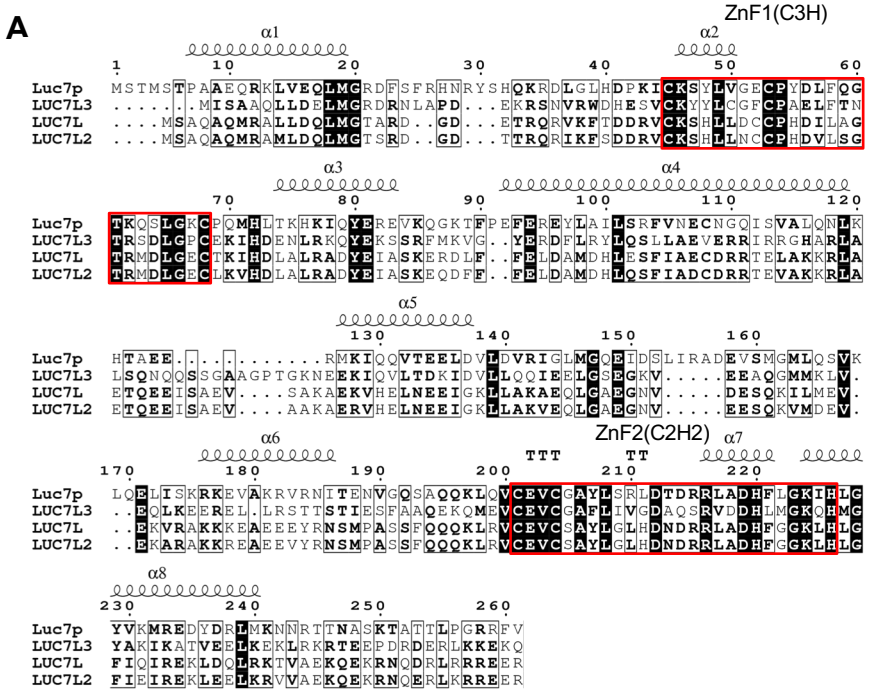
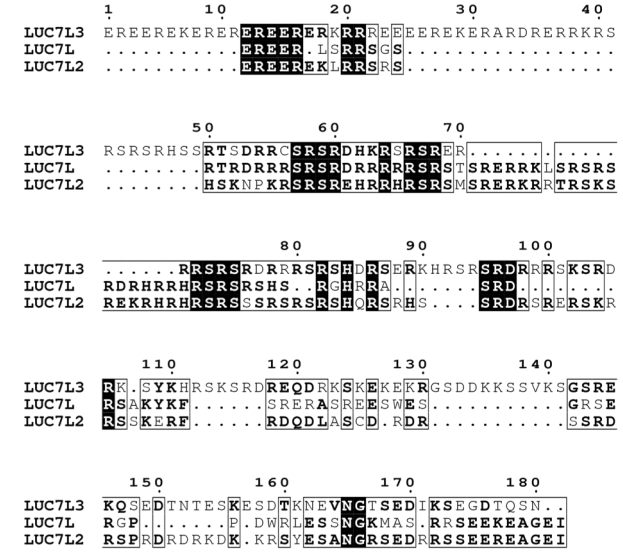
Cell Reports, Volume *35*

Supplemental information

Functional analyses of human LUC7-like proteins involved in splicing regulation and myeloid neoplasms

Noah J. Daniels, Courtney E. Hershberger, Xiaorong Gu, Caroline Schueger, William M. DiPasquale, Jonathan Brick, Yogen Saunthararajah, Jaroslaw P. Maciejewski, and Richard A. Padgett


A
ZnF1(C3H)
α1
α2
α3
α4
α5
α6
α7
α8
ZnF2(C2H2)
Luc7p
LUC7L3
LUC7L
LUC7L2


B
LUC7L3
LUC7L
LUC7L2

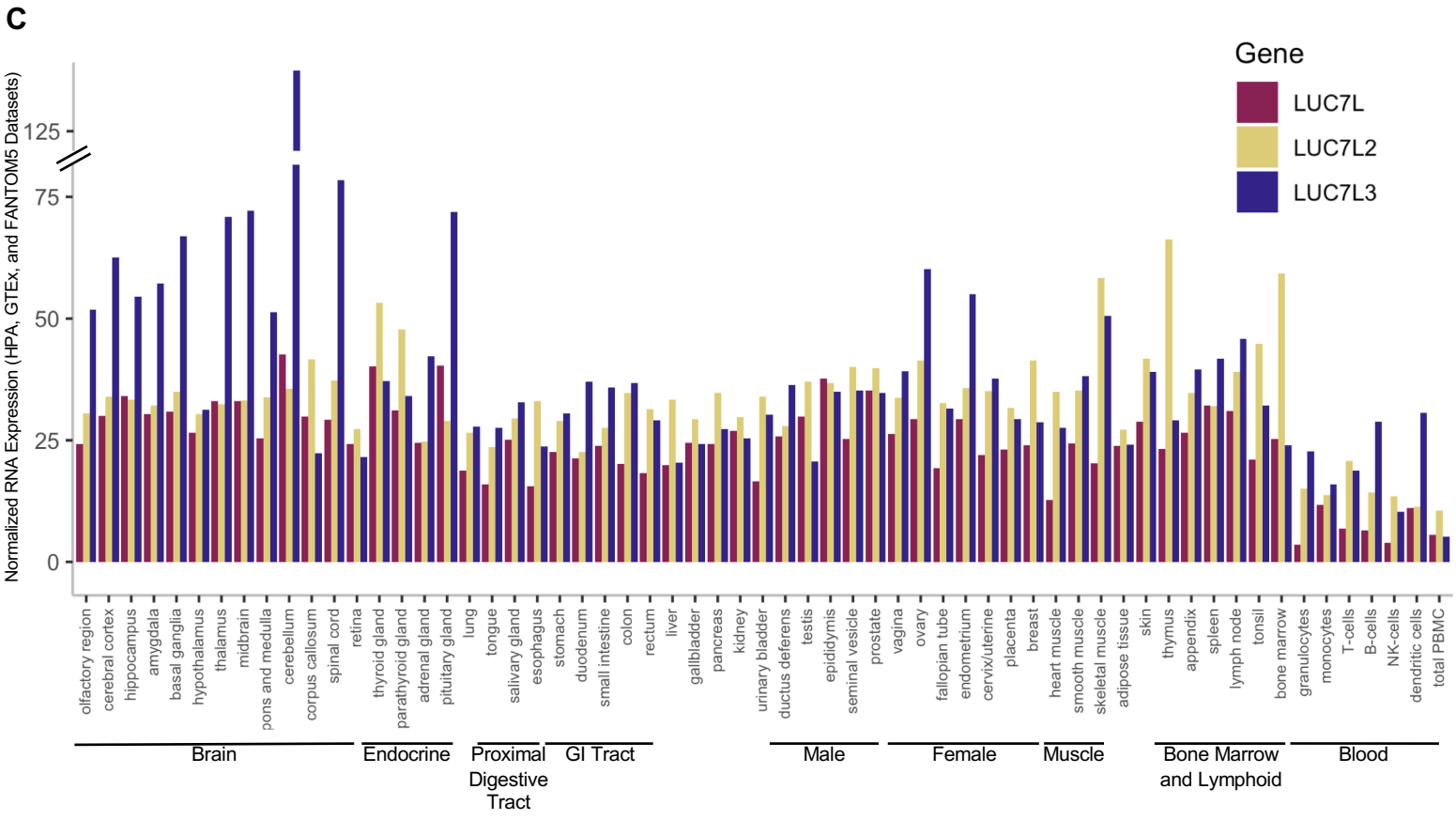

C
Gene
LUC7L
LUC7L2
LUC7L3
Normalized RNA Expression (HPA, GTEx, and FANTOM5 Datasets)
125
75
50
25
0
olfactory region
cerebral cortex
hippocampus
amygdala
basal ganglia
hypothalamus
thalamus
midbrain
pons and medulla
cerebellum
corpus callosum
spinal cord
retina
thyroid gland
parathyroid gland
adrenal gland
pituitary gland
lung
tongue
salivary gland
esophagus
stomach
duodenum
small intestine
colon
rectum
liver
gallbladder
pancreas
kidney
urinary bladder
ductus deferens
testis
epididymis
seminal vesicle
prostate
vagina
ovary
fallopian tube
endometrium
cervix/uterine
placenta
breast
heart muscle
smooth muscle
skeletal muscle
adipose tissue
skin
thymus
appendix
spleen
lymph node
tonsil
bone marrow
granulocytes
monocytes
T-cells
B-cells
NK-cells
dendritic cells
total PBMC
Brain
Endocrine
Proximal Digestive Tract
GI Tract
Male
Female
Muscle
Bone Marrow and Lymphoid
Blood

Figure S1: Amino acid conservation and tissue specific expression of the LUC7-like proteins. Related to Figure 1.

(a) Amino acid alignment of the N-terminal domains of the mammalian LUC7-like proteins along with the paralogous yeast protein Luc7p using CLUSTAL OMEGA 1.2.4 multiple sequence alignment tool and ESPript 3. Depicted are the conserved N-terminal α-helix, ZnF1 (CH3 type containing three cysteines, one histidine) and ZnF2 (C2H2 type containing two cysteines, two histidines). The structure of the coiled-coil domain is located between the two ZnFs. White letters with black background represent 100% conservation among the four proteins. Black letters with black frame represent conservation among three proteins. α and β-turns are depicted as TT and TTT. **(b)** Amino acid alignment of the arginine-glutamic acid rich (RE), arginine-serine rich (SR), and arginine rich (R) domains of the mammalian LUC7-like proteins following the second zinc finger. White letters with black background represent 100% conservation among the three proteins. Black letters with black frame represent conservation between two proteins. **(c)** RNA Expression (TMP) of the *LUC7*-like family across tissues from HPA, GTEX, and FANTOM databases [Uhlen et al., 2015].

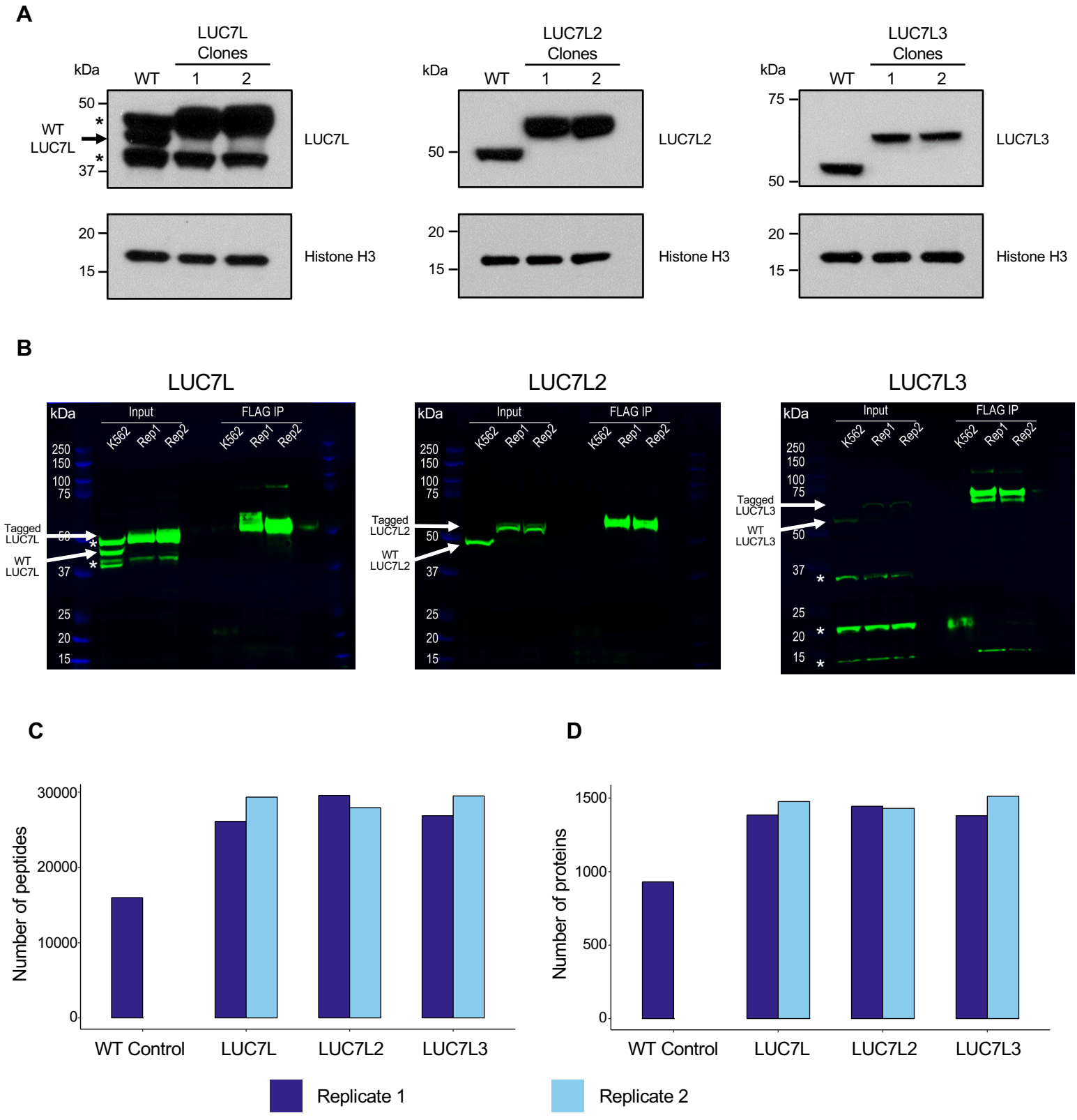


Figure S2: Co-Immunoprecipitation of the LUC7-like proteins. Related to Figure 1.
(a) LUC7-like genes homozygously CRISPR-tagged with V5, FLAG, HA in individual K562 clones shown by western blot (WB) using LUC7L, LUC7L2, and LUC7L3 antibodies, respectively. Epitope tagged LUC7L protein runs at the same size as the upper nonspecific band. Asterisks depict nonspecific bands. **(b)** Co-IP'd CRISPR-tagged LUC7-like cell lines using LUC7L, LUC7L2, and LUC7L3 antibodies respectively. WT and tagged proteins are depicted by arrows. Asterisks depict nonspecific bands. **(c)** Number of peptides analyzed in the Co-IP mass spectrometry experiments. **(d)** Number of proteins with at least one unique peptide identified from the Co-IP mass spectrometry experiments.

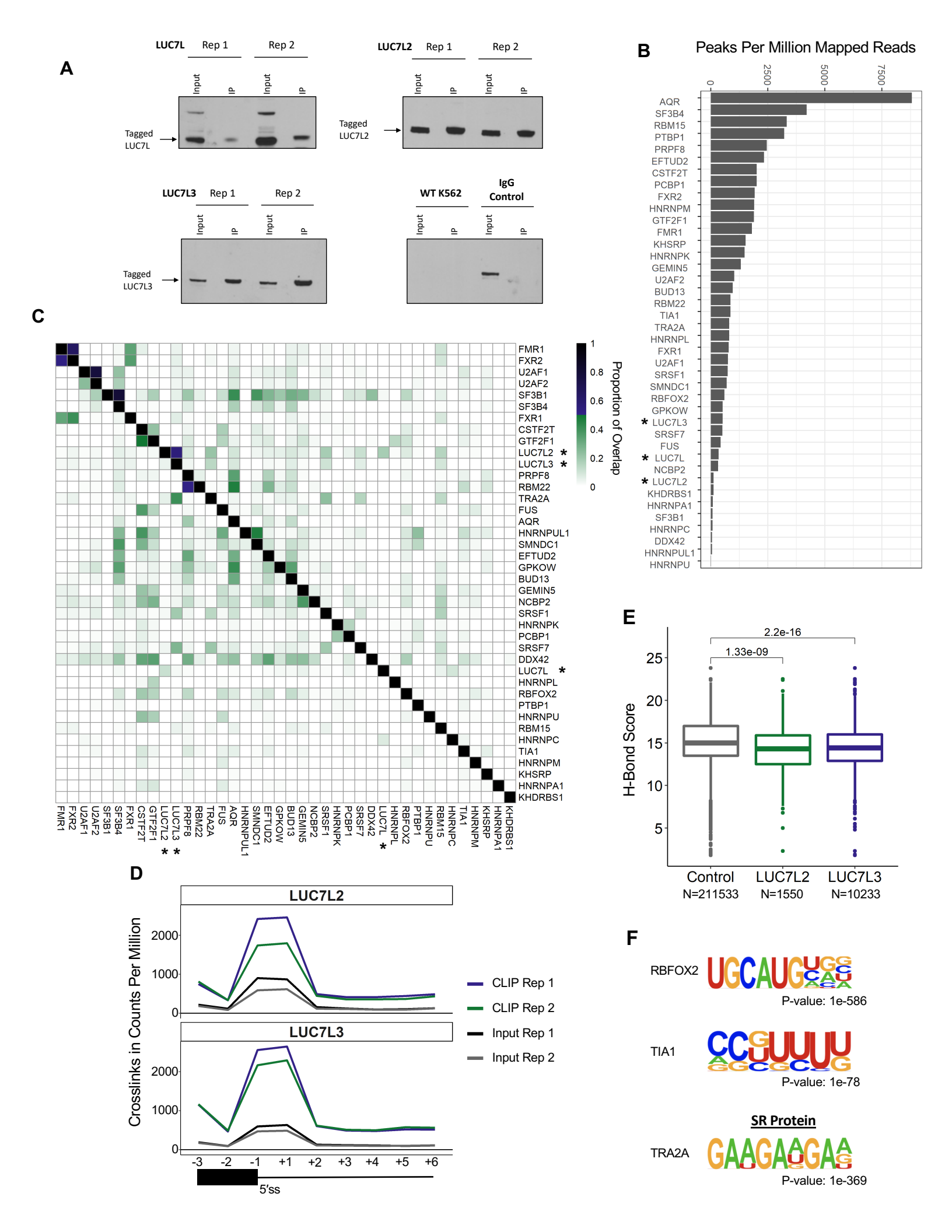

A
LUC7L
Rep 1
Rep 2
Input
IP
Tagged LUC7L
LUC7L2
Tagged LUC7L2
LUC7L3
Tagged LUC7L3
WT K562
IgG Control
B
Peaks Per Million Mapped Reads
0
2500
5000
7500
AQR
SF3B4
RBM15
PTBP1
PRPF8
EFTUD2
CSTF2T
PCBP1
FXR2
HNRNPM
GTF2F1
FMR1
KHSRP
HNRNPK
GEMIN5
U2AF2
BUD13
RBM22
TIA1
TRA2A
HNRNPL
FXR1
U2AF1
SRSF1
SMNDC1
RBFOX2
GPKOW
* LUC7L3
SRSF7
FUS
* LUC7L
NCBP2
* LUC7L2
KHDRBS1
HNRNPA1
SF3B1
HNRNPC
DDX42
HNRNPUL1
HNRNPU
C
Proportion of Overlap
1
0.8
0.6
0.4
0.2
0
FMR1
FXR2
U2AF1
U2AF2
SF3B1
SF3B4
FXR1
CSTF2T
GTF2F1
LUC7L2 *
LUC7L3 *
PRPF8
RBM22
TRA2A
FUS
AQR
HNRNPUL1
SMNDC1
EFTUD2
GPKOW
BUD13
GEMIN5
NCBP2
SRSF1
HNRNPK
PCBP1
SRSF7
DDX42
LUC7L *
HNRNPL
RBFOX2
PTBP1
HNRNPU
RBM15
HNRNPC
TIA1
HNRNPM
KHSRP
HNRNPA1
KHDRBS1
D
LUC7L2
LUC7L3
Crosslinks in Counts Per Million
2000
1000
0
-3
-2
-1
+1
+2
+3
+4
+5
+6
5'ss
CLIP Rep 1
CLIP Rep 2
Input Rep 1
Input Rep 2
E
2.2e-16
1.33e-09
H-Bond Score
25
20
15
10
5
Control
N=211533
LUC7L2
N=1550
LUC7L3
N=10233
F
RBFOX2
P-value: 1e-586
TIA1
P-value: 1e-78
SR Protein
TRA2A
P-value: 1e-369

Figure S3: seCLIP-Seq on CRISPR-tagged LUC7-like family members. Related to Figure 2.
(a) CLIP-Seq: Input and V5-IP'd lysates in CRISPR-tagged cell lines and controls. **(b)** The number of reproducible and significantly enriched peaks per million mapped reads using the biological replicate containing the fewest uniquely mapped reads (log2 fold-change ≥ 3, -log10 p-value ≥ 3, IDR ≤ 0.01) in CLIP-Seq experiments. **(c)** Proportion of significantly enriched peaks that overlap between the CLIP experiments. The LUC7-like experiments are depicted with a black asterisk. **(d)** CLIP crosslinking sites normalized to mapped library size in counts per million at a meta-5′ splice site containing all 5′SS in the human genome. 5′ splice site nucleotides are depicted as -3 – +8. Shown are CLIP and input replicates for LUC7L2 and LUC7L3. **(e)** U1 snRNA/5′ splice site hydrogen bonding score using H-Bond tool. Control group contains all 5′ splice sites used in CLIP-Seq mapping. LUC7L2 and LUC7L3 groups contain enriched 5′ splice sites where there is an enriched crosslinking site at either position -1 or +1. A Wilcoxon rank-sum test was performed to determine significance. **(f)** Binding motifs enriched in significant CLIP-peaks identified in RBFOX3, TIA1, and the SR protein TRA2A.

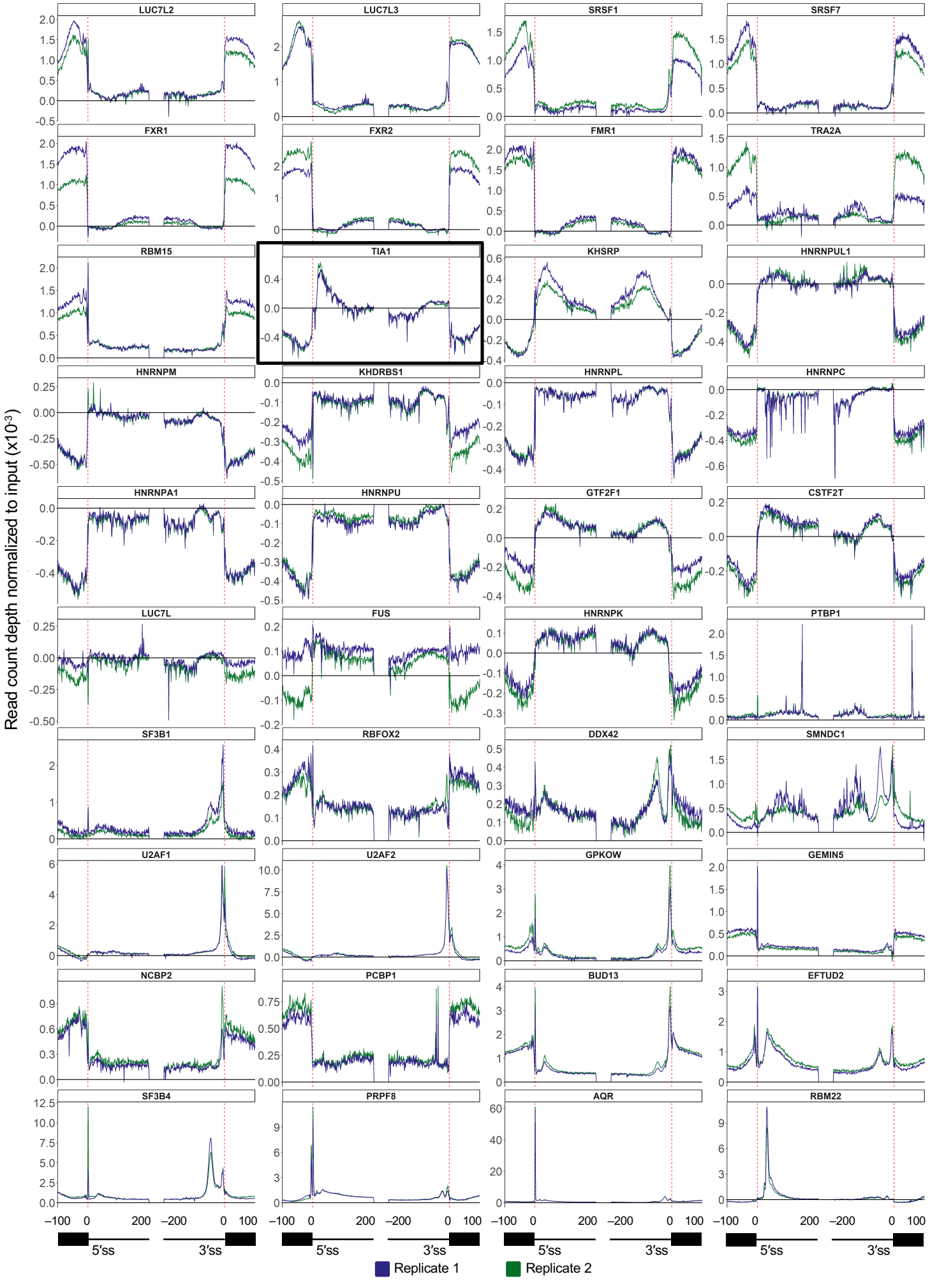
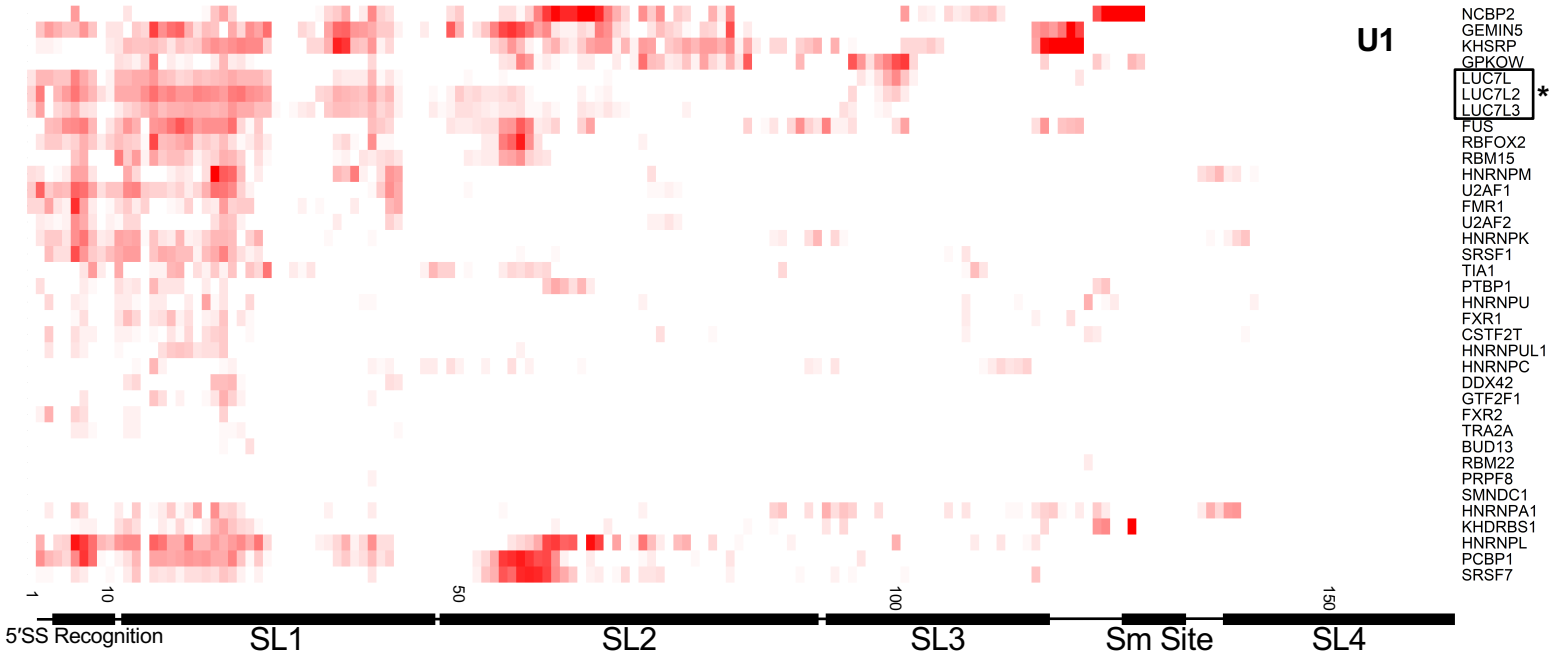
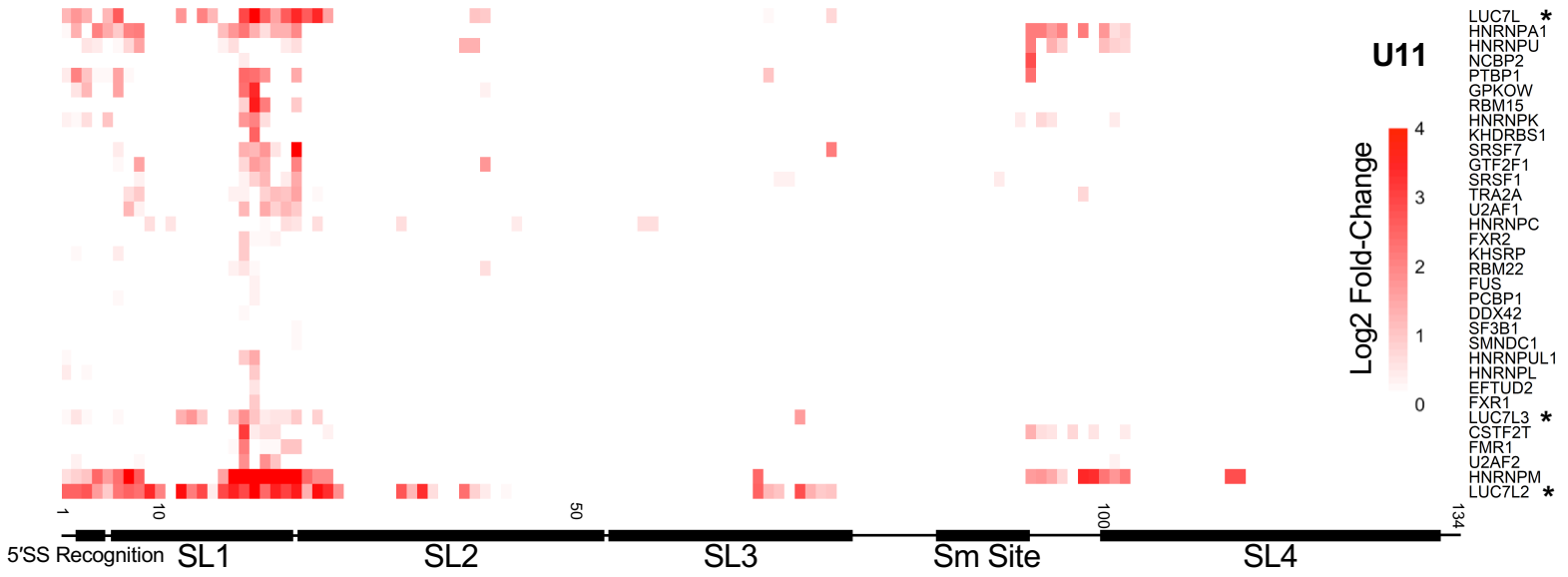

LUC7L2
LUC7L3
SRSF1
SRSF7
FXR1
FXR2
FMR1
TRA2A
RBM15
TIA1
KHSRP
HNRNPUL1
HNRNPM
KHDRBS1
HNRNPL
HNRNPC
HNRNPA1
HNRNPU
GTF2F1
CSTF2T
LUC7L
FUS
HNRNPK
PTBP1
SF3B1
RBFOX2
DDX42
SMNDC1
U2AF1
U2AF2
GPKOW
GEMIN5
NCBP2
PCBP1
BUD13
EFTUD2
SF3B4
PRPF8
AQR
RBM22
Read count depth normalized to input (x10-3)
−100 0 200 −200 0 100
5′ss
3′ss
Replicate 1
Replicate 2

Figure S4: RBP binding profiles on generic 5′SS and 3′SS metagenes. Related to Figure 2.
The number of crosslink sites, normalized to input crosslink sites at each base pair of all annotated human splice junctions. Depicted are binding data for 100 nucleotides into the exon and 200 nucleotides into the intron downstream and upstream of the 5′SS and 3′SS, respectively. Anything above the 0-y-axis threshold depicted by a bold black line is enriched binding over the input control.


U1
NCBP2
GEMIN5
KHSRP
GPKOW
LUC7L
LUC7L2
LUC7L3
*
FUS
RBFOX2
RBM15
HNRNPM
U2AF1
FMR1
U2AF2
HNRNPK
SRSF1
TIA1
PTBP1
HNRNPU
FXR1
CSTF2T
HNRNPUL1
HNRNPC
DDX42
GTF2F1
FXR2
TRA2A
BUD13
RBM22
PRPF8
SMNDC1
HNRNPA1
KHDRBS1
HNRNPL
PCBP1
SRSF7
1
10
50
100
150
5′SS Recognition
SL1
SL2
SL3
Sm Site
SL4


U11
LUC7L
*
HNRNPA1
HNRNPU
NCBP2
PTBP1
GPKOW
RBM15
HNRNPK
KHDRBS1
SRSF7
GTF2F1
SRSF1
TRA2A
U2AF1
HNRNPC
FXR2
KHSRP
RBM22
FUS
PCBP1
DDX42
SF3B1
SMNDC1
HNRNPUL1
HNRNPL
EFTUD2
FXR1
LUC7L3
*
CSTF2T
FMR1
U2AF2
HNRNPM
LUC7L2
*
Log2 Fold-Change
4
3
2
1
0
1
10
50
100
134
5′SS Recognition
SL1
SL2
SL3
Sm Site
SL4

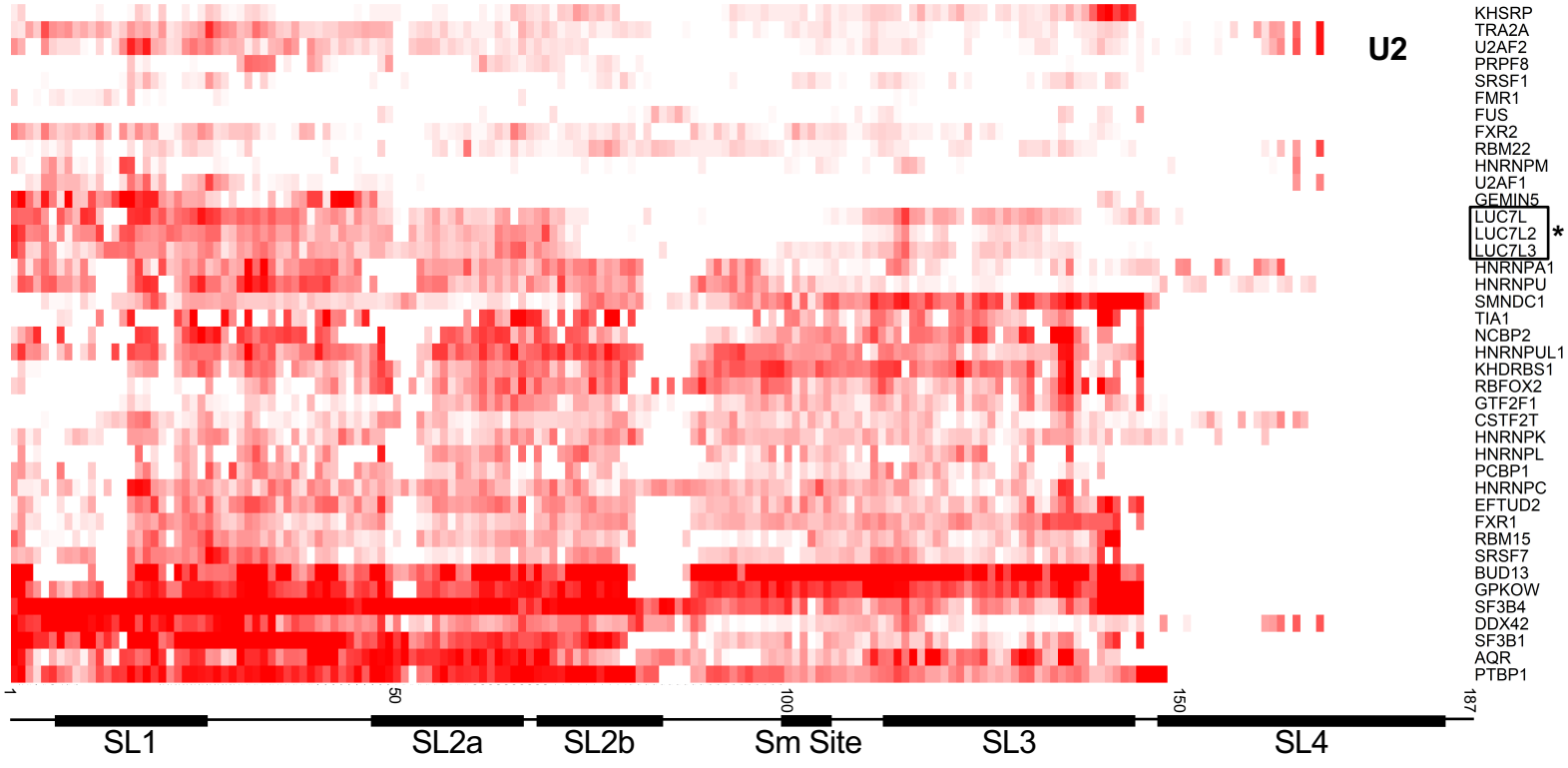

U2
KHSRP
TRA2A
U2AF2
PRPF8
SRSF1
FMR1
FUS
FXR2
RBM22
HNRNPM
U2AF1
GEMIN5
LUC7L
LUC7L2
*
LUC7L3
HNRNPA1
HNRNPU
SMNDC1
TIA1
NCBP2
HNRNPUL1
KHDRBS1
RBFOX2
GTF2F1
CSTF2T
HNRNPK
HNRNPL
PCBP1
HNRNPC
EFTUD2
FXR1
RBM15
SRSF7
BUD13
GPKOW
SF3B4
DDX42
SF3B1
AQR
PTBP1
1
50
100
150
187
SL1
SL2a
SL2b
Sm Site
SL3
SL4

Figure S5: RBP crosslinking sites on snRNAs. Related to Figure 3.
Single nucleotide resolution crosslinking maps on U1, U11, and U2 snRNAs shown as the averaged replicate log2 fold-change enrichment over the input control.

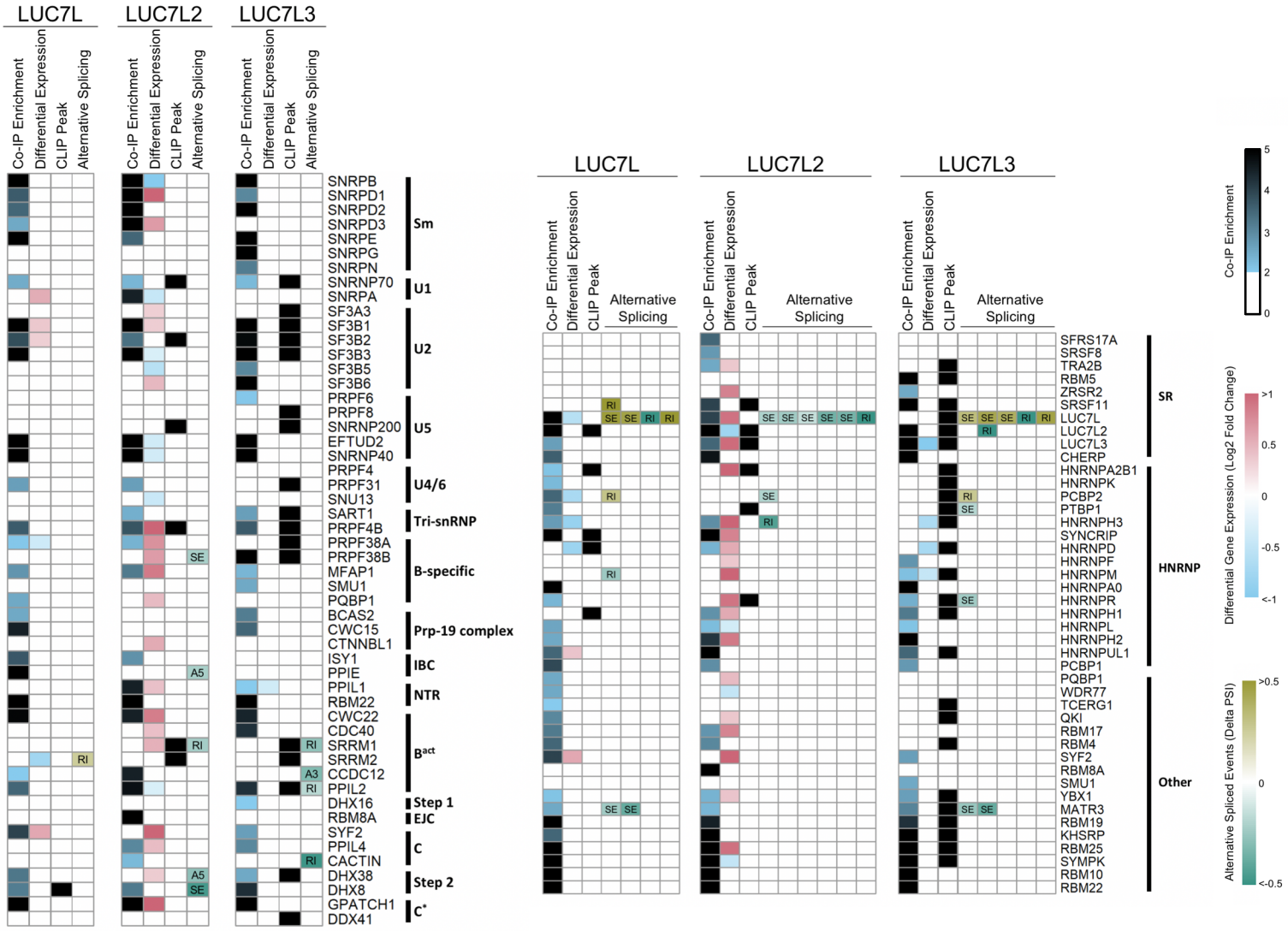


Figure S6: LUC7L2 knockdown reveals a complex interplay with its protein interactors. Related to Figures 1 and 6.

Extension of Figure 1g that includes the significant Co-IP enrichment, differential expression (FDR ≤ 0.05), RNA binding, and alternative splicing events (ΔPSI 10%, q-value ≤ 0.05) of SFs ordered by their appearance in sub-spliceosomal complexes (left) as well as factors involved in alternative splicing (right) split by each LUC7-like protein. A black cell in the CLIP-Peak column depicts whether there is at least 1 significant CLIP-Peak (log2 fold-change ≥ 3, -log10 p-value ≥ 3, IDR ≤ 0.01) on the gene in question.