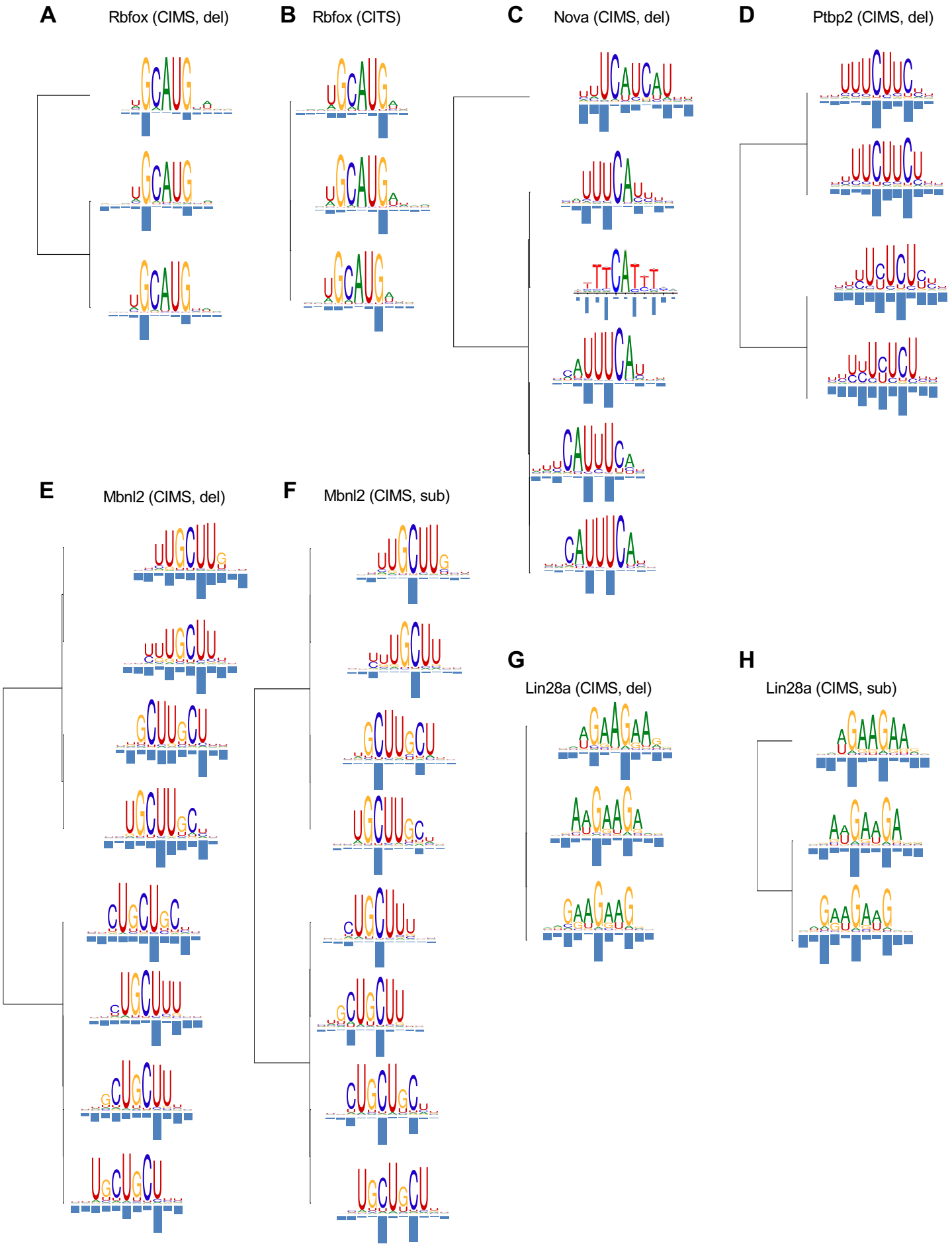

A
Rbfox (CIMS, del)
B
Rbfox (CITS)
C
Nova (CIMS, del)
D
Ptbp2 (CIMS, del)
E
Mbnl2 (CIMS, del)
F
Mbnl2 (CIMS, sub)
G
Lin28a (CIMS, del)
H
Lin28a (CIMS, sub)

Figure S1: The complete list of motifs discovered by mCross for RBPs with well characterized binding specificity. Related to Figure 1. (A, B) Rbfox. (C) Nova. (D) Ptbp2. (E, F) Mbnl2. (G, H) Lin28a. For each panel, the method used to infer crosslink sites is indicated. Motifs are clustered based on the similarity of the PWMs.

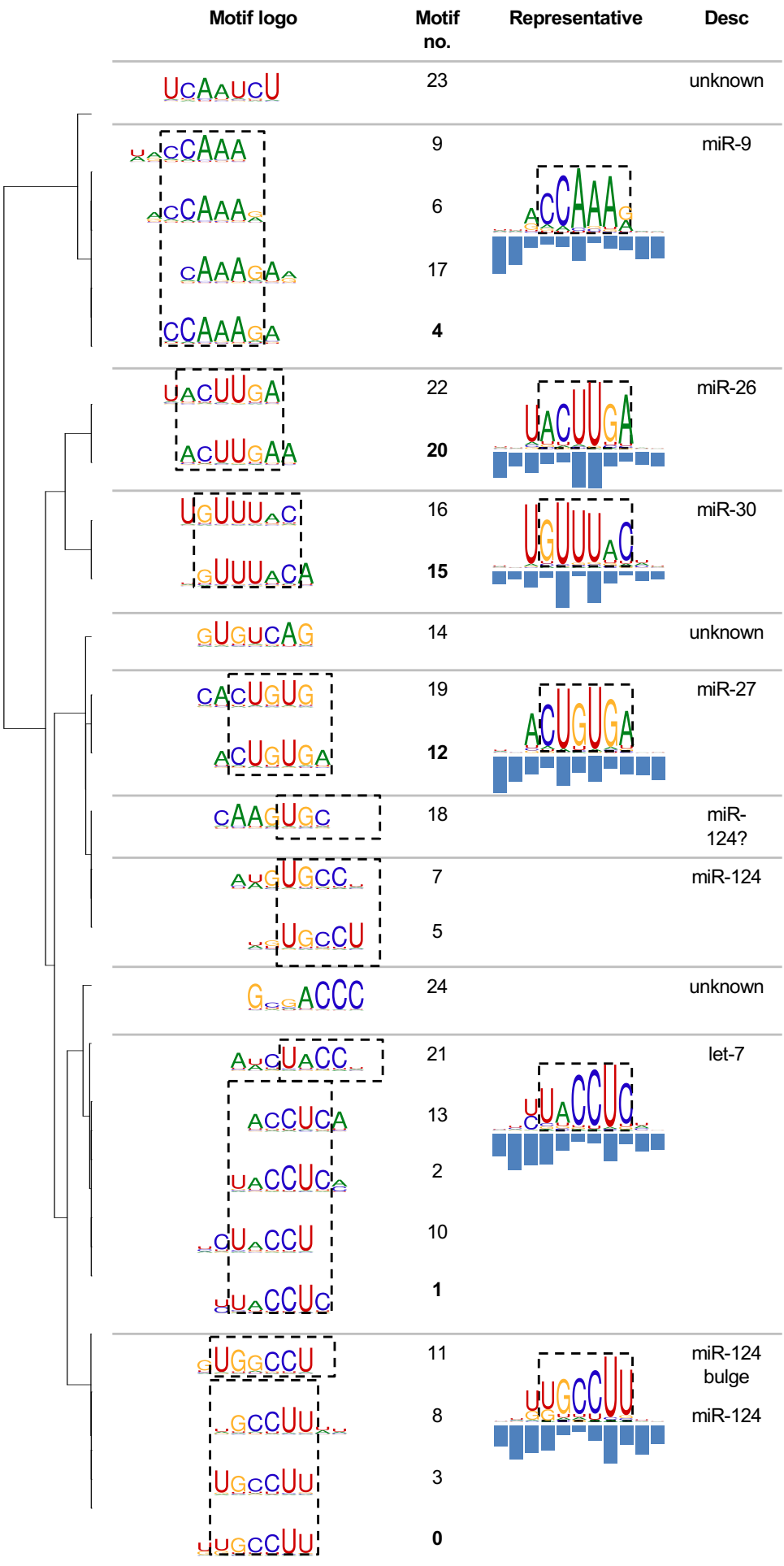

Motif logo
Motif no.
Representative
Desc
23
unknown
9
6
17
4
miR-9
22
20
miR-26
16
15
miR-30
14
unknown
19
12
miR-27
18
miR-124?
7
5
miR-124
24
unknown
21
13
2
10
1
let-7
11
8
3
0
miR-124 bulge
miR-124

Figure S2: mCross accurately identified miRNA seed match sequences from Ago CLIP data. Related to Figure 1. Motifs were clustered based on the similarity of the PWMs. Groups of motifs corresponding to the same miRNAs are indicated. A representative motif (with the highest motif score) and the crosslink sites are provided for each miRNA.

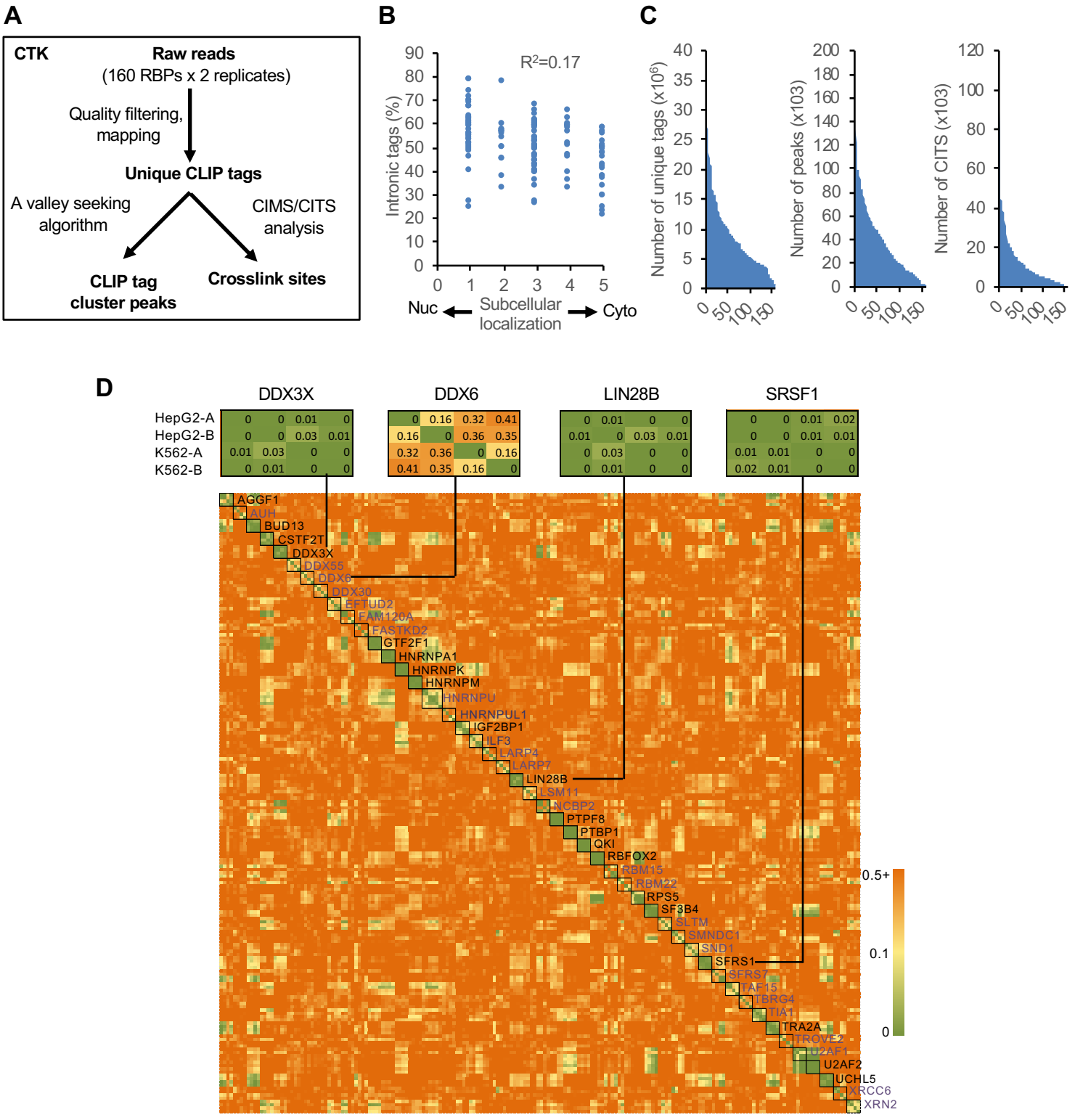


Figure S3: Summary of eCLIP data analyzed in this study. Related to Figure 2. (A) Schematic illustrating the pipeline of eCLIP data analysis using CTK. (B) Correlation between RBP subcellular localization quantified from immmunostaining and percentage of intronic CLIP tags in HepG2 cells. Subcellular localization based on immunostaining (x-axis) was derived by visual examination of images to assign a score between 1 (exclusively nuclear) and 5 (exclusively cytoplasmic). (C) Summary of eCLIP experiments with regard to the total number of unique tags, significant peaks and CITS. (D) Pairwise disconcordance score of top 7-mer enrichment for RBPs that were assayed in both K562 and HepG2 cells. Four representative RBPs with a varying degree of binding specificity are highlighted.

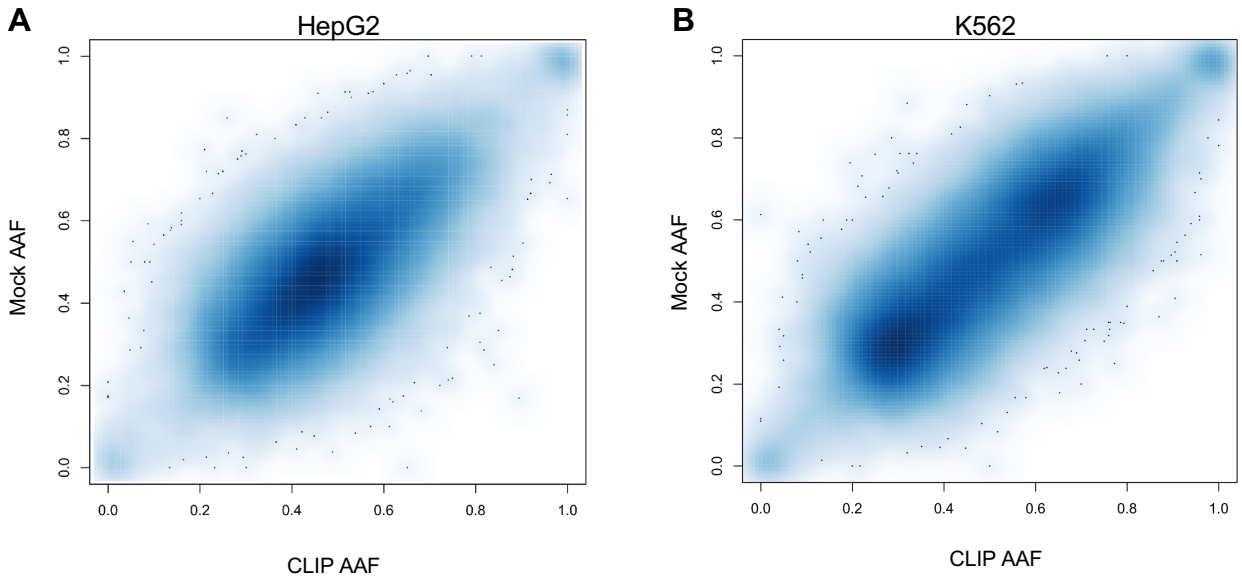


Figure S4: Identification of allelic protein-RNA interaction sites from eCLIP data. Related to Figure 4. (A,B) Correlation of alternative allele frequency (AAF) estimated from CLIP and mock (input) data for HepG2 (A) and K562 (B) cells. Note the mode of SNPs with AAF at 1/3 and 2/3 in K562 cells, due to triploidy in many genomic regions in this cell line.

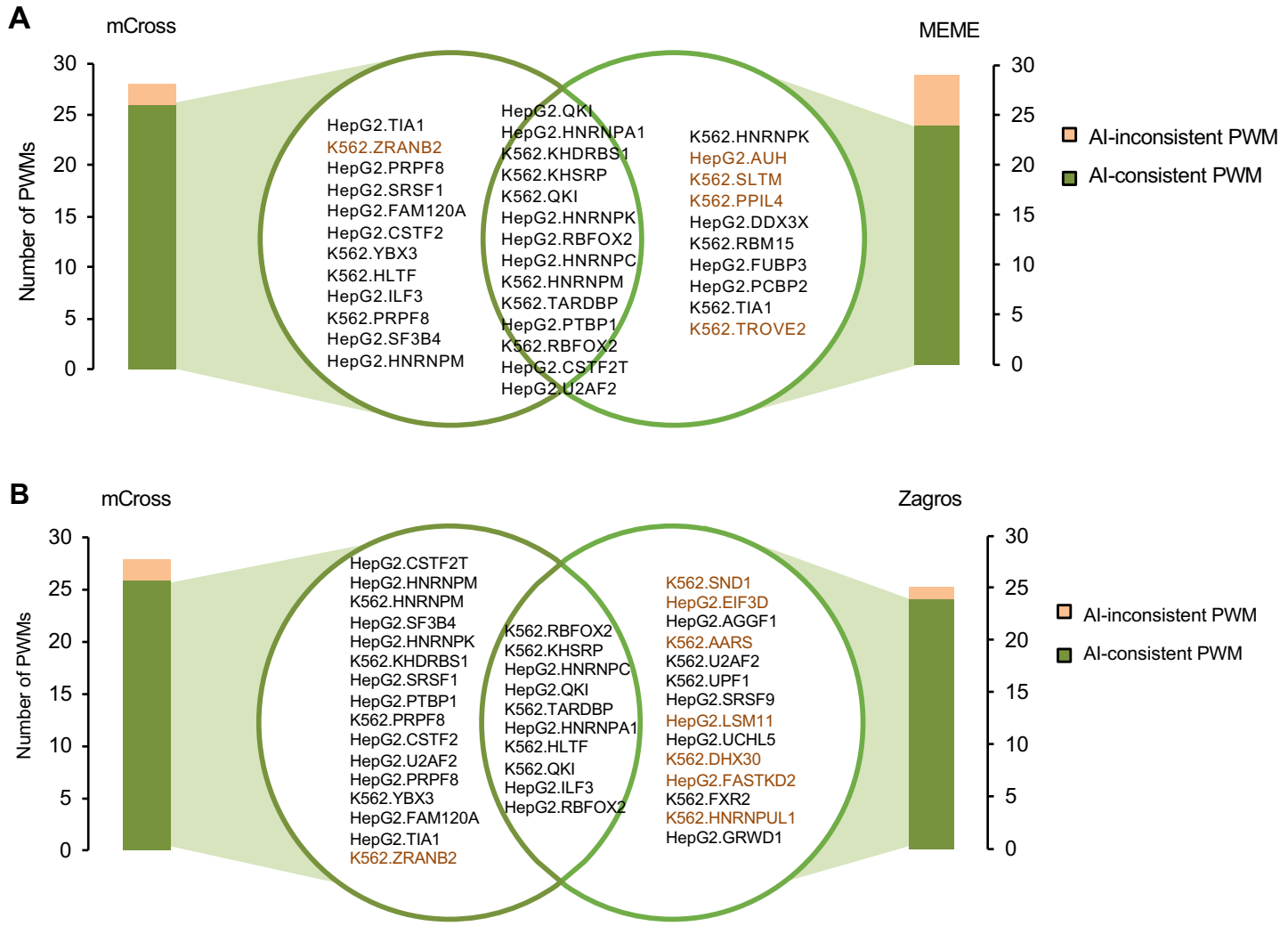
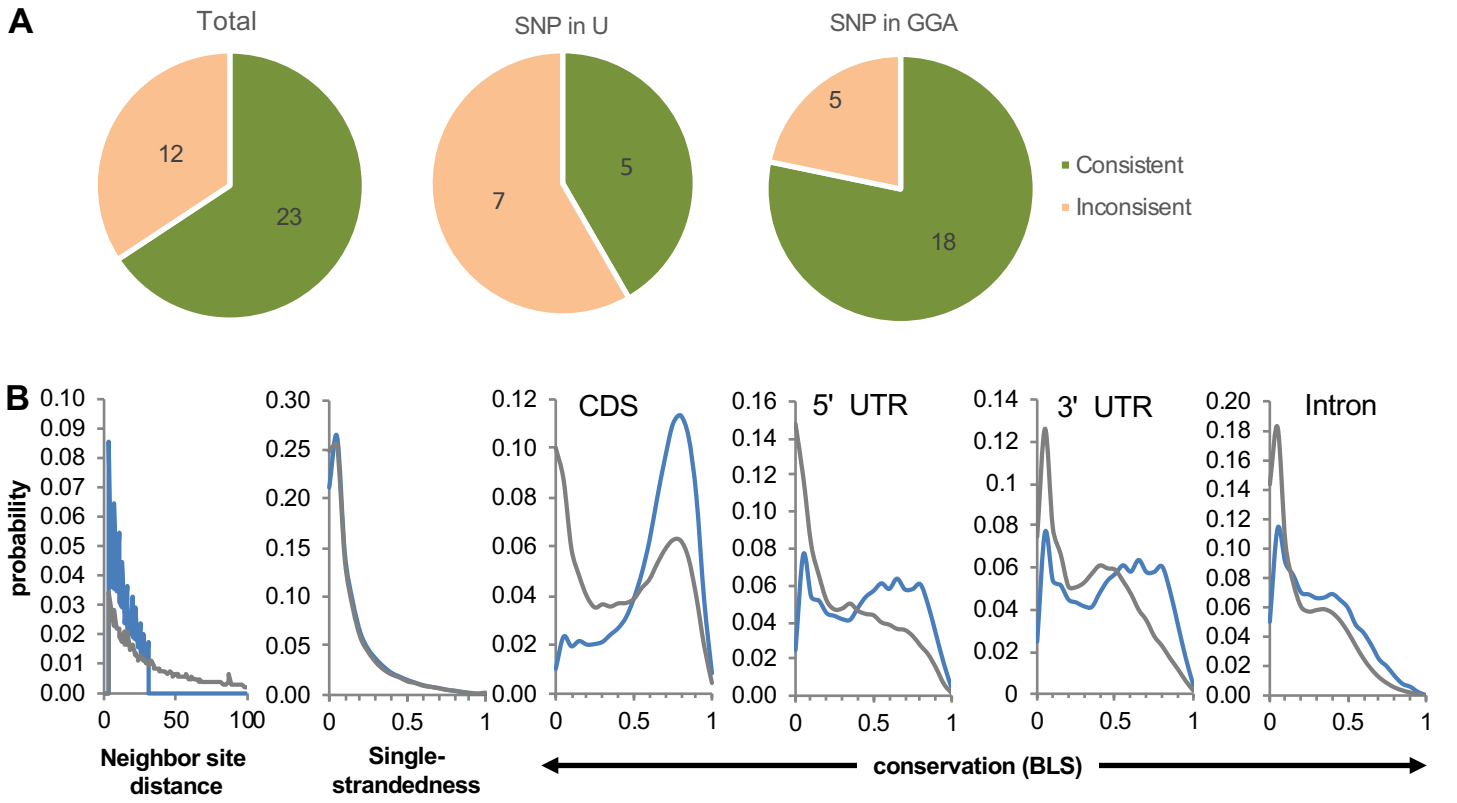
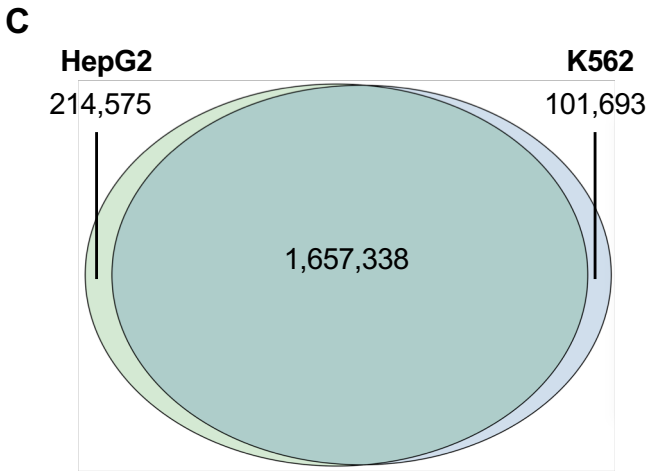
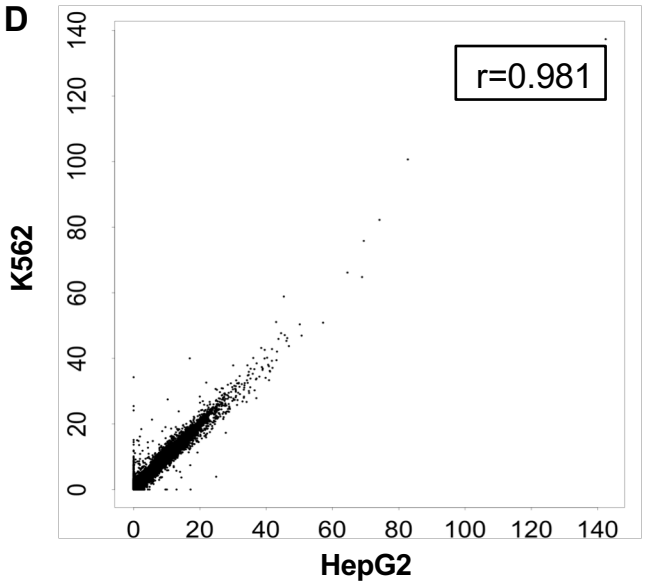
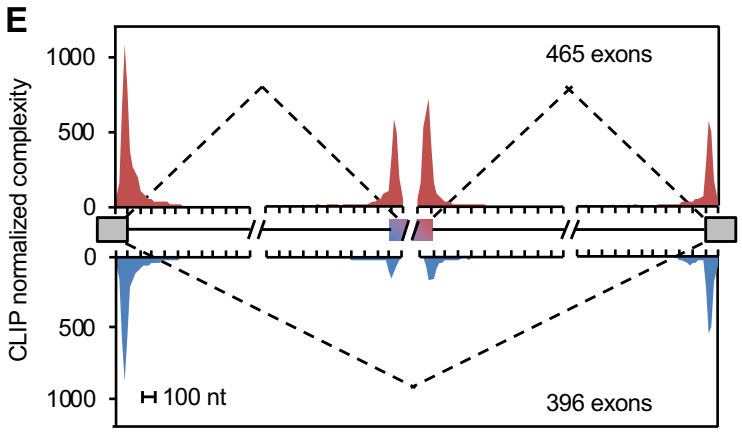


Figure S5: Comparison of mCross and other *de novo* motif discovery programs using eCLIP data and allelic interaction analysis. Related to Figure 4. (A) The overlap of AI-consistent PWMs identified by mCross and MEME. RBPs with D>0.05 between replicates (indicating a lack of reproducible binding specificity) are highlighted in red. (B) Similar to (A) but comparison of mCross and Zagros.


A
Total
SNP in U
SNP in GGA
12
23
7
5
5
18
Consistent
Inconsisent
B
probability
Neighbor site distance
Single-strandedness
CDS
5' UTR
3' UTR
Intron
conservation (BLS)


C
HepG2
K562
214,575
101,693
1,657,338


D
r=0.981
K562
HepG2


E
CLIP normalized complexity
465 exons
100 nt
396 exons

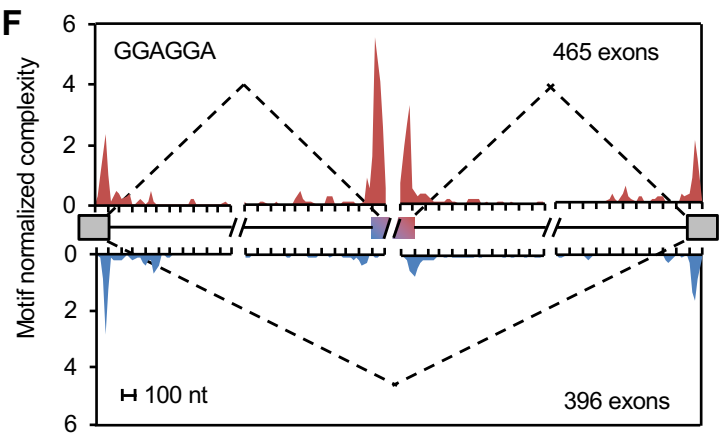

F
Motif normalized complexity
GGAGGA
465 exons
100 nt
396 exons

Figure S6: Prediction of SRSF1-binding GGA clusters and SRSF1-dependent splicing. Related to Figure 6. (A) The number of AI sites overlapping with the different positions of the UGGA motif. There are significantly more consistent AI sites than inconsistent AI sites, when the SNP overlaps with GGA ($p=0.005$; Binomial test), but not when the SNP overlaps with the first U. (B) Emission probabilities of the HMM trained by mCross using HepG2 eCLIP data. (C) The overlap of predicted GGA clusters using the models trained in HepG2 and K562 eCLIP data. (D) Correlation of motif site scores for exactly matched GGA clusters predicted using the models trained in HepG2 and K562 eCLIP data. (E) The normalized complexity RNA map using HepG2 CLIP tags. (F) The normalized complexity RNA map using conserved GGAGGA motif sites weighted by BLS.

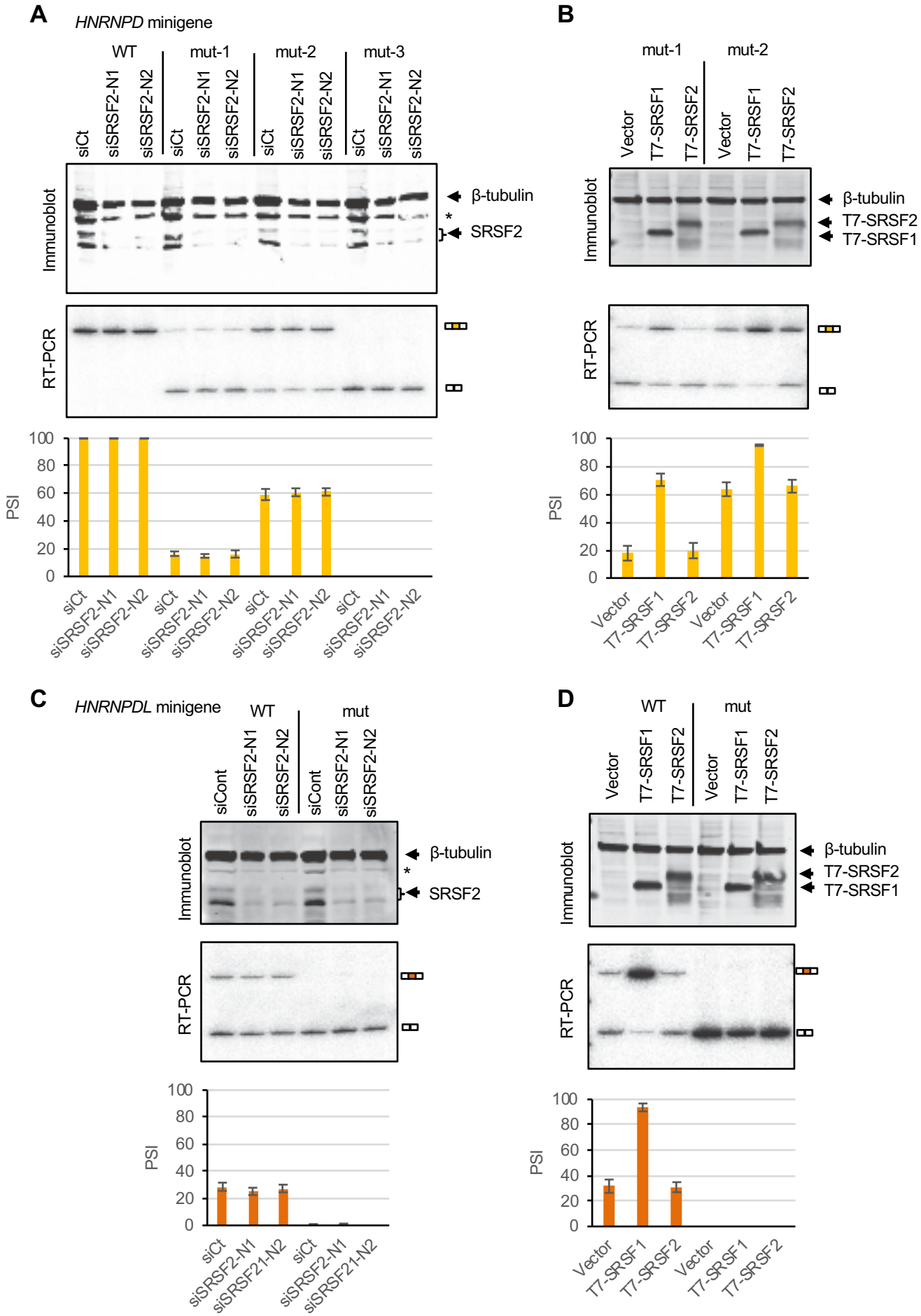

A
HNRNPD minigene
WT
mut-1
mut-2
mut-3
siCt
siSRSF2-N1
siSRSF2-N2
Immunoblot
β-tubulin
SRSF2
RT-PCR
PSI
B
mut-1
mut-2
Vector
T7-SRSF1
T7-SRSF2
T7-SRSF2
T7-SRSF1
C
HNRNPDL minigene
WT
mut
siCont
siSRSF21-N2
D

Figure S7: SRSF1 but not SRSF2 regulates *HNRNPD* and *HNRNPDL* exon inclusion. Related to Figure 7. (A,C) Splicing of the *HNRNPD* (A) and *HNRNPDL* (C) minigenes in HeLa cells treated with siRNAs targeting SRSF2 (siSRSF2-N1 and siSRSF2-N2). siRNA targeting luciferase (siCt) was used as control. Knockdown efficiency in each sample is shown by immunoblotting (top) and splicing products are shown using radioactive RT-PCR analysis (middle). Average exon inclusion level (percent spliced in or PSI) in each condition was quantified using three independent replicates (bottom). Error bars represent standard deviation (SD). An asterisk in immunoblots against SRSF2 indicates a band, the origin of which was not confirmed. (B,D) Splicing of the *HNRNPD* (A) *and HNRNPDL* (C) minigenes in HeLa cells treated with cDNAs expressing SRSF1 or SRSF2. cDNA expression in each sample is shown by immunoblotting (top) and splicing products are shown using radioactive RT-PCR analysis (middle). Average exon inclusion level (PSI) in each condition was quantified using three independent replicates (bottom). Error bars represent SD.