SUPPLEMENTAL INFORMATION

Coordinate regulation of alternative pre-mRNA splicing events by the human RNA chaperone proteins hnRNPA1 and DDX5

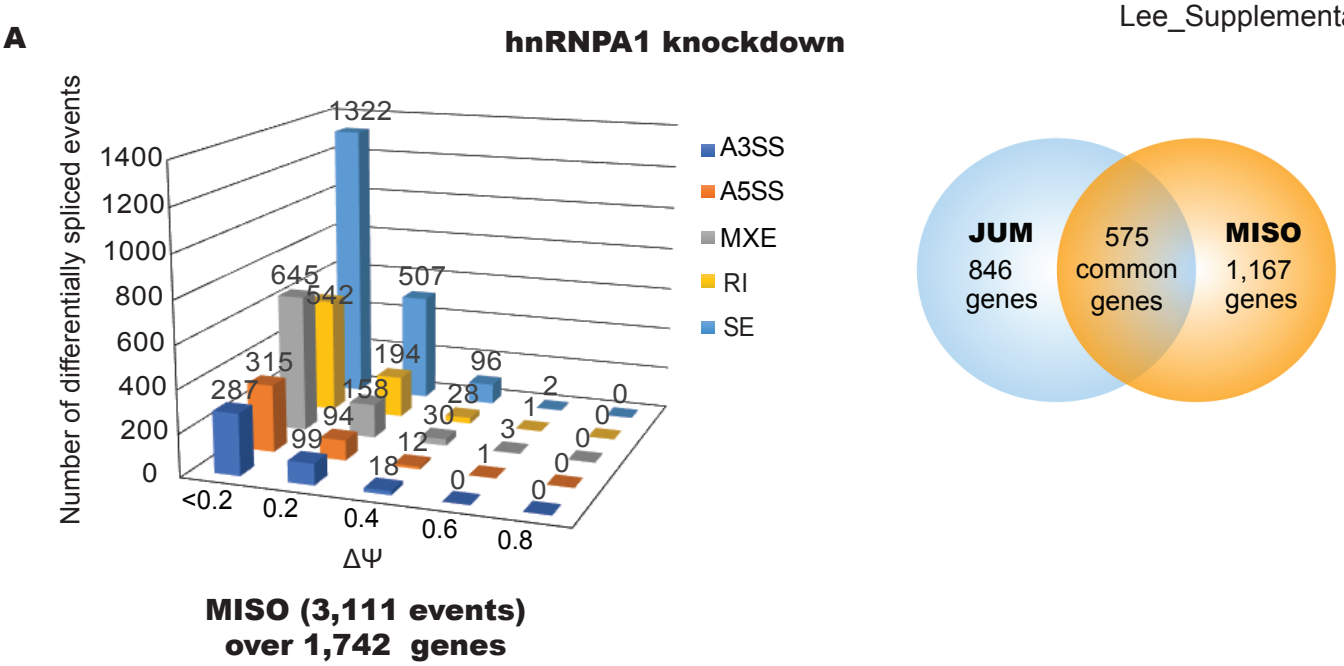
Yeon J. Lee[1,2,3], Qingqing Wang[1,2,3], Donald C. Rio[1,2,3,] ✚

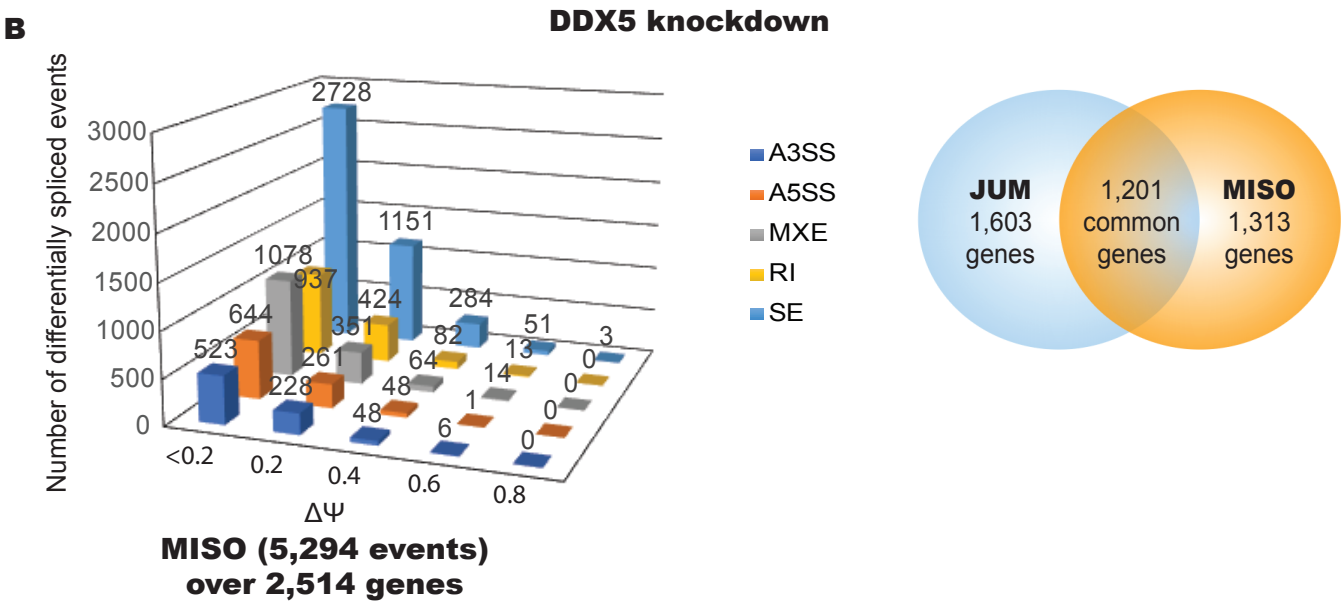
[1] Department of Molecular and Cell Biology, University of California, Berkeley, California, 94720, USA

[2] Center for RNA Systems Biology, University of California, Berkeley, California, 94720, USA

[3] California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, California, 94720, USA

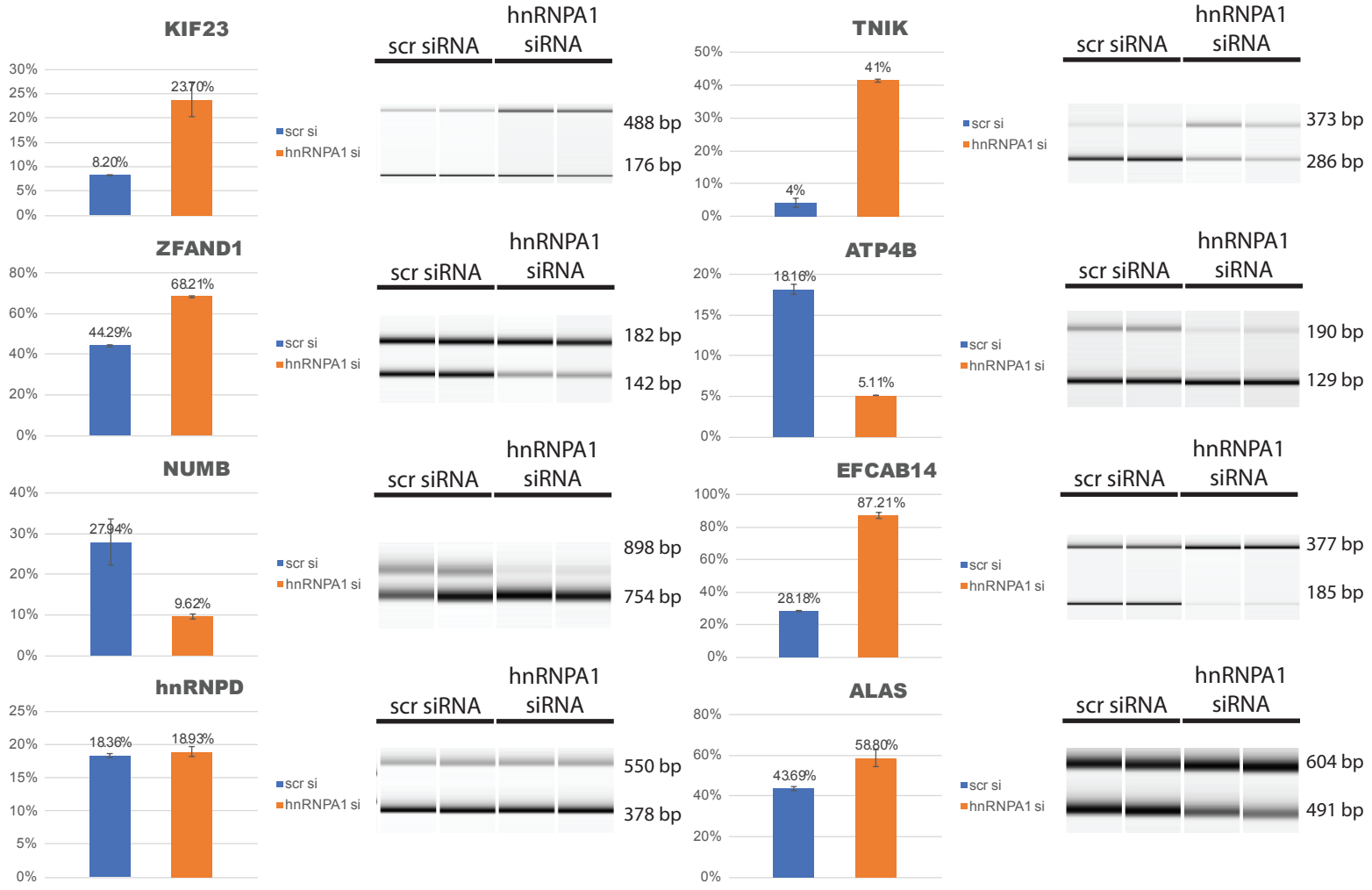
Content





Supplemental Figure S1. Quantitation of differential alternative splicing events controlled by hnRNPA1 and DDX5 using MISO. (S1A) Left panel: MISO splicing analysis revealed 3,111 differential splicing events over 1,742 genes upon hnRNPA1 knockdown. Pairwise comparison of differential splicing levels was computed between nonspecific control scr siRNA and hnRNPA1 siRNA knockdown samples (Bayes factor = 10). Events were filtered based on the significance and order of magnitude of expression change (ΔΨ). Right panel: The overlap in total number of splicing changes upon hnRNPA1 knockdown between MISO and JUM splicing targets are shown in a Venn diagram. (S1B) Left panel: Upon DDX5 knockdown, a total of 5,294 statistically significant differential splicing events over 2,514 genes were detected using MISO. Right panel: The overlap in total number of splicing changes upon DDX5 knockdown between MISO and JUM splicing targets are shown in a Venn diagram.

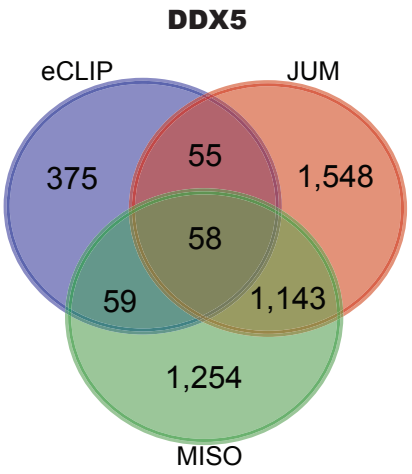
Note: We also detected splicing targets that are unique to either JUM or MISO. There are several reasons for this finding. There are differences in the statistical methods and algorithms underlying JUM and MISO that account for some of the differences in calling splicing event changes. Even for initial mapping of the short RNA-seq reads to the human genome, different tools were used (Tophat for MISO; STAR for JUM, see methods). Most importantly, MISO requires the use of a pre-annotated library of splice junctions, so splicing events existing in the RNA-seq data, but missing from the splice junction library, will be overlooked during the analysis. JUM does not require a previously annotated library of splice junctions and so JUM can identify novel and more complex splicing events, which are put into the “composite” category. JUM also only utilizes splicing junctions that are detected in all experimental replicates and requires that splice junction reads appear in each and every experimental replicate (with 5 junction reads per replicate), resulting in higher confidence detection of differential splice junction use.



NOTE: ΔPSI value is calculated as control - knockdown											
	MISO (hg19)	start	end	strand	ΔPSI	**JUM (hg38)**	start	end	strand	ΔPSI	**RT-PCR ΔPSI**
hnRNPD	chr4	83276554	83278048	-	-0.04	chr4	82355401	82356794	-	-0.04	-0.01
ALAS2	chrX	55047485	55051273	-	-0.14	chrX	55021274	55024716	-	-0.06	-0.15
KIF23	chr15	69732647	69737380	+	-0.34	chr15	69440487	69444788	+	-0.32	-0.16
ZFAND1	chr8	82627222	82630459	-	-0.33	chr8	81715114	81718180	-	-0.28	-0.24
NFE2L1	chr17	46133748	46134864	+	0.30	chr17	48056598	48057342	+	0.37	0.18
NUMB	chr14	73741918	73754022	-	0.38	chr14	73277293	73282357	-	0.39	0.18
TIMM17B	chrX	48752321	48754141	-	0.25	chrX	48895101	48896757	-	0.38	0.00
TNIK	chr3	170855980	170858298	-	-0.36	chr3	171138379	171140397	-	-0.40	-0.37
ATG4B	chr2	242590679	242593025	+	0.40	chr2	241651335	241653510	+	0.41	0.13
EFCAB14	Not found					chr1	46686870	46689585	-	-0.29	-0.59
ST6GALNAC1	chr17	74622088	74622871	-	0.28	chr17	76626095	76626649	-	0.28	0.27
FN1	chr2	216243853	216247048	-	0.17	chr2	215379317	215382210	-	0.16	0.15
CD97	chr17	14501736	14508616	+	0.12	Not found					0.08

Supplemental Figure S2. RT-PCR validation of several pre-mRNA splicing target cassette exon events confirms splicing changes in hnRNPA1-dependent, alternative splicing events detected in RNA-seq data analysis. K562 cells were electroporated with nonspecific control siRNAs (scr si) or hnRNPA1 duplex siRNA duplexes. After a second round of siRNA transfection, the cells were harvested for total RNA isolation and subjected to RT-PCR analysis. Replicate RT-PCR reactions for cassette exons were separated on Agilent Bioanalyzer chips and ratios of the skipped or included exons were calculated (plotted at the left of the electrophoresis traces). Cassette exons from the KIF23, ZFAND1, NUMB, hnRNPD, TNIK, ATP4B, EFCAB14 and ALAS genes are shown as examples. Table at the bottom shows coordinates of splice junctions tested and the PSI/ values determined from the MISO and JUM analyses.

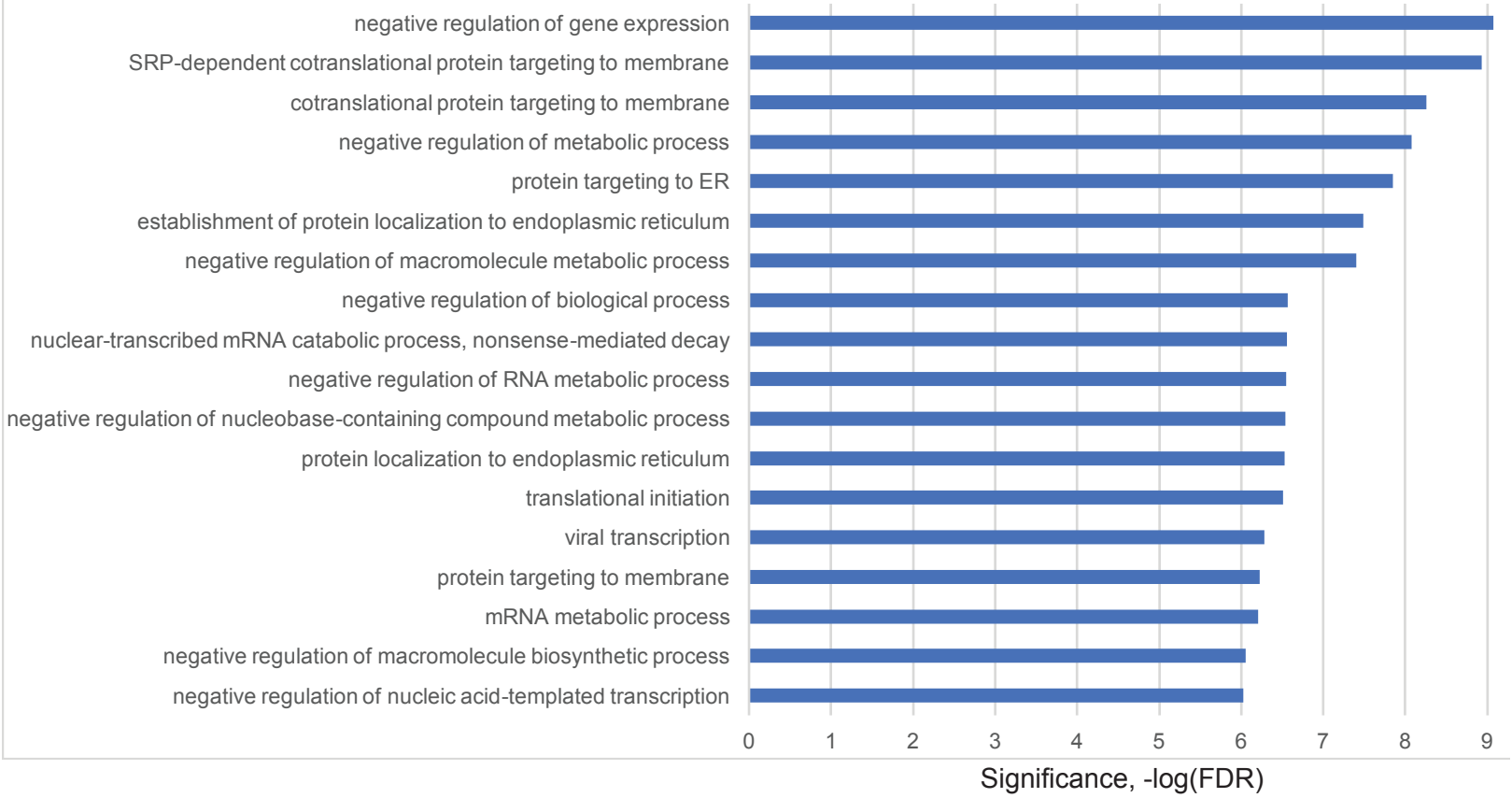
A



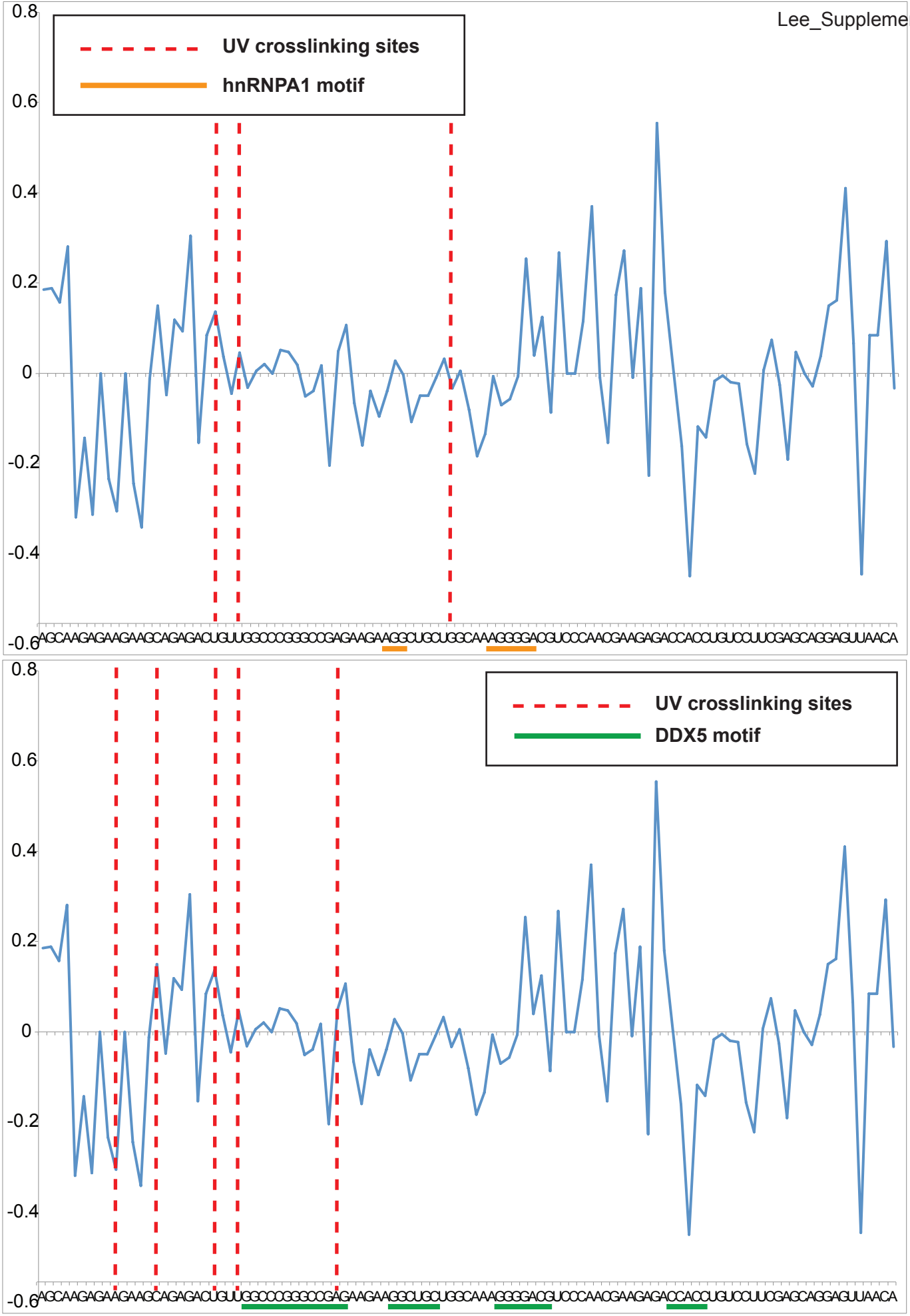
B

	total	elements
DDX5 differential splicing target with eCLIP tags	58	RPL7A, PSMG4, EIF1, PTP4A2, GSTP1, HBG2, RPL4, RPS21, PHB2, GABRE, WSB1, VEGFA, GTPBP2, HNRNPH1, SNHG12, HNRNPU, UQCRH, RPL10A, CDC42, DKC1, CREBZF, SRSF4, PTBP1, PTMA, THAP7, RPS12, VPS28, RPL13, SOCS2, EIF4A2, CIRBP, POLR2C, UBE2J2, HMGN1, PTMS, PRSS21, RPL10, TPT1, CALM2, BUB3, HNRNPD, NARF, ZNF706, CFL1, HSP90AB1, RHOT2, RPL8, SNHG1, TPTEP1, BEX4, HNRNPH3, RAN, ATP5SL, DDIT3, SRSF2, FAM160B2, LARP4B, HMGA1

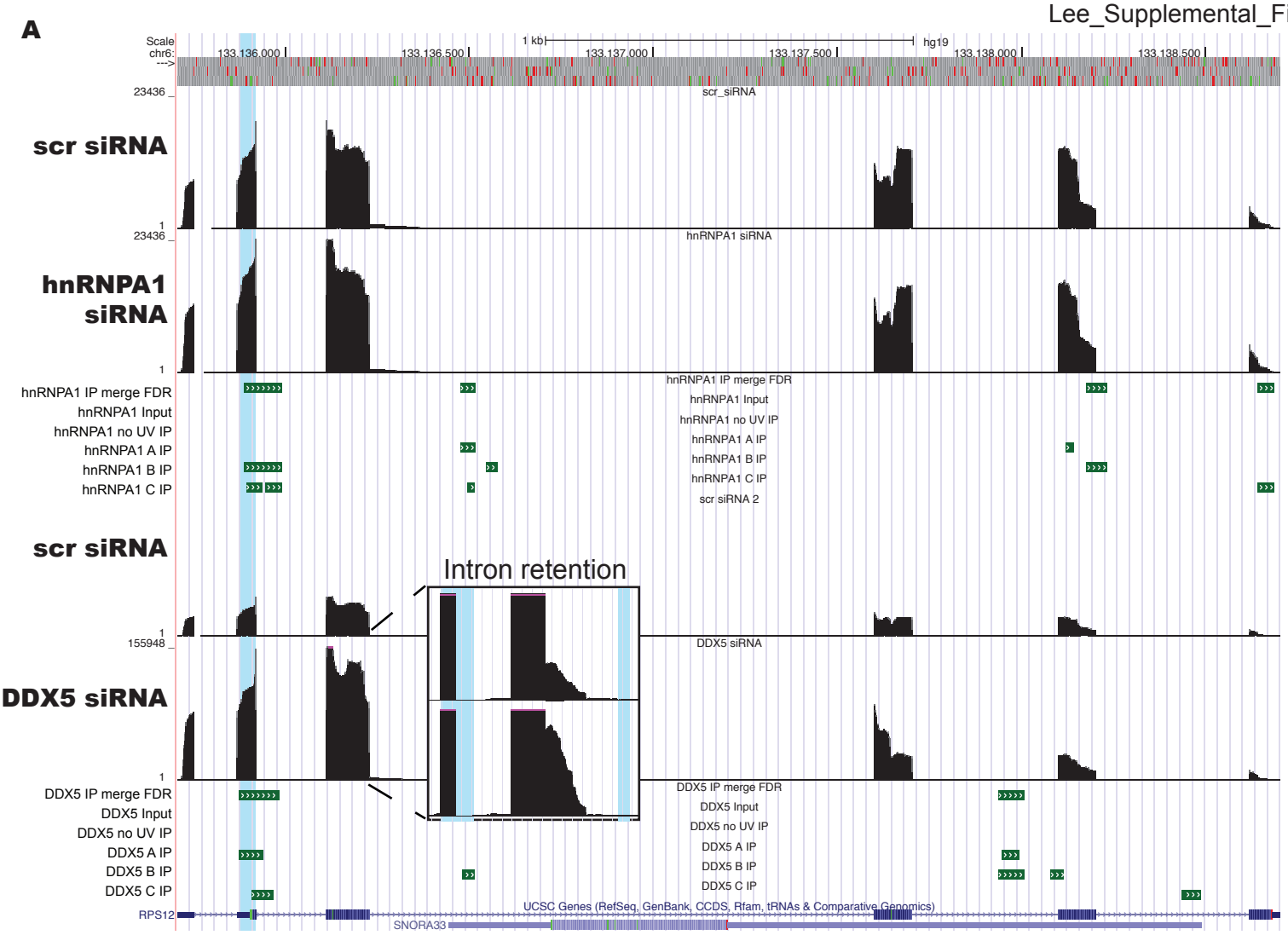
C

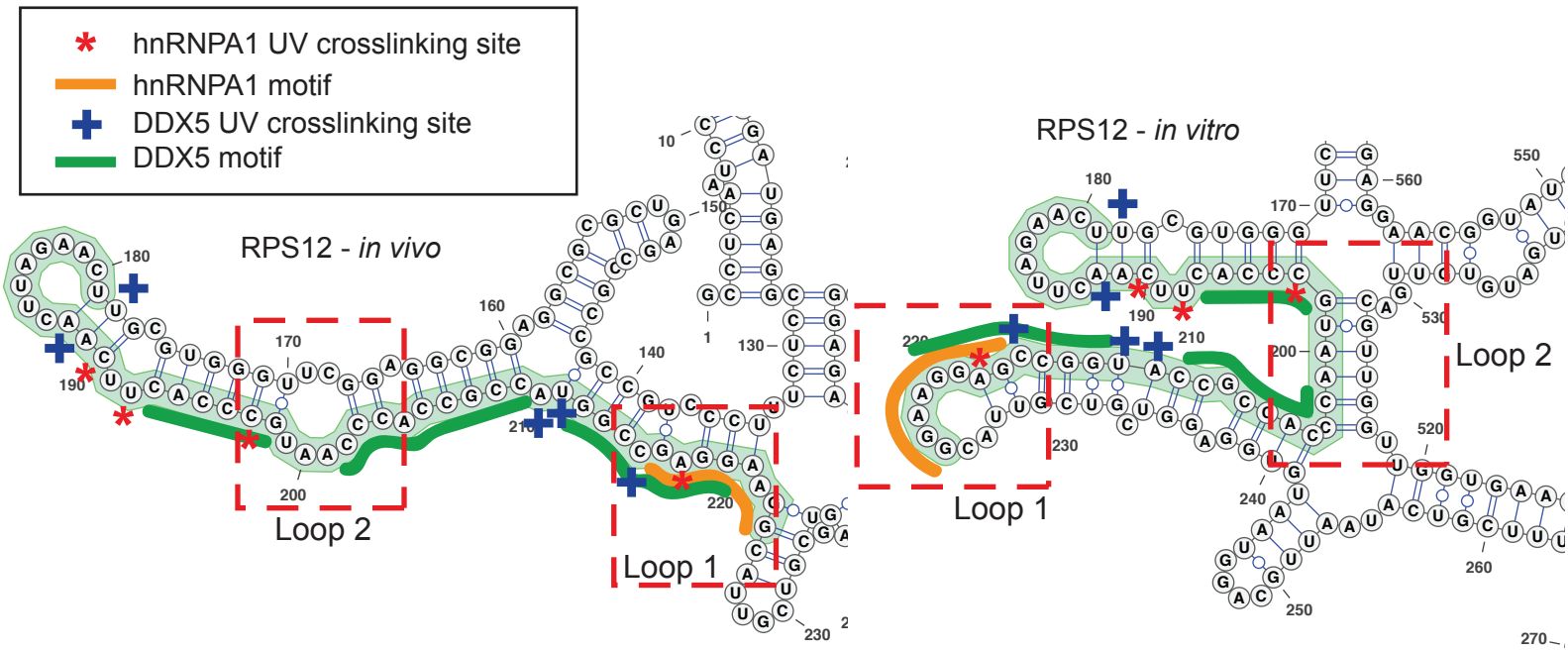


Supplemental Figure S3. DDX5 differentially spliced transcripts with nuclear eCLIP peaks in human K562 cells. (S3A) The overlap between the differential splicing target RNAs and transcripts with nuclear DDX5 eCLIP tag clusters in K562 cells. (S3B) List of common differential splicing target pre-mRNAs that contain nuclear DDX5 eCLIP targets. (S3C) Gene Ontology (GO) enrichment of DDX5 splicing target RNAs containing eCLIP tags revealed that these target RNAs are involved in cellular processes including regulation of gene expression, mRNA metabolic process, protein localization to the endoplasmic reticulum and transcriptional regulation.

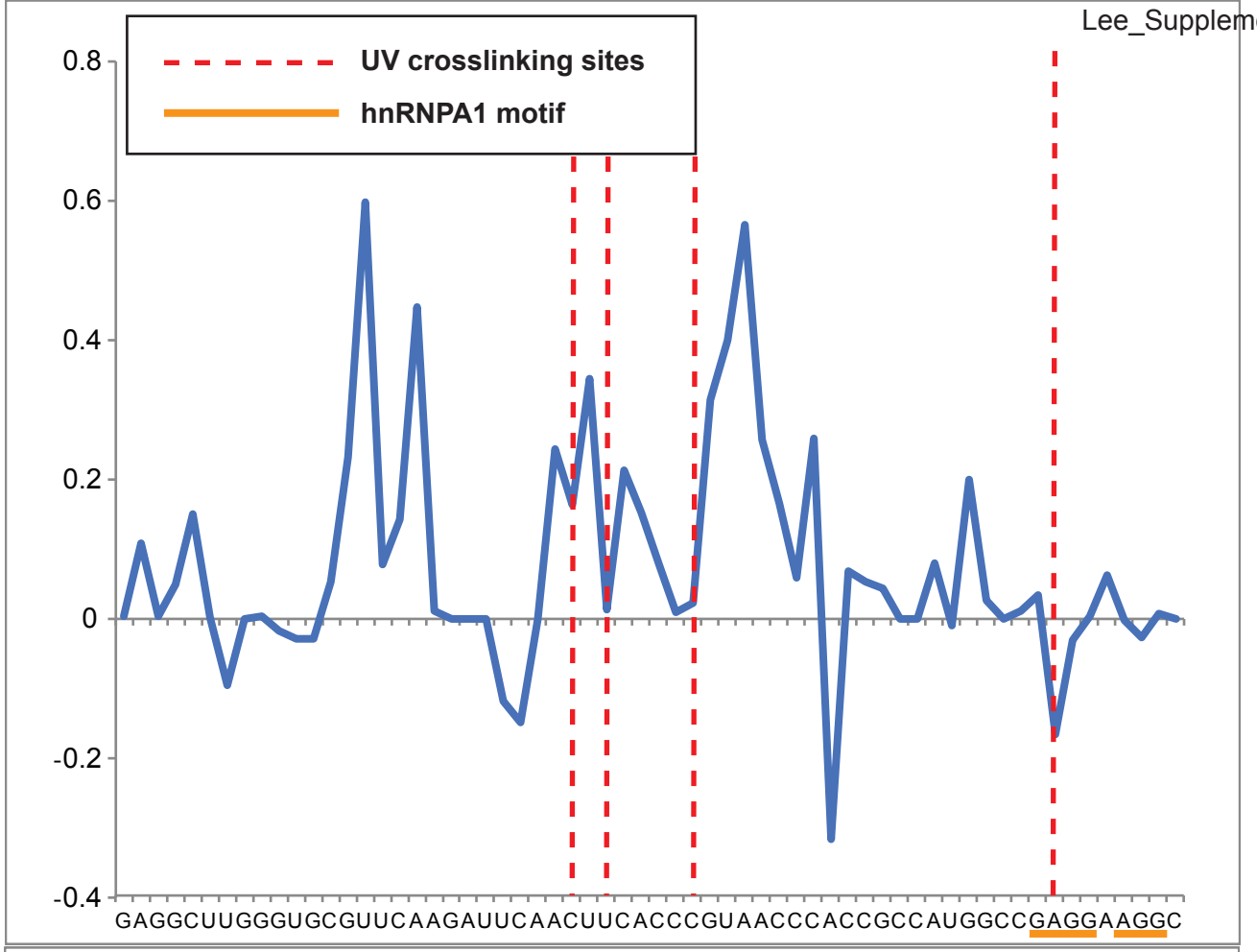


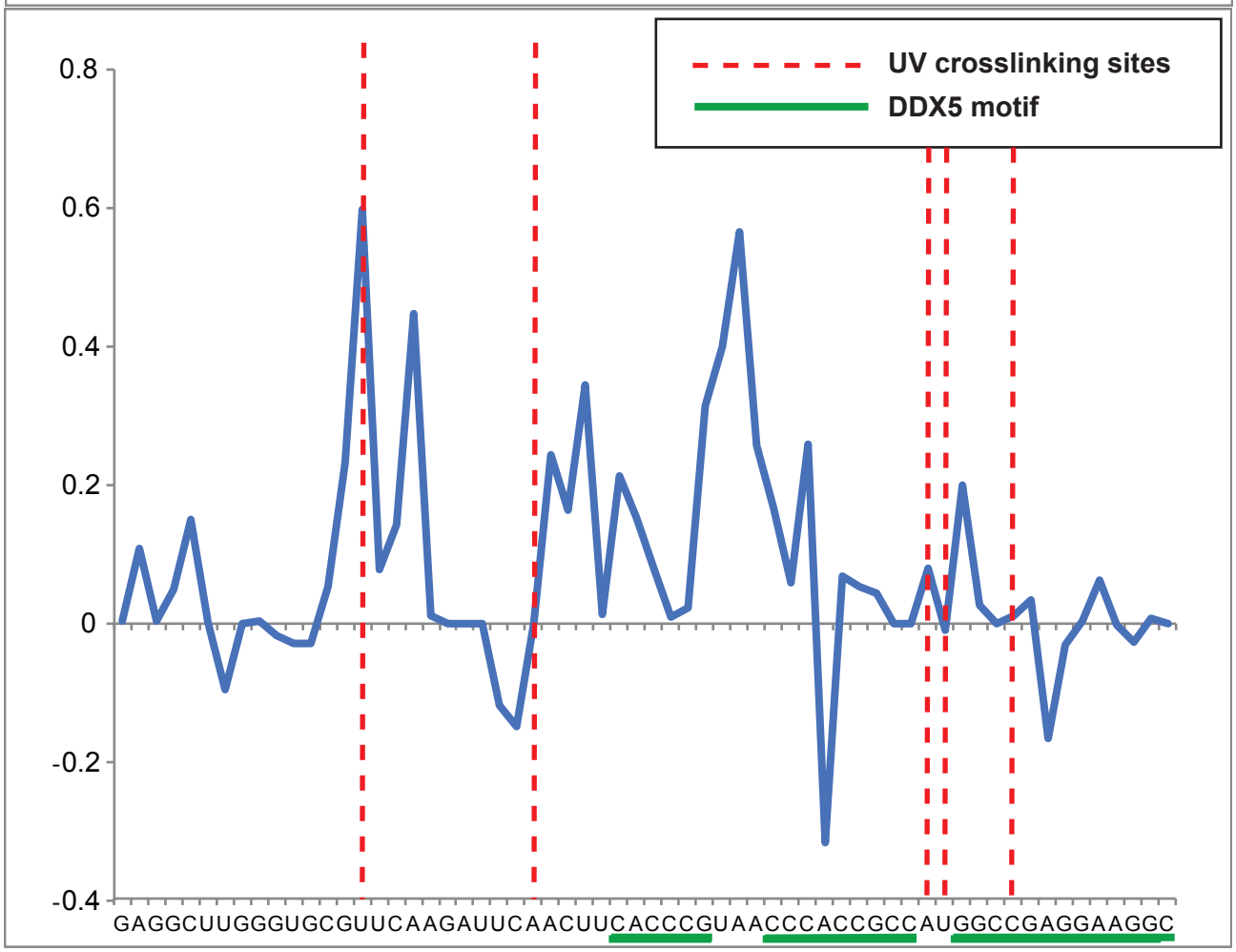
Supplemental Figure S4. The differential icSHAPE profiles between *in vivo* versus *in vitro* reactivities (VTD) can be used to map protein binding sites on target RNAs in conjunction with the eCLIP data. The '*vivo–vitro* difference' (VTD) of icSHAPE profiles of the RPL7A RNA target near the nuclear hnRNPA1 (top panel) and DDX5 (bottom panel) eCLIP peak regions are shown. Areas below 0 indicate less reactivity at those sites in the *in vivo* sample relative to the *in vitro* probed RNA. In the plot, red dashed lines indicate eCLIP crosslink sites. hnRNPA1 motifs are indicated by orange lines in the top panel and DDX5 motifs are indicated in dark green lines in the bottom panel.



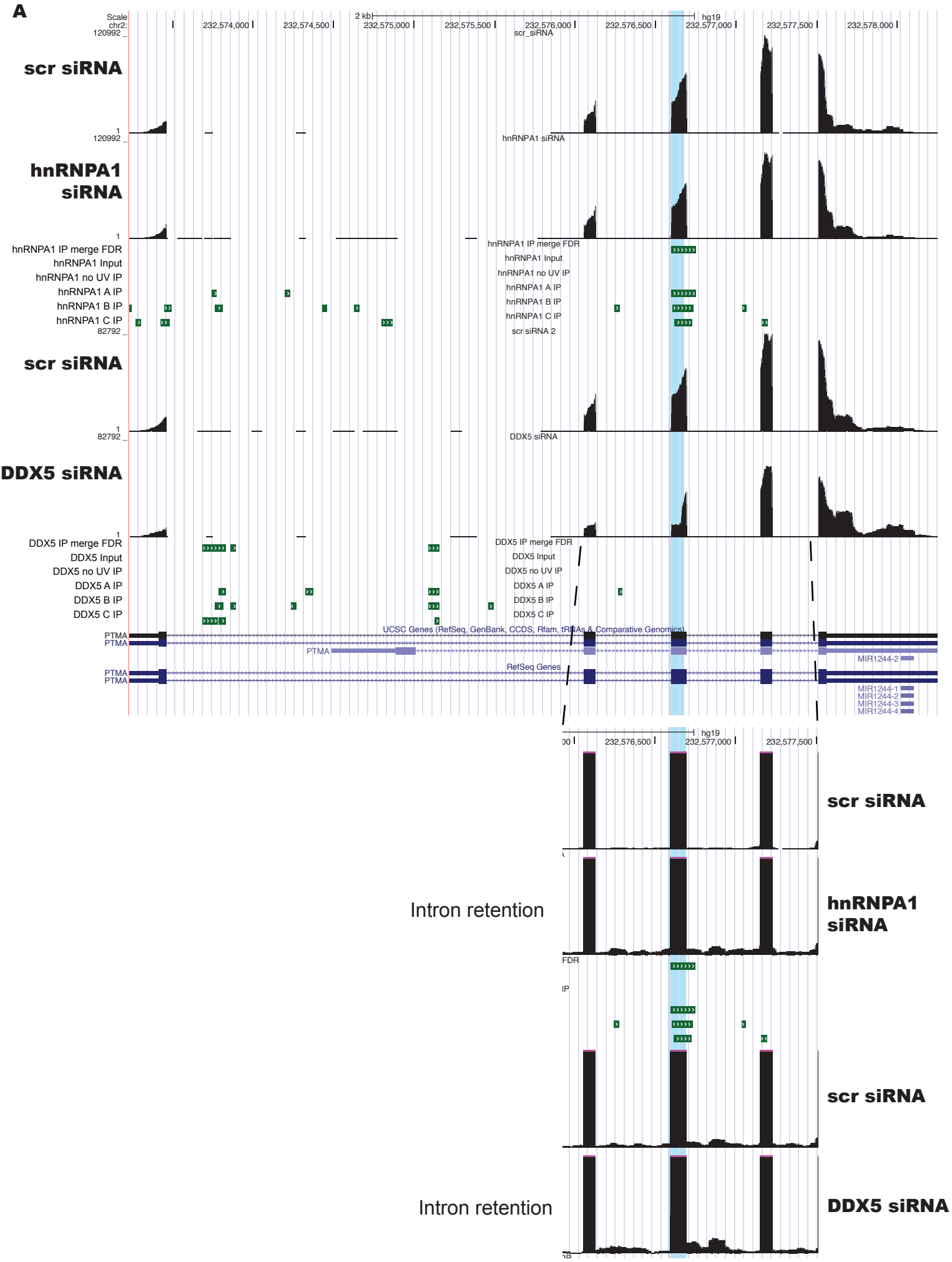


Supplemental Figure S5. Correlating icSHAPE-seq RNA structural information with hnRNPA1/DDX5 binding sites on the alternative splicing target from the RPS12 mRNA. (S5A) Genome browser shot of the RPS12 gene with RNA-seq data upon hnRNPA1/DDX5 knockdown and nuclear hnRNPA1/DDX5 eCLIP peak cluster regions shown. Gene annotation at the bottom with exons indicated. eCLIP tag clusters shown in green. (S5B) icSHAPE-constrained *in vivo*/*in vitro* secondary structures for RPS12 generated using VARNA. The nuclear eCLIP cluster/peak region is highlighted in green and the hnRNPA1 crosslink sites are marked with red asterisks and hnRNPA1 binding motifs are indicated by an orange line. The DDX5 crosslinking sites are indicated by blue crosses and DDX5 binding motifs indicated by green lines. Dotted red rectangles indicate regions of the RNA with secondary structural changes between the *in vivo* and *in vitro* icSHAPE constraint predictions.

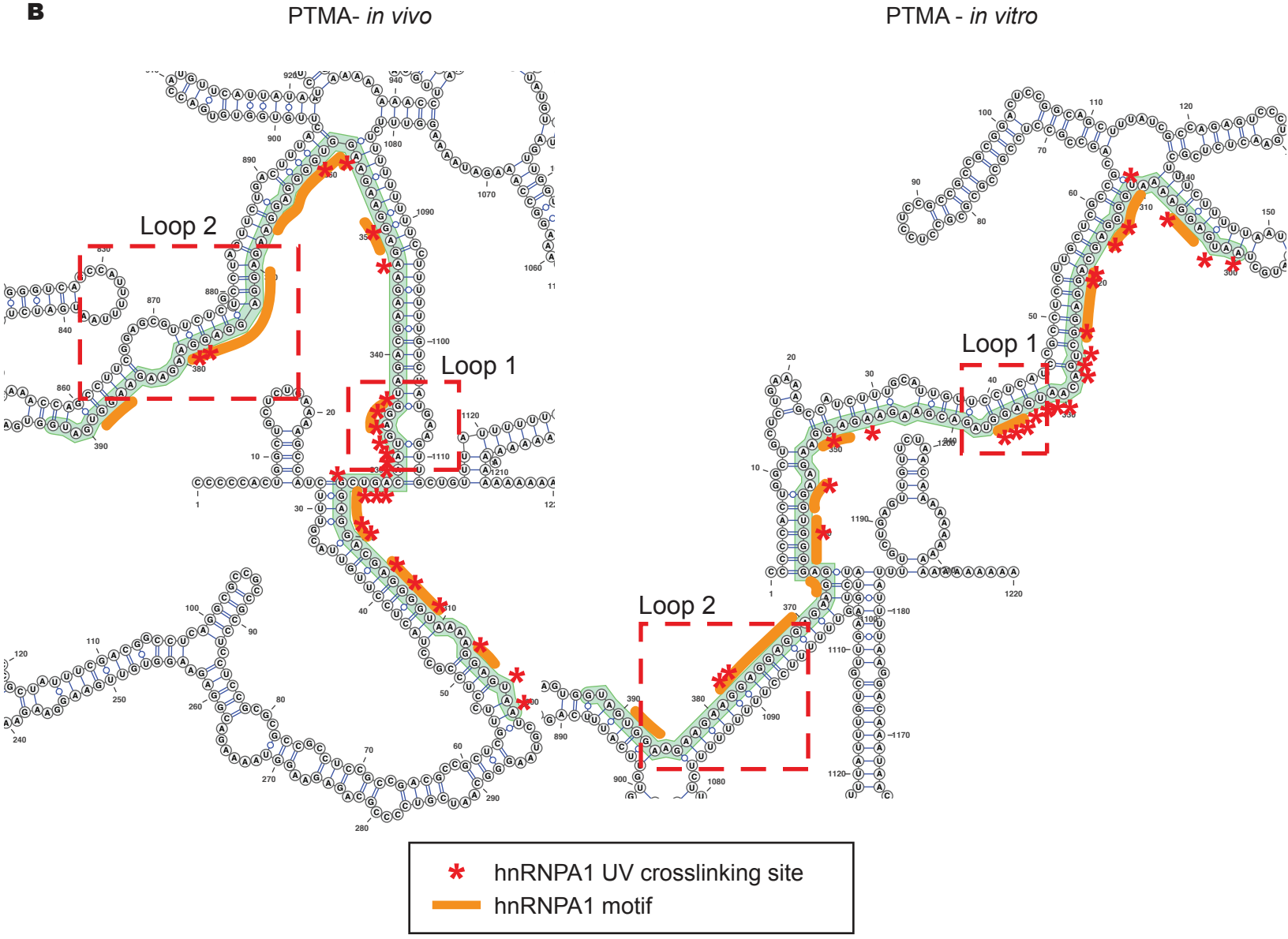




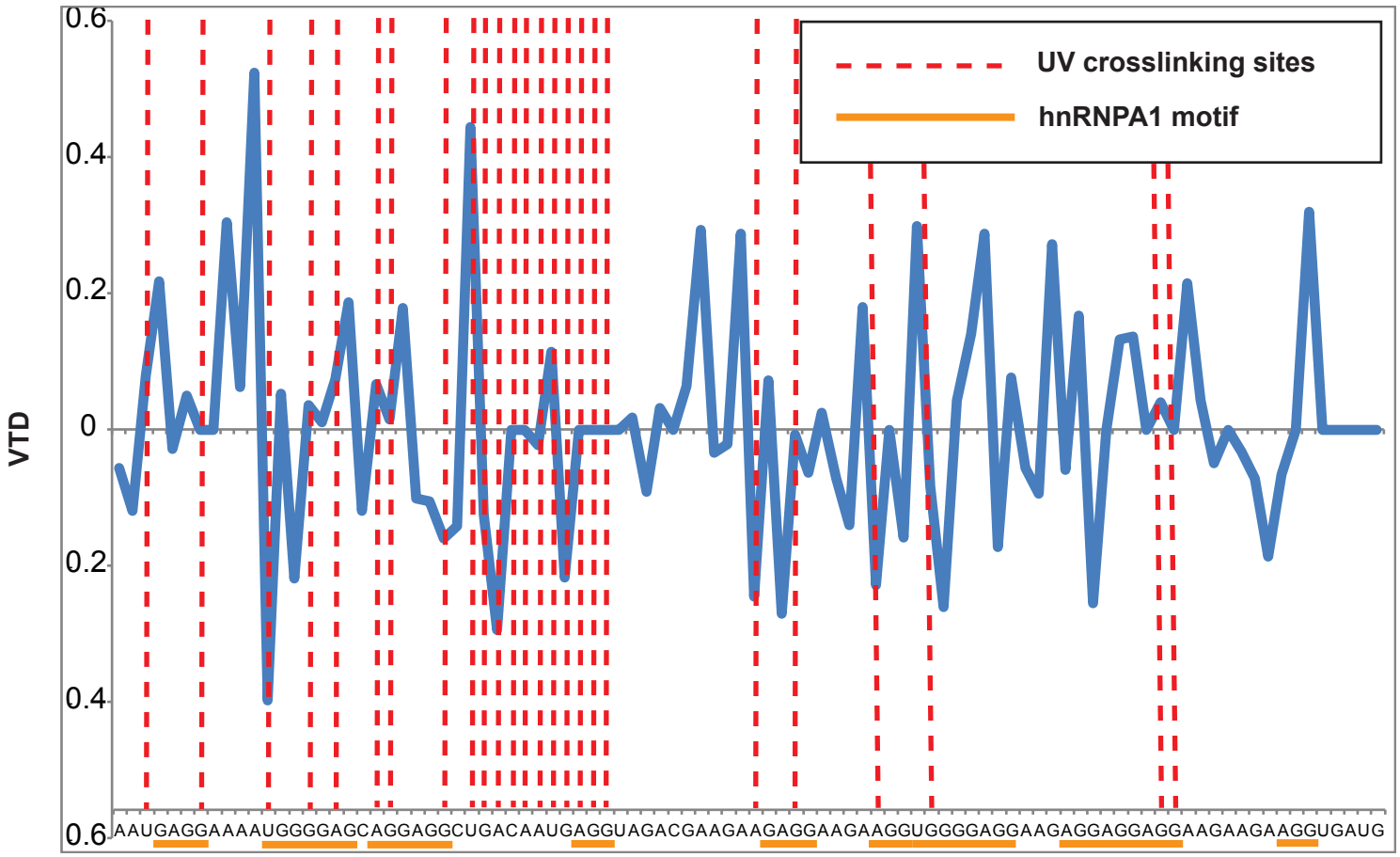
Supplemental Figure S6. The '*vivo–vitro* difference' VTD plots of icSHAPE profiles for a region of the RPS12 target RNA. Locations of the nuclear hnRNPA1 eCLIP peak region with hnRNPA1 motifs indicated in orange (top) and DDX5 eCLIP peak region with DDX5 peaks indicated in green (bottom). These differential icSHAPE profiles between *in vivo* versus *in vitro* probing can be used to map protein binding sites on target mRNAs. In the plot, red dash lines indicate eCLIP crosslink sites. Often, nuclear hnRNPA1/DDX5 eCLIP peak regions reside near areas of negative VTD, indicative of less chemical reactivity, either because the RNA is less structured *in vivo* structure compared to *in vitro* or that there is reduced chemical reactivity *in vivo* due to protein binding to that region of the RNA.


A
scr siRNA
hnRNPA1 siRNA
scr siRNA
DDX5 siRNA
hnRNPA1 IP merge FDR
hnRNPA1 Input
hnRNPA1 no UV IP
hnRNPA1 A IP
hnRNPA1 B IP
hnRNPA1 C IP
DDX5 IP merge FDR
DDX5 Input
DDX5 no UV IP
DDX5 A IP
DDX5 B IP
DDX5 C IP
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RefSeq Genes
PTMA
MIR1244-1
MIR1244-2
MIR1244-3
MIR1244-4
Intron retention
scr siRNA
hnRNPA1 siRNA
scr siRNA
Intron retention
DDX5 siRNA

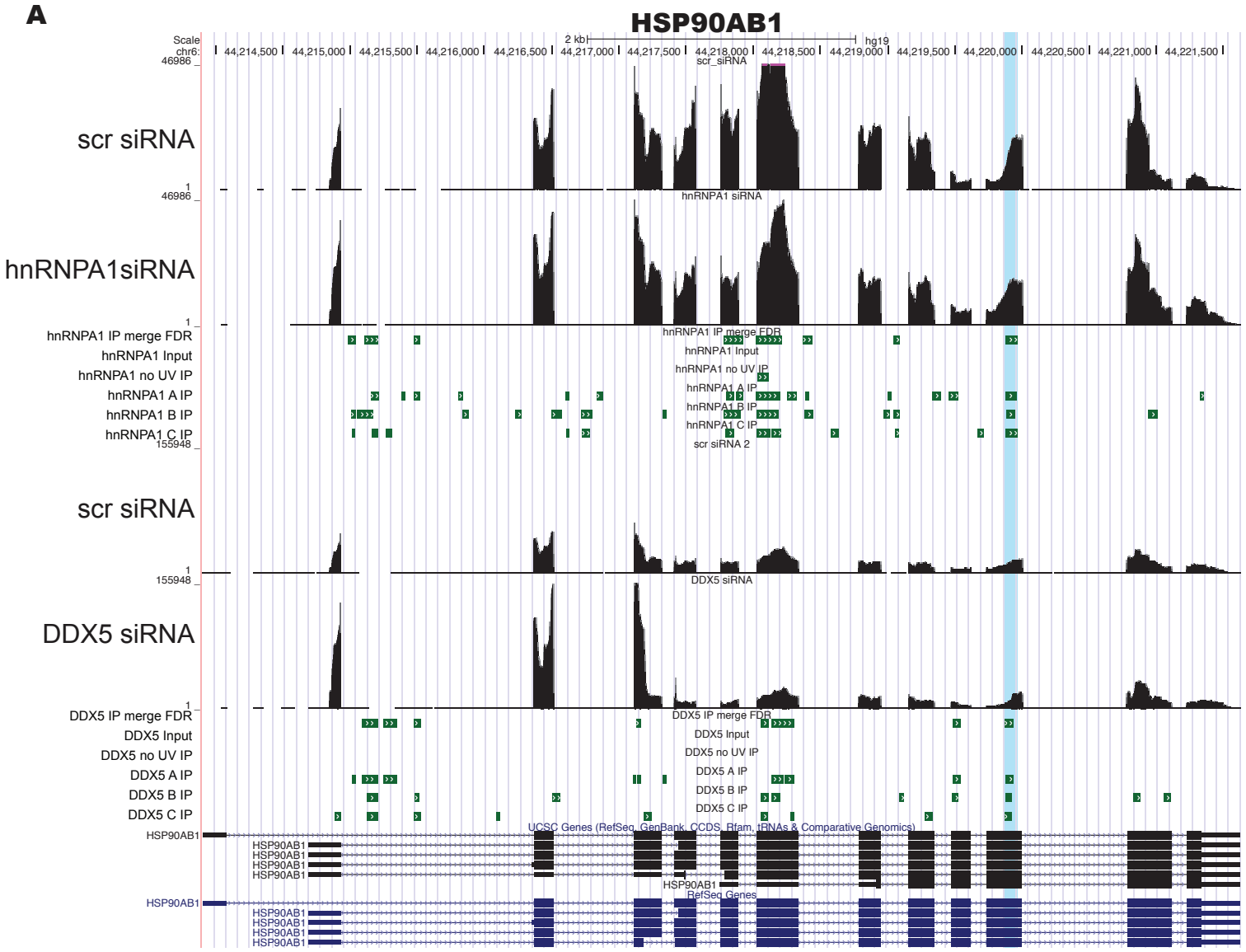
B



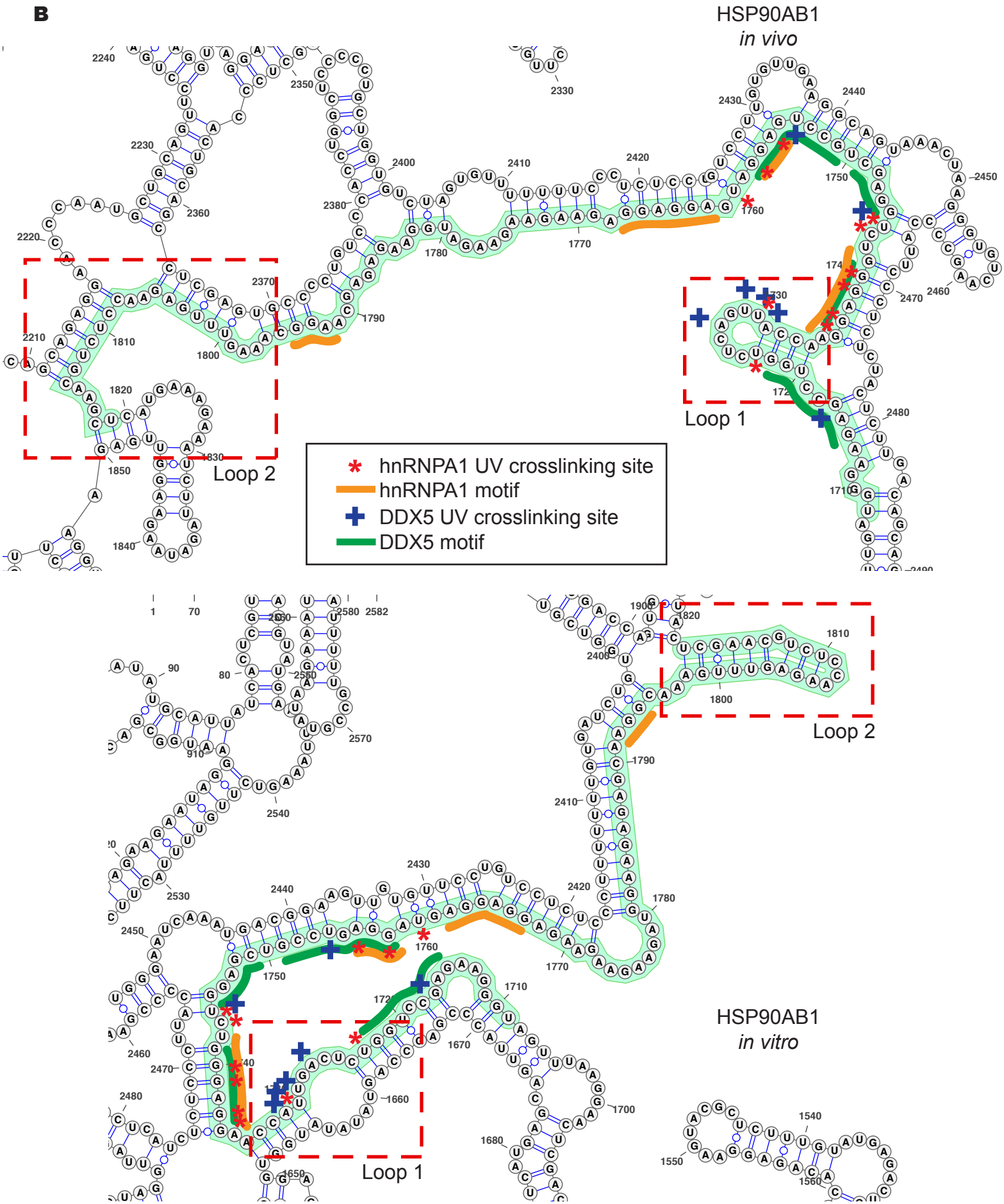
Supplemental Figure S7. Correlating binding of hnRNPA1 with the icSHAPE constrained structure of the Prothymosin alpha (PTMA) pre-mRNA. (S7A) Genome browser view of the PTMA gene and its pre-mRNA that was shown to be differentially spliced upon the knockdown of both hnRNPA1 and DDX5, but these proteins bind to two different regions of PTMA mRNA. The visualized exonic region of PTMA is highlighted in light blue on the genome browser with hnRNPA1 and DDX5 nuclear eCLIP peak cluster regions indicated by green bars. (S7B) icSHAPE-constrained *in vivo* and *in vitro* RNA secondary structures for the PTMA RNA. The nuclear eCLIP cluster/peak region is highlighted in orange and the hnRNPA1 crosslink sites are marked with red asterisks. Dotted red rectangle indicate regions of the RNA with secondary structural changes between the *in vivo* and *in vitro* icSHAPE constraints.



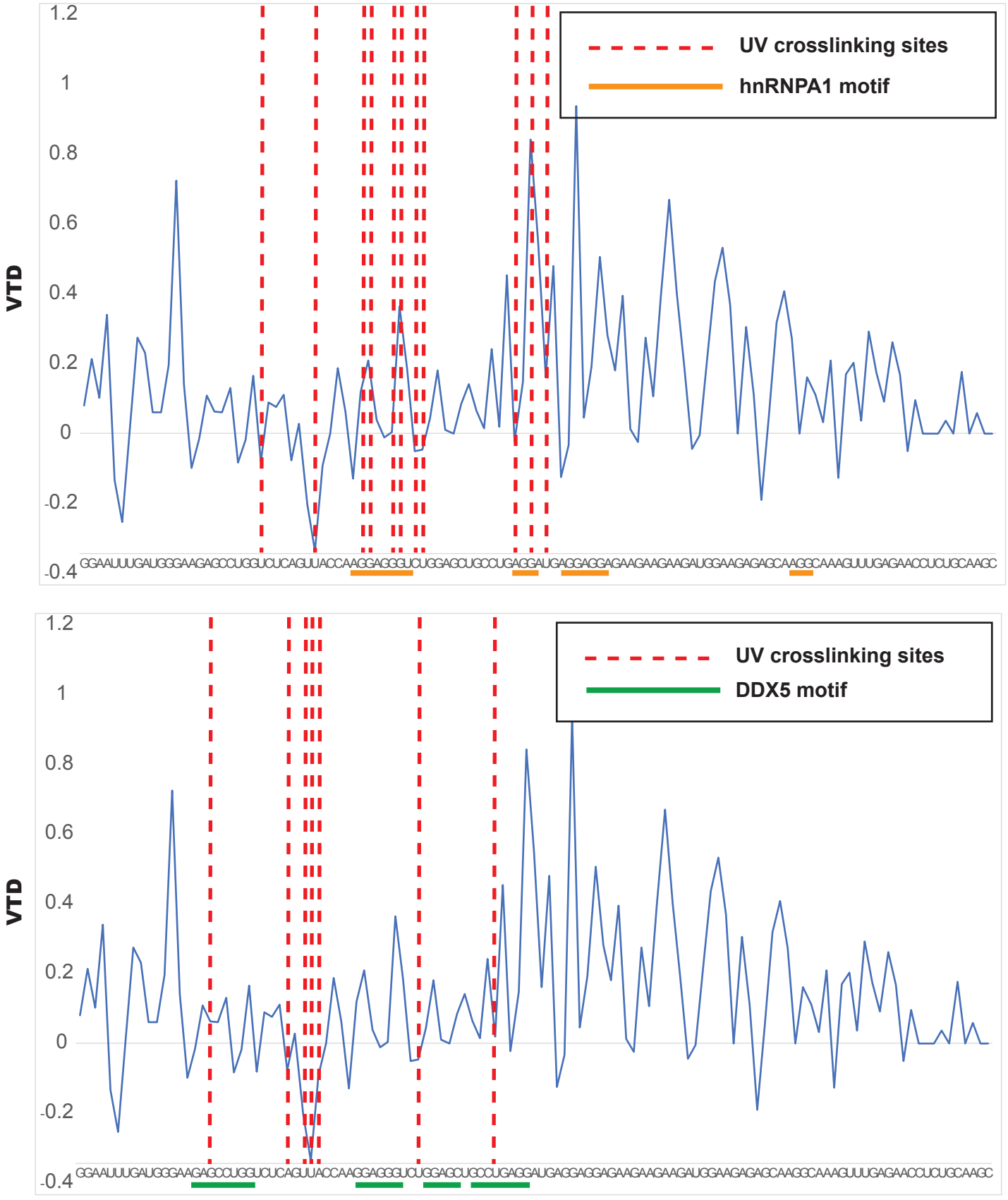
Supplemental Figure S8. The '*vivo–vitro* difference' VTD of icSHAPE profiles of the VTMA pre-mRNA target near the nuclear hnRNPA1 eCLIP tag region. Red dash lines indicate eCLIP crosslink sites for hnRNPA1. Nuclear hnRNPA1 eCLIP peak regions reside in the negative VTD areas, indicative of less chemically reactive regions in the *in vivo* RNA sample compared to *in vitro* probed RNA, most likely due to binding of the hnRNPA1 protein to the VTMA RNA.



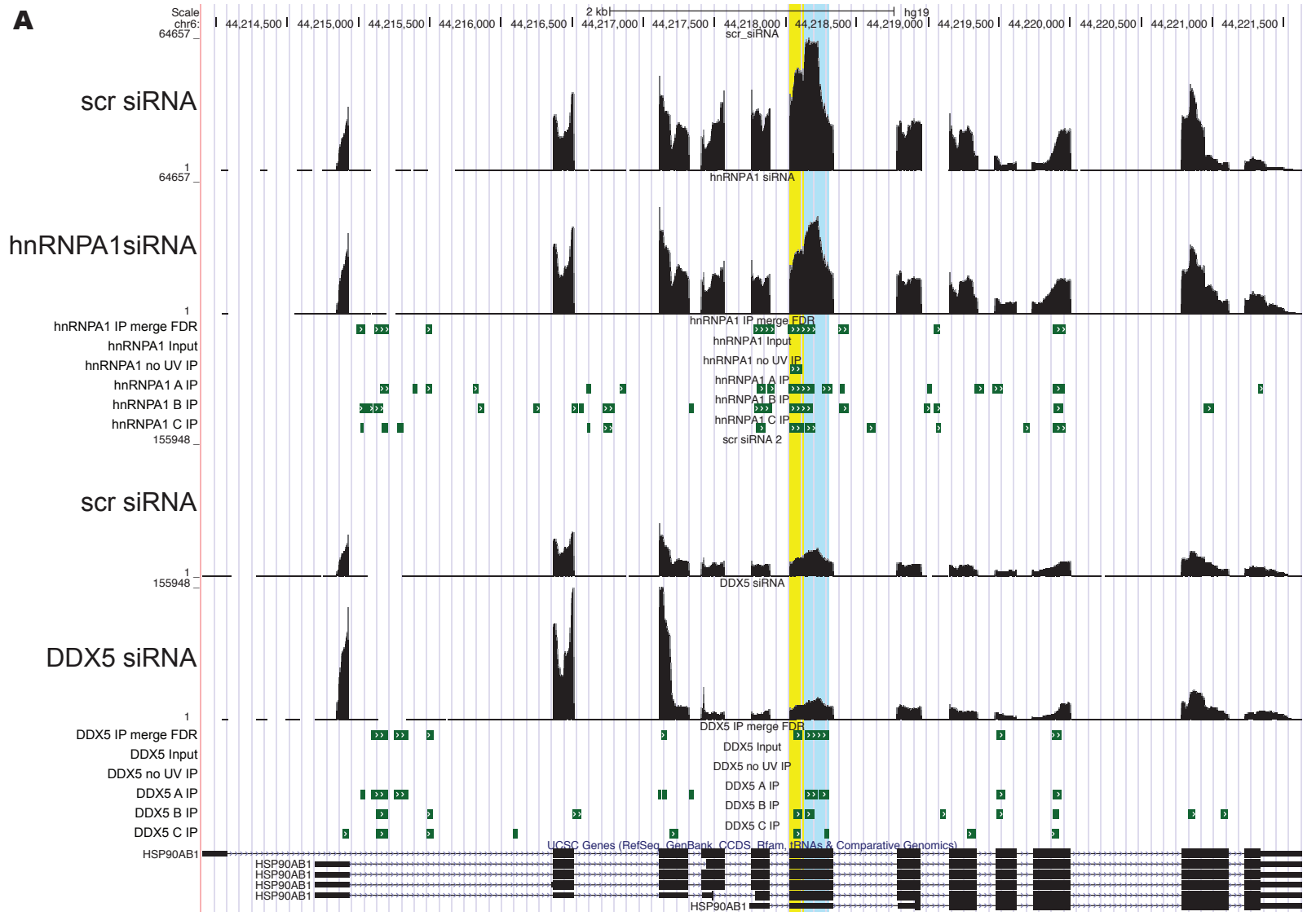
Supplemental Figure S9. The Heat Shock Protein 90 Alpha Family Class B Member 1, HSP90AB1, pre-mRNA cytosolic 90kDa is differentially spliced upon knockdown of hnRNPA1 and DDX5. (S9A) Genome browser of shot of the HSP90AB1 gene and RNA transcripts. The hnRNPA1 and DDX5 proteins share 5 common binding sites on the HSP90AB1 pre-mRNA. The region visualized by RNA structure and Vienna is highlighted in light blue on the genome browser.



Supplemental Figure S9B. icSHAPE-constrained *in vivo* and *in vitro* secondary structures of a region of the HSP90AB1 pre-mRNA. The nuclear eCLIP cluster region is highlighted in light green. The hnRNPA1 crosslink sites are indicated by red asterisks and the hnRNPA1 binding motifs by orange lines. The DDX5 crosslink sites are indicated by blue crosses and the enriched GC-rich binding motifs for DDX5 are indicated by dark green lines. Dotted red rectangle areas indicate regions of the RNA with secondary structural changes between the *in vivo* and *in vitro* icSHAPE constraints.

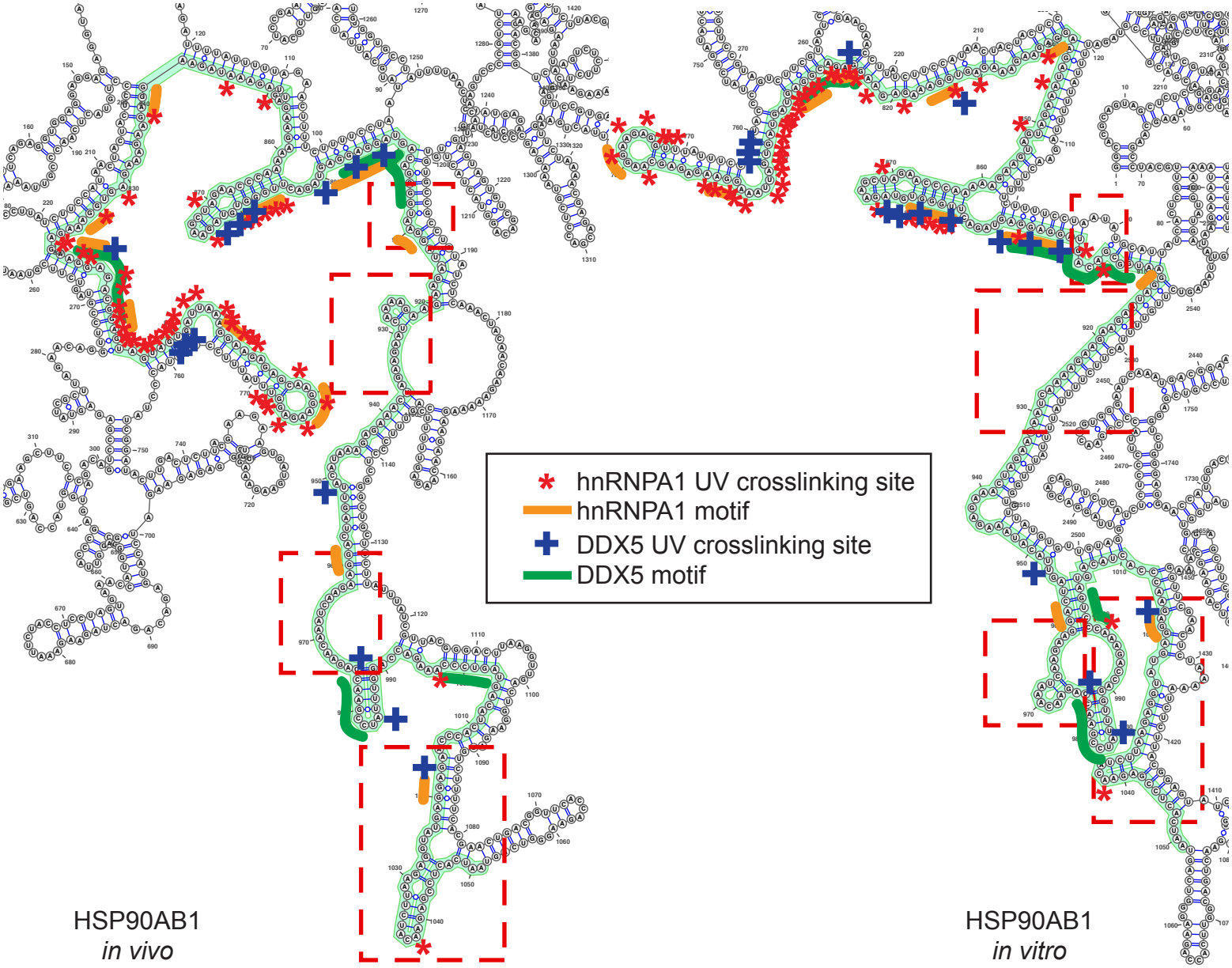


Supplemental Figure S9C. The differential icSHAPE profiles between *in vivo* versus *in vitro* reactivities (VTD) correlate with the positions of protein binding sites on target RNAs mapped using eCLIP data. The '*vivo–vitro* difference' (VTD) profiles of icSHAPE reactivities for the HSP90AB1 RNA target near the nuclear hnRNPA1 (top panel) and DDX5 (bottom panel) eCLIP peak regions are shown. In the plot, red dashed lines indicate eCLIP crosslink sites, and hnRNPA1 motifs are indicated by orange lines in the top panel and DDX5 motifs are indicated dark green lines in the bottom panel.

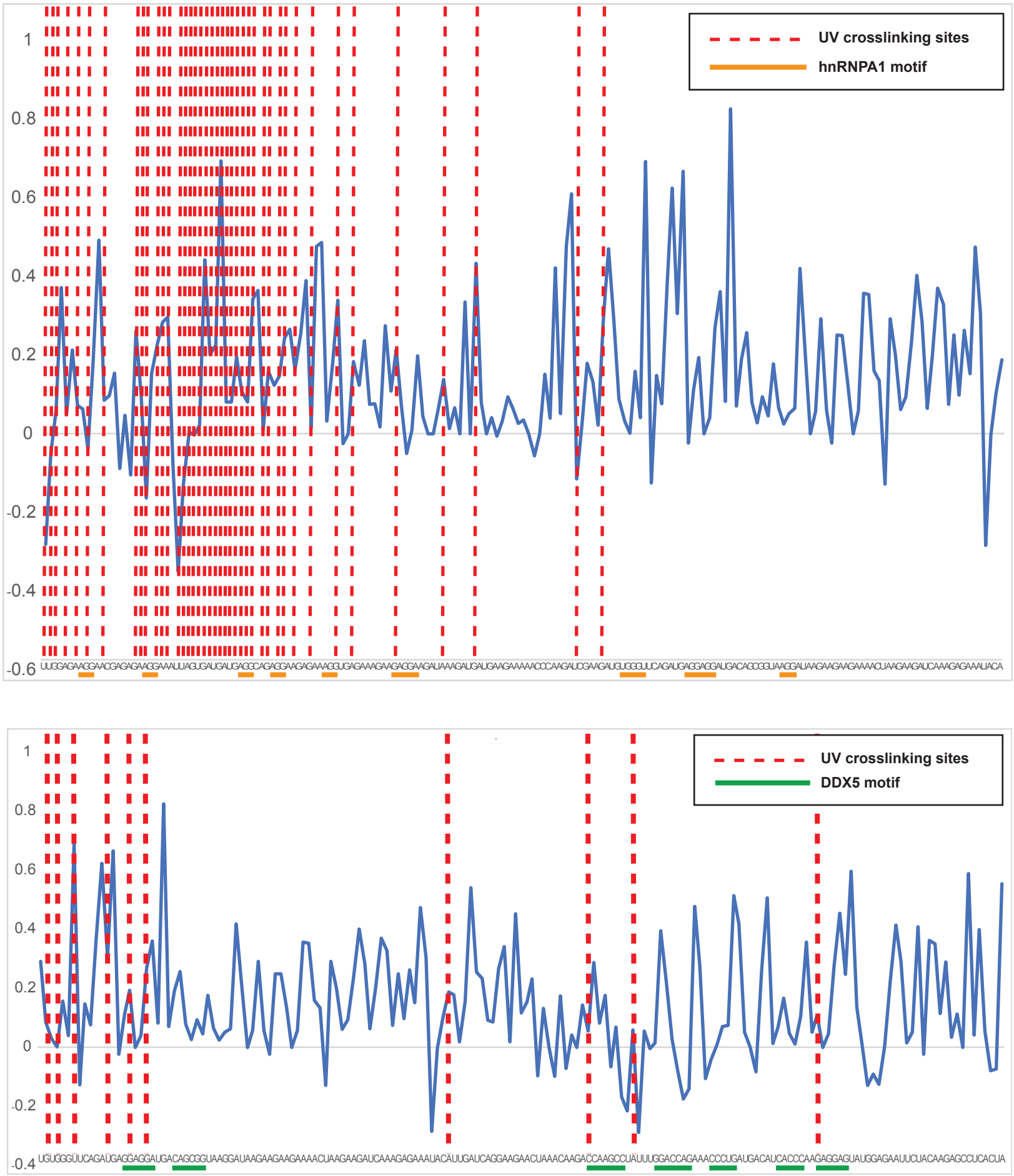


Supplemental Figure S10. The Heat Shock Protein 90 Alpha Family Class B Member 1, HSP90AB1, pre-mRNA cytosolic 90kDa heat-shock protein is common splicing target for hnRNPA1 and DDX5. (S10A) Genome browser of view of HSP90AB1 gene and RNA transcripts. hnRNPA1 and DDX5 share 5 common binding sites on the HSP90AB1 pre-mRNA. The region visualized by RNA structure and Vienna is highlighted on the genome browser in yellow for the hnRNPA1 eCLIP region and in blue for the DDX5 eCLIP region.

B



Supplemental Figure S10B. icSHAPE-constrained in vivo and in vitro predicted RNA secondary structures for a region of the HSP90AB1 pre-mRNA. The hnRNPA1 UV crosslinking sites are indicated by red asterisks and the DDX5 UV crosslinking sites are indicated by blue crosses. The nuclear eCLIP cluster/peak region is indicated by light green shading. The hnRNPA1 binding motifs are indicated by orange lines and for hnRNPA1 and the DDX5 binding motifs are indicated by green lines. The dotted red rectangles denote regions of the RNA with secondary structural changes between the *in vivo* and *in vitro* icSHAPE constraints.



Supplemental Figure S10C. The '*vivo–vitro* difference' (VTD) of icSHAPE profiles of the HSP90AB1 RNA target near the nuclear hnRNPA1 (top panel) and DDX5 (bottom panel) eCLIP peak regions are shown below. In the plot, red dashed lines indicate eCLIP crosslink sites, and hnRNPA1 motifs are indicated by orange lines in the top panel and the DDX5 motifs are indicated by dark green lines in the bottom panel. Multiple hnRNPA1 UV crosslinking sites were detected near the 5' end of the exon, with nearby hnRNPA1 binding motifs indicated by orange lines (top panel). The GC-rich regions resembling DDX5 binding motifs (green line) were detected near the negative VTD region (bottom panel).