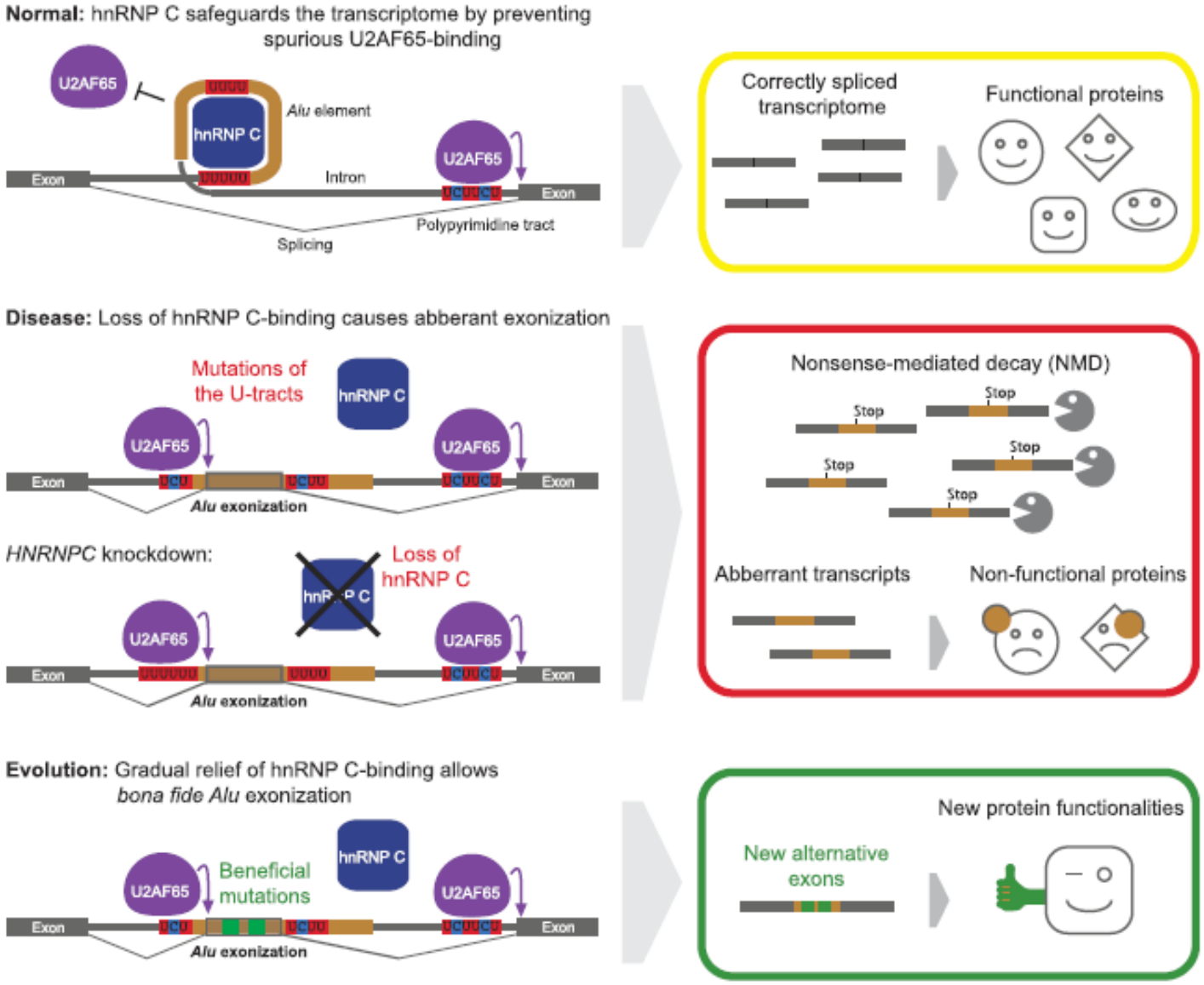
Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements

Kathi Zarnack, Julian König, Mojca Tajnik, Inigo Martincorena, Sebastian Eustermann, Isabelle Stevant, Alejandro Reyes, Simon Anders, Nicholas M. Luscombe and Jernej Ule

Jclub in Bioinformatics

Tarmo Puurand

10.06.2014


Normal: hnRNP C safeguards the transcriptome by preventing spurious U2AF65-binding
U2AF65
hnRNP C
Alu element
Exon
Intron
U2AF65
Exon
Polypyrimidine tract
Splicing
Correctly spliced transcriptome
Functional proteins
Disease: Loss of hnRNP C-binding causes abberant exonization
Mutations of the U-tracts
hnRNP C
U2AF65
U2AF65
Exon
Exon
Alu exonization
HNRNPC knockdown:
Loss of hnRNP C
U2AF65
U2AF65
Exon
Exon
Alu exonization
Nonsense-mediated decay (NMD)
Stop
Stop
Stop
Stop
Stop
Abberrant transcripts
Non-functional proteins
Evolution: Gradual relief of hnRNP C-binding allows *bona fide Alu* exonization
hnRNP C
U2AF65
Beneficial mutations
U2AF65
Exon
Exon
Alu exonization
New protein functionalities
New alternative exons

Results

- hnRNP C and U2AF65 Bind at 3’ Splice Sites
- hnRNP C Competes with U2AF65 Binding
- hnRNP C Blocks U2AF65 from Continuous U-Tracts In Vitro
- The hnRNP C-U2AF65 Competition Leads to Exon Repression
- hnRNP C Prevents the Aberrant Exonization of Alu Elements
- The U-Tracts Facilitate Strong hnRNP C Binding to Alu Elements
- The Competition between hnRNP C and U2AF65 Controls Alu Exonization
- Mutations Disrupting hnRNP-C-Dependent Repression of Alu Exons Can Cause Disease

Nextgen sequencing

hnRNP C iCLIP	Replicate 1 (LUjh03[1])	Replicate 2 (LUjh25)	Total
Total reads	17,016,384	20,525,828	**37,542,212**
Total crosslink events	4,869,687	9,488,133	**14,357,820**
Total crosslink sites	4,412,114	8,131,878	**11,899,326**
Binding sites (Flank 10; FDR 0.05)	ND	ND	**438,360**
Binding sites in annotated genes	ND	ND	**392,818**

U2AF65 iCLIP (Ctrl)	Replicate 1 (LUjh23_1)	Replicate 2 (LUjh23_2)	Total
Total reads	2,478,269	15,555,645	**18,033,914**
Total crosslink events	1,781,117	10,135,903	**11,917,020**
Total crosslink sites	1,488,737	7,209,165	**8,214,397**
Binding sites (Flank 5; FDR 0.05)	ND	ND	**518,794**

U2AF65 iCLIP (KD1)	Replicate 1 (LUjh21_1)	Replicate 2 (LUjh21_2)	Total
Total reads	4,378,005	6,228,475	**10,606,480**
Total crosslink events	3,077,838	4,384,385	**7,462,223**
Total crosslink sites	2,626,609	3,596,131	**5,778,770**
Binding sites (Flank 5; FDR 0.05)	ND	ND	**330,674**

U2AF65 iCLIP (KD2)	Replicate 1 (LUjh22_1)		Total U2AF65
Total reads	11,330,534		**39,970,928**
Total crosslink events	8,227,587		**27,606,830**
Total crosslink sites	6,254,382		**16,994,031**
Binding sites (Flank 5; FDR 0.05)	363,844		**1,122,142**
Binding sites in annotated genes	ND		**996,784**

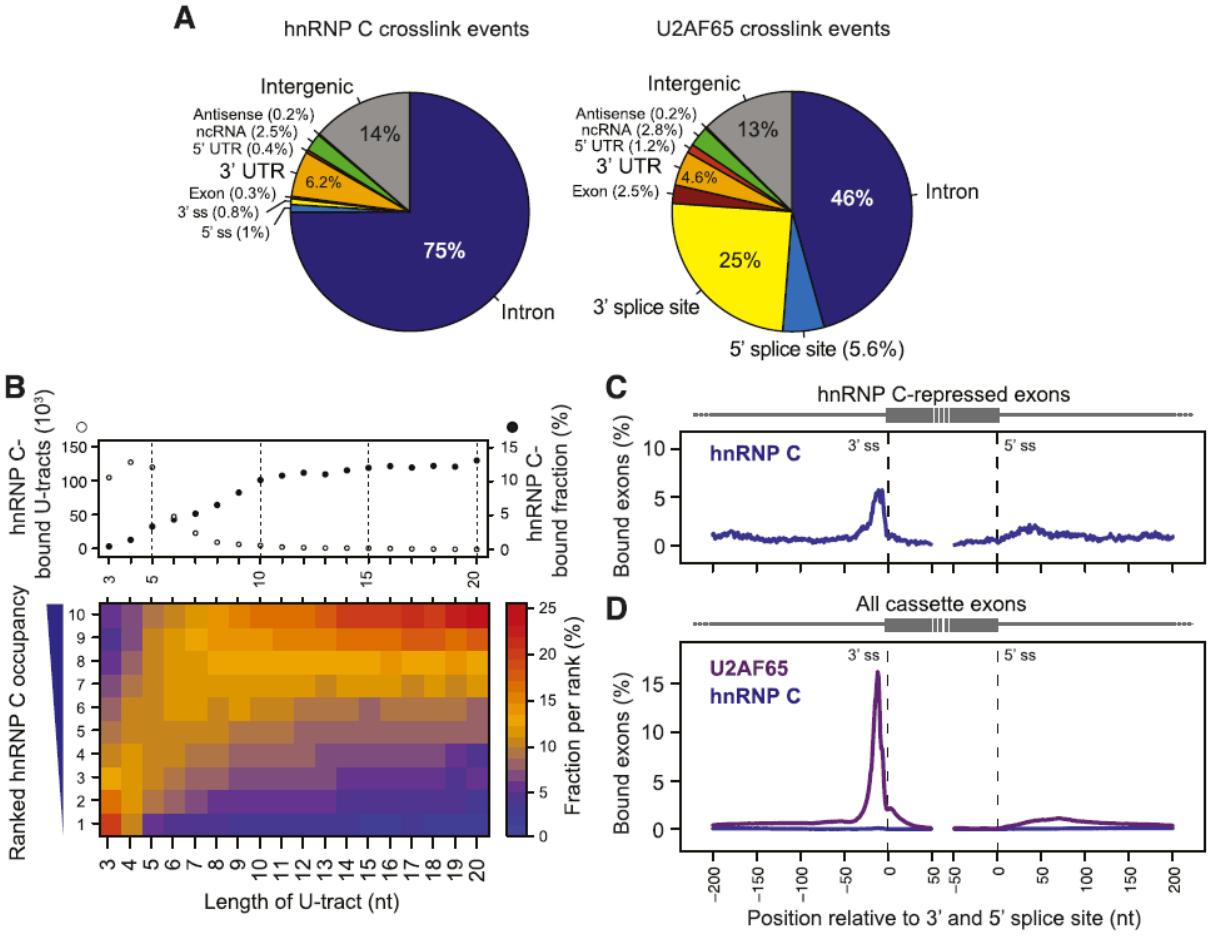
Nextgen sequencing

RNA-seq (Ctrl)	Replicate 1	Replicate 2	Total
Total reads	120,269,114	127,222,962	**247,492,076**
Mapped	102,679,885	101,888,778	**204,568,663**
Mapped (% of total reads)	85.4	80.1	**82.7**
Uniquely mapped	99,206,481	98,420,834	**197,627,315**
Uniquely mapped (% of mapped)	96.6	96.6	**96.6**
Junction-spanning	20,687,792	20,210,793	**40,898,585**

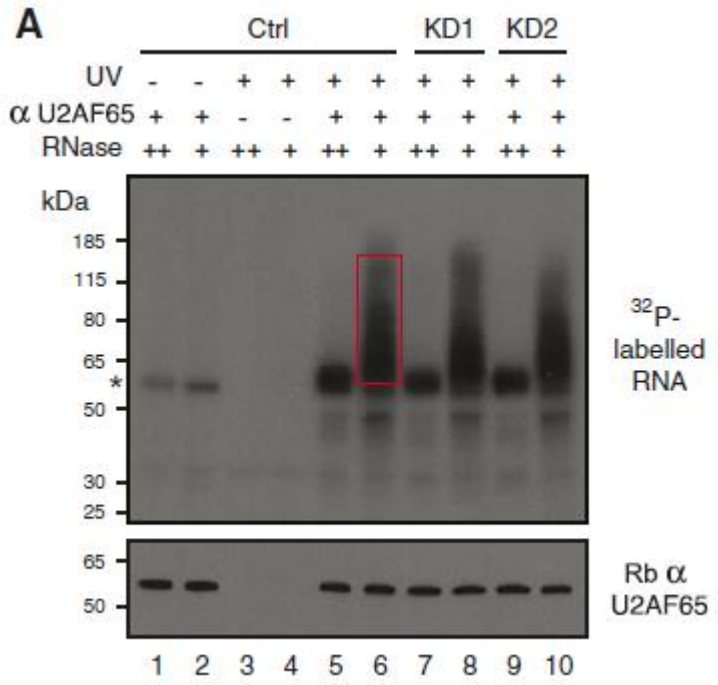
RNA-seq (KD1)	Replicate 1	Replicate 2	Total
Total reads	59,483,704	65,331,134	**124,814,838**
Mapped	51,049,794	55,786,169	**106,835,963**
Mapped (% of total reads)	85.8	85.4	**85.6**
Uniquely mapped	49,387,650	53,933,424	**103,321,074**
Uniquely mapped (% of mapped)	96.7	96.7	**96.7**
Junction-spanning	9,046,955	9,221,022	**18,267,977**

RNA-seq (KD2)	Replicate 1	Replicate 2	Total
Total reads	63,752,362	64,691,484	**128,443,846**
Mapped	55,201,140	54,369,665	**109,570,805**
Mapped (% of total reads)	86.6	84	**85.3**
Uniquely mapped	53,281,319	52,506,281	**105,787,600**
Uniquely mapped (% of mapped)	96.5	96.6	**96.5**
Junction-spanning	9,755,630	10,275,003	**20,030,633**

hnRNP C and U2AF65 Bind at 3’ Splice Sites

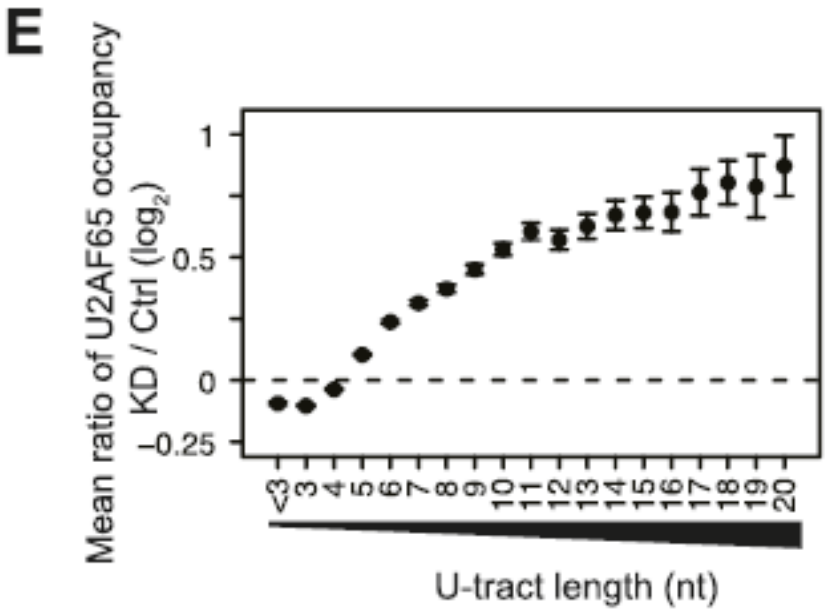


hnRNP C Competes with U2AF65 Binding

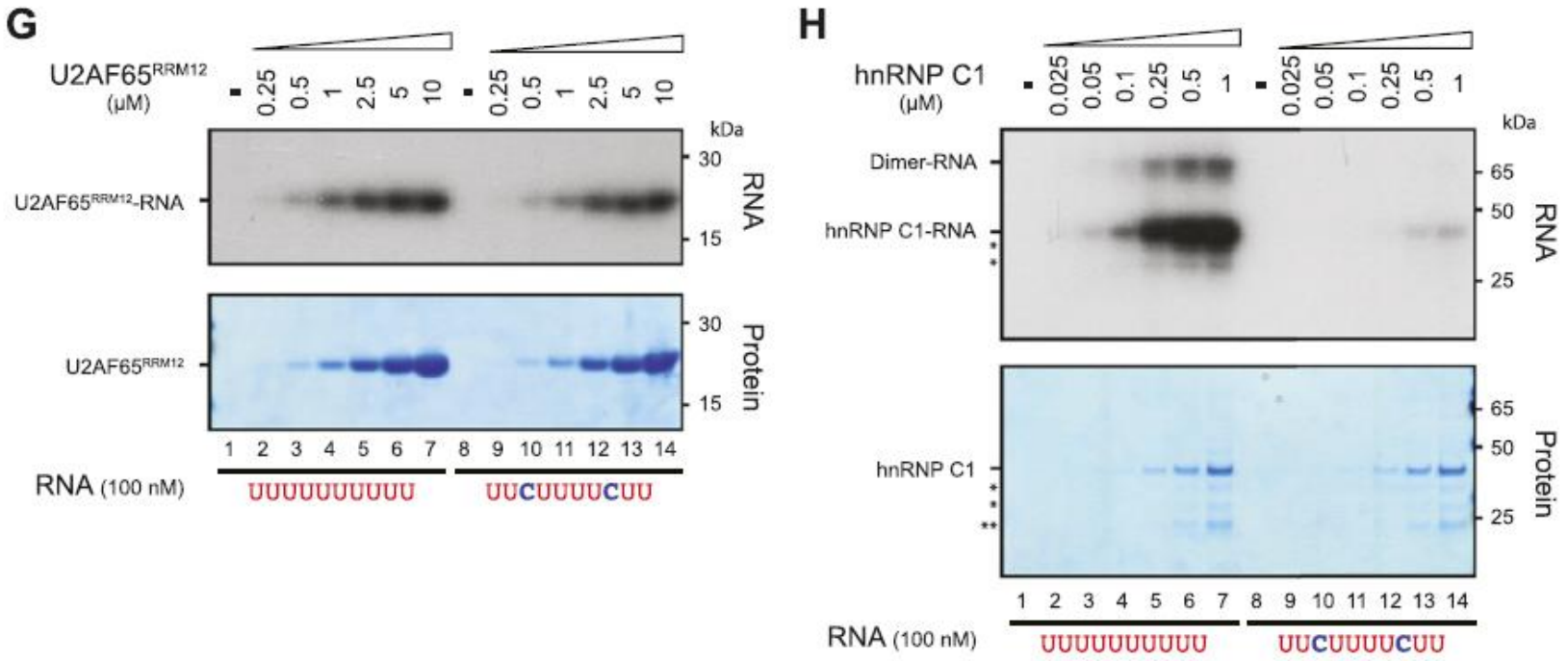


(A) The total amount of crosslinking of U2AF65 to RNA is not altered in the absence of hnRNP C. Analysis of crosslinked U2AF65-RNA complexes using denaturing gel electrophoresis and membrane transfer. Protein extracts were prepared from UV-crosslinked (UV+) control HeLa cells (Ctrl) and *HNRNPC* knockdown cells (KD1 and KD2), and RNA was partially digested using low (+) or high (++) concentrations of RNase. U2AF65-RNA complexes were immuno-purified with a mouse antibody against U2AF65 (α U2AF65). To allow visualization of the protein-RNA complexes, the 5′ ends of the RNAs were radioactively labeled. The complexes were size-separated using denaturing gel electrophoresis and transferred to a nitrocellulose membrane. The upper panel shows the autoradiograph of this membrane. U2AF65-RNA complexes are shifted upward from the size of the protein (53 kilo Dalton, kDa; lane 6: the red box indicates the region that was extracted for subsequent analyses). This shift is focused when high RNase concentrations are used (lane 5). A similar pattern with comparable intensity is observed for the *HNRNPC* knockdown cells (lanes 7-10), indicating that crosslinking is not generally affected. As a control, no signal is observed in experiments where the antibody was omitted during immunoprecipitation (lanes 3 and 4). Importantly, when omitting UV irradiation, no shifted U2AF65-RNA complexes can be observed. The remaining radioactive signal at the size of U2AF65 (marked by *) in these samples indicates that part of the protein is labeled under the used conditions. The lower panel shows the Western blot analysis of the same immunoprecipitations with a rabbit antibody against U2AF65 (Rb α U2AF65).

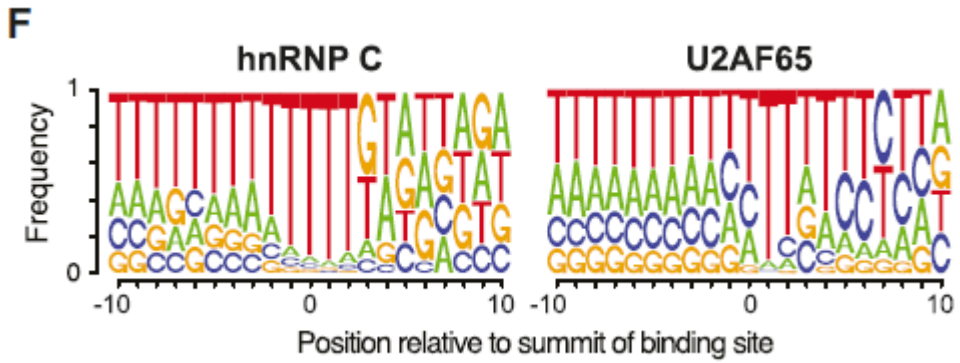
hnRNP C Competes with U2AF65 Binding



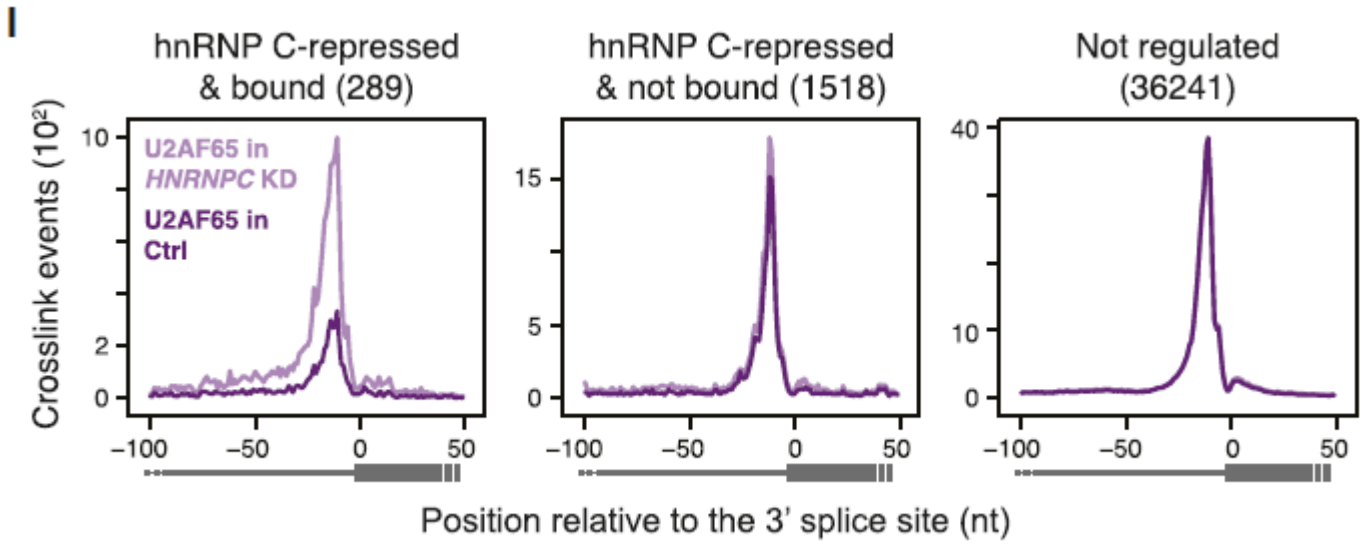
(E) U2AF65-binding sites that overlap with a long U-tract show increased U2AF65 occupancies in the HNRNPC knockdown. The average ratio of U2AF65 occupancy from HNRNPC knockdown (KD) over control HeLa cells (Ctrl) is shown for U2AF65-binding sites that overlap with U-tracts of varying lengths. Error bars indicated the 95% confidence interval of the mean.



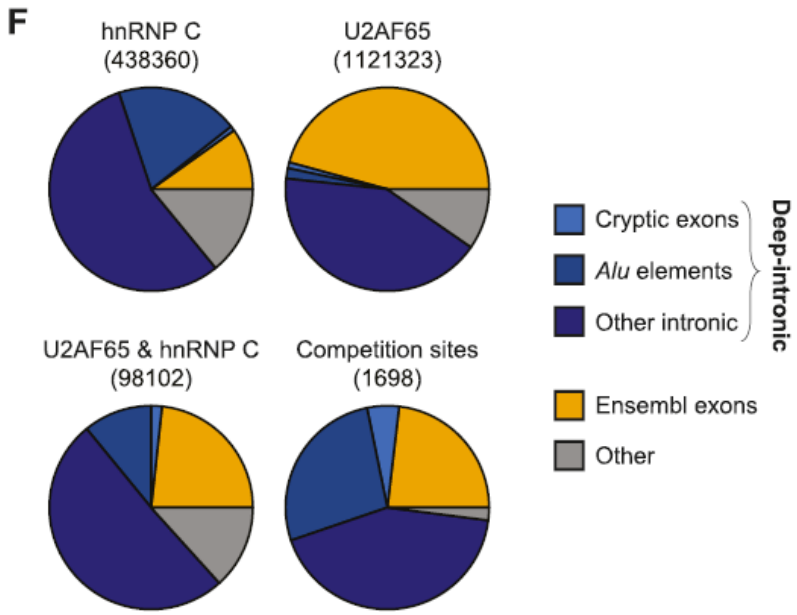
(G) Recombinant U2AF65RRM12 shows comparable crosslinking to the wild-type (U_{10}) and mutant ($U_2CU_4CU_2$) RNA oligonucleotides that resemble the upstream U-tract of the Alu exon in the NUP133 minigene (Figure S6). Increasing concentrations of U2AF65RRM12 (21 kDa, concentration indicated above in mM) were incubated with radioactively labeled wild-type or mutant RNA oligonucleotide (100 nM), UV crosslinked and analyzed by denaturing gel electrophoresis. The radioactive signal from the RNA crosslinked to the protein can be observed in the autoradiograph (U2AF65RRM12-RNA, top panel). Coomassie staining of the same gel serves as loading control (lower panel).
(H) Recombinant hnRNP C1 crosslinking is drastically reduced to the mutant RNA oligonucleotide. Experimental setup as in (G) but using increasing concentrations of hnRNP C1 (33 kDa, concentration indicated above). Note that in addition to the signal derived from hnRNP C1 crosslinked to RNA (hnRNP C1-RNA), an additional crosslinking signal is visible at about 65 kDa. This signal is likely derived from two hnRNP C1 proteins crosslinked to one RNA molecule and therefore labeled as dimer-RNA. Impurities are indicated by asterisks on the left (* C-terminal truncations of hnRNP C1; ** GST).



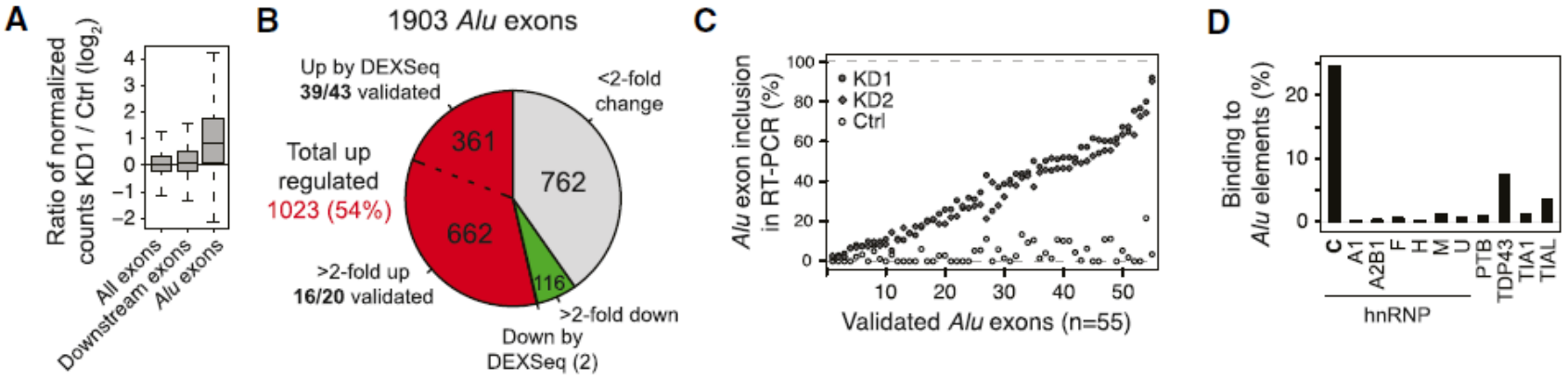
(F) Weblogos showing the relative nucleotide frequency around the summits (position 0) of hnRNP-C- and U2AF65-binding sites.



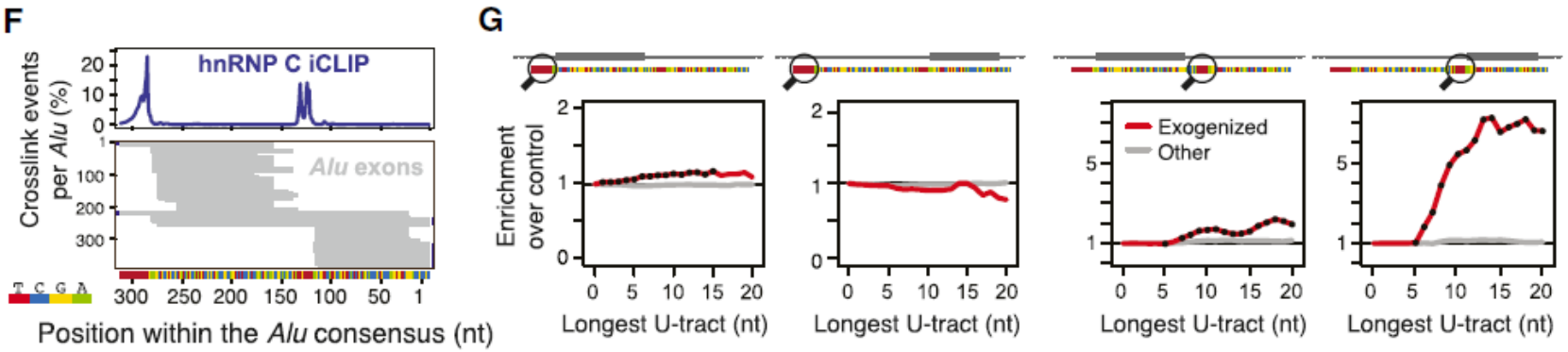
(I) RNA maps showing the total number of crosslink events of U2AF65 in control HeLa (light purple) and HNRNPC knockdown cells (dark purple) relative to the 30 splice sites of all exons that (i) are repressed and bound by hnRNP C (left), (ii) are repressed but not bound by hnRNP C (middle), and (iii) are not subject to any regulation in the HNRNPC knockdown (fold change < 1.1; right). The number of exons in each category is indicated above.



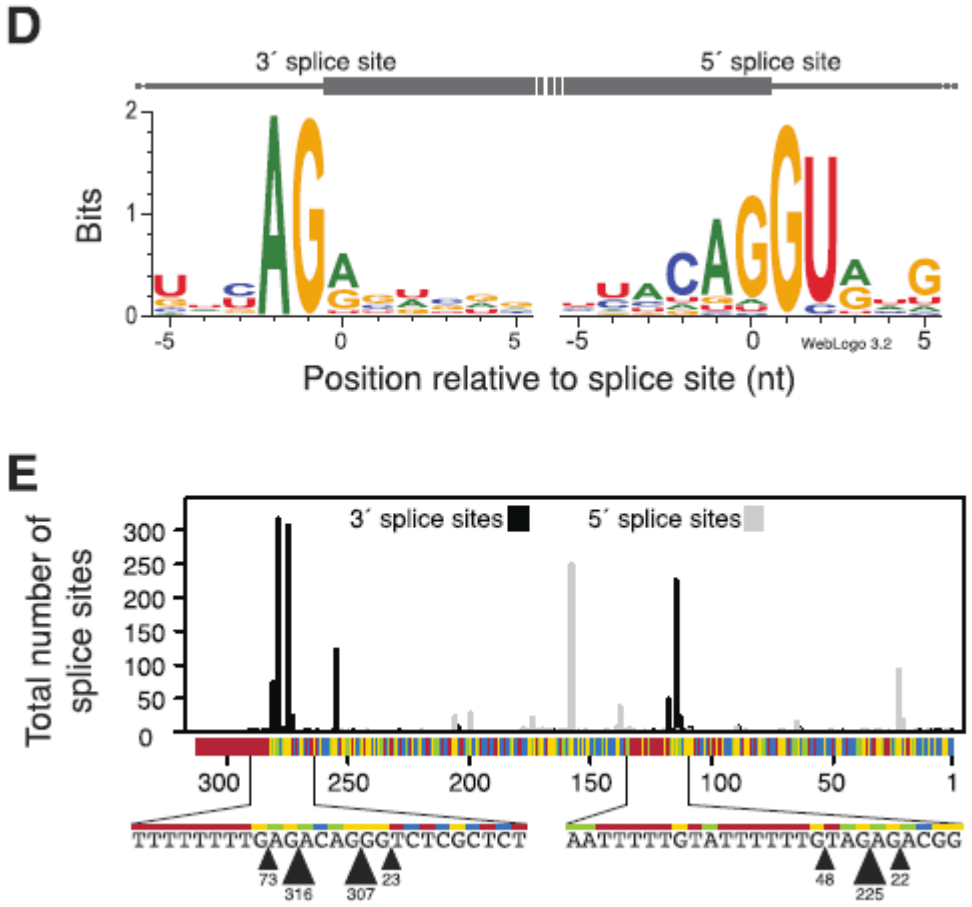
(F) Sites of hnRNP C-U2AF65 competition are strongly enriched at deep-intronic locations (p value < 0.001 compared with all other U2AF65-binding sites, hypergeometric test), in particular at cryptic exons and within Alu elements. Pie charts showing the fraction of binding sites located at annotated Ensembl exons as well as at deep-intronic positions (subdivided into positions at cryptic exons, within Alu elements and other). Graphs are shown from left to right for: all hnRNP C, all U2AF65 binding sites, U2AF65 binding sites that overlap with hnRNP C binding, and U2AF65 binding sites that show an at least 4-fold increase in occupancy in the HNRNPC knockdown and overlap with hnRNP C (competition sites).



(A) Box plots summarizing the change in normalized expression of Alu exons compared to downstream non-Alu exons as well as all exons.
(B) Pie chart summarizing the regulation of all 1,903 Alu exons detect from our RNA-seq data. Upregulated and downregulated exons are further subdivided into
those called by DEXSeq or displaying a more than 2-fold change in the HNRNPC knockdown.

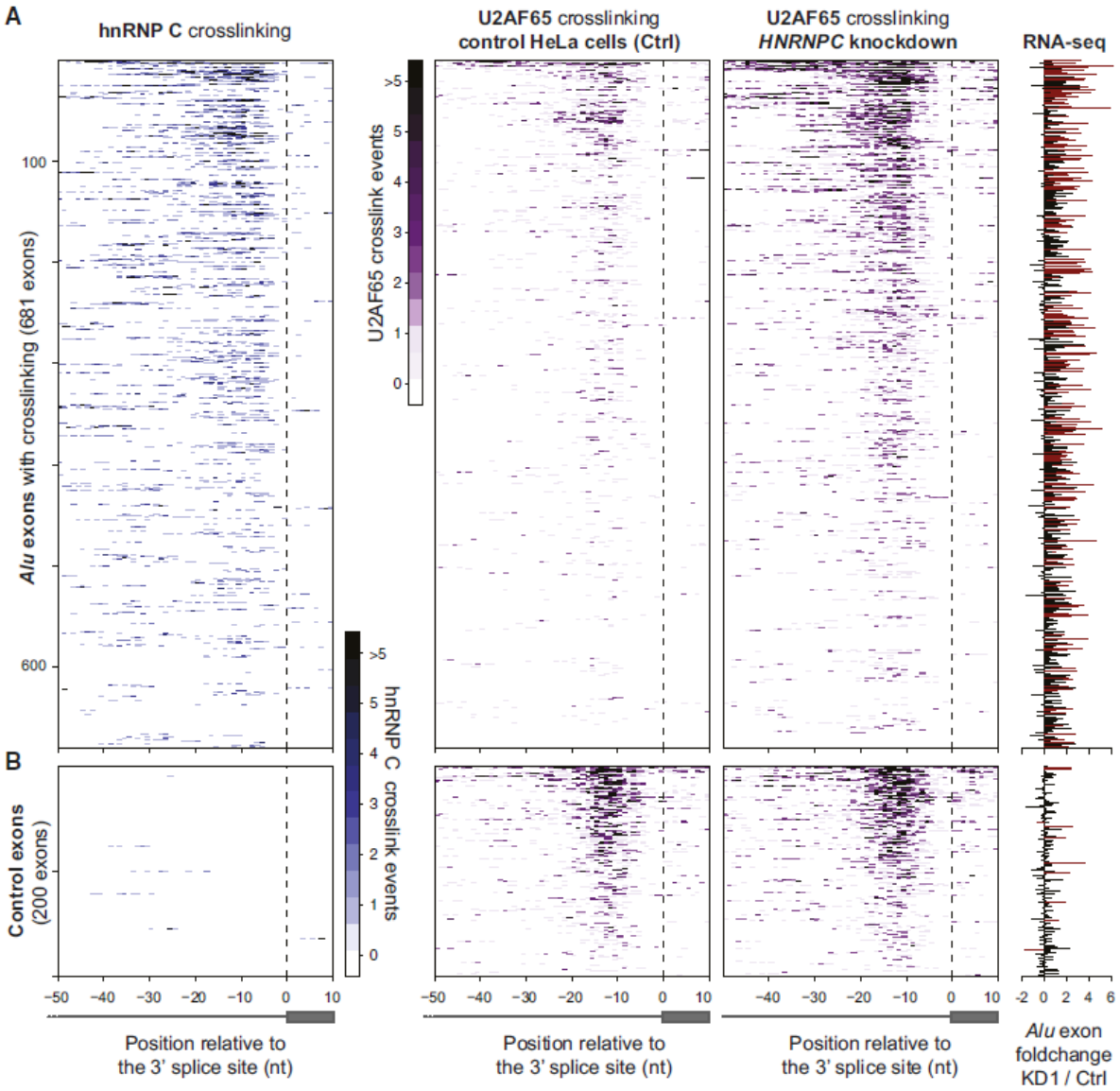


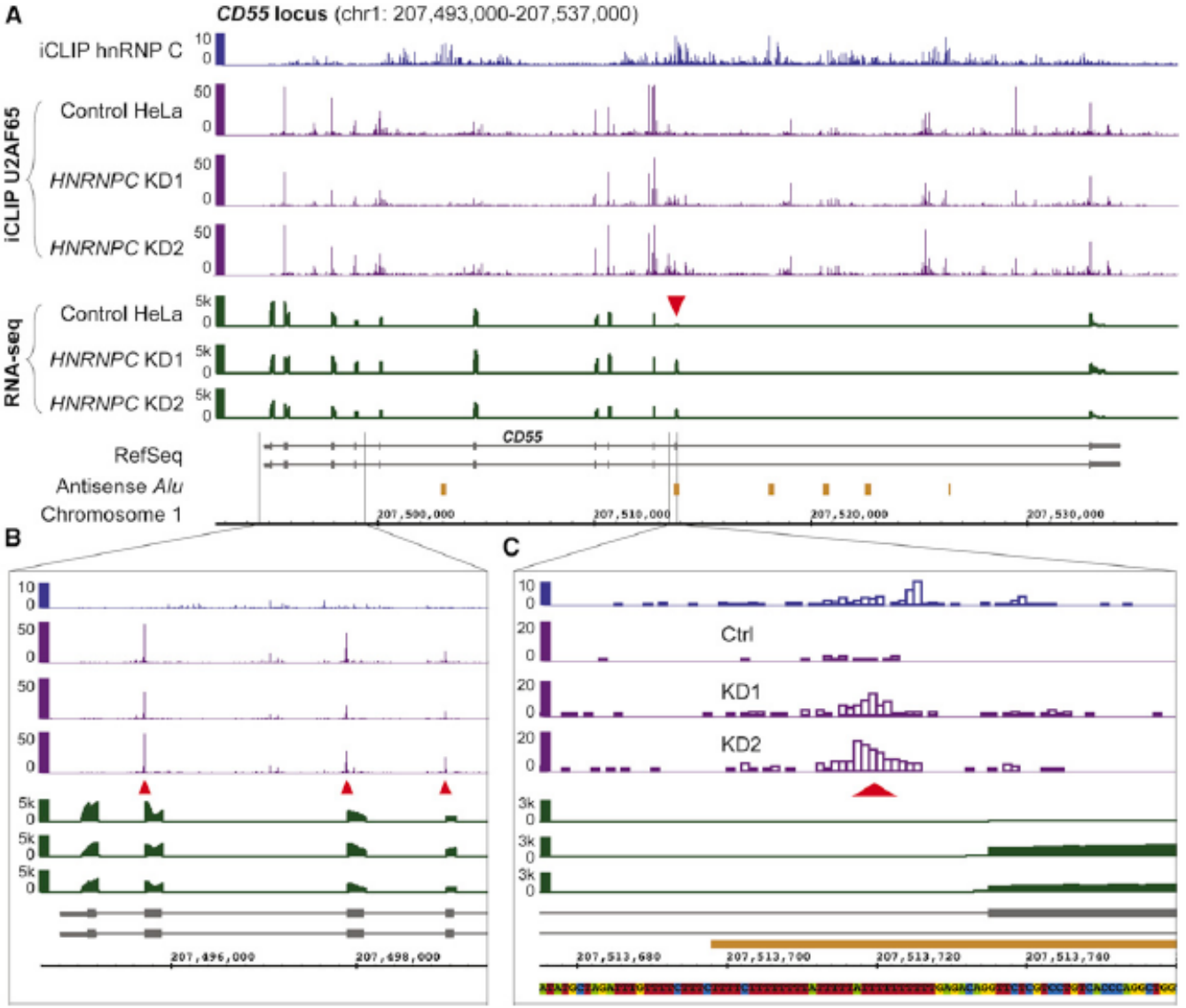
(F) Schematic representations of hnRNP C crosslink events per nucleotide (top) and of Alu exon locations (bottom) along the Alu consensus sequence. Exons that extend beyond the Alu element end with a blue dash.
(G) Plots depicting the ratio of the cumulative frequencies of U-tracts of a given length (e.g., at least five uridines) in exonized Alu elements (red line) compared to nonexonized Alu elements within the same genes. Analyses are separately shown for Alu exons from the first or second arm of the Alu element (gray rectangle above) as well as for the upstream and linker U-tracts (magnifier icon). Nonexonized Alu elements of all other genes (gray line) serve as control. Black dots, p value < 0.05 (Pearson's chi-square test).



(D) The Alu exons arise from canonical splice sites. Weblogos indicating the consensus sequence at the 30 and 50 splice sites of the Alu exons.
(E) Bar diagram summarizing the usage of different 30 (black) and 50 (gray) splice sites in the Alu consensus sequence (color coding as for enlarged sequence below). The region around the most commonly used 30 splice sites is enlarged below (black arrowheads; the number of exons using each splice site is given below the arrowhead).

The Competition of hnRNP C with U2AF65 at 30 Splice Sites Represses Alu Exon Inclusion



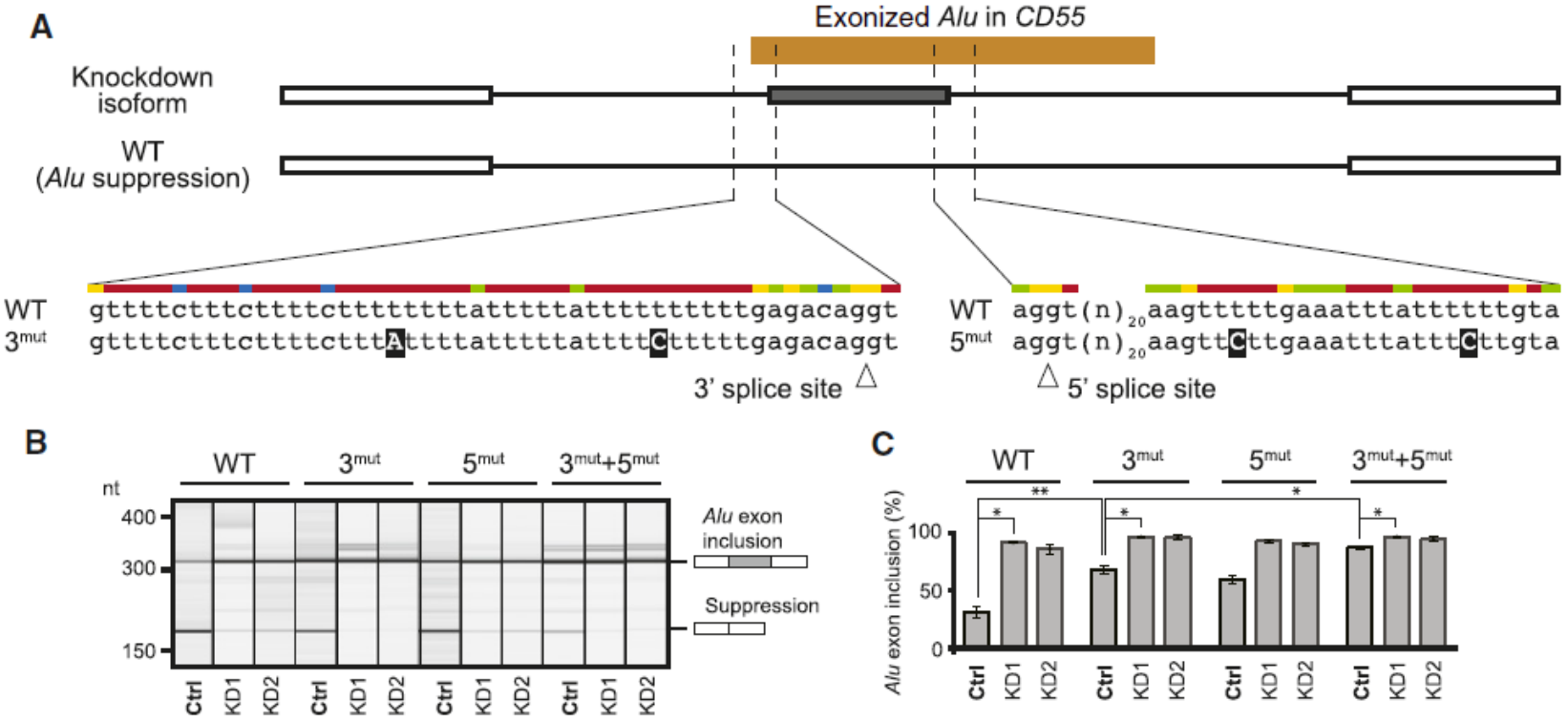


Examples of U2AF65 and hnRNPCBinding at the 3' Splice Sites of Constitutive or hnRNP-C-Repressed Exons within the CD55 Gene

(A) Genome browser view of the CD55 gene displaying the iCLIP data (crosslink events per nucleotide) of hnRNP C (blue) and U2AF65 (purple) as well as the RNAseq data (overlapping reads per nucleotide; green) from control and HNRNPC knockdown HeLa cells. The red arrowhead marks the hnRNP-C-repressed alternative Alu exon. RefSeq transcript annotations (gray) and Alu elements in antisense orientation to the shown strand (orange) are depicted below.

(B) Enlargement of the genomic region containing the 50 UTR and the first four exons. Red arrowheads mark U2AF65 peaks at 30 splice sites.

(C) Enlargement of the region around the 3' splice site of the hnRNP-C-repressed Alu exon (marked in A) including the underlying genomic sequence. The red arrowhead marks the site of increased U2AF65 occupancy in the HNRNPC knockdown



(A) Schematic overview of the minigene including the Alu exon (gray square), intronic regions (black lines), and two flanking exons (white squares) from the CD55 gene. The original sequence (WT) as well as the mutated sequence surrounding the 3' and 5' splice sites (3mut and 5mut, respectively; splice sites marked by arrowheads) are depicted below. Introduced point mutations are highlighted in black.
(B) RT-PCR monitoring inclusion or suppression of the Alu exon in the minigenes with wild-type (WT) or mutated sequences (3mut, 5mut) in HNRNPC knockdown (KD1 and KD2) and control HeLa cells (Ctrl). The corresponding capillary electrophoresis data is given in a gel-like representation with Alu exon inclusion and suppression indicated schematically on the right.
(C) Average Alu exon inclusion in percent from three replicate RT-PCR experiments. Lines indicate relevant comparisons with asterisks representing different levels of significance (*p value < 0.05; **p < 103; Student's t test). Error bars represent SDM.

hnRNP C Repression of Alu Exonization in the PTS Gene Is Relevant for Disease