	188
MYOIB	AAAAGCGCTGTAAGGAAGCA	TTCTTTCCAGCATTGGCTCT	277	103
NEXN	AACGCAGAATTGAGCAGGAT	TCCTTGAGAGATGGTCGTTG	286	244

NF1	CGAAGTGTGTGCCACTGTTT	GGTGAGACAATGGCAGGATT	179	116
NONO	CTCCGAGGAGATACCAGTCG	GTTGCCCATTTGCAGGTATT	346	232
NRCAM	TCAAACCATACAGCAGAAGCA	ACTTGCATTGCCTTCTGGAG	160	103
OSBPL6	TCAAAGAAAGACAAGCGGGT	AAGCGAACTGGTGACATGGT	185	110
PICALM	GCCCAATGATCTGCTTGATT	ATTAAGGCCAGCTGAAGGGT	292	142
PLCXD2	TTGGATCATCTAAACCGGAAA	AACTTCACGCAGGAGCTGTT	315	186
PPP1R7	AGTCGCAGGAGATGATGGAG	TCCCTATGCGATAGTGATTCAA	237	108
PTPN3	TGGAAATCCTGTGTTGAGCA	TCCGTTTCCGTGATGTAGGT	338	203
RBM15	CACTGGCCAAATCTGAAGAAG	AGGCCCATGTAAACTCCACA	256	145
RBM6	GGGATTCTCGACCTGCTAAC	TGCTGGGTTTCTCCTCACTT	1359	80
SLC15A4	CGTTAGGTGGCATTGCCTAT	GGAATCCTCAAATGAAGACTCTGT	430	261
SLC35B3	AAGTCCTGTGGCTGGTACCTT	AATTTGGTGCAGTTGTGCTG	346	191
SOX6	GCTGCTGCTTCTGGACTCA	CATCTTTGCTCCAGGTGACA	213	90
SPAG9	GCTCTTATGGTGTCTTGTAATCG	GGACGATTTCCTGGTACACCT	149	110
STRAP	GTTGATTTGGCCTTCAGTGG	GCGTGAAATCCACAGTCTTG	277	144
SYNCRIP	CGATACCATCGGACAGGATT	TTGAAGAACTGCCAATGCAC	290	130
TMEM53	TGGACTACACCATCGAGATCC	CTCTGGCCAGGACTACTTCG	663	541
TTC7A	CTGTCGGAGGCTTTTGTCAT	TTCCTCCTCCCTCTCTGTCA	151	93
USPL1	ACGTTGCAACTAGGGTGGAG	CAAGCAGGGCAATACTCATCT	294	127
WDR85	CTGGTCGAGGTCCAAAGAAA	TCCTCCCCAACCCTCTCTAT	295	201
WDHD1	TCTTGTGGATCCTCAGTGGC	CATCATCATCCAAGTCTTCCC	193	100
ZCCHC17	AAGGCCTGAGACCATGGAA	GCCCCATAGTCTGTCACCAT	151	81

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell culture and transfections

Human 293T cells were obtained from ATCC and cultured in DMEM supplemented with 10% FBS at 37˚C and 5% CO_2. Cells were transfected at 40% confluency using 5µl/ml Lipofectamine-2000 (Invitrogen) and siRNAs at a final concentration of 100nM. For hnRNP A1, A2/B1, H1, M and U, equal mixtures of multiple siRNA sequences were used and for hnRNP F only one siRNA sequence was used (See Table S3 for all siRNA sequences). Media was changed 4-6 hrs after the initial transfection, and the transfection was repeated 48 hrs later. Cells intended for RNA analysis were suspended and lysed in TRIzol (Invitrogen) 72 hrs after the first transfection. Human primary fibroblasts were grown at 37˚C and 5% CO_2 in DMEM supplemented with 10% FBS, NEAA, and L-glutamine. At 90% confluency, the cells were washed with PBS and transfected with antisense oligonucleotides (ISIS Pharmaceuticals) at a final concentration of 12.5nM using 5µl/ml Lipofectamine-2000 diluted in OptiMEM (See Table S3 for antisense oligonucleotide (ASO) sequences). Growth media was added 4 hrs after, followed by a full media change the next day. Cells were harvested 48 hours after transfection.

Western blot analysis

Cells intended for protein analysis were suspended in NP-40 lysis buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.01% NP-40, Roche Complete Protease Inhibitors EDTA Free), sonicated for 30 seconds in 4˚C water bath sonicator, and then centrifuged at 14,000 RPM for 20 minutes at 4˚C. The lysate supernatant was used for western blotting. Western blots were performed using 25 µg of the obtained protein lysate separated on 10% Bis-Tris gels (Invitrogen) in 1X MOPS Buffer (Invitrogen) and transferred to Hybond-P membrane (Amersham Biosciences). Membranes were incubated overnight with anti-GAPDH (Abcam), anti-hnRNP A1 (Novus

Biologicals), anti-hnRNP A2/B1 (Santa Cruz Biotechnologies), anti-hnRNP F (Santa Cruz Biotechnologies), anti-hnRNP M (Aviva Systems Biology), anti-hnRNP H1 (Bethyl Laboratories), anti-hnRNP U (Bethyl Laboratories), diluted to 1:5000, 1:1000, 1:200, 1:200, 1:2000, 1:2000 and 1:2000, respectively. Secondary antibodies (GE Healthcare) were diluted to 1:7000 and chemiluminescence reagents (Pierce) were used according manufacturers recommendations.

Microarray analysis for splicing changes

Microarray data analysis was performed as previously described (Polymenidou et al., 2011). For each microarray condition, the $\log_2$ ratio of skipping intensities to inclusion intensities was estimated using least-squares analysis. Significantly changing splicing events between hnRNP depletion and control were identified using a q-value < 0.05 and an absolute separation score > 0.5.

RT-PCR and qRT-PCR validations

cDNA was generated by reverse transcribing 1 µg of the obtained total RNA extracted from the control and hnRNP-depleted cells using oligo (dT) primer and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. To test candidate splicing targets, PCR amplification was preformed for 32-36 cycles using ~30ng cDNA template with primers falling in exons flanking the alternate cassette exons (all primer sequences are listed in Table S4). Products were separated on 2% agarose gels followed by staining with SYBR safe (Invitrogen). Quantification of the isoforms was performed with ImageJ software. Intensity ratios between products including the cassette exon and skipping the cassette exon were averaged for 3 biological replicates for the hnRNP depletion and control experiments. Quantitative RT-PCR (qRT-PCR) was carried out on 7900HT Fast Real-time PCR System (Applied Biosystems) using approximately 25ng of cDNA with Power SYBR-Green PCR Master mix (Applied Biosystems) in triplicate. Gene expression values were normalized to GAPDH levels and are shown as fold change relative to sample control. Primer pairs used for qRT-PCR are listed in Table S4 and verified for a single product by a melting curve analysis.

Crosslinking immunoprecipitation followed by high-throughput sequencing (CLIP-seq)

Three 10cm plates of confluent, untreated human 293T cells were subjected to UV cross-linking on ice. CLIP-seq libraries for hnRNP A1, A2/B1, F, M, and U were constructed as previously described (Yeo et al., 2009), using anti-hnRNP A1 (Novus Biologicals), anti-hnRNP A2/B1 (Santa Cruz Biotechnologies), anti-hnRNP F (Santa Cruz Biotechnologies), anti-hnRNP M (Aviva Systems Biology), and anti-hnRNP U (Bethyl Laboratories). Libraries were subjected to standard Illumina sequencing protocols.

Human (hg18) gene structure annotation

The human genome sequence (hg18) and annotations for protein-coding genes were obtained from the University of California, Santa Cruz (UCSC) Genome Browser. Known human genes (knownGene containing 66,803 entries) and 26,570 known isoform clusters (knownIsoforms) with annotated exon alignments to the human genomic sequence were processed as follows. All mRNAs that were aligned to hg18 and were greater than 300 nt were clustered together with the known isoform clusters. A total of 4,162,937 spliced ESTs were mapped onto the high-quality gene clusters to annotate AS events. Final annotated gene regions were clustered together so that any overlapping portion of these databases was defined by a single genomic

position. To identify 5′ and 3′ untranslated regions we relied on the coding annotation in UCSC known genes that we extended 1.5kb downstream or upstream the start and stop codons, respectively.

CLIP-seq data processing and cluster generation

CLIP-seq reads were processed as previously described (Polymenidou et al., 2011). Briefly, reads were trimmed to remove sequencing adaptors and homopolymeric runs >10nt, and mapped to the human genome (hg18) using Bowtie (version 0.12.2 with parameters –q –l 20 –m 5 –k 5 —best). Multiple sequencing runs for a single CLIP experiment were combined, collapsing reads from the same library preparation with identical sequence into one unique read to eliminate redundancies caused by PCR amplification. For the purposes of defining significant clusters, all reads falling at the exact position in the genome were collapsed and considered a single read. Significant clusters were calculated as previously described (Polymenidou et al., 2011).

Motif Analysis

De novo motif finding was implemented using hnRNP clip derived clusters (CDCs), which were extended equally from the center such that all CDCs were 150nt in length. For an appropriate background sequence set, all repeat-masked pre-mRNA sequences were extracted from the human hg18 genome. The Homer algorithm was used for this analysis (findMotifs.pl with parameters –len 5 –homer1 –chopify –norevopp –rna –fasta) (Heinz et al., 2010).

RNA splicing maps

CLIP-seq reads were mapped to the set of all 4,829 splicing array-detected cassette exons at each nucleotide position in four regions, –50 nt to +400 nt near the upstream exon, –400 nt to +50 nt at the 3′ end of the cassette exon, –50 nt to +400 nt at the 5′ end of the cassette exon, and –400 nt to +50 nt near the downstream exon. If at least one CLIP-seq read overlapped a given nucleotide position, the cassette exon was counted as having binding at that position. For each hnRNP, cassette exons were separated into three groups, repressed, activated and unchanged. Using R (version 2.11.1), we plotted at each nucleotide position the fraction of activated cassette exons with binding, the fraction of repressed cassette exons with binding, and the fraction of unchanged cassette exons with binding. Significant peaks around regulated exons compared to unchanged exons were computed using the chisq.test() function in R and a Bonferonni corrected p-value cutoff of 0.05.

Tissue specific exon analysis

The unprocessed single-end RNA-seq Illumina human BodyMap data was downloaded from http://sra.dnanexus.com. Using an in-house analysis pipeline, we mapped the reads to the human genome (hg18) using the GSNAP short-read alignment algorithm (Wu and Nacu, 2010). Next, we ran the MISO algorithm (Katz et al., 2010) on all mapped reads for each dataset in order to compute percent spliced in (PSI) values for each exon. We next applied a Z-score statistic to identify exons that were specifically enriched in the 16 tissues ($|Z\text{-score}| > 2$). We defined sets of exons that were specific to each of the 16 tissues, taking all exons that passed the Z-score cutoff in only one tissue (the exon is differentially spliced only in one of the 16 tissues). For each tissue specific exon, we counted hnRNP CDCs within 2kb. We also generated a randomly distributed set of clusters (RDCs), taking the same size and number of CDCs and randomly shuffling their location in a gene, and counted the number of RDCs within 2kb of all tissue specific exons as a background count. Comparisons of the number of exons with nearby CDCs with the number of exons with nearby RDCs

determined whether the set of exons were enriched for hnRNP binding. Using a Fisher exact test and a p-value cutoff of 0.001, no set of exons were identified as significantly enriched for hnRNP binding. We repeated this same enrichment analysis for all ACEscan exons, downloaded from the UCSC genome browser.

Gene ontology analysis

We used the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID Bioinformatic Resources 6.7; http://david.abcc.ncifcrf.gov/) for all genes with hnRNP binding. The target gene set for each hnRNP was analyzed individually, as well as the intersection of all hnRNP target genes using RefSeq identifiers for each gene.

RNA sequencing preparation (RNA-seq)

Strand-specific RNA-seq libraries were prepared based on the method described in Parkhomchuk et al., 2009, with a few modifications. 10 µg of total RNA was DNase treated (Turbo DNase, Ambion) and then subjected to 2 rounds of polyadenylation selection using Oligo $(dT)_{25}$ Dynabeads (Invitrogen) according to manufacturer's recommendations. Eluted mRNA was then reverse transcribed with a mixture of random hexamers and oligo-dT primers. Unincorporated nucleotides were then removed using G-25 spin columns (GE Healthcare) and the second strand of the cDNA was then synthesized in the presence of dUTP instead of dTTP. This reaction was then brought to 350 µl and sonicated on high for 1 hour of 30 seconds on, 30 seconds off sonication bursts. Fragments ranging from 150-225 bp were then selected from a 6% polyacrylamide gel, before the ends of each fragment were repaired using T4 DNA Polymerase, Klenow DNA Polymerase, and T4 PNK (NEB) at 20°C for 1 hour. Enzymes, buffers, and nucleotides in all reactions were removed using Qiagen PCR Purification MinElute columns. Following the end repair, a single dATP was added to each 3′ end using Klenow Exo minus (NEB) at 37°C for 1 hr. The adaptors were then ligated using 2000 units per µl T4 DNA ligase (NEB) at room temperature for 30 minutes. The dUTPs were then removed using Uracil N-Glycosylase (Fermentas) at 37°C for 30 minutes. One half of the library was then subjected to 15 cycles of PCR amplification using Phusion High Fidelity DNA Polymerase (NEB). The library was then separated on a 6% polyacrylamide gel and fragments between 250-325 nucleotides were selected. 8 pM of the resulting amplified libraries were subjected to standard Illumina sequencing protocols.

RNA-seq data processing and gene expression analysis

Strand-specific RNA-seq reads from control siRNA treatment, and each hnRNP depletion experiment were processed as previously described (Polymenidou et al., 2011). An average of 70% of reads mapped uniquely to our gene structure database, using Bowtie (version 0.12.2, with parameters –l 20 –m 5 –k 5 —best —un —max –q). Multiple sequencing runs of the same condition were combined, collapsing reads from the same library preparation with identical sequence into one unique read to eliminate redundancies caused by PCR amplification. For each condition, the mRNA RPKM (reads per kilobase of million mapped reads) was computed for each gene as a metric for expression. Significantly up- and down-regulated genes upon hnRNP depletion were identified as previously described (Polymenidou et al., 2011), using a local Z-score analysis with cutoffs of $Z > 2$ for upregulated and $Z < -2$ for downregulated.

RBP and cancer target analysis

RNA binding proteins were computationally filtered from PFAM database (http://pfam.sanger.ac.uk/), selecting all proteins in the genome that contained any of the 81 known RNA binding domains. This list was further filtered to remove proteins containing the RNA binding domains specific for tRNAs, snoRNAs and rRNAs, with the remaining 443 genes containing RNA binding domains predicted to bind pre-mRNA and mRNA sequences. This list of 443 RNA binding proteins were filtered for those bound by at least one of the 6 hnRNP proteins assessed in this study (as indicated by CLIP-seq). Genes with binding were filtered further for those significantly regulated at the expression level (from RNA-seq) or those that contained a significantly regulated alternative cassette exon (from splicing arrays) in at least one hnRNP depletion experiment. We also repeated this analysis for a set of 457 cancer-associated genes that were obtained from the Sanger Cancer Gene Census (http://www.sanger.ac.uk/genetics/CGP/Census/), of which 368 could be cross-referenced to our gene annotations. Genes that were bound and regulated according to the above criteria were loaded into Cytoscape (http://www.cytoscape.org/) to visualize the hnRNP regulation network.