iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution

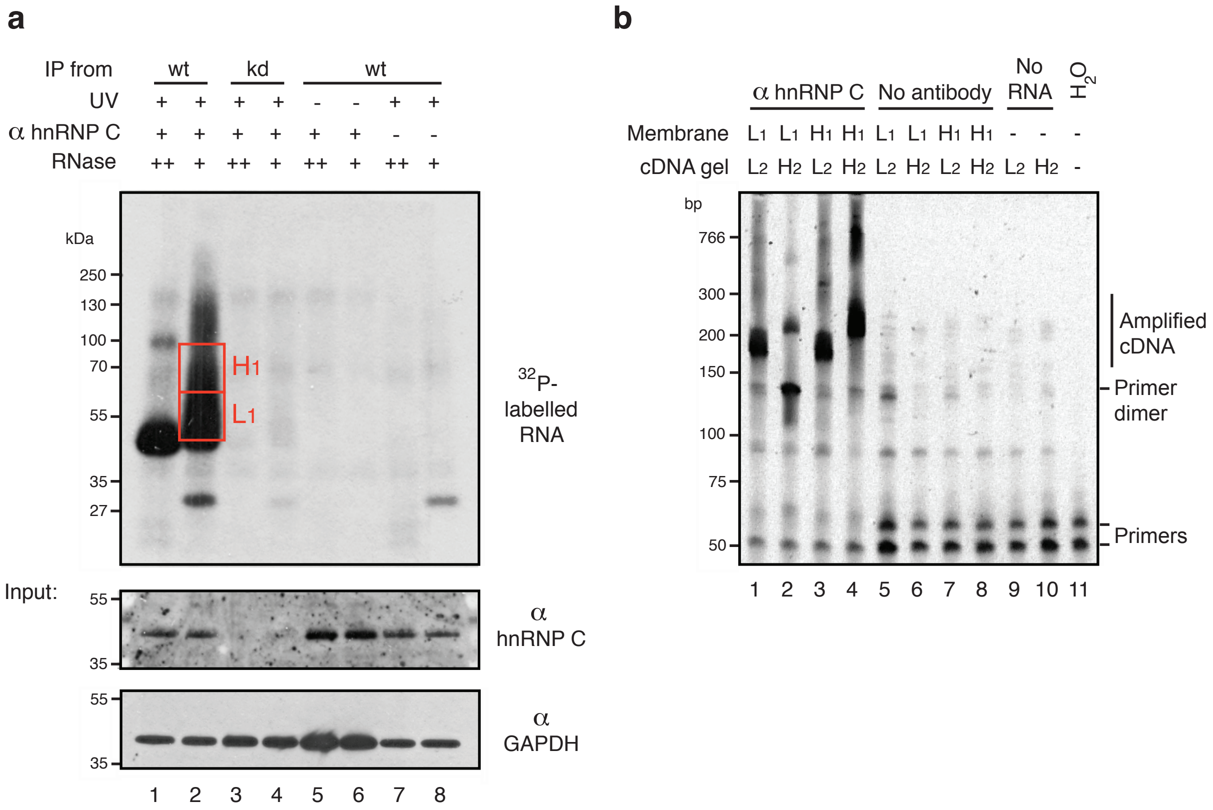
Julian König[1*], Kathi Zarnack[2*], Gregor Rot[3], Tomaž Curk[3], Melis Kayikci[1], Blaž Zupan[3], Daniel J. Turner[4], Nicholas M. Luscombe[2,5], Jernej Ule[1]

[1] MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, UK. [2] EMBL - European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [3] Faculty of Computer and Information Science, University of Ljubljana, Tržaška 25, SI-1000, Ljubljana, Slovenia. [4] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. [5] European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Meyerhofstraße 1, 69117 Heidelberg, Germany.

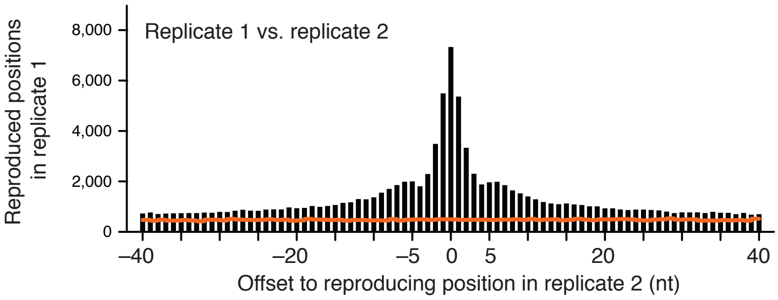
Correspondence should be addressed to J.U. (jule@mrc-lmb.cam.ac.uk).

[*] These authors contributed equally to this work.

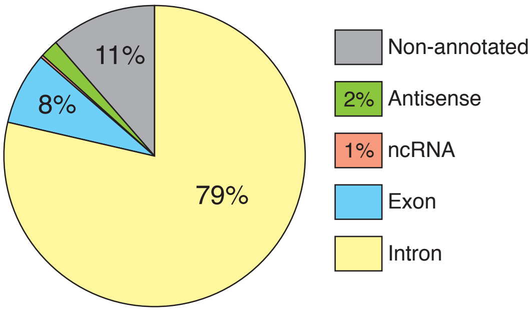
Supplementary Information



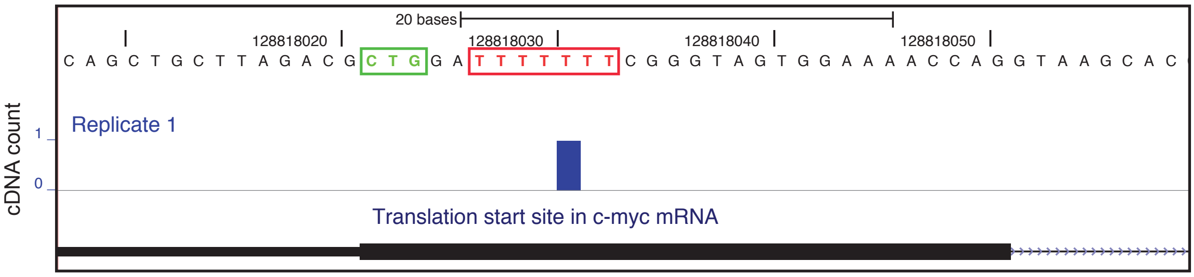
Supplementary Figure 1 iCLIP experiments. **(a)** Analysis of cross-linked hnRNP C–RNA complexes using denaturing gel electrophoresis and western blotting. Protein extracts were prepared from UV-cross-linked and control HeLa cells, and RNA was partially digested using low (+) or high (++) concentration of RNase. hnRNP C–RNA complexes were immuno-purified (IP) from cell extracts using an antibody against hnRNP C (α hnRNP C). The RNA adapter was ligated to the 3` ends of RNAs before radioactively labeling the 5` ends. Complexes were size-separated using denaturing gel electrophoresis and transferred to a nitrocellulose membrane. The upper panel shows an autoradiogram of this membrane. hnRNP C–RNA complexes shifting upwards from the size of the protein (40 kDa) can be observed (lane 2). The shift is less pronounced when high concentrations of RNase were used (lane 1). The radioactive signal disappears when hnRNP C is knocked down (lane 3 and 4), cells were not cross-linked (lane 5 and 6) or no antibody was used in IP (lane 7 and 8). The two red boxes (L1 and H1) mark regions of the membrane that were cut out for subsequent purification steps. The two lower panels show western blot analyses of protein extracts used as input for the IPs above. Antibody against hnRNP C visualizes knock-down efficiency, and antibody against GAPDH (α GAPDH) documents equal protein amounts in input extracts. (**b)** Analysis of PCR-amplified iCLIP cDNA libraries using denaturing gel electrophoresis. RNA recovered from membrane regions L1 and H1 (see above) was reverse transcribed and size-purified using denaturing gel electrophoresis (not shown). Two size fractions of cDNA (L2, 100 – 175 nt, and H2, 175 – 350 nt) were recovered, circularized, linearized and PCR-amplified. PCR products of different sizes can be observed according to different size combinations of input fractions (lane 1 – 4; L1 and H2 recovered from the protein membrane; L2 and H2 recovered from the cDNA gel). PCR products are absent when no antibody was used for the IP (lane 5 – 8) or no RNA was added to the reverse transcription reaction.



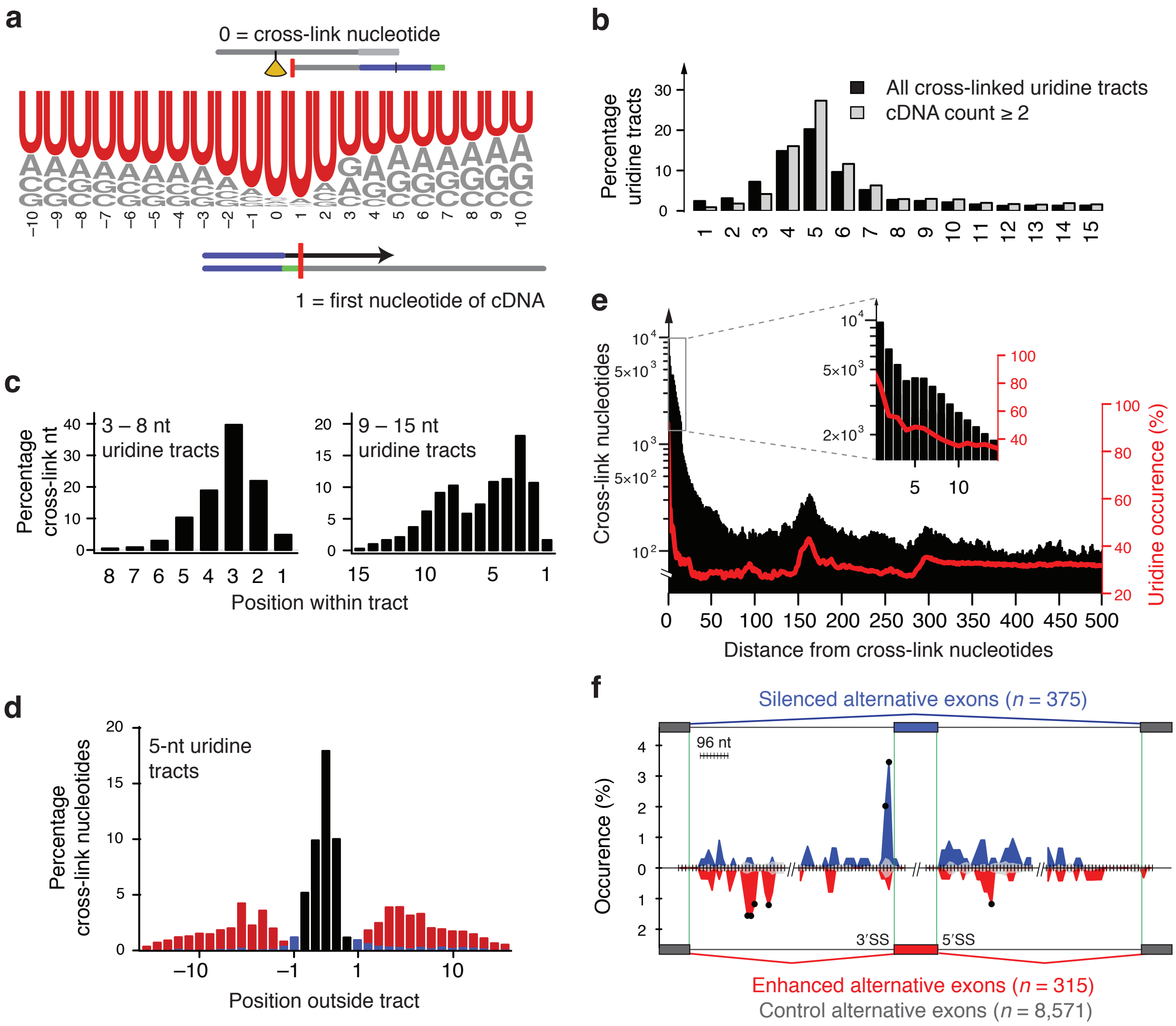
Supplementary Figure 2 Reproducibility analysis comparing replicate 1 with replicate 2. Black bars show the number of cross-link nucleotides in replicate 1 that are reproduced in replicate 2 with a given offset. An offset of 0 nt indicates the number of cross-link nucleotides in replicate 1 that were reproduced by a crosslink nucleotide at exactly the same position in replicate 2. Negative or positive offset values indicate whether the reproducing position in replicate 2 is located upstream or downstream of the cross-link nucleotide in replicate 1, respectively. For example, the bar of height 5,266 at offset +1 nt shows that 5,266 cross-link nucleotides of replicate 1 were reproduced by a cross-link nucleotide 1 nt downstream in replicate 2. The orange curve depicts results of the same analysis upon randomization of cross-link nucleotide positions in replicate 2.



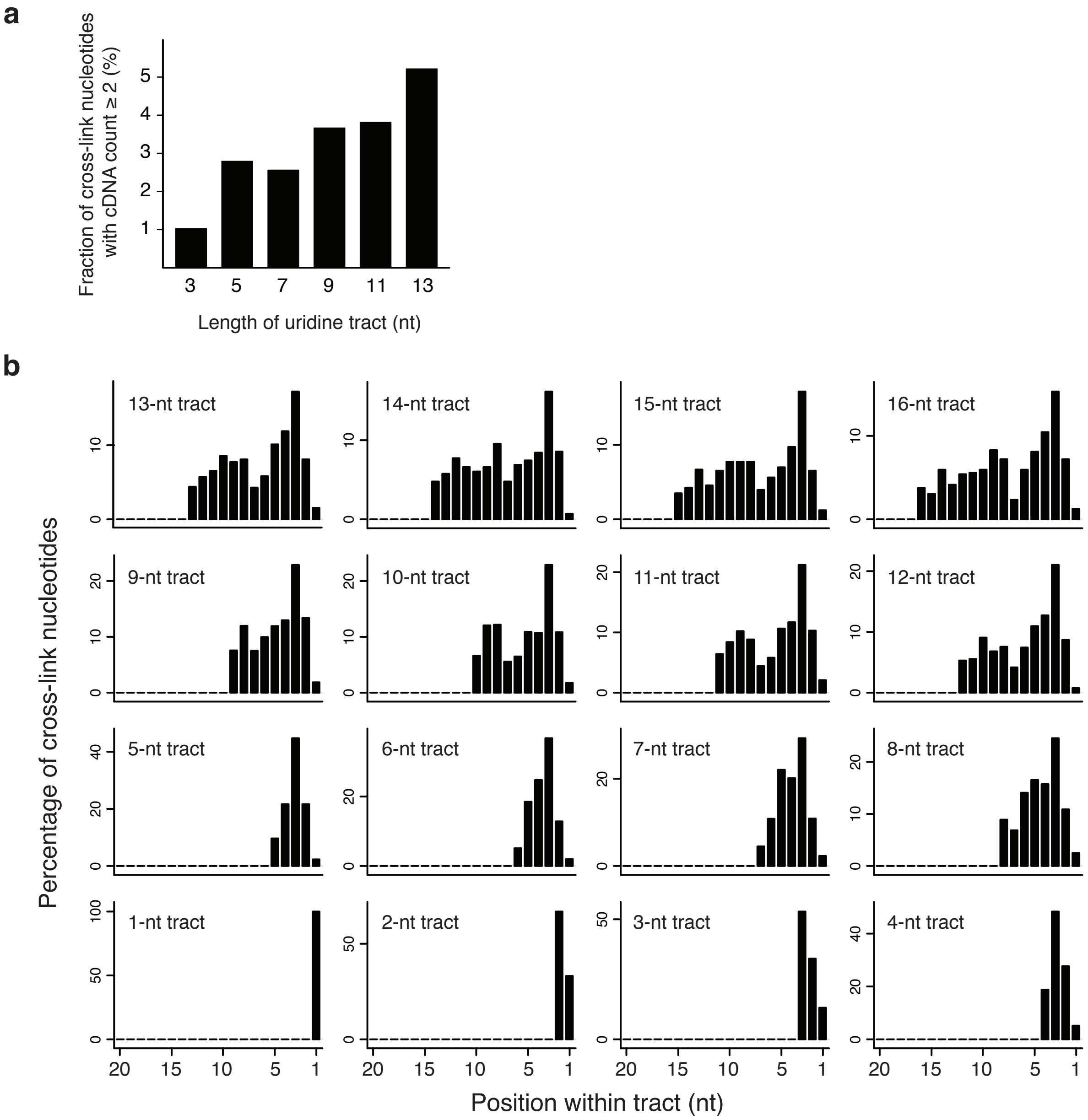
Supplementary Figure 3 Genomic location of hnRNP C cross-link nucleotides. A pie chart depicting the fraction of cDNA sequences that map to different genomic regions (as given on the right; gene annotations based on UCSC hg18.knownGene).


20 bases
128818020
128818030
128818040
128818050
C A G C T G C T T A G A C G C T G G A T T T T T T T C G G G T A G T G G A A A A C C A G G T A A G C A C
cDNA count
1
0
Replicate 1
Translation start site in c-myc mRNA

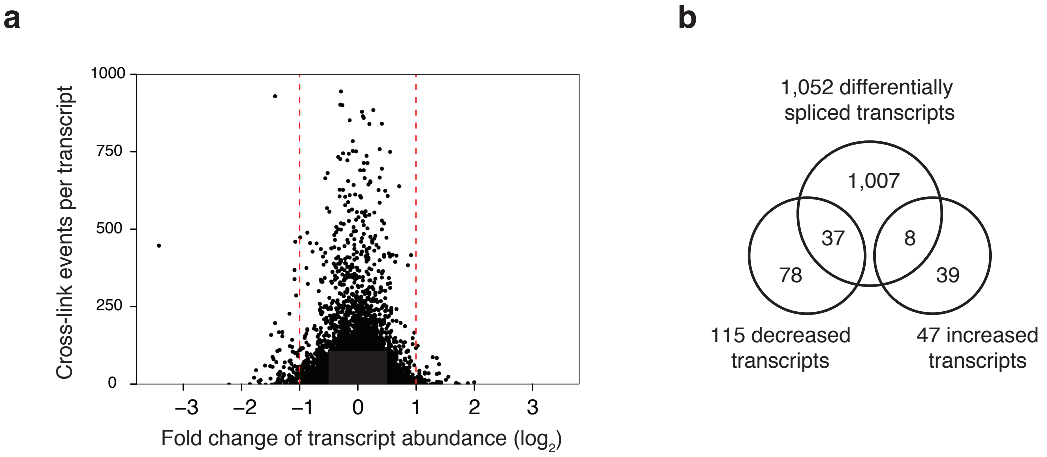
Supplementary Figure 4 hnRNP C cross-linking to the regulatory element in c-myc mRNA. A hnRNP C cross-link nucleotide locates to a seven nucleotide uridine tract within the c-myc mRNA. The corresponding genomic locus on chromosome 8 (nucleotides 128,818,008 to 128,818,059; modified UCSC genome browser image) surrounding the respective thymine tract (red) is shown. A cross-link nucleotide within the shown locus was only found in replicate 1. Binding of hnRNP C to this element within the internal ribosomal entry site (IRES) was shown to regulate alternative usage of an upstream start codon (CTG, green).



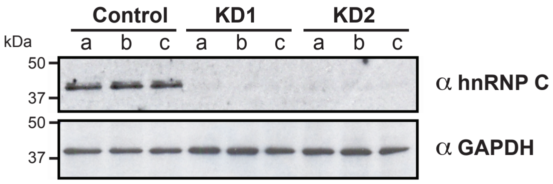
Supplementary Figure 5 Analyses of hnRNP C binding based on the clustered cross-link nucleotides dataset. **(a)** Weblogo showing base frequencies of clustered cross-link nucleotides and 20 nucleotides of surrounding genomic sequence. Labeling as in **Fig. 3a**. Uridine represented 91% of cross-link nucleotides. **(b)** Length distribution of uridine tracts harboring clustered cross-link nucleotides. Analysis and labeling as in **Fig. 3b**. 83% of cross-link nucleotides were part of contiguous tracts of four or more uridines. **(c)** Positioning of clustered cross-link nucleotides within uridine tracts. Analysis and labeling as in **Fig. 3c**. Longer tracts contain two peaks at a defined spacing of 5 – 6 nucleotides. **(d)** Binding neighborhood of five nucleotide uridine tracts. Analysis and labeling as in **Fig. 3d**. Clustered cross-link nucleotides within 5 nt uridine tracts are commonly associated by flanking cross-link nucleotides again residing in uridin tracts. **(e)** Long-range spacing of clustered cross-link nucleotides. Analysis and labeling as in **Fig. 3e**. Increased occurrence of clustered cross-link nucleotides coincided with peaks in uridine density at 165 and 300 nucleotides distance. **(f)** The RNA map of clustered cross-link nucleotides within regulated pre-mRNAs. Analysis and labeling as in **Fig. 4a**. Silenced alternative exons show strong enrichment of cross-link nucleotides proximal to the 3′ splice sites (3′SS).



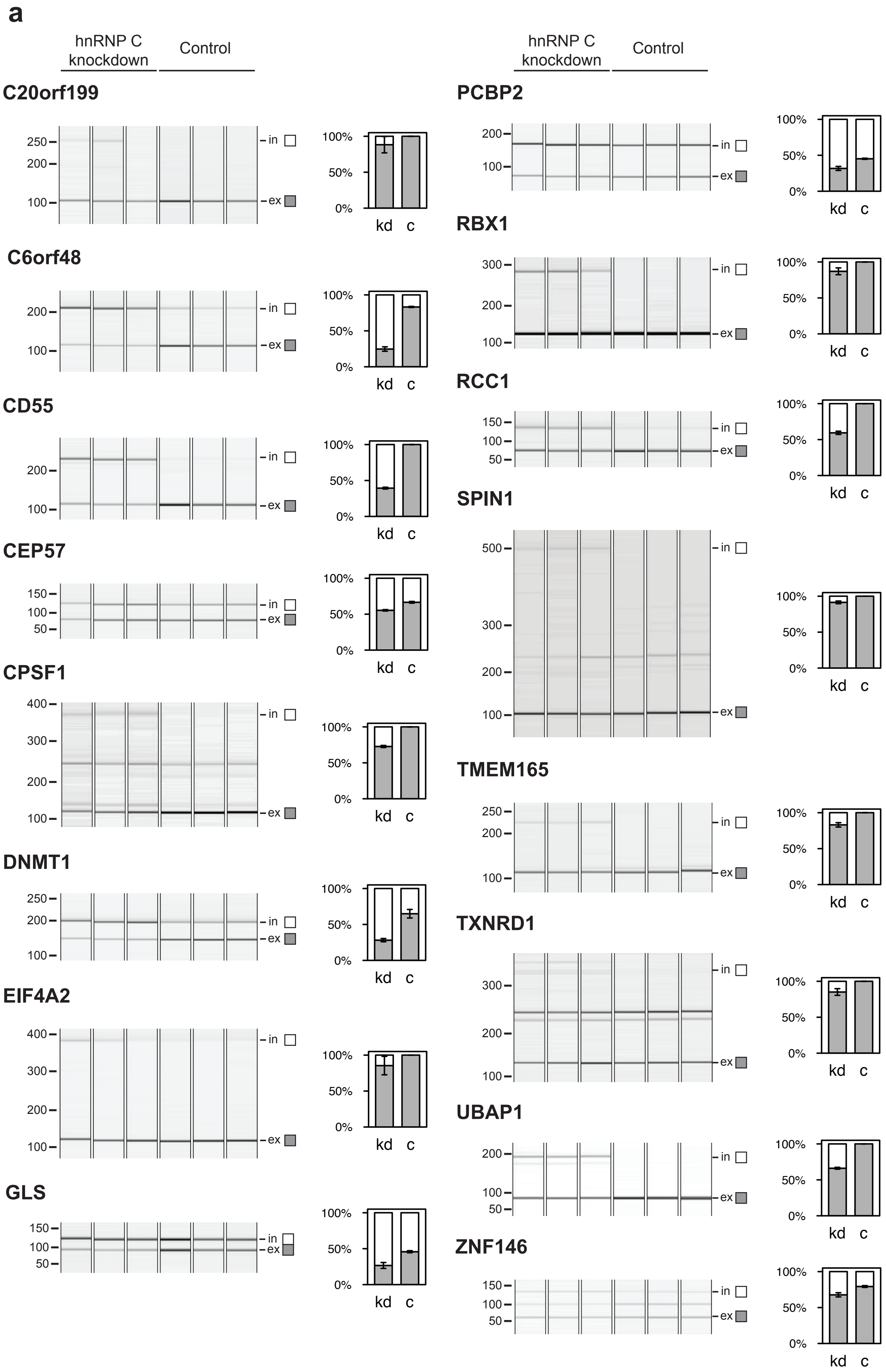
Supplementary Figure 6 The dual pattern of hnRNP C cross-linking on uridine tracts. **(a)** Fraction of cross-link nucleotides with a cDNA count of at least two on the third position from the 3` end of uridine tracts of different lengths (as given below). With increasing tract length from 3 nt to 13 nt, cross-link nucleotides with a cDNA count of at least two represent an increasing proportion of all cross-link nucleotides (p value < 10^{-5} by Wilcoxon rank sum test comparing tracts of 5 and 13 uridines). **(b)** Distribution of cross-link nucleotides over uridine tracts of different length. The number of cross-link nucleotides locating to each position is given as a fraction of all cross-link nucleotides locating to tracts of a given length. Cross-linking predominantly occurred on the third position from the 3` end. In addition, tracts of more than eight uridines display a second peak at a constant distance of five or six nucleotides from the downstream peak.

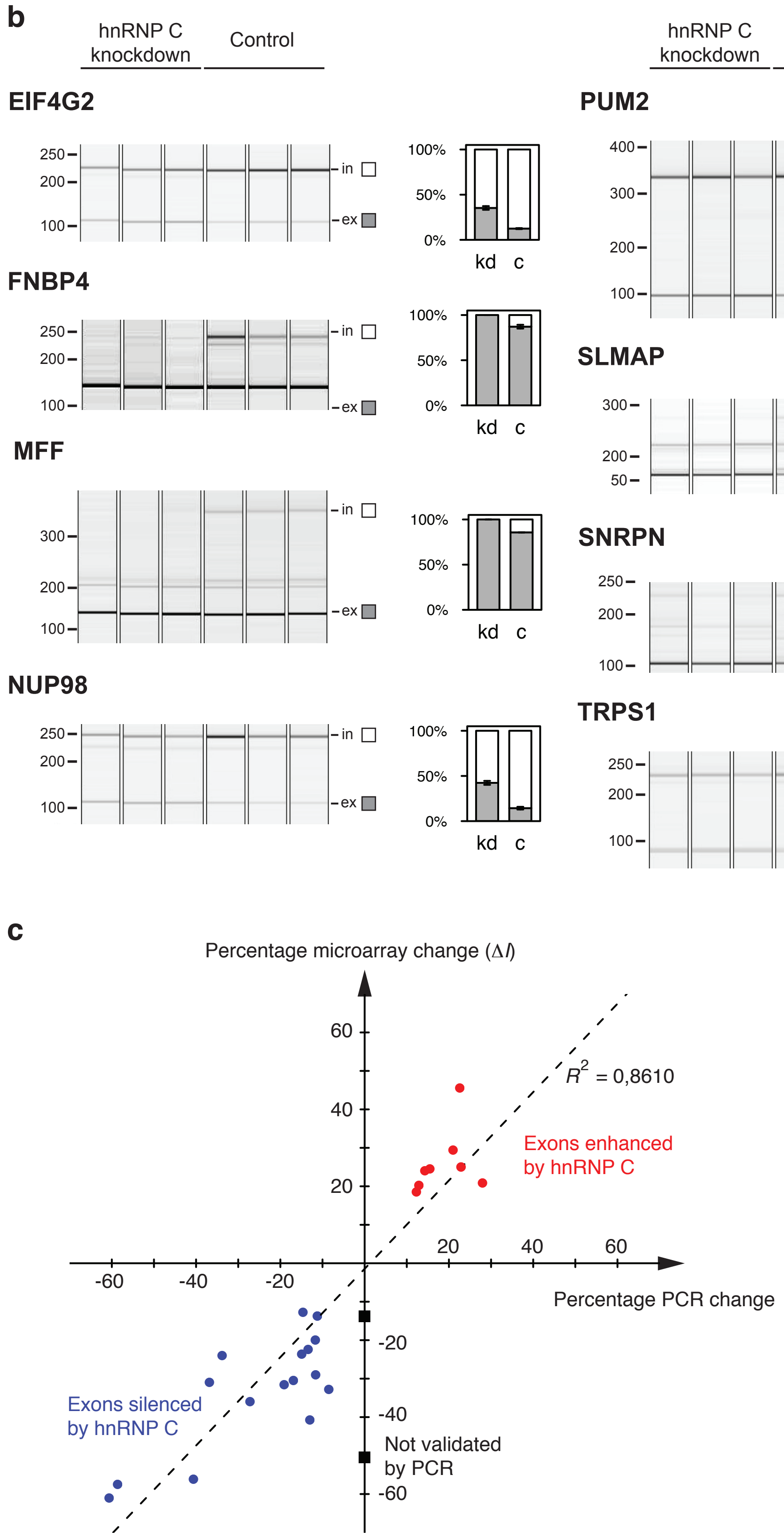


Supplementary Figure 7 Analyses of differentially expressed transcripts in hnRNP C knockdown cells. **(a)** Scatter plot comparing the change in expression level in the hnRNPC knockdown with the total number of hnRNP C cross-link events per transcript. The red dashed lines indicate a change in transcript abundance by a factor of 2. We did not observe an apparent correlation between cross-linking and differential regulation (Pearson correlation coefficient 0.099 and 0.106 for decreased and increased transcripts, respectively). **(b)** Venn diagram depicting the significant overlap between differentially expressed transcripts (162 in total, including 115 decreased and 47 increased transcripts) and those that show a change in at least one alternative splicing event upon hnRNP C knockdown (1,052 transcripts harboring a total of 1,340 differentially spliced exons). 4.3% of the transcripts with at least one splicing change (45/1,052) also showed differential expression, which was significantly more common than among all transcripts (0.7%, 162 out of 24,571 transcripts; p value = 2.0×10^{-24} by to hypergeometric distribution). Vice versa, 27.7% of transcripts with changes in expression levels (45/162) harboured at least one differentially spliced exon, which was also significantly more common than among all transcripts (4.3%, 1,052 out of 24,571 transcripts; p value = 2.0×10^{-24} by hypergeometric distribution).



Supplementary Figure 8 Western analysis of hnRNP C knockdown and control HeLa cells prepared for microarray and RT-PCR analyses. Protein extracts from HeLa cells transfected with two different siRNAs (KD1 and KD2) were compared to control samples (Control). For each condition Western analysis is shown in triplicates (a, b and c). The upper panel was probed with an hnRNP C antibody (α hnRNP C), while the lower panel controls for loading using a GAPDH antibody (α GAPDH). Numbers on the left refer to the sizes of a protein standard in kDa.


a
hnRNP C knockdown
Control
C20orf199
250
200
100
in
ex
100%
50%
0%
kd
c
C6orf48
CD55
CEP57
150
50
CPSF1
400
300
DNMT1
EIF4A2
GLS
PCBP2
RBX1
RCC1
SPIN1
500
TMEM165
TXNRD1
UBAP1
ZNF146



Supplementary Figure 9 Quantification of splicing changes using RT-PCR and capillary electrophoresis. **(a)** (previous page) **(b)** Quantification of alternative splicing in hnRNP C knockdown (kd) and control (c) HeLa cells. Capillary electrophoresis image and signal quantification are shown for each validated gene. Quantified transcripts including (in) or excluding (ex) the regulated alternative exon are marked on the right. Average quantification values of exon inclusion (white) and exclusion (grey) are given as a fraction of both. Error bars represent standard deviation of three replicate experiments. (a) and (b) show results for exons that are silenced and enhanced by hnRNP C, respectively. **(c)** Graph comparing the percent change values determined by quantitative PCR and splice-junction microarray analyses (Δ*I* values as determined with ASPIRE3). Silenced (blue) and enhanced (red) alternative exons that were reproduced by quantitative PCR are shown as circles. Exons that displayed no change in quantitative PCR are depicted as black squares. Changes in 24 of 26 analyzed alternative exons could be reproduced.

Supplementary Table 1 Genomic mapping of iCLIP sequence reads.

	Replicate 1	Replicate 2	Replicate 3	Total
hnRNP C iCLIP experiments:				
Initial sequence reads[a]	6,544,506	6,544,506	6,544,506	6,544,506
After experiment separation	2,610,554	2,292,169	1,376,258	6,278,981 (96%)[b]
After mapping to the human genome	1,595,604	1,624,238	942,970	4,162,812 (66%)
After random barcode evaluation	309,489 (19%)	216,295 (13%)	115,566 (12%)	641,350 (15%)
Cross-link nucleotides	302,692	212,098	113,920	614,740[c]
No-antibody iCLIP controls:				
Initial sequence reads	5,782,612	12,597,621	12,597,621	18,380,233
After experiment separation	91,310	122,957	71,044	285,311 (2%)
After mapping to the human genome	6,589	11,055	15,244	32,888 (11%)
After random barcode evaluation	386 (6%)	551 (5%)	843 (6%)	1,780 (5%)
Cross-link nucleotides	384	520	803	1,707

[a] Number of sequence reads from Illumina GA2 before data analyses.

[b] Numbers in brackets indicate fraction relative to entry above.

[c] The total number of crosslink nucleotides is smaller than the sum of replicates 1 – 3, since reproduced positions were counted only once.

Supplementary Table 2 Quantification of alternative mRNA isoforms using RT-PCR.

Gene symbol	Gene description	Exon	Spliced region	Exon coordinates	Strand	% Microarray change	% PCR change	Forward primer	Reverse primer	Product sizes in bp (in/ex)
Exons silenced by hnRNP C										
AL590482	n.a.	E2	chr6:166282734-166283392	chr6:166258138-166321007	-	-13.8	no change	AACTCGAAATGAAGCGGAAA	GCCTCCCTGTGAATTCTCTC; TGGCTATTTTTGTTGATGATAGGA	124/87
C20orf199	Uncharacterized protein C20orf199	E7	chr20:47330430-47330514	chr20:47329153-47338989	+	-19.9	-11.7	TTGGAAGAGGGAGTCACCAC	TCCAGAGGGCTCCTCTCATA	257/109
C6orf48	Protein G8	E13	chr6:31912181-31912273	chr6:31910957-31912992	+	-57.5	-58.6	GTTCATCGCCGTGTTATCCT	GGGGGAGATTCCAAACCTTA	214/120
CD55	Complement decay-accelerating factor Precursor	E15	chr1:205580360-205580476	chr1:205579386-205599514	+	-61.0	-60.6	CCAGGACAACCAAGCATTTT	GGAATCATCTTAAGTGTCCATCAA; CAAGCAAACCTGTCAACGTG	407/114
CEP57	Centrosomal protein of 57 kDa	E3	chr11:95168328-95168371	chr11:95163555-95172044	+	-13.7	-11.2	CGGCTTCTGGTTCTCACTTG	GATGAAGAATGCCGAACCAT	123/79
CPSF1	Cleavage and polyadenylation specificity factor subunit 1	E5	chr8:145599070-145599193	chr8:145597875-145605207	-	-36.0	-27.2	CTACGTGTACCGCCTCAACC	CGTTGCCAAAGAAGGAGAAG	374/119
DNMT1	DNA (cytosine-5)-methyltransferase 1	E5	chr19:10151863-10151910	chr19:10149043-10152026	-	-31.0	-36.8	GAAGCCCGTAGAGTGGGAAT	GCCTGGTGCTTTTCCTTGTA	193/145
EIF4A2	Eukaryotic initiation factor 4A-II	E6	chr3:187985179-187985445	chr3:187985179-187985445	+	-12.7	-14.6	CCTTCCGCTATTCAGCAGAG	CAACTGTTGCAGGATGGAAA	385/120
GLS	Glutaminase kidney isoform, mitochondrial Precursor	E17	chr2:191505688-191508062	chr2:191504609-191526536	+	-31.5	-19.1	CCTCGAAGAGAAGGTGGTGA	CCTCATTTGACTCAGGTGACA; CGAAGTGCAGACACATCTCC	124/86
NTNG1	Netrin-G1 Precursor	E19	chr1:107774895-107775062	chr1:107762790-107824756	+	-50.4	no change	CCCAAAGGCACTGCAAATAC; GCACAACTGGACGATGAGAA	AGCTCGTTGTCGCAGACATT	188/71
PCBP2	Poly(rC)-binding protein 2	E14	chr12:52144811-52144903	chr12:52142618-52145983	+	-22.4	-13.4	GTCATCTTTGCAGGTGGTCA	GCTTGGTCAAATCTGGCTGT	166/73
RBX1	RING-box protein 1	E4	chr22:39681237-39681396	chr22:39679583-39689997	+	-40.7	-13.0	TGCAGGAACCACATTATGGA	CGAGAGATGCAGTGGAAGTG	285/125
RCC1	Regulator of chromosome condensation	E6	chr1:28708001-28708004	chr1:28708001-28715824	+	-56.1	-40.6	GATCTGCACTTCGCATTTTG	CCCTGGGATCTCTGATCTCTC; AAAAATCAGTTTACCTACTCTCCCTTC	129/80
SPIN1	Spindlin-1	E6	chr9:90221380-90221501	chr9:90193274-90223513	+	-32.8	-8.5	CCGTGGGCCTGTGGACTG	TCTGGTTAATCCACCATCCAA	495/105
TMEM165	Transmembrane protein 165	E3	chr4:55964161-55964262	chr4:55957321-55972538	+	-30.4	-16.9	TAGCCACCGGAACAAAGAAC	GAACTGGAGCTGCTGGTGTA	222/119
TXNRD1	Thioredoxin reductase 1, cytoplasmic	E12	chr12:103208902-103209012	chr12:103205018-103229198	+	-23.6	-14.9	TTTTCTTCACTCCGGCATTT	TCAGGGCCGTTCATTTTTAG	358/136
UBAP1	Ubiquitin-associated protein 1	E5	chr9:34193380-34193533	chr9:34169239-34210906	+	-24.0	-33.8	CACCTTTCGGCTTCTGAGAC	CATGAAAATCTGCACCCAACT	205/83
ZNF146	Zinc finger protein OZF	E4	chr19:41400864-41400937	chr19:41397938-41411490	+	-29.0	-11.6	CCGAGTGGACATTTTGGTCT	TTCTTGCTTCAACAGAGGATCA	132/58
Exons enhanced by hnRNP C										
EIF4G2	Eukaryotic translation initiation factor 4 gamma 2	E20	chr11:10779784-10779897	chr11:10779210-10780172	-	25.1	22.9	ATCGCAGTTTGGAGAGATGG	TATCTGGGGCTGAAGCTTTG	226/112
FNBP4	Formin-binding protein 4	E19	chr11:47703866-47703964	chr11:47702907-47709502	-	20.3	12.9	TTGCCAAACAGACCTTGAAA	GGAGGGTCCAGAATGGAGTA	250/150
MFF	Mitochondrial fission factor	E12	chr2:227920186-227920344	chr2:227913340-227928637	+	24.1	14.3	GAAGAAATCCGAGCAGTTGG	TGACGTTCCTTCAATGGTTG	357/138
NUP98	Nuclear pore complex protein Nup98-Nup96 Precursor	E13	chr11:3722316-3722455	chr11:3713131-3731122	-	20.9	28.0	TAAACCAGCACCTGGGACTC	ATTGATGTGCTGCTGGAGAA	253/112
PUM2	Pumilio homolog 2	E14	chr2:20341825-20342061	chr2:20326702-20346189	-	24.5	15.5	GGGTGCTGCTATAGGCTCAG	CTCCAGGTGCTGCAGAGATA	330/93
SLMAP	Sarcolemmal membrane-associated protein	E14	chr3:57826009-57826059	chr3:57825483-57832403	+	45.6	22.6	GGAGCTCCAGGCAAAAATAG	TTGGTTAGATGCCCTTCGAC	270/168
SNRPN	Small nuclear ribonucleoprotein-associated protein N	E43	chr15:22798965-22799128	chr15:22778677-22815267	+	18.6	12.3	GTGATGTCCAGGAAGGAGGA	TGATTCCATTTGCAGGTCAG	229/107
TRPS1	Zinc finger transcription factor Trps1	E3	chr8:116705004-116705160	chr8:116701463-116749946	-	29.4	21.0	CGAGGGTGTTCTTGACGATT	CCTTCACTTGCAACGTTTCTC	236/78

Supplementary Table 3 Quantification of predicted splicing changes using RT-PCR.

Gene symbol	Gene description	Exon	Alternative exon coordinates	Coordinates of skipped region	Strand	% PCR change	p value	Forward primer	Reverse primer	Product sizes in bp (in/ex)
Exons silenced by hnRNP C										
C12orf23	UPF0444 transmembrane protein C12orf23	E7	chr12:105885192-105885309	chr12:105885089-105889013	+	–19.9	3.9×10^{5}	CCTTAATGATGAACCACCAGAA	AAGATACCCCCAGTCACACG	87/206
MTRF1	Peptide chain release factor 1, mitochondrial Precursor	E3	chr13:40734548-40734617	chr13:40734468-40735621	–	–23.5	1.7×10^{2}	TTCCGACCTCAGTAAAGAGAGC	CCAAACACACAGGTGACGAT	79/150
PRKAA1	5'-AMP-activated protein kinase catalytic subunit alpha-1	E6	chr5:40810788-40810831	chr5:40807723-40811269	–	–6.2	1.8×10^{3}	TGTCTCAGGAGGAGAGCTATTTG	GACGCCGACTTTCTTTTTCA	71/116
TBL1XR1	F-box-like/WD repeat-containing protein TBL1XR1	E3	chr3:178361354-178361470	chr3:178299024-178397603	–	–17.1	1.0×10^{3}	GTTGGAGGCCACCGTTTC	TGCAACTGAATATCCGGTCA	70/188
ZNF195	Zinc finger protein 195	E7	chr11:3347164-3347308	chr11:3340409-3348781	–	–18.5	1.0×10^{3}	AGCCCTGGAATGTGAAGAGA	CTGGCAGAAGGTCTTGGGTA ACGCAGCAATCACACTTCTG	81/185
no change observed										
BRD2	Bromodomain-containing protein 2	E16	chr6:33054846-33054933	chr6:33054144-33055583	+	-3.3	0.2	TGGACCTTCTGGAGGAAGTG	CTGTAGGCAGGGCAGGTG	74/179
CHD2	Chromodomain-helicase-DNA-binding protein 2	E3	chr15:91227852-91228012	chr15:91227778-91229749	+	-4.8	0.11	GGTTTGGGCGACCAGGAG	CAGAACCAACAGCAACCAAA TGAAACGTAGTCAGGGTTCCA	86/142
FLNB	Filamin-B	E30	chr3:58102626-58102663	chr3:58099297-58103417	+	1.5	0.23	TCCTAACAGCCCCTTCACTG	CAGGCCGTTCATGTCACTC	70/142
IQWD1	Nuclear receptor interaction protein	E16	chr1:166258851-166258909	chr1:166240656-166274233	+	-3.1	0.14	TCTGTTGAGGCATCTGGACA	GTTCACCTGTCCCTGGTTTG	85/145

Supplementary Methods:

iCLIP protocol. HeLa cells grown in a 10 cm plate were covered with ice-cold PBS buffer and subjected to UV-C irradiation (100 mJ/cm^2, Stratalinker 2400). Upon removal of PBS buffer, cells were scraped off and transferred into microtubes (2 ml each). Cells were precipitated by centrifugation for 1 min at 14,000 rpm and shock-frozen on dry ice.

For magnetic bead preparation, 50 µl of protein A-coated Dynabeads (Invitrogen) were washed 2× with 900 µl lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 1 % NP-40, 0.1 % SDS, 0.5 % Na-Deoxycholate). Dynabeads were resuspended in 200 µl lysis buffer, and 10 µg of hnRNP C antibody (Santa Cruz H-105) were added. After rotation at room temperature for 30 – 60 min, Dynabeads were washed 2× with lysis buffer and kept in last wash until addition of cross-linked lysate.

Pellets were resuspended in 1 ml lysis buffer and sonicated. For partial RNase digestion, RNase I (Ambion) was diluted 1:50 and 1:100 in lysis buffer for high and low RNase treatment, respectively. 10 µl RNase I dilution and 5 µl Turbo DNase (Ambion) were added to the cross-linked lysate and incubated for 3 min at 37°C and 800 rpm. Cells were precipitated by two rounds of centrifugation at 4°C and 14,000 rpm for 10 min followed by careful collection of the supernatant. The supernatant was added to Dynabeads and incubated for 1 h or overnight at 4°C and 800 rpm. Dynabeads were washed 2× with high-salt wash buffer (50 mM Tris-HCl pH 7.4, 1 M NaCl, 1 mM EDTA, 0,1 % SDS, 0.5 % Na-Deoxycholate, 1 % NP-40) and 1× with PNK wash buffer (20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 0,2 % Tween-20).

For dephosphorylation of 3′ ends, Dynabeads were resuspended in 2 µl 10× Shrimp alkaline phosphatase buffer (Promega), 17.5 µl H_2O and 0.1 µl Shrimp alkaline phosphatase (Promega) and incubated at 37°C for 10 min with intermittent shaking (10 sec at 700 rpm followed by 20 sec pause). Samples were washed 2× with high-salt wash buffer, 1× with 900 µl PNK wash buffer and 1× with 50 µl 1× RNA ligase buffer (NEB, freshly prepared from frozen stock). For RNA linker ligation, Dynabeads were resuspended in 15 µl L3 ligation mix (5 µl L3 RNA linker [5′-phosphate-UGAGAUCGGAAGAGCGGTTCAG-3′-Puromycin, 20 pM], 1.5 µl 10× RNA ligase buffer, 7.75 µl H_2O, 0.5 µl RNasin [Promega], 0.25 µl RNA ligase [NEB]) and incubated overnight at 16°C. Samples were mixed with 5 µl NuPAGE loading buffer (Invitrogen), incubated for 5 min at 70°C and placed on a magnetic stand to collect the eluate.

Samples were run on 9-well or 10-well Novex NuPAGE 4-12% Bis-Tris gels (Invitrogen) with 1× MOPS running buffer (Invitrogen). After gel electrophoresis, protein and covalently bound RNAs were transferred to a nitrocellulose membrane (Whatman) using a Novex wet transfer apparatus (Invitrogen). The nitrocellulose membrane was rinsed with 1× PBS, wrapped into cling film and exposed to a BioMax XAR Film (Kodak) at –80°C.

For isolation of cross-linked RNAs, 2 mg/ml proteinase K (Roche) was pre-incubated in PK buffer (100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA) for 5 min at 37°C. In order to recover different size fractions of RNAs, two fragments were cut out of the nitrocellulose membrane at different heights above the molecular weight of the protein (40 kDa). 200 µl proteinase K solution was added to each fragment and incubated for 30 min at 55°C. Incubation was repeated after addition of 130 µl PK/7 M urea buffer (100 mM Tris-HCl pH 7,5, 50 mM NaCl, 10 mM EDTA, 7 M urea). Samples were cooled to 37°C, mixed with 170 µl H_2O and 600 µl RNA phenol/$CHCl_3$ (Ambion) and incubated for 5 min at 37°C and 1,100 rpm. After centrifugation for 10 min at 13,000 rpm and room temperature, 450 µl of the aqueous phase were transferred into a new microtube and again subjected to centrifugation. 400 µl of supernatant were mixed with 0.5 µl Glycoblue (Ambion), 40 µl 3 M sodium acetate pH 5.5 and 1 ml 100 % EtOH and incubated overnight at –20°C. RNAs were precipitated by centrifugation for 30 min at 15,000 rpm and 4°C, washed with 500 µl 80 % EtOH and resuspended in 12 µl H_2O.

For reverse transcription, 1 µl RT primer (2 pmol/µl; the following three primers were used for replicates 1 to 3: 5′-phosphate–NNN**CA**AGATCGGAAGAGCGTCGTGGATCCT GAACCGCTC-3′; 5′-phosphate-NNN**GA**AGATCGGAAGAGCGTCGTGGATCCTGAACCG

C-3′; 5′-phosphate-NNN**TG**AGATCGGAAGAGCGTCGTGGATCCTGAACCGCTC-3′; NNN represents 3-nt random barcode and bold nucleotides mark 2-nt barcode used as an experiment identifier) and 1 ml 10 mM dNTP mix were added to the RNA, preheated for 5 min to 70°C and then held at 42°C. Once 6 µl RT mix (5 µl 5× RT buffer [Invitrogen], 1 µl 0.1 M DTT, 0.5 µl Superscript III reverse transcriptase [Invitrogen], 0.5 µl RNasin) were added and mixed by pipetting, reverse transcription was performed with the following program: 10 min at 42°C, 40 min at 50°C, 20 min at 55°C, and hold at 4°C. To remove RNA, samples were heated for 2 min to 95°C, mixed with 1 µl RNase A (Ambion) and incubated for 20 min at 37°C. cDNAs were precipitated by addition of 80 µl TE buffer, 0.5 µl Glycoblue, 10 µl 3 M sodium acetate pH 5.3 and 250 µl 100 % EtOH, incubation for 1 h on dry ice or overnight at –20°C, and centrifugation for 30 min at 4°C and 15,000 rpm. Pellets were washed with 500 µl 80 % EtOH, dried for 3 min at room temperature and resuspended in 6 µl H_2O.

For size separation, cDNAs were mixed with 2 µl 2× TBE-urea loading buffer (Invitrogen) and incubated for 3 min at 70°C. Samples were run on a 6 % TBE urea gel (Invitrogen) in 1× TBE buffer for 40 min at 180 V. In order to recover different size fractions, two bands were cut from the gel corresponding to a cDNA size of 100 – 175 nt and 175 – 350 nt. Gel fragments were mixed with 400 ml TE buffer, crushed with a 1 ml syringe plunger and incubated for 2 h at 37°C and 1,100 rpm. A Costar SpinX column (Corning Incorporated) was prepared by addition of two 1 cm glass wool pre-filters (Whatman 1823-101) and centrifugation for 1 min at 13,000 rpm. After transfer of the supernatant to the column, 40 µl 3 M sodium acetate pH 5.5 and 0.5 µl glycogen were added. Columns were vortexed before adding 1 ml 100 % EtOH and incubating overnight at –20°C. Columns were washed by addition of 500 µl 80 % EtOH and centrifugation for 10 min at 15,000 rpm and 4°C. Pellets were dried for 3 min at room temperature and resuspended in 12 µl H_2O.

In order to circularize the cDNAs, samples were mixed with 1.5 µl 10× CircLigase buffer II (Epicentre), 0.75 µl 50 mM $MnCl_2$ and 0.75 µl CircLigase II (Epicentre) and incubated for 1h at 60°C. For subsequent linearization, a primer (5′-GTTCAGGATCCA CGACGCTCTTCAAAA-3′) complementary to the BamHI restriction site in the RT primer was annealed by adding 26 µl H_2O, 5 µl FastDigest buffer (Fermentas) and 1 µl 10 µM primer and incubation with the following program: 2 min at 95°C, 70 cycles starting for 1 min at 95°C and reducing the temperature with every cycle by 1°C. BamHI cleavage was performed by adding 3 µl Fastdigest BamHI (Fermentas) and incubating for 30 min at 37°C. Samples were mixed with 50 µl TE buffer, 0.5 µl Glycoblue, 10 µl 3 M sodium acetate pH 5.5 and 250 µl 100 % EtOH and incubated for 1 h on dry ice or overnight at –20°C. cDNAs were precipitated by centrifugation for 30 min at 15,000 rpm and 4°C, washed with 500 µl 80 % EtOH, dried for 3 min at room temperature and resuspended in 9 µl H_2O.

For high-throughput sequencing, cDNAs were PCR-amplified by adding 0.3 µl Illumina paired-end primer mix (10 µM each; 5′-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3′; 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′, oligonucleotide sequences © 2006 and 2008 Illumina, Inc. All rights reserved) and 10 µl 2× Immomix (Bioline) and incubation with the following program: 10 min at 95°C, 35 cycles of [10 sec at 95°C, 10 sec at 65°C, 20 sec at 72°C], 3 min at 72°C. In order to desalt the PCR products, a Microspin G-25 column (GE Healthcare) was resupended by vortexing and spinned for 1 min at 735 × g. Upon sample application, PCR products were re-eluted by centrifugation for 2 min at 735 × g, and sequenced on an Illumina GA2 flow cell.

High-throughout sequencing and mapping. High-throughput sequencing of iCLIP cDNA libraries from three replicate experiments was performed on one lane of an Illumina GA2 flow cell with 54 nt run length. Sequence reads included a 2-nt barcode as unique experiment identifier plus a 3-nt random barcode that were introduced during cDNA synthesis. The obtained 6,544,506 sequence reads were separated per experiment based on the 2-nt barcode. In order minimize misassignments due to sequencing errors in the 2-nt barcode, cDNAs from different replicates starting at the same cross-link nucleotide and having the same 3-nt random barcode sequence were assigned to the replicate with the higher occurrence of this random barcode. Thus, replicates were actually separated based on 5-nt information at individual positions. The three expected 2-nt barcodes together represented 96% of all sequences (TG, 2,610,554; TC, 2,292,169; CA, 1,376,258; total, 6,278,981). The three no-antibody control samples were sequenced on Illumina GA2 flow cells with 54 nt run length (replicates 2 and 3 were sequenced together in one lane). The respective 2-nt or 3-nt barcodes as experiment identifiers were CA, ACT and AAG and identified 91,310, 122,957 and 71,044 reads, respectively (out of 5,782,612, 12,597,621 and 12,597,621 reads that were generated in total on the respective lanes).

Before mapping to the human genome, adapter sequences were removed from both ends of the sequence reads. In all hnRNP C and no antibody control experiments, the majority of sequences did not contain 3′ adapter sequences (hnRNP C: replicate 1, 74%; replicate 2, 75%; replicate 3, 84%; control: replicate 1, 91%; replicate 2, 37%; replicate 3, 75%), indicating that the respective inserts were longer than 49 nt.

Mapping of sequence reads was performed against the human genome (version Hg18/NCBI36) using bowtie version 0.10.1[1]. After allowing one mismatch and 10 multiple hits (bowtie parameters –v 1 –m 10 –a), single hits were extracted by post-processing. Genomic annotations were assigned based on gene annotations given by UCSC (hg18.knownGene; 29,413 genes; **Supplementary Fig. 3**).

Randomization of iCLIP cross-link nucleotide positions. As a control for bioinformatic analyses, iCLIP cross-link nucleotide positions were randomized as follows: In order to account for potential differences in transcript abundance, cross-link nucleotides were assigned to transcript regions that are expected to have a common expression level. To this end, exons were separated from introns, non-coding RNA genes within introns from the rest of intronic regions, and untranslated regions from coding sequence based on gene annotations given by UCSC (hg18/NCBI36). Since exons are generally small, all exons of a given gene were concatenated into one region. Randomization was performed within these regions considering cDNA counts, such that e.g. for a position of cDNA count = 2 within an intron, two positions were randomly selected within the same intron during randomization.

Reproducibility analyses. Reproducibility of cross-link nucleotides at single nucleotide resolution (**Fig. 1b**) was determined by counting the number of cross-link nucleotides with a given cDNA count that were present in two or three replicates. If reproducing cross-link nucleotides harbored identical cDNA count values, all except one were excluded from the count. Thereby, the resulting total number of cross-link nucleotides with a given cDNA count reflects equal the total number of positions in the genome that were identified with that cDNA count. The number of cross-link nucleotides of a given cDNA count that were reproduced in at least two or all three replicates is given as a fraction of the total number of cross-link nucleotides with that cDNA count identified within the genome.

In order to determine the offset of reproducing positions (**Supplementary Fig. 2**), cross-link nucleotides of hnRNP C iCLIP replicate 1 were compared against replicate 2. For each cross-link nucleotide in replicate 1, we summarized the offset of all surrounding cross-link nucleotides in replicate 2 up to a distance of 40 nt. Positive or negative offset values indicate whether the reproducing position in replicate 2 locates downstream or upstream of the cross-link nucleotide in replicate 1, respectively. In order to assess the expected background distribution, replicate 1 was also compared against a randomized version of replicate 2. Randomization was performed as described above.

Evaluation of significance of hnRNP C cross-link nucleotides. In order to determine the false-discovery rate (FDR) for each position, we applied a strategy similar to the approach used by Yeo and coworkers[2] performing the following steps: (i) Cross-link nucleotides were assigned to transcript regions as described for randomization above. Both coding and non-coding genes were included (in case of overlapping genes, the cross-link nucleotide was assigned to the shorter gene). Cross-link nucleotides in antisense orientation to the associated gene or locating to non-annotated genomic regions were removed. (ii) Cross-link nucleotides were extended by 15 nt to both directions. Subsequently, we calculated the height at each cross-link nucleotide as the total number of overlapping extended cross-link nucleotides at this position by adding up their cDNA counts. (iii) The distribution of heights was defined as follows: The height h at a cross-link nucleotide position lies within the interval $[1,H]$, where H is the maximum observed height within a given region. n_h and N donate the number of cross-link nucleotides with height h and of total cross-link nucleotides within the same region, respectively. The resulting distribution of heights is $\{n_1, n_2, \ldots n_h, \ldots n_{H-1}, n_H\}$. Thus, the probability of observing a height of at least h is $P_h = \Sigma\ n_i(i = h,...,H)/N$. (iv) The background frequency was computed by 100 iterations of randomization as described above. The modified FDR for a cross-link nucleotide with height h was computed as FDR(h) = (μ_h + σ_h)/P_h, where μ_h and σ_h are the average and standard deviation, respectively, of $P_{h,\mathrm{random}}$ across the 100 iterations. This identified 33,991 cross-link nucleotides as part of significant hnRNP C binding clusters which were referred to as clustered cross-link nucleotides (FDR < 0.05).

Knockdown of hnRNP C. In order to knockdown hnRNP C in HeLa cells, we independently used two different HNRNPC Stealth Select RNAi™ siRNAs (KD1 and KD2 refer to siRNAs HSS179304 and HSS179305 from Invitrogen, respectively) at a final concentration of 5 nM. The siRNAs were transfected using Lipofectamine™ RNAiMAX (Invitrogen) according to the manufacturer's instructions (protocol for forward transfection). Control samples were generated using Stealth RNAi™ siRNA Negative Control (Invitrogen) following the same procedure. Knockdown efficiency was controlled by Western blot analyses using hnRNP C-specific antibodies (**Supplementary Fig. 8**). For microarray analysis, KD1a, KD1b and KD2a were used whereas for RT-PCR analyses KD1c, KD2b and KD2c were used.

Splice-junction microarrays. mRNA from hnRNP C knockdown and control HeLa cells was purified using the RNeasy MinElute Cleanup Kit (Qiagen) combined with the_RiboMinus$^{\mathrm{TM}}$ Eukaryotic Kit for RNAseq (Invitrogen). Labeled sense cDNA for microarray hybridization was prepared using GeneChip® WT Sense Target Labeling and Control Reagents (Affymetrix) according to the manufacturer's instructions, but replacing the included Superscript II with Superscript III (Invitrogen). Labeled samples were hybridized to the non-commercial human exon-junction microarray (HJAY, Affymetrix).

PCR validations. In order to validate the splicing changes identified in our splice-junction microarray analyses, we performed quantitative PCR measurements (**Supplementary Tables 2, 3; Fig. 5b; Supplementary Fig. 9**) using BIOTaq polymerase (Bioline) under the following conditions: 95°C for 5 minutes, 40 cycles of [95°C for 15 seconds, 60°C for 15 seconds, 72°C for 30 seconds], then finally 72°C for 3 minutes. A QIAxcel capillary gel electrophoresis system was used to visualize the PCR products. A photomultiplier detector converted the emission signal into a gel image and an electropherogram that allowed visualization and quantification of each PCR product, respectively. All measurements were performed in three replicates.

ASPIRE3 algorithm. The high-resolution splice-junction microarray was produced by Affymetrix, monitoring 260,488 exon-exon junctions (each with 8 probes) and 315,137 exons (each with 10 probes). cDNA samples were prepared using the GeneChip WT cDNA Synthesis and Amplification Kit (Affymetrix). Analysis of microarray data was done using version 3 of ASPIRE (Analysis of SPlicing Isoform REciprocity). ASPIRE3 predicts splicing changes from reciprocal sets of microarray probes that recognize either inclusion or skipping of an alternative exon. The primary difference in version 3 of ASPIRE software relative to the previous versions is that background detection levels are experimentally determined for each probe, allowing to subtract the background in a probe-specific manner. By analysing the signal of reciprocal probe sets, ASPIRE3 was able to monitor 53,632 alternative splicing events.

The following nomenclature is used:
TA – estimated absolute transcript abundance (arbitrary value)
ΔT – fold change in transcript abundance
ΔT rank – modified t-test to sort the genes based on ΔT significance
I – estimated percentage of exon inclusion
ΔI - estimated change in percentage of exon inclusion
ΔI rank – modified t-test to sort the exons based on ΔI significance

The analysis includes the following basic steps:
1. All probe sets were mapped to human transcripts (positional gene annotations given by Affymetrix) and linked to the *x*/*y* coordinates of the individual probes on the microarray. Detected exons were categorized as constitutive or alternative. For the former, probes were combined into reciprocal groups that detect exon inclusion (E_{in}) or exon skipping (E_{ex}). Constitutive exons were only monitored by E_{in} probes.

2. For each probe, background percentiles were experimentally determined by hybridizing the microarray with labeled 33 nt and 34 nt random oligonucleotides. Background detection probes were grouped according to their GC content, and for

each group the background signal percentiles (5%, 17.5%, 32.5%, 47.5%, 62.5%, 75%, 84%, 91%, and 97%) were calculated. Each probe on the microarray was then assigned to its specific group of background detection probes that shared the same GC content. This allows determination of background values for each probe based on a subset of background detection probes with equal GC content that should detect a similar background signal.

3. Data from CEL files were normalized by background values. To this end, replicate-specific percentile values were calculated for each group of background detection probes and subtracted from the signal values of the respective probes with the same GC content. Resulting values < 0 were set to 0. Finally, values for each experiment were normalized by total signal on the microarray to correct for inter-replicate variations. The resulting values represent the fold-enrichment of signal relative to background.

4. Upon removal of outliers with high variation, signal values were weighted according to their signal intensity and variation. To this end, the probe weight (NUM) was determined by first calculating value X as the quotient of average and standard deviation within each set of reference (1) and experimental (2) samples. X values > 5 were set to 5. Value Y was then determined according to the higher of the two average values: if average <50, Y was set to average/50, if $50 <$ average < 1000, Y was set to 1, and if average > 1000, Y was set to 1000/average. Finally, NUM was calculated as the product of Y and the average of both X values. Probes with $NUM < 1$ were excluded from further analyses.

5. Abundance and change of each transcript cluster were assessed by collecting all probes from probe sets categorized as constitutive within each transcript cluster. If this gained less than 15 non-filtered probe values, also probe sets categorized as alternative were taken into account. For each replicate, weighted average values were calculated for each considered probe ($VAL_1 \ldots VAL_x$, where x stands for the number of considered probes in the transcript cluster) within each transcript cluster and integrated into a value of transcript cluster abundance (TA): $TA = ((VAL_1 \times NUM_1) +...+ (VAL_x \times NUM_x)) / n$, where n is the sum of all respective probe weights (NUM_1

+...+ NUM_x). Then, the probe ratio (R) was determined for each probe as the quotient of median probe values for reference (1) and experimental (2) samples. Finally, the transcript cluster change (ΔT) was calculated as follows: $M(\log_2(R)) = ((\log_2(R_1) \times NUM_1) +...+ (\log_2(R_x) \times NUM_x)) / n$, and $\Delta T = 2^{M(\log2(R))}$.

6. Probe values were normalized relative to the transcript cluster change to account for gene-specific changes in transcription and RNA degradation, allowing to specifically analyze changes in alternative splicing. To this end, all probe values in reference or experimental samples were divided or multiplied, respectively, by the square root of ΔT for the corresponding transcript cluster. Based on the assumption that all probes within a probe set should detect the same transcript isoform and should thus have the same average signal, each probe set value was divided by its own average in all replicates and then multiplied by the average value of all probes within the given probe set over all replicates. This resulted in normalized probe values that were used in all subsequent steps (except for ranking the significance of transcript changes).

7. Exon abundance (A) and percentage of exon inclusion (I) were determined by first calculating a weighted average over all probes within a probe set for each replicate (VAL_1 ...VAL_x): $A = ((VAL_1 \times NUM_1) +...+ (VAL_x \times NUM_x)) / n$. For reciprocal sets of both E_{in} and E_{ex}, the percentage of exon inclusion (I) was calculated as $I = A_{Ein} \times 100 / (A_{Ein} + A_{Eex})$. For all E_{in} probes without reciprocal E_{ex} probes, the replicate with the highest exon abundance was taken as 100% and I was calculated by dividing each exon abundance value by the respective value of this replicate. Finally, changes in exon inclusion (ΔI) were detected by evaluating the difference of the averages over all I values of the two sets of samples.

8. Reciprocal probe set pairs were re-analyzed to rank exons by the predicted splicing change (ΔI rank). To this end, the significance of the difference in average probe values within a probe set was assessed as follows: The weighted average of all probe values in the probe set was determined as $AV = ((VAL_1 \times NUM_1) +...+ (VAL_x \times NUM_x)) / n$, and S calculated as the square root of the sum of squared standard deviations of probes in sample sets 1 and 2. If $4 \times S$ was smaller than a quarter of the

average of *AV* values of sample set 1 and 2, *S* was set to the latter value. Value *Test* was then calculated as the difference of individual averages of sample sets 1 and 2 multiplied by the square root of *N* minus 1 and divided by *S*, where *N* stands for the number of probes with non-filtered values in the probe set (this should be 8 probes in an exon-exon border set and 10 in an exon probe set, if none of the probe values were filtered out). Finally, ΔI rank was calculated as $\Delta I \times Test_{Ein} / 400$, if only E_{in} probes were available for the exon as it is the case for constitutive exons. When E_{in} and E_{ex} probe sets detect the reciprocal signal change, their *Test* values will have opposite signs, therefore subtracting them will rank the exon higher in significance. If the absolute value of $Test_{Ein}$ is smaller than the absolute value of $Test_{Eex}$, ΔI rank was calculated as $\Delta I \times (2 \times Test_{Ein} - Test_{Eex}) / 200$, or as $\Delta I \times (Test_{Ein} - 2 \times Test_{Eex}) / 200$, if the opposite is true, since doubling the value of the probe set with the smaller *Test* value gives a stronger weight to the reciprocity of the change. Exons with ΔI rank > 1 were predicted as enhanced or silenced in the experimental sample set.

9. In order to rank transcripts by the predicted trancript cluster change (ΔT rank), we first normalized all corresponding probe values to their average values over all replicates and within the complete set following the assumption that they detect the same transcript (normalized probe value = probe value × average value of all probes corresponding to this transcript cluster over all replicates / average value of the given probe over all replicates). Then, the two sets of probe values were compared (all probes and all replicates of one experiment within the same transcript cluster). To this end, *AV* and *S* values were calculated as described in 8. and integrated into ΔT rank = $(\log_2 \Delta T \times ((AV_{\mathrm{Sample1}} - AV_{\mathrm{Sample2}}) \times \surd(N\text{-}1) / S) / 20$, where *N* is the number of probes with non-filtered values in the transcript cluster (most transcript clusters contain more than 100 probes).

RNA map. In order to analyze the impact of hnRNP C positioning on splicing regulation, we assessed the positioning of hnRNP C cross-link sites at exon-intron boundaries of alternative exons and flanking constitutive exons (as annotated for the Affymetrix microarray), including 45 nt of exonic and 315 nt of intronic sequence (**Fig. 4a, b**). In addition, 348 nt of exonic and 372 nt of intronic sequence were analyzed at the exon-intron boundaries of constitutive exons (**Fig. 4c**). When introns or exons were shorter than two times the length of the analyzed area, analysis was restricted up to the middle of this intron or exon, respectively. For all RNA maps, regions were divided into non-overlapping windows of 12 nucleotides. For each window, the number of cross-link nucleotides was counted as 1 if at least one cross-link nucleotide resided within this window. Thus, the resulting occurrence value reflects the number of exons with at least one cross-link nucleotide within this window. When positioning of particles was analyzed (**Fig. 4b, c**), only cross-link nucleotides with a spacing of 160 – 170 nt as well as all intervening nucleotides were taken into account. For all RNA maps, percentages were calculated by dividing the number of exons that have at least one cross-link nucleotide within a given window by the total number of exons analyzed at this window.

Supplementary References:

1. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
2. Yeo, G.W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**, 130-137 (2009).