Supplemental Material

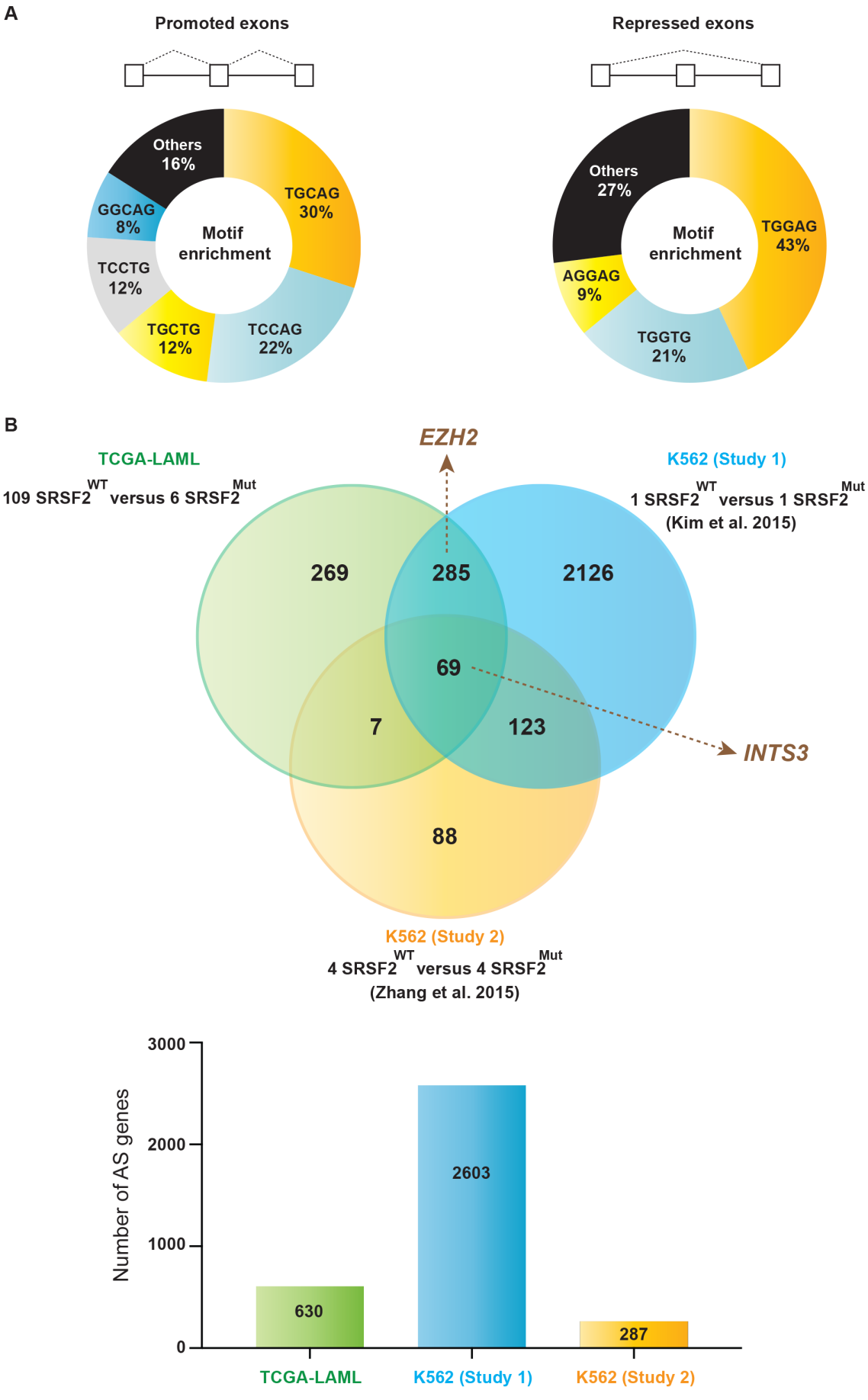
Recurrent SRSF2 Mutations in MDS Affect Both Splicing and NMD

Mohammad Alinoor Rahman[1], Kuan-Ting Lin[1], Robert K. Bradley[2,3], Omar Abdel-Wahab[4], and Adrian R. Krainer[1,*]

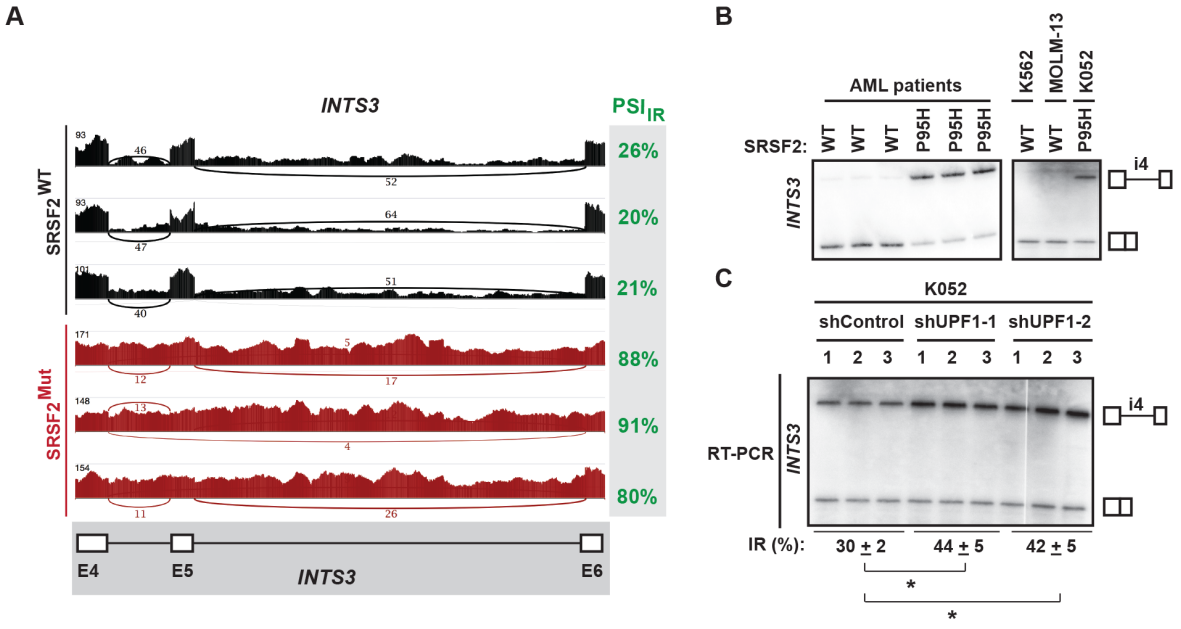
[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[2]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[3]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[4]Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
*Corresponding author: krainer@cshl.edu

Supplemental Material includes:

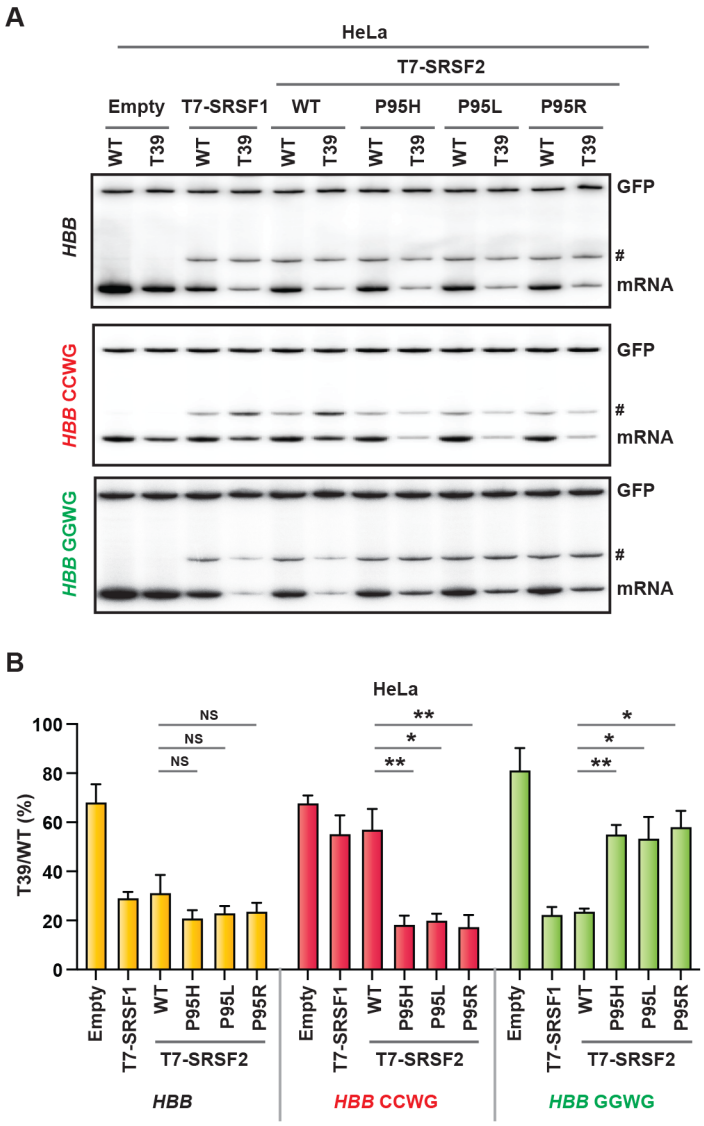
Supplemental Figures: S1-S7
Supplemental Tables: S1-S5 (Table S1 is uploaded separately as excel file)
Additional materials and methods


A
Promoted exons
Others 16%
TGCAG 30%
GGCAG 8%
Motif enrichment
TCCTG 12%
TGCTG 12%
TCCAG 22%
Repressed exons
Others 27%
TGGAG 43%
Motif enrichment
AGGAG 9%
TGGTG 21%
B
EZH2
TCGA-LAML
109 SRSF2WT versus 6 SRSF2Mut
K562 (Study 1)
1 SRSF2WT versus 1 SRSF2Mut
(Kim et al. 2015)
269
285
2126
69
7
123
INTS3
88
K562 (Study 2)
4 SRSF2WT versus 4 SRSF2Mut
(Zhang et al. 2015)
Number of AS genes
3000
2000
1000
0
630
2603
287
TCGA-LAML
K562 (Study 1)
K562 (Study 2)

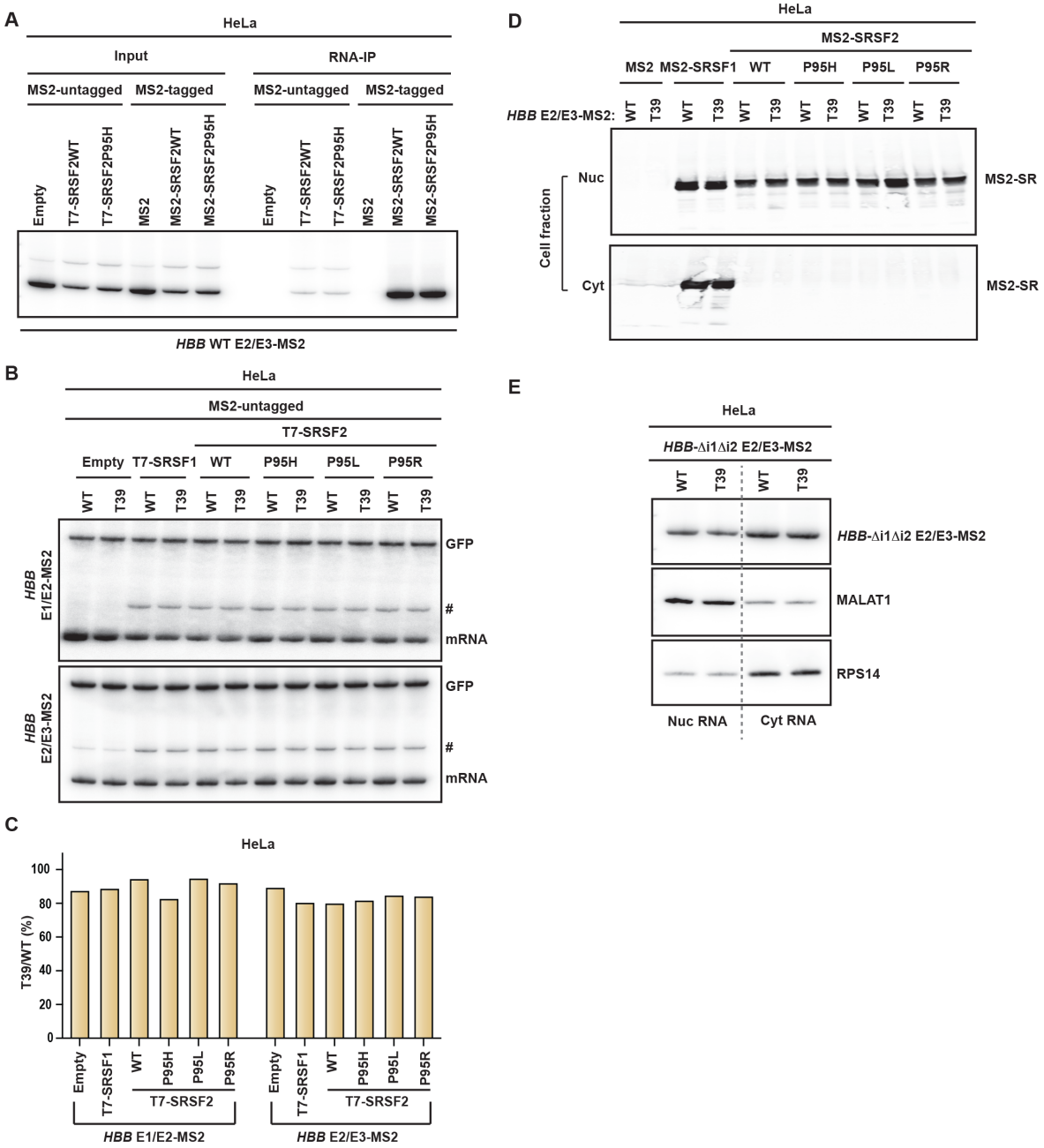
Supplemental Figure S1. Altered splicing events in acute myeloid leukemia patients and isogenic cells with oncogenic mutations in Pro95 of SRSF2, Related to Fig. 1. (*A*) Pie charts showing the distribution of different motifs enriched in promoted or repressed exons in $SRSF2^{Mut}$ patients in the TCGA-LAML RNA-seq data set. (*B*) Top, Venn diagram comparing genes with altered splicing due to Pro95 mutations of SRSF2 in TCGA-LAML RNA-seq data and two previous RNA-seq studies in K562 cells (Kim et al. 2015; Zhang et al. 2015). Bottom, bar graph showing number of altered splicing genes in each RNA-seq data set.


A
INTS3
PSI IR
SRSF2 WT
SRSF2 Mut
26%
20%
21%
88%
91%
80%
E4
E5
E6
INTS3
B
AML patients
K562
MOLM-13
K052
SRSF2:
WT
P95H
i4
INTS3
C
K052
shControl
shUPF1-1
shUPF1-2
RT-PCR
INTS3
i4
IR (%):
30 ± 2
44 ± 5
42 ± 5
*
*
D
HeLa
Total cell
Nuclear fraction
Cytoplasmic fraction
Empty
T7-SRSF1
T7-SRSF2
WT
P95H
P95L
P95R
Tubulin
T7-tagged protein
E
HeLa
T7-SRSF2
Empty
T7-SRSF1
WT
P95H
P95L
P95R
EZH2:
CCWG
GGWG
GFP
inc
EZH2
skp

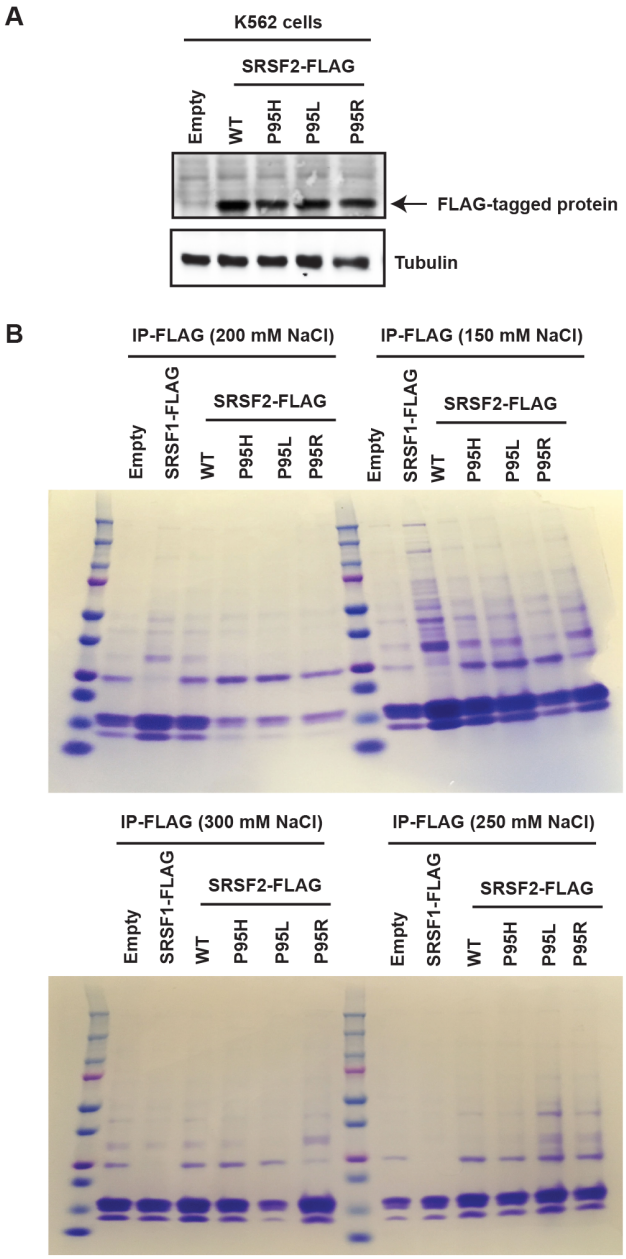
Supplemental Figure S2. Validation of key AS-NMD alterations promoted by mutant SRSF2, Related to Fig. 1. (*A*) Representative Sashimi plots showing intron retention events in *INTS3* in AML samples with the genotypes indicated. (*B*) Radioactive RT-PCR of intron 4 retention event in *INTS3* in an independent set of SRSF2 WT and mutant AML samples, and in three different human leukemia cells: K562, MOLM-13 and K052. (*C*) Radioactive RT-PCR of intron 4 retention event in *INTS3* in SRSF2P95H mutant K052 cells, with or without UPF1 knockdown (mean ± SD, n=3, *p<0.05, t test). (*D*) Western-blotting analysis of T7-tagged SRSF1, SRSF2 WT, and SRSF2 Pro95 mutants (P95H, P95L, P95R) transfected HeLa cells in total cell lysate, nuclear or cytoplasmic fractions, using anti-T7-tag and anti-β-Tubulin (Tubulin) antibodies. (*E*) Radioactive RT-PCR analysis of *EZH2* reporters (shown in Figure 1H) in HeLa cells co-transfected with the indicated T7-tagged empty vector or SR protein. Note that experiments in panels (*B*) and (*C*) were done from the same samples as described in Fig. 1E and 1F, respectively. Therefore, control gels in Fig. 1E and 1F are also comparable as control for Supplemental Fig. S2B and S2C, respectively.



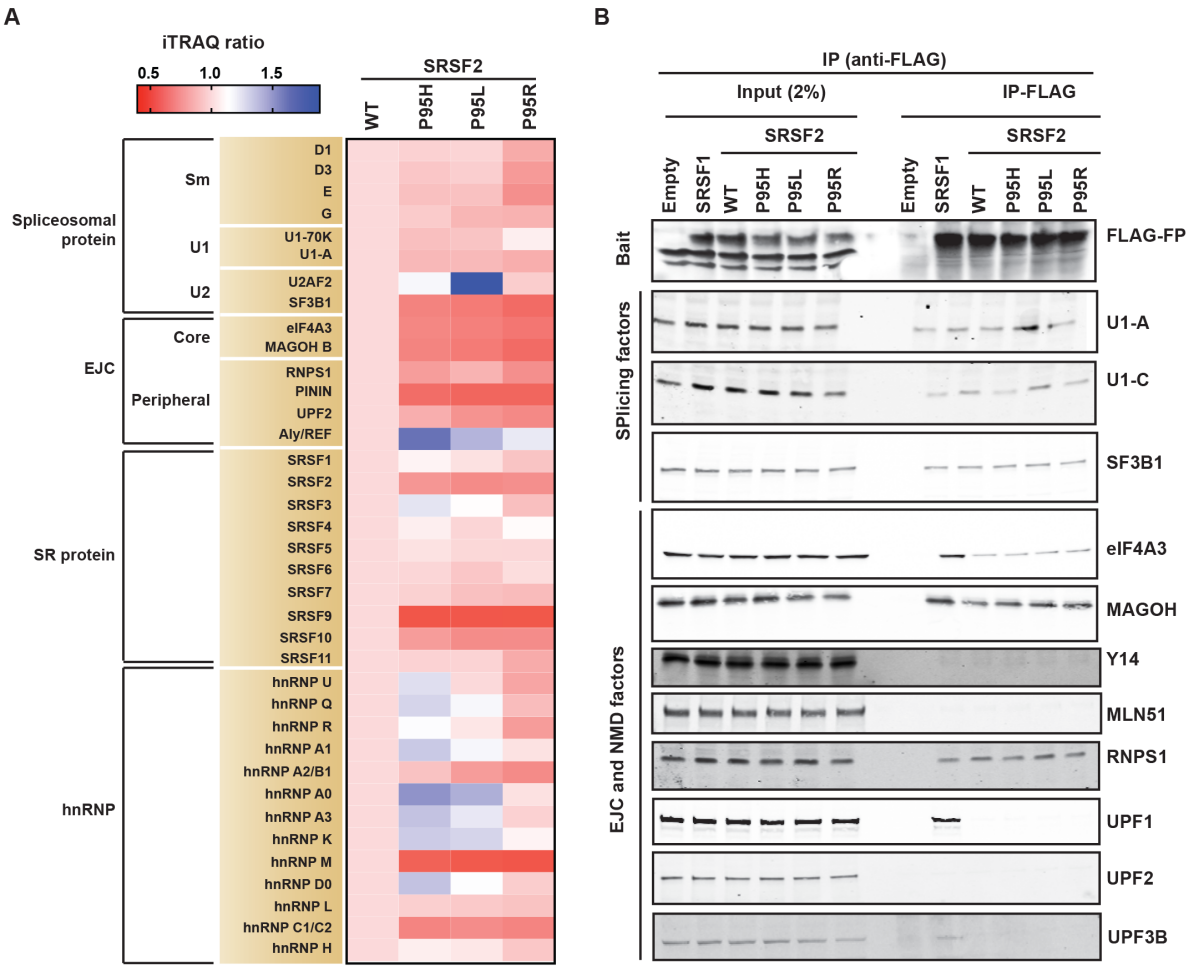
Supplemental Figure S3. SRSF2-promoted NMD is dependent on sequence-specific RNA-binding, Related to Fig. 2. (*A*) Radioactive RT-PCR of the *HBB* reporters co-transfected with the indicated T7-tagged cDNAs into HeLa cells (see Fig. 2 for the constructs' design). An intron 1 retained transcript variant is marked by a hash (#). (*B*) mRNA bands in panel (*A*) were quantified, normalized to GFP, and plotted as T39/WT (%) (mean ± SD, n=3, *p<0.05, **p<0.01, NS: not significant, t test).


A
HeLa
Input
RNA-IP
MS2-untagged
MS2-tagged
MS2-untagged
MS2-tagged
Empty
T7-SRSF2WT
T7-SRSF2P95H
MS2
MS2-SRSF2WT
MS2-SRSF2P95H
Empty
T7-SRSF2WT
T7-SRSF2P95H
MS2
MS2-SRSF2WT
MS2-SRSF2P95H
HBB WT E2/E3-MS2
B
HeLa
MS2-untagged
T7-SRSF2
Empty
T7-SRSF1
WT
P95H
P95L
P95R
WT T39 WT T39 WT T39 WT T39 WT T39 WT T39
HBB E1/E2-MS2
GFP
#
mRNA
HBB E2/E3-MS2
GFP
#
mRNA
C
HeLa
T39/WT (%)
100
80
60
40
20
0
Empty
T7-SRSF1
WT
P95H
P95L
P95R
T7-SRSF2
HBB E1/E2-MS2
Empty
T7-SRSF1
WT
P95H
P95L
P95R
T7-SRSF2
HBB E2/E3-MS2
D
HeLa
MS2-SRSF2
MS2
MS2-SRSF1
WT
P95H
P95L
P95R
HBB E2/E3-MS2:
WT T39 WT T39 WT T39 WT T39 WT T39 WT T39
Cell fraction
Nuc
MS2-SR
Cyt
MS2-SR
E
HeLa
HBB-Δi1Δi2 E2/E3-MS2
WT T39 WT T39
HBB-Δi1Δi2 E2/E3-MS2
MALAT1
RPS14
Nuc RNA
Cyt RNA

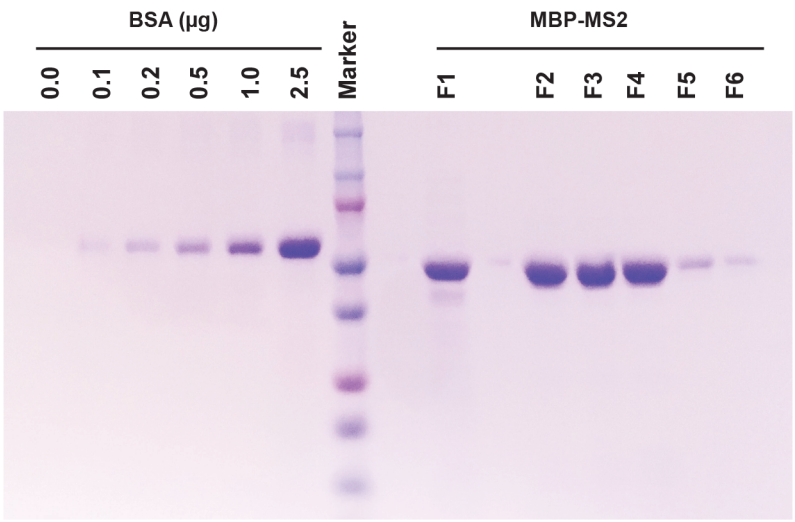
Supplemental Figure S4. SRSF2 induces NMD when tethered downstream of a PTC, an effect coupled with intron splicing, Related to Fig. 3. (*A*) Radioactive RT-PCR of RNA extracted from IP (RNA-IP). HeLa cells were co-transfected with the *HBB* WT E2/E3-MS2 reporter, along with either MS2-tagged or MS2-untagged cDNA. Note that all cDNAs harbor a T7-tag. Cells were processed for IP using anti-T7-tag antibody, followed by extraction of RNA. Radioactive RT-PCR was performed to amplify the *HBB* WT E2/E3-MS2 reporter-specific transcript to assess enrichment with MS2-tagged versus MS2-untagged proteins. (*B*) Radioactive RT-PCR analysis of splicing of the *HBB* E1/E2-MS2 and E2/E3-MS2 reporters in HeLa cells co-transfected with the indicated T7-tagged empty vector or SR protein. An intron 1 retained transcript variant is marked by a hash (#). (*C*) mRNA bands in panel (*B*) were quantified, normalized to GFP, and plotted as T39/WT (%). (*D*) Western-blotting of nuclear or cytoplasmic fractions from HeLa cells co-transfected with MS2-fused T7-tagged empty vector or SR protein and *HBB*-E2/E3-MS2 reporters (WT or T39), using anti-T7-tag antibody. Note that MS2 cDNA does not have a T7-tag. (*E*) Radioactive RT-PCR of the RNA extracted from the nuclear (nuc) and cytoplasmic (cyt) fractions of HeLa cells transfected with *HBB*-Δi1Δi2 E2/E3-MS2 reporters. MALAT1 lncRNA and ribosomal protein S14 (RPS14) RNA served as internal controls, due to their exclusive expression in the nucleus and cytoplasm, respectively.



Supplemental Figure S5. Co-IP to purify interacting partners of SRSF2 WT versus Pro95 mutants, Related to Fig. 4. (*A*) Western-blotting using anti-FLAG antibody of isogenic K562 cells expressing an empty vector or SRSF2 (WT/P95H/P95L/P95R) with a C-terminal FLAG-tag, generated by lentiviral transduction. (*B*) Coommassie staining of purified proteins immunoprecipitated with anti-FLAG antibody from the indicated isogenic K562 cells, in the presence of Benzonase nuclease and RNase A+T1 cocktail. IPs were done using binding buffers with the indicated salt concentrations, to determine optimal conditions for downstream iTRAQ mass spectrometry. Immunopurified proteins from the buffer with 200 mM NaCl were used in iTRAQ mass spectrometry.



Supplemental Figure S6. Compendium of interacting partners of SRSF2 WT versus Pro95 mutants, Related to Fig. 4 and supplemental Fig. S5. (*A*) Heat map showing interaction of selected proteins (related to splicing and NMD) with SRSF2 WT versus Pro95 mutants (P95H/P95L/P95R) in isogenic K562 cells in the presence of Benzonase nuclease and RNase A+T1 cocktail analyzed by iTRAQ mass spectrometry. (*B*) IP-westerns to validate several interactions from iTRAQ mass spectrometry in panel (*A*) in the presence of Benzonase nuclease and RNase A+T1 cocktail using the indicated antibodies. FLAG-FP denotes FLAG-fusion protein. Among the splicing factors, the interactions between SRSF2 (both WT and mutants) and U1-A and SF3B1 were confirmed. U1-C was not detected by mass spectrometry, but was detected by IP-western. Among the EJC factors, interactions of SRSF2 (both WT and mutants) were confirmed with eIF4A3, MAGOH, and RNPS1, but were not detected with Y14 or MLN51. In contrast, interactions of SRSF2 (WT or mutants) with any UPF proteins were not detected by IP-western, although UPF2 was enriched by mass spectrometry.



Supplemental Figure S7. Purification of recombinant MBP-MS2, Related to Fig. 5. Coomassie staining of purified recombinant MS2-tagged maltose binding protein (MBP-MS2). Proteins were aliquoted in several consecutive fractions (F1-F6). Different concentrations of bovine serum albumin (BSA) were run in parallel as control.

Table S2: Expression level of EJC and NMD factors (SRSF2 mutant vs. SRSF2 WT) in RNA-seq data of TCGA-LAML and K562 cells (Zhang et al. 2015). Related to Fig. 4 and 5.

Gene Symbol	TCGA-LAML		K562	
	Log_2(Fold Change)	p-value	Log_2(Fold Change)	p-value
EIF4A3	0.5280	0.4865	0.0328	0.7179
RBM8A	0.3473	0.6290	-0.2241	0.0379
MAGOH	0.4067	0.6483	-0.0625	0.7500
CASC3	0.5161	0.4408	0.0546	0.6871
RNPS1	0.4837	0.5399	-0.1361	0.1252
UPF1	0.2195	0.7548	0.0056	0.9052
UPF2	0.6949	0.1978	0.2077	0.3066
UPF3A	1.0926	0.1651	0.0930	0.4541
UPF3B	0.3268	0.5970	-0.1056	0.4407
SMG1	0.4002	0.4125	0.2110	0.3786
SMG5	-0.0328	0.9593	0.0599	0.7048
SMG6	0.4010	0.5434	0.0297	0.7825
SMG8	-0.0138	0.9844	0.2643	0.2863
SMG9	0.4675	0.5482	-0.1269	0.3649
DHX34	0.2968	0.6447	-0.3111	0.0919

Table S3: List of primers

Name	Sequence
HBB RT-F	TGCACCTGACTCCTGAGGAGAA
HBB RT-R	GGACTTAGGGAACAAAGGAACCT
HBB qRT-F	ACTTCAGGCTCCTGGGCAAC
HBB qRT-R	CAGCAAGAAAGCGAGCTTAGTG
SRSF3-F	TGGAACTGTCGAATGGTGAA
SRSF3-R	GGGTGGTGAGAAGAGACATGA
18*S* rRNA-F	GTAACCCGTTGAACCCCATT
18*S* rRNA-R	CCATCCAATCGGTAGTAGCG
EZH2 RT-F	TTTCATGCAACACCCAACACT
EZH2 RT-R	CCCTGCTTCCCTATCACTGT
EZH2 INC-F	CAGCATTTGCCACTCCTACC
EZH2 INC-R	AGAGCAGCAGCAAACTCCTTT
EZH2 SKP-F	CAGCATTTGGAGGGAGCA
EZH2 SKP-R	GCTGGGCCTGCTACTGTTATT
INTS3-F	TGAGTCGTGATGGCATGAAT
INTS3-R	TCTTCACCAGTTCCCGTACC
GFP-F	AAGTTCATCTGCACCACCG
GFP-R	TGTACAGCTCGTCCATGCCGAGAGTG
RPS14-F	GTGTCTGCCATATCTTTGCATCC
RPS14-R	GGTGAGGATTCATCTCGGTCTG
MALAT1-F	TGTGTGCCAATGTTTCGTTT
MALAT1-R	AGGAGAAAGTGCCATGGTTG

Table S4: List of antibodies

Name	Company	Catalog Number/Reference
Anti-T7-tag	Cold Spring Harbor Laboratory	T7-KLH 42 1-87
Anti-eIF4A3	Cold Spring Harbor Laboratory	Aznarez et al. 2018
Anti-MAGOH	ABCAM	ab38768
Anti-Y14	SIGMA ALDRICH	Y1253
Anti-RNPS1	Dr. Akila Mayeda, Fujita Health University, Toyoake, Japan	Mayeda et al. 1999
Anti-CASC3 (MLN51)	Bethyl Lab. Inc.	A302-472A
Anti UPF1	Cell Signaling	9435S
Anti-UPF2	ABCAM	ab157108
Anti-eRF1	ABCAM	ab30928
Anti-eRF3	ABCAM	ab126090
Anti-U1A (snRNP A)	Thermo Fisher Scientific	PA5-27474
Anti-U1C (snRNP C)	SIGMA ALDRICH	SAB4200188
Anti-β-Tubulin III	Genscript	A01203
Anti-SF3B1	Bethyl Lab. Inc.	A300-996A

Table S5: List of siRNAs

Name	Sequence	Reference
eiF4A3	CGAGCAAUCAAGCAGAUCA	Lee et al. 2009
MAGOH	AGAGGCUUAUGUACAUAAA	Muromoto et al. 2009
Y14	CGCTCTGTTGAAGGCTGGA	Gehring et al. 2003
UPF1	GAUGCAGUUCCGCUCCAUU	Lee et al. 2009
Luciferase	GCCAUUCUAUCCUCUAGAGGAUG	Aznarez et al. 2018

Additional materials and methods

mRNA decay rate assay

HeLa cells were transfected with 1 μg of plasmid mixture (0.1 μg of GFP, 0.3 μg of SR protein expression plasmid or empty vector, and 0.6 μg of NMD reporter construct), per well of a 6-well plate. Six hours after transfection, cells were washed once with PBS and transcription was halted by resuspending in fresh medium containing 5 μg/ml actinomycin D. Cells were harvested at different time points to assess the decay rate of reporter mRNA, as described (Nomakuchi et al. 2016). Reporter mRNAs were quantified by RT–qPCR, normalized to transfected GFP RNA levels, and expressed as a percentage of the levels at time 0. Data were plotted using GraphPad Prism 7 software.

iTRAQ mass spectrometry

Tryptic digestion and iTRAQ labeling. The beads were reconstituted with 20 μL of 5% sodium dodecyl sulfate (SDS), 50 mM triethylammonium bicarbonate buffer (TEAB). RapiGest was added to a final concentration of 0.1% and tris(2-carboxyethyl)phosphine (TCEP) was added to final concentration of 5 mM. Samples were then heated to 55 °C for 20 min, allowed to cool to room temperature, and methyl methanethiosulfonate (MMTS) was added to a final concentration of 10 mM. Samples were incubated at room temperature for 20 min to complete blocking of free sulfhydryl groups. Samples were then digested using S-Trap Micro sample prep cartridges (Protifi) according to the manufacturer's instructions. After digestion, peptides were dried under vacuum and reconstituted in 50 μL of 0.5 M TEAB/70% isopropanol and labeled with 8-plex iTRAQ reagent for 1 hour at room temperature, as described (Ross et al. 2004). Labeled samples were then acidified to pH 4.0 using formic acid, combined and concentrated under vacuum until ~10.0 μL remained. Peptides were then fractionated using a Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo Scientific) according to the manufacturer's instructions, with slight modifications. Briefly, peptides were reconstituted in 150 μL of 0.1% TFA, loaded onto the spin column and centrifuged at 3000 x g for 2 min. The column was washed with water, and peptides were eluted with the following percentages of acetonitrile (ACN) in 0.1% triethylamine (TEA): 5%, 10%, 15%, and 50%. Each of the 4 fractions was then separately injected into the mass spectrometer, using capillary reverse phase LC at low pH.

Mass spectrometry (MS). An Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific), equipped with a nano-ion spray source was coupled to an EASY-nLC 1200 system (Thermo

Scientific). The LC system was configured with a self-pack PicoFrit™ 75-µm analytical column with an 8-µm emitter (New Objective, Woburn, MA) packed to 25 cm with ReproSil-Pur C18-AQ, 1.9 µM material (Gilar et al. 2005). Mobile phase A consisted of 2% acetonitrile; 0.1% formic acid, and mobile phase B consisted of 90% acetonitrile; 0.1% formic Acid. Peptides were then separated using a time span of 84 min linear gradient starting at 6% B and reaching 30 % B.

Eluted peptides were directly electrosprayed into the Fusion Lumos mass spectrometer with the application of a distal 2.3 kV spray voltage and a capillary temperature of 300 °C. Full-scan mass spectra (Res=60,000; 400-1600 m/z) were followed by MS/MS using the "Top Speed" method for selection. High-energy collisional dissociation (HCD) was used with the normalized collision energy set to 35 for fragmentation, the isolation width set to 1.2, and a duration of 10 sec was set for the dynamic exclusion with an exclusion mass width of 10 ppm. We used monoisotopic precursor selection for charge states 2+ and greater, and all data were acquired in profile mode.

Database searching. Peaklist files were generated by Mascot Distiller (Matrix Science). Protein identification and quantification were carried out using Mascot 2.4 (Perkins et al. 1999) against the UniProt human sequence database (92,919 sequences; 36,868,442 residues). Methylthiolation of cysteine and N-terminal lysine iTRAQ modifications were set as fixed modifications, and methionine oxidation and deamidation (NQ) as variable. Trypsin was used as cleavage enzyme, with one missed cleavage allowed. Mass tolerance was set at 30 ppm for intact peptide mass, and 0.3 Da for fragment ions. Search results were rescored to give a final 1% FDR using a randomized version of the same Uniprot Human database. Protein-level iTRAQ ratios were calculated as intensity weighted, using only peptides with expectation values < 0.05. As this was a protein IP experiment, no global ratio normalization was applied. Protein enrichment was then calculating by dividing the true sample protein ratios by the corresponding control sample ratios.

Extraction of nuclear and cytoplasmic RNA

RNA was isolated from nuclear and cytoplasmic fractions using RNA Subcellular Isolation Kit (Active Motif) according to manufacturer's instructions.