Supplementary Information

Supplementary Methods

SEPT2 and *AKAP8* shRNA constructs

Lentiviral shRNA vectors (pLKO.1 backbone) targeting *AKAP8* (TRCN0000218647 and TRCN0000229896, denoted as sh-A and sh-B respectively) and *SEPT2 (*TRCN0000062153 and TRCN0000062154, denoted as sh-C and sh-D respectively) were obtained from Mission Sigma.

Real-Time Quantitative Polymerase Chain Reaction

The knockdown confirmation of *AKAP8* and *SEPT2* was performed by measuring the mRNA expression level by real-time quantitative PCR. Reactions were performed on LightCycler 96 Real-Time PCR system (Roche) using TaqMan assay probes (*AKAP8*: Hs00935915_m1; *SEPT2*: Hs01565417_m1) obtained from ThermoFisher Scientific. *B2M* (*β2-microglobulin*) was used as endogenous control (Hs99999907_m1, ThermoFisher Scientific) and relative mRNA expression levels were calculated by using $\Delta\Delta C_T$ method.

Cell Growth Assay

On day 7 of culture, transduced cells were seeded in 96 well plates (2000 cells). Cell viability was assessed by using CellTiter-Glo® Luminescent Cell Viability Assay (Promega) from day 7 to day 13 of differentiation. Luminescent signal was detected using a SpectraMax i3x plate reader (molecular devices).

Flow Cytometry

Erythroid differentiation on day 11 and day 14 was evaluated using flow cytometry. Differentiation of erythroid cells was assessed by using antibodies against CD36 (PE anti-human CD36, clone 5-271, Biolegend), CD71 (PE-Cyanine7, clone OTK-9, eBiosciences) and CD235a (APC, clone HIR2, eBiosciences). DAPI (Sigma Aldrich) was used as viability dye. Cells were incubated with antibodies for 30 mins at 4°C in dark. Further, cells were washed with FACS buffer (PBS with 1% Bovine serum albumin) and resuspended in FACS buffer with DAPI for flow cytometry analysis. DAPI negative (viable) cells were used for subsequent analysis.

Cell cycle analysis was performed on ethanol-fixed cells. Fixed cells were washed with PBS and treated with RNase (40µg/ml) and PI (10µg/ml) for 30 mins at room temperature. Cell cycle analysis using Flow cytometry.

All flow cytometry was performed using an LSRII flow cytometer (BD Biosciences) and data analyzed by using FlowJO VX software.

Colony-Forming Cell Assay

Colony-forming cell assay was performed by plating 3000 transduced cells on day 7 of differentiation on methylcellulose (MethoCult H4434 Classic, Stemcell Technologies) containing 0.65µg/ml puromycin according to the manufacturer's protocol. Plates were incubated in a humidified incubator at 37°C and 5% CO_2. Colonies were counted after 14 days of incubation.

Hierarchical clustering using aberrantly spliced events identified in SFmut MDS

Hierarchical clustering was performed using rMATS-calculated inclusion levels of the 245, 236 and 287 aberrantly spliced events identified in *SF3B1*mut, *SRSF2*mut and *U2AF1*mut MDS cases. Hierarchical clustering heatmap plots were generated using the ClustVis tool (https://biit.cs.ut.ee/clustvis/).

Annotation of NMD sensitive and insensitive aberrantly spliced events identified in SFmut MDS

rMATS output files were converted to bed files using the DASEResultConvertor.jar tool available from the Alternative Splicing Encyclopedia (ASpedia; http://combio.snu.ac.kr/aspedia/). Bed files were subsequently used to map splicing events to Ensembl transcript ID using the multiple AS event query in ASpedia. Finally, transcript IDs were annotated as NMD sensitive or insensitive using the Ensembl BioMart tool (https://www.ensembl.org/biomart/martview).

Gene Ontology Analysis

To identify splicing factor mutant specific effects, we investigated the overlapping significant GO themes for each of the splicing factor (*SF3B1*, *SRSF2* and *U2AF1*) mutant MDS versus healthy controls, and versus splicing factor wildtype MDS.

REVIGO (http://revigo.irb.hr/)[1], which removes redundant GO terms, treemap function was used to visualize the important biological process (BP) ontology themes affected by the splicing factor mutations.

Analysis of genes involved in heme metabolism and iron processing

For the investigation of aberrantly spliced genes involved in heme metabolism and iron processing in SFmut MDS, a total of 200 genes involved in heme metabolism and 150 genes involved in iron homeostasis and transport were analyzed. The lists of genes were obtained from relevant gene sets within the Molecular Signatures Database (MSigDB; http://software.broadinstitute.org/gsea/msigdb/index.jsp). Specifically, for the genes involved in heme metabolism, the HALLMARK_HEME_METABOLISM gene set was used. For the genes involved in iron homeostasis and transport, genes from Lane et al (2015)[2] and from the following gene sets were used: GO_CELLULAR_IRON_ION_HOMEOSTASIS, GO_IRON_COORDINATION_ENTITY_TRANSPORT, GO_IRON_ION_IMPORT, GO_IRON_ION_TRANSPORT, and GO_IRON_ION_HOMEOSTASIS.

Clinical association of splicing factor mutation status in MDS

To identify the respective association of splicing factor mutations with continuous (hemoglobin, Hb; white blood cell counts, WBC; absolute neutrophil count, ANC; platelet count, Plt; BM blasts; and age) and categorical (gender, IPSS and transfusion dependency) clinical variables, a two-tailed Mann-Whitney non-parametric test and Fisher's exact test was performed.

Pathway analysis heatmaps

We investigated the overlapping significant pathways for each of the splicing factor (*SF3B1*, *SRSF2* and *U2AF1*) mutant MDS versus healthy controls, and versus splicing factor wildtype MDS. Significant pathways were determined by a -log p-value $\geq$ 1.3 (p-value $<$ 0.05). Heatmaps from collapsed IPA-generated -log p-values were created using using R (http://www.R-project.org) and graphical CRAN package pheatmap (https://cran.r-project.org/package=pheatmap).Values were collapsed by the comparison (vs Healthy control or vs SFWT) that had the lowest -log p-values (highest p value) and only significant values were plotted. To identify the converging pathways with the lowest p-value across all the splicing factor mutant MDS (*SF3B1, SRSF2* and *U2AF1*), pathways were ranked by the

minimum enrichment -log p-values (highest p value) identified in the *SF3B1*, *SRSF2* and *U2AF1* mutant MDS IPA analysis.

Bioinformatics analysis

Computational analysis and statistical testing was conducted using the R statistical programming language (http://www.R-project.org), with the following packages used: ggplot2 (https://cran.r-project.org/package=ggplot2), survminer (https://cran.r-project.org/package=survminer), biomaRt,[3] RWebLogo (https://github.com/omarwagih/RWebLogo), DESeq2[4] and UpSetR.[5]

AKAP8 and *SEPT2* downregulation splicing factor mutant MDS

Down-regulation of *AKAP8* in SRSF2 mutant MDS, and *SEPT2* in SF3B1 mutant MDS was determined using the R language Bioconductor package DESeq2. Differential expression was determined using the default Wald test and adjusted for multiple hypothesis testing using Benjamini-Hochberg correction across all genes tested. Specific gene expression boxplots plots for *AKAP8* and *SEPT2* expression were generated using ggplot2.

Validation of aberrant splicing events identified using semi-quantitative PCR

Total RNA was extracted using TRIzol (Thermo Scientific, UK) with a linear acrylamide carrier, treated with DNase I (Life technologies) and purified using Agencourt RNAClean XP beads (Beckman Coulter). DNase-free RNA was converted to cDNA using a High-Capacity cDNA Reverse Transcription Kit (Thermo-Fisher). Long and short isoforms of *AKAP8*, *SEPT2*, *PFKM* and *METTL17* were amplified using primers flanking the region of interest in a PCR reaction using maxima hotstart PCR mastermix (Thermo-Fisher). PCR products were separated using DNA 1000 chips (Agilent). PCR long:short ratios were calculated using band intensities as quantified by the 2100 Expert software (Agilent). Statistical significance was determined by Kruskal-Wallis one-way analysis of variance followed by Dunn's post hoc test. Primers used are presented in the Table below:

Gene	**Primer direction**	**Sequence (5′-3′)**
SEPT2	Forward	GAGTCATACCTGGAGCAGCA
	Reverse	TAGCTTGACCCCTCGCTCTT
AKAP8	Forward	AGGGTGAGGATGAACTCTGC
	Reverse	GCATACAGAACAGGCAAACTGA

PFKM	Forward	ATCATTGTGGCTGAGGGTGC
	Reverse	TCCCCTCCAAAAGTGCCATC
METTL17	Forward	TCAGGTGCAAACACTGACCA
	Reverse	TGCTCCACGAAGTTTCTCCTC

Analysis of 3'splice site and RI properties

For analysis of 3' splice site (ss) properties, we collected data sets from the A3SS rMATS output on the basis of FDR <0.05. For events with IncLevelDifference >0.1 (>0.15 in the case of *SF3B1* mutant analysis) the upstream A3SS was classified as cryptic and the downstream A3SS as associated canonical A3SS. Likewise, for events IncLevelDifference <-0.1 (<-0.15 in the case of *SF3B1* mutant analysis), the downstream A3SS was classified as cryptic A3SS and the upstream one associated canonical.

Human sequences around A3SS and upstream introns were retrieved from UCSC (hg19, Feb. 2009) using R and the Bioconductor packages: Genomic Ranges, Genomic Features, biomaRt and BSgenome.Hsapiens.UCSC.hg19. Graphical outputs were generated ggplot2. Statistical analysis comparing sequences properties between data sets were done using Two-tailed Mann-Whitney test in R.

Sequence logos were produced using RWebLogo. For that 35nt upstream and 3nt downstream were extracted for each A3SS as mentioned above and used as input for the program.

Density plots: For each pair of A3SS, the distance between was calculated as the difference of its chromosomal coordinates, and the log2 of the difference was plot using density plot in ggplot2.

Survival effects of isoform expression

MDS: Kaplan-Meier survival curves were constructed using the surviner and survival packages in R. Statistical testing of differences between survival curves used Cox proportional hazards multivariate modelling. Initial exploratory survival analysis used a full dataset with covariates including splice factor mutation status, karyotype and confounding mutations – none of these covariates were identified as statistically significant and a minimal model was used for definitive testing with age, gender, IPSS, and isoform expression high or low (median split) as covariates.

TCGA: We analyzed expression levels of isoforms in the publicly available RNA sequencing data generated by the TCGA Research Network (http://cancergenome.nih.gov/). Filtered and

log 2 normalized RNA isoform expression data along with all available clinical data were downloaded from the GDAC firehose database (run: stddata__2015_06_01) for each gene of interest from the AML dataset. Survival analysis was performed using the survminer and survival R packages. Kaplan-Meier estimated survival curves were constructed using the TCGA clinical data. Statistical testing of differences between survival curves used Cox proportional hazards modelling as above.

Supplementary References

1. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
2. Lane, D.J., Merlot, A.M., Huang, M.L., Bae, D.H., Jansson, P.J., Sahni, S., Kalinowski, D.S. & Richardson D.R. Cellular iron uptake, trafficking and metabolism: Key molecules and mechanisms and their roles in disease. *Biochim Biophys Acta* **1853**, 1130-1144 (2015).
3. Durinck, S., Spellman, P.T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).
4. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
5. Conway, J.R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940 (2017).

Table S1: Clinical data and splicing factor mutation status of 82 MDS patients.

ID	SF3B1	SRSF2	U2AF1	ZRSR2	Disease Subtype	Dead or alive at last observation	Days from sample to last observation	IPSS	Karyotype details	Sex	Age	Hb	WBC	Neut	Platelet	BM blast %	Transfusion dependency
A112	1	0	0	0	RARS	ALIVE	1758	Low	46XX	F	61	9.5	3.84	2.2272	226	1	no
A114	0	0	1	0	RA	ALIVE	1254	na	46XX	F	49	na	na	na	na	na	na
A116	1	0	0	0	RARS	ALIVE	0	Low	46XY	M	71	7.9	7.1	5.538	430	1	yes
A117	0	0	0	0	RAEB2	ALIVE	258	Int-1	46XX,del(5q)	F	74	8.8	na	0.9968	228	na	no
A118	1	0	0	0	RA	ALIVE	1056	Low	46XX,del(5q)	F	65	10	6.34	2.7262	347	3	yes
A119	0	1	0	0	RAEB1	DEAD	168	Int-1	46XY	M	52	11.9	3.81	2.5908	85	7	no
A120	0	0	0	0	RARS	ALIVE	300	Int-1	46XY	M	77	8.1	1.93	0.91	52	4	no
A121	0	0	0	0	RA	ALIVE	36	Int-1	46XY	M	87	10.4	3.25	1.3	89	2	no
A122	0	1	0	0	RAEB2	ALIVE	1040	Int-2	46XX	F	79	9.3	2	0.54	63	11	no
A123	1	0	0	0	RARS	ALIVE	1596	Low	46XX	F	83	9.3	4.41	2.8665	227	2	no
A124	0	0	0	0	RA	ALIVE	1125	Low	46XX	F	76	10.1	5.38	2.5286	389	3	no
A125	1	0	0	0	RARS	ALIVE	264	Int-1	46XY	M	66	7.9	2.7	1.62	114	1	no
A126	0	0	0	0	RA	ALIVE	1986	Low	46XX	F	85	9.4	4	2.8	224	1	yes
A127	1	0	0	0	RARS	ALIVE	1097	Low	46XX	F	59	9	5.5	2.64	305	1	no
A128	1	0	0	0	RARS	ALIVE	975	Int-1	46XX,del(20q)	F	65	9.6	na	1.019	139	4	yes
A129	0	0	0	0	RAEB1	ALIVE	2549	Int-1	46,XX,del(5)(q14q34)[20]; 46,XX[5]	F	56	9	3.43	2.13	349	8	na
A130	0	0	0	0	RAEB2	ALIVE	184	Int-2	46XX	F	34	10.5	1.5	0.39	87	17	no
A131	0	0	0	0	RAEB1	ALIVE	147	Int-1	46XY	M	55	8.6	4.46	1.87	196	5	no
A132	0	0	0	0	RA	ALIVE	2130	Low	46XX	F	40	10.7	na	2.6	77	1	no
A134	0	0	0	0	RAEB2	DEAD	988	Int-2	46,XY	M	67	10.4	na	0.8	336	12	no
A136	0	0	0	0	RAEB2	DEAD	228	Int-2	46XY	M	57	5.4	1.9	0.95	45	12	yes
A138	0	0	0	0	RARS	DEAD	458	Int-1	46XY,del(17p)	M	62	6.6	3.25	1.0725	278	3	yes
A139	0	0	0	0	RARS	DEAD	1348	Int-1	46XY;del(11q),del(17p)	M	63	9.5	4.68	2.9952	218	3	no
A140	0	0	0	0	RA	ALIVE	1350	Int-1	46,XX,del(5q)/9[25]	F	67	12.1	na	1.5	85	2	yes
A141	0	0	0	0	RA	ALIVE	2116	Int-1	complex	M	74	na	na	na	na	1	no
A142	1	0	0	0	RARS	ALIVE	1231	Low	46XY	M	66	10.7	2.5	0.96	171	0	yes
A143	0	0	0	0	RA	ALIVE	520	Low	46XY	M	71	10.3	5.4	2.43	318	2	no
A145	0	0	0	0	RA	ALIVE	981	Int-1	46XX, del(12p), del(5q)	F	74	8.2	na	2.16	292	3	no
A147	1	0	0	0	RARS	ALIVE	1158	Low	46XX	F	53	7	na	3.55	577	2	yes
A148	0	0	0	0	RAEB2	DEAD	539	High	46,XX	F	69	11.1	na	1.7	10	14	yes
A149	1	0	0	0	RARS	ALIVE	693	Int-1	46XY,del(7q),del(20q)	M	75	8.7	5.2	3.016	251	4	no
A150	1	0	0	0	RARS	ALIVE	525	Int-1	46XY,del(9q),del(13q)	M	62	9.5	na	2.855	180	4	no
A151	1	0	0	0	RARS	ALIVE	409	Low	46XY	M	71	10.4	4.74	3.0336	227	3	no
A153	1	0	0	0	RARS	ALIVE	1316	Low	46XY	M	69	11.2	na	3.022	461	2	no
A154	1	0	0	0	RARS	DEAD	1119	Int-1	46XY	M	57	8.4	na	1.005	305	0	no
A155	0	0	0	0	RA	ALIVE	1352	Int-1	46XY	M	26	11	4	0.16	51	2	no
A156	0	1	0	0	RAEB1	DEAD	444	Int-2	47XY,+8	M	73	8.6	1.7	0.44	148	7	no
A158	1	0	0	0	RARS	ALIVE	1534	Low	46XX	F	48	10.1	7.91	4.9042	543	3	no
A161	0	0	1	0	RAEB1	ALIVE	21	Int-1	46XY	M	66	7.4	9.39	1.6902	18	8	yes
A163	0	0	0	0	CMML	na	na		46XX,del(5q)	F	82	na	na	na	232	0.2	na
A164	0	0	0	0	RA	ALIVE	410	Int-1	46XY	M	64	9.7	2.79	1.4508	105	3	yes
A165	0	0	0	0	RAEB2	ALIVE	313	Int-2	46XY	M	57	4.5	1.22	0.1708	43	15	no
A166	1	0	0	0	RARS	ALIVE	2233	Int-1	46XX,del(12p)	F	79	8.1	3.83	2.298	594	1	no
A167	1	0	0	0	RA	ALIVE	1643	Int-1	47XY +19	M	74	8.5	5.8	2.7	160	0	yes
A168	0	0	0	0	RAEB2	ALIVE	430	Low	hyperdyploidy	F	59	8.8	2	0.88	77	11	no
A170	1	0	0	0	RARS	ALIVE	1625	Low	46XY	M	65	9.8	6.57	3.4821	104	3	no
A171	0	0	0	0	RAEB1	ALIVE	22	Int-2	complex	M	74	8.3	2.88	1.7568	16	6	yes
A172	1	0	0	0	RAEB1	DEAD	407	Int-1	47XX, +8, del17p	F	67	8.6	6.87	3.7098	538	6	yes
A173	0	0	0	0	CMML	ALIVE	1690	Int-1	46,XY	M	35	11	na	8	36	15.5	no
A175	0	0	0	0	RAEB1	DEAD	185	Int-2	46XY,del(12p)	M	71	9.3	6.47	2.98	93	6	yes
A176	0	0	0	0	CMML	DEAD	663		46,XX	F	56	11.4	na	na	336	6	no
A177	0	1	0	0	RARS	ALIVE	25	Low	46XY,del(20q)	M	60	11.5	2.67	1.36	187	2	no
A178	1	0	0	0	RARS	ALIVE	1440	Int-1	46XX	F	81	10.7	7	4.8	504	2	no
A179	0	0	0	0	RA	ALIVE	14	Int-2	46XX, -7	F	72	13	na	1.4098	86	3	no
A180	0	0	1	0	RAEB1	ALIVE	49	Int-1	46XY	M	77	7.5	3.3	1.58	20	6	no
A181	1	0	0	0	RARS	ALIVE	1107	Low	46XX,del(20q)	F	78	9.1	3.63	2.4684	166	1	no
A182	0	0	0	0	RAEB1	ALIVE	19	Int-1	46XY	M	69	12	2.25	0.4275	42	8	no
A183	1	0	0	0	RARS	DEAD	747	Int-1	46XX	F	69	9.5	5.6	3.7	362	4.5	no
A184	0	0	0	0	RAEB1	DEAD	1635	Int-1	46 XY	M	67	10	0.4	0.1	83	7	yes
A185	1	0	0	0	RARS	ALIVE	223	Int-1	48XY,+8,+21	M	58	6.9	na	2.854	252	0	yes
A186	0	1	0	0	CMML	DEAD	930	na	46,XY	M	58	11.1	na	15.2	45	10.5	no
A187	0	0	1	0	RAEB2	ALIVE	1755	Int-2	46 XY	M	76	9.9	2.6	0.5	103	13.5	yes
A188	0	0	0	0	RARS	DEAD	1057	Low	46,XY	M	64	7.9	6.5	4.4	244	2	yes
A189	0	0	0	0	RA	ALIVE	1833	Low	46XX	F	65	7.4	5.8	2.9	382	4	yes
A190	1	0	0	0	RARS	ALIVE	1603	Low	46XY	M	69	10.8	4.7	2.6	349	2	no
A191	0	0	1	0	RAEB1	ALIVE	80	Int-1	46XX	F	46	11.9	2.8	1.204	120	7	no
A192	0	1	0	0	RAEB2	ALIVE	154	Int-2	47XY,+21	M	61	8.3	5.1	2.4	314	13	yes
A193	0	0	0	0	RA	DEAD	412	Low	46XY	M	77	10.1	4.3	1.4	183	3	no
A194	0	0	0	0	RA	ALIVE	1650	Int-1	46XX, del(5)(q13q33)	F	66	9.5	9.1	6.1	305	4	no
A195	0	0	0	0	RAEB2	ALIVE	0	Int-2	46XY	M	63	9.6	1.82	0.8372	34	16	no
A196	1	0	0	0	RARS	ALIVE	1221	Low	46,XY	M	70	10.8	na	2.4	348	3	no
A197	1	0	0	0	RA	na	na	Low	46XX	F	58	10	na	2.5	500	2	no
A198	0	0	0	0	RAEB1	DEAD	695	Int-1	46XX	F	72	9.9	5.5	3.025	465	7	yes
A199	1	0	0	0	RARS	DEAD	124	Int-1	47XY,+8	M	72	8.2	6.33	2.7852	300	1	yes
A200	0	0	1	0	RAEB2	ALIVE	63	High	46XY;46XY,t(X;3),del(17p)	M	78	8.3	5.51	2.1	63	18	no
A203	0	0	0	0	RAEB2	DEAD	366	High	5q- + complex (X;12, del5, der12)	M	73	13.4	2.82	1.02	91	16	na
A204	0	0	0	0	RAEB1	DEAD	1040	Int-1	46XX	F	77	13.1	3.1	0.94	100	8	no
A205	1	0	0	0	RAEB1	ALIVE	466	Int-1	46XX,t(3;3),t(4;6)	F	81	8.8	na	3.6	496	6	no
A206	0	0	0	0	RAEB1	DEAD	953	Int-1	46XX, riarr(1p)	F	76	11.9	2.74	0.7672	120	9	no
A207	0	1	0	0	RAEB2	ALIVE	193	Int-2	47XY,+8	M	63	12.1	30	20.1	59	11	no
A208	0	1	0	0	RARS	ALIVE	1832	Low	46XY	M	62	7.5	11.73	4.8093	106	4	yes
A209	0	0	0	0	RA	DEAD	2057	Low	46XY,del(5q)	M	63	10.3	na	3.9	99	1	no

Table S2: The splicing factor mutations identified in the CD34+ MDS cohort.

	Specific Splicing factor Mutations in cohort			
ID	**SF3B1**	**SRSF2**	**U2AF1**	**ZRSR2**
A153	E622D			
A142	E622D			
A190	E622D			
A178	E622D			
A166	H662Q			
A123	H662Q			
A205	K666N			
A172	K700E			
A199	K700E			
A118	K700E			
A149	K700E			
A125	K700E			
A127	K700E			
A116	K700E			
A150	K700E			
A185	K700E			
A128	K700E			
A147	K700E			
A181	K700E			
A170	K700E			
A158	K700E			
A197	K700E			
A112	K700E			
A196	K700E			
A167	K700E			
A183	K700E			
A154	R625L			
A151	R625L			
A119		P95H		
A192		P95_R102delPPDSHHSR		
A156		P95H		
A207		P95H		
A208		P95H		
A122		P95H		
A186		P95H		
A177		P95L		
A180			Q157P	
A161			Q157P	
A200			Q157R	
A187			R156H	
A191			S34F	
A114			S34F	
A160		P95H		H191Y
A174		P95R		V253fs*36

Table S3: Clinical details and mutation status of the 11 patients used in the study of bone marrow monocytic, granulocytic and erythroid precursors.

			Cell population present (1=yes, 0 =no)			Mutation status (1=yes, 0 =no)								
ID	**Disease Subtype**	**Age**	**MON**	**GRA**	**ERY**	**SF3B1**	**SRSF2**	**TET2**	**CBL**	**IDH2**	**ASXL1**	**JAK2**	**PHF6**	**NPM1**
MDS142	RA	76	1	0	0	1	0	1	0	0	0	0	0	0
MDS152	RARS	64	1	0	1	0	1	0	1	1	0	0	0	0
MDS155	MDS/AML	72	1	1	1	1	0	0	0	0	1	1	0	0
MDS163	MDS/AML	76	1	1	1	1	0	0	0	0	0	0	0	0
MDS166	RAEB2	44	1	1	1	1	0	0	0	0	0	0	1	0
MDS168	MDS/AML	71	1	0	1	0	1	0	0	1	0	0	0	1
MDS177	RAEB2	81	1	0	1	1	0	0	0	0	0	0	0	0
MDS178	RARS	63	1	1	1	1	0	1	0	0	0	0	0	0
MDS189	RARS	81	1	1	1	1	0	0	0	0	0	0	0	0
MDS191	RA	83	1	1	0	0	1	0	0	0	0	0	0	0
MDS218	RA	87	1	1	1	0	1	1	0	0	1	0	0	0

Table S4: Top 40 significant aberrant splicing events in SF3B1mut, SRSF2mut and U2AF1mut MDS.

SF3B1 mutant MDS						*SRSF2 mutant MDS*						*U2AF1 mutant MDS*					
Event ID	*Gene*	*Event type*	*Chr*	*Strand*	*Start - End position*	*Event ID*	*Gene*	*Event type*	*Chr*	*Strand*	*Start - End position*	*Event ID*	*Gene*	*Event type*	*Chr*	*Strand*	*Start - End position*
7506	TMEM91	A3SS	19	+	41889619-41889988	743	PPIL3	MXE	2	-	201747064-201747158	33174	ANKRD12	SE	18	+	9204473-9204542
7507	TMEM91	A3SS	19	+	41889465-41889988	19440	MX1	SE	21	+	42802459-42802535	27516	RAD51C	SE	17	+	56783849-56783969
518	ERCC3	RI	2	-	128046912-128047400	20608	RIMKLB	SE	12	+	8932630-8932708	17455	CRYZ	SE	1	-	75172786-75172888
4728	NICN1	RI	3	-	49462381-49462871	23987	TIA1	SE	2	-	70455475-70455594	1940	AC084018.1	RI	12	-	122233172-122234148
2362	SNRPN	A3SS	15	+	25219434-25219603	26349	C6orf25	SE	6	+	31691918-31692009	6163	RNF216	A3SS	7	-	5778906-5779252
3211	QTRT1	RI	19	+	10822836-10823304	28447	LST1	SE	6	+	31555725-31555743	3301	AP1G2	RI	14	-	24031712-24032847
5420	RPRD1A	RI	18	-	33605560-33607038	28455	LST1	SE	6	+	31555720-31555743	4380	ZWINT	RI	10	-	58117198-58118220
5429	TMEM91	RI	19	+	41888676-41889988	34068	GGCT	SE	7	-	30540151-30540297	3880	SLC35B1	A3SS	17	-	47783565-47783696
5375	COASY	RI	17	+	40714168-40714505	36368	C8orf59	SE	8	-	86131464-86131592	30413	APLP2	SE	11	+	130007150-130007186
1520	SNRPN	A5SS	15	+	25212175-25212387	36371	C8orf59	SE	8	-	86131464-86131589	17704	AC005154.6	SE	7	-	30591715-30591795
2359	SNRPN	A3SS	15	+	25219457-25219603	1758	LST1	MXE	6	+	31555417-31555510	4147	GUCY1A3	A3SS	4	+	156617907-156618274
2896	AMT	RI	3	-	49454210-49455151	28445	LST1	SE	6	+	31555417-31555510	30841	PCBP2	SE	12	+	53861588-53861627
6153	ABCC5	A3SS	3	-	183703091-183703243	34270	GMPR2	SE	14	+	24703312-24703447	13212	MRPS28	SE	8	-	80940929-80941031
4975	GPR108	A3SS	19	-	6730997-6731122	28	S100A13	RI	1	-	153598795-153600074	12249	INTS3	SE	1	+	153733495-153733585
3305	AP1G2	RI	14	-	24031170-24031624	2457	RNF34	SE	12	+	121840544-121840610	3122	RRP12	RI	10	-	99118293-99118767
33241	DPH5	SE	1	-	101458192-101458296	17455	CRYZ	SE	1	-	75172786-75172888	14132	YPEL5	SE	2	+	30371110-30371407
6131	SUGP1	A3SS	19	-	19414532-19414852	37277	MPI	SE	15	+	75189351-75189560	1076	DHRS1	RI	14	-	24761389-24761990
2380	ABTB1	RI	3	+	127396307-127396686	1389	CRYZL1	A3SS	21	-	34971455-34971591	20633	GUCY1A3	SE	4	+	156617907-156618001
2437	SEPT6	A3SS	X	-	118759297-118759359	14189	RP11-354E11.2	SE	10	-	20013855-20013910	2872	ZNF83	RI	19	-	53115630-53118050
30228	ABCC5	SE	3	-	183703091-183703166	29831	SLC25A26	SE	3	+	66286967-66287124	4410	SCYL1	RI	11	+	65305740-65306175
30229	ABCC5	SE	3	-	183703091-183703243	5801	LST1	A3SS	6	+	31556294-31556686	4427	C1orf43	SE	1	-	154192311-154192413
1579	SERBP1	A3SS	1	-	67890570-67890660	6154	VTI1A	SE	10	+	114208139-114208247	27002	ARID2	SE	12	+	46298124-46298181
6376	UXS1	A3SS	2	-	106781191-106781255	2935	MAP2K2	RI	19	-	4094450-4095447	2935	MAP2K2	RI	19	-	4094450-4095447
3372	DYNLL1	RI	12	+	120933915-120934356	5601	IMMP1L	SE	11	-	31477806-31477933	30796	PTCHD3P1	SE	10	+	29747202-29747449
7563	PPOX	A3SS	1	+	161137128-161137276	11281	TSEN2	SE	3	+	12574160-12574272	6164	RNF216	A3SS	7	-	5780603-5781446
4127	THOC1	A3SS	18	-	224083-224200	5602	IMMP1L	SE	11	-	31484718-31484852	13912	HMGXB4	SE	22	+	35659974-35660089
5460	TCEA2	A3SS	20	+	62703210-62703294	35758	FES	SE	15	+	91434211-91434421	24105	BAZ1A	SE	14	-	35255331-35255427
14165	MZB1	SE	5	-	138725015-138725109	17810	DGUOK	SE	2	+	74166036-74166149	32492	NICN1	SE	3	-	49463315-49463429
6157	ERGIC3	A3SS	20	+	34144725-34144880	28446	LST1	SE	6	+	31555417-31555510	22516	AP1G2	SE	14	-	24035770-24035895
11475	SNRPN	SE	15	+	25212175-25212299	35711	TEX30	SE	13	-	103419622-103419828	3500	PNPLA7	RI	9	-	140355085-140356072
11476	SNRPN	SE	15	+	25212175-25212387	649	GPBP1L1	SE	1	-	46151247-46151292	22809	TRA2A	SE	7	-	23561750-23562051
1071	CDC27	A3SS	17	-	45229171-45229302	2391	BDH2	RI	4	-	104006505-104007697	14563	LAT2	SE	7	+	73629151-73629429
28927	ICA1	SE	7	-	8267267-8267481	17539	NHP2L1	SE	22	-	42083725-42084331	9625	NR2C2AP	SE	19	-	19313278-19313384
697	METTL5	A3SS	2	-	170668966-170669034	29515	PFKM	SE	12	+	48529073-48529166	16107	ELK3	SE	12	+	96591803-96591920
4277	LARP4	A3SS	12	+	50822699-50822873	2560	WDR75	A3SS	2	+	190313970-190315694	29817	POLE2	SE	14	-	50121501-50121556
4070	ZNF410	A3SS	14	+	74360478-74360635	32534	TAMM41	SE	3	-	11886364-11886570	22811	TRA2A	SE	7	-	23561739-23562051
4415	ANAPC5	A3SS	12	-	121766118-121766300	25161	C12orf73	SE	12	-	104347191-104347312	3852	MBD2	SE	18	-	51714083-51714207
4520	FOXRED1	A3SS	11	+	126143210-126143349	8830	BRD8	SE	5	-	137502206-137502416	22808	TRA2A	SE	7	-	23561972-23562051
4363	SEPT2	A3SS	2	+	242275373-242275513	331	SMOX	RI	20	+	4162735-4163495	24415	POLDIP3	SE	22	-	42997975-42998062
252	PSME2	RI	14	-	24614918-24615449	27873	CASP8	SE	2	+	202139611-202139676	25132	GOLGA2	SE	9	-	131035063-131035144

Table S5: Biological pathways identified by IPA as significant in SF3B1mut MDS versus Healthy control and Splicing factor wildtype MDS (SFWT).

	Versus Healthy Control		Versus SFWT	
Ingenuity Canonical Pathways	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**
Oxidative Phosphorylation	3.82	0.441 (30)	8.01	0.522 (36)
Sirtuin Signaling Pathway	3.47	0.35 (63)	4.01	0.33 (60)
Mitochondrial Dysfunction	2.47	0.35 (41)	6.08	0.407 (48)
Protein Ubiquitination Pathway	4.37	0.362 (71)	2.13	0.284 (57)
TCA Cycle II (Eukaryotic)	2.24	0.526 (10)	2.04	0.474 (9)
Heme Biosynthesis II	2.51	0.75 (6)	1.89	0.625 (5)
Heme Biosynthesis from Uroporphyrinogen-I	1.88	1 (3)	2.04	1 (3)
BER pathway	1.83	0.6 (6)	2.09	0.6 (6)
EIF2 Signaling	1.76	0.312 (49)	4.74	0.352 (57)
Regulation of eIF4 and p70S6K Signaling	1.65	0.322 (37)	1.8	0.295 (36)
Methionine Degradation I (to Homocysteine)	1.37	0.467 (7)	1.47	0.438 (7)
Spermine and Spermidine Degradation I	1.36	0.75 (3)	1.51	0.75 (3)

Table S6: Biological pathways identified by IPA as significant in SRSF2mut MDS versus Healthy control and Splicing factor wildtype MDS (SFWT).

	Versus Healthy Control		Versus SFWT	
Ingenuity Canonical Pathways	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**
EIF2 Signaling	4.22	0.39 (60)	4.87	0.312 (50)
Sirtuin Signaling Pathway	3.37	0.362 (64)	3.59	0.28 (51)
Mitochondrial Dysfunction	3.15	0.386 (44)	5.14	0.345 (40)
Oxidative Phosphorylation	2.8	0.418 (28)	4.29	0.377 (26)
Heme Biosynthesis II	3.45	0.875 (7)	2.24	0.625 (5)
Heme Biosynthesis from Uroporphyrinogen-I	1.82	1 (3)	2.27	1 (3)
Protein Ubiquitination Pathway	1.77	0.317 (60)	1.54	0.229 (46)
Hypoxia Signaling in the Cardiovascular Syste	1.44	0.364 (20)	1.82	0.298 (17)
Antigen Presentation Pathway	3.47	0.577 (15)	1.36	0.321 (9)

Table S7: Biological pathways identified by IPA as significant in U2AF1mut MDS versus Healthy control and Splicing factor wildtype MDS (SFWT).

	Versus Healthy Control		Versus SFWT	
Ingenuity Canonical Pathways	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**
Sirtuin Signaling Pathway	3.24	0.366 (64)	3.37	0.282 (51)
Heme Biosynthesis II	3.38	0.875 (7)	2.19	0.625 (5)
Oxidative Phosphorylation	1.71	0.373 (25)	3.32	0.353 (24)
Mitochondrial Dysfunction	1.68	0.342 (39)	3.31	0.308 (36)
EIF2 Signaling	1.65	0.327 (51)	2.03	0.256 (41)
Colanic Acid Building Blocks Biosynthesis	1.95	0.571 (8)	1.57	0.429 (6)
Protein Ubiquitination Pathway	1.49	0.314 (60)	1.53	0.234 (47)
Tetrapyrrole Biosynthesis II	1.79	0.8 (4)	1.36	0.6 (3)
Galactose Degradation I (Leloir Pathway)	1.79	0.8 (4)	1.36	0.6 (3)
Estrogen Receptor Signaling	2.71	0.398 (35)	1.3	0.253 (23)

Table S8: Aberrantly spliced genes involved in suppression/regulation of R-loop formation and in the DNA damage response in SFmut MDS.

	Aberrantly spliced (1=yes) versus SF-WT MDS and Healthy Control		
Gene	*SF3B1mut*	*SRSF2mut*	*U2AF1mut*
AID	0	0	0
AQR	0	0	0
AQR	0	0	0
ASF	0	0	0
ATM	0	0	0
ATR	1	1	0
BRCA1	0	0	0
BRCA2	0	0	0
CHEK1	0	1	0
ERCC1	0	0	0
ERCC1	0	0	0
ERCC2	0	0	0
ERCC3	1	0	0
ERCC4	0	0	0
ERCC4	0	0	0
ERCC5	0	0	0
ERCC5	0	0	0
ERCC6	0	0	0
ERCC8	0	0	1
FANCA	0	0	0
FANCC	0	0	0
FANCD2	0	0	0
FANCE	0	0	0
FANCF	0	0	0
FANCG	0	0	0
FANCI	1	0	0
FANCL	0	0	0
FANCM	0	0	1
FIP1L1	0	0	0
FTO	0	0	0
MUS81	0	0	0
PIF1	0	0	0
RAD51	0	0	0
RAD52	0	0	0
RANBP2	0	0	0
RRM3	0	0	0
S100A9	0	0	0
SETX	1	1	0
TOP1	0	0	0
TOP3B	0	0	0
TPR	0	0	0

Table S9: Differences of clinical variables between SFmut MDS and SFwt MDS patients in our cohort.
Details of statistical tests performed are provided in the Supplementary methods. * = p<0.05, ** = p< 0.01, *** = p< 0.001

	Continuous variables					***Categorical variables***		
Group	*Hb*	*WBC*	*ANC*	*Plt.*	*Blast %*	*Gender*	*IPSS*	*Transfusion Dependency*
SF3B1mut	ns	**	***	***	***	ns	*	ns
SRSF2mut	ns	ns	ns	ns	ns	ns	ns	ns
U2AF1mut	ns	ns	ns	ns	ns	ns	ns	ns

Table S10: Multivariate Cox proportional hazard modeling for 14 aberrantly spliced isoforms that predict survival in MDS.

		p-value							***Hazard Ratio***						
Gene	*Transcript*	*Transcript level Low*	*Age*	*Sex Male*	*IPSS High*	*IPSS Int-1*	*IPSS Int-2*	*IPSS Low*	*Transcript level Low*	*Age*	*Sex Male*	*IPSS High*	*IPSS Int-1*	*IPSS Int-2*	*IPSS Low*
AHSA2	ENST00000493310	0.029	0.407	0.103	0.979	0.027	0.172	0.006	-1.90	-0.02	0.88	-0.03	-2.31	-1.63	-3.13
CAP1	ENST00000424977	0.004	0.472	0.089	0.152	0.207	0.703	0.022	1.70	0.02	0.89	1.67	-1.07	-0.36	-2.24
CD46	ENST00000480003	0.039	0.791	0.149	0.077	0.582	0.879	0.097	-1.19	-0.01	0.77	2.16	-0.50	0.16	-1.65
CRTC2	ENST00000487235	0.035	0.669	0.046	0.368	0.120	0.667	0.021	1.13	0.01	1.04	0.99	-1.30	-0.40	-2.24
DPH5	ENST00000477293	0.026	0.510	0.103	0.255	0.190	0.701	0.021	-1.38	-0.02	0.84	1.27	-1.10	0.39	-2.22
FCGR2A	ENST00000482233	0.016	0.259	0.019	0.130	0.215	0.888	0.036	1.44	0.03	1.37	1.67	-1.04	0.14	-2.07
IFI44	ENST00000446486	0.012	0.542	0.062	0.347	0.245	0.796	0.021	-1.38	0.01	0.98	1.03	-0.98	-0.25	-2.27
IFI44L	ENST00000462041	0.009	0.689	0.066	0.865	0.021	0.242	0.003	-1.60	0.01	0.96	-0.22	-2.37	-1.32	-3.58
MECR	ENST00000464915	0.022	0.951	0.109	0.265	0.265	0.920	0.030	-1.27	0.00	0.83	1.23	-0.95	-0.10	-2.10
MECR	ENST00000493928	0.022	0.632	0.081	0.179	0.225	0.561	0.019	-1.29	0.01	0.96	1.47	-1.01	-0.59	-2.31
NASP	ENST00000472408	0.014	0.605	0.041	0.475	0.112	0.655	0.017	-1.42	0.01	1.15	0.82	-1.44	-0.47	-2.41
PABPC4	ENST00000525669	0.036	0.646	0.085	0.319	0.181	0.974	0.030	-1.10	0.01	0.90	1.09	-1.12	-0.03	-2.10
PFDN5	ENST00000552341	0.042	0.597	0.115	0.132	0.637	0.872	0.171	-1.51	-0.01	0.85	1.70	-0.43	0.17	-1.42
PPOX	ENST00000466452	0.031	0.409	0.129	0.116	0.406	0.740	0.073	-1.24	0.02	0.81	1.81	-0.71	-0.34	-1.76
PTPRC	ENST00000413409	0.050	0.697	0.192	0.828	0.046	0.392	0.006	-1.11	0.01	0.72	0.27	-1.87	-0.90	-3.00
PTPRC	ENST00000571847	0.050	0.697	0.192	0.828	0.046	0.392	0.006	-1.11	0.01	0.72	0.27	-1.87	-0.90	-3.00
PTPRC	ENST00000530727	0.009	0.300	0.021	0.439	0.257	0.806	0.016	-1.51	0.03	1.26	0.87	-0.97	-0.23	-2.37
PTPRC	ENST00000575923	0.009	0.300	0.021	0.439	0.257	0.806	0.016	-1.51	0.03	1.26	0.87	-0.97	-0.23	-2.37

Table S11: Biological pathways identified by IPA as significant in ZRSR2 mutant MDS (with co-mutation of SRSF2) versus Healthy control and SRSF2mut MDS.

	Versus Healthy Control		Versus SRSF2mut MDS	
Ingenuity Canonical Pathways	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**	**-log(p-value)**	**Genes found/ Genes in pathway (No. Genes found)**
Sirtuin Signaling Pathway	2.2	0.221 (38)	1.73	0.198 (34)
Galactose Degradation I (Leloir Pathway)	1.59	0.6 (3)	1.67	0.6 (3)
Protein Ubiquitination Pathway	1.5	0.2 (37)	1.56	0.191 (35)
Colanic Acid Building Blocks Biosynthesis	1.35	0.357 (5)	2.3	0.462 (6)

Figure S1: Mutational landscape of MDS patient samples. A, Mutations identified in the cohort of 84 MDS cases. Of these, 28 had *SF3B1* mutations, 8 had *SRSF2* mutations, 6 had *U2AF1* mutations and 2 had *ZRSR2* mutations (with co-mutation of *SRSF2*). **B**, Mutations identified in the 11 patients used for the study of granulocytic, monocytic and erythroid precursors.

Figure S2: Gene expression levels of nonsense-mediated mRNA decay (NMD)-sensitive and NMD-insensitive events associated with splicing factor mutant MDS. A-D, Tukey box plots of the expression levels of NMD-sensitive and NMD-insensitive aberrant splicing events identified in *SRSF2*mut MDS (**A**), *U2AF1*mut MDS (**B**), and *SF3B1*mut MDS (**C**, all event types and **D**, A3SS events).

Figure S3: GO characteristics in splicing factor mutant MDS. A, REVIGO treemap of overlapping significant BP GOs identified in *SF3B1* mutant and *SRSF2* mutant CD34+ MDS patients. **B,** REVIGO treemap of BP GOs identified as significant in only the *SF3B1* mutant CD34+ MDS patients. REVIGO panel sizes are inversely proportional to enrichment p values.

Figure S4: Sirtuin pathway dysregulation in splicing factor mutant MDS. Example Sirtuin pathway map and genes found to be aberrantly spliced in *SF3B1* mutant MDS. The pathway map was generated from IPA and the aberrantly spliced genes are denoted by red fills.

Figure S5: Aberrant splice site usage in splicing factor mutant MDS. A, Sequence logos for upstream and downstream cryptic 3′ splice sites (top left and lower right sequence logos respectively) along with their associated canonical sites (top right and lower left sequence logos respectively) in SF3B1 mutant MDS. **B**, Density plot showing the distance (log2) between pairs of competing 3′ splice sites. **C**, Sequence logos showing the 5′SS and 3′SS usage in U2AF1 mutant patients harbouring S34 (top) and R156/Q157 variants (bottom). **D**, Sequence logos showing the 5′SS and 3′SS usage in SRSF2 mutant MDS. **E**, kmer frequency plot showing showed enrichment of CCNG motifs in cassette exons upregulated in SRSF2 mutant cells, while GGNG motifs were enriched in downregulated exons.

Figure S6: Examples of decreased intron retention in *SF3B1* mutants. **A**, NICN1 intron 4. **B**, ERCC3 intron 10 shows decreased intron retention in association with significant upregulation of an upstream cryptic A3SS. **C**, DOM3Z intron 3 shows decreased retention of an intron using an upstream A3SS. In this case, the change in A3SS usage was non-significant. All intron retention events are shown 5′ to 3′ in reverse orientation (from right to left)

Figure S7: Additional gene isoforms predictive of MDS patient survival. A-F, Kaplan-Meier survival plots of isoforms from 6 genes: AHSA2 (**A**), DPH5 (**B**), CAP1 (**C**), IFI44 (**D**), MECR (**E**), and NASP (**F**).

Figure S8: Aberrant splicing events and dysregulated pathways in CD34$^+$ cells of *ZRSR2*mut MDS patients. **A**, Venn Diagram showing the aberrant splicing events identified in *ZRSR2* mutant MDS patients (with co-mutation of *SRSF2*) versus healthy controls and versus *SRSF2*mut MDS patients. **B**, Ranked heatmap, as determined by significance across both comparisons of *ZRSR2*mutant MDS patients versus healthy controls and versus *SRSF2*mut MDS patients, showing the significant dysregulated pathways.

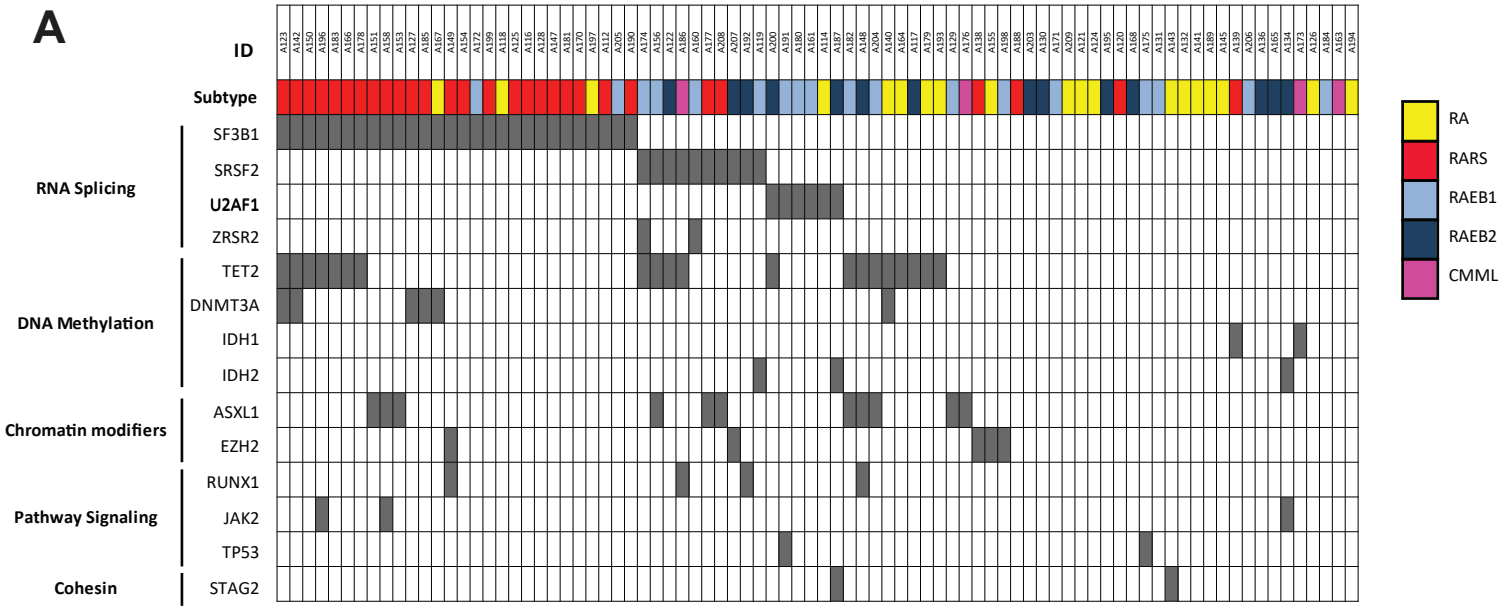
Figure S9: Principal component analysis (PCA) of bone marrow cell populations of SFmut MDS. Monocytic (MON), granulocytic (GRA), and erythroid (ERY) precursors isolated from the bone marrow of *SF3B1* mutant and *SRSF2* mutant MDS patients and healthy controls showing strong clustering by precursor cell population.

Figure S10: Aberrant splicing in BM cell populations of SFmut MDS. A-B, UpSet plots showing the overlap of aberrant splicing events identified in CD34$^+$ cells, and in monocyte (MON), granulocyte (GRA) and erythroid (ERY) precursor cell populations isolated from *SF3B1* (**A**) and *SRSF2* (**B**) mutant MDS patient samples.

Figure S11: Validation of aberrant splicing events identified in splicing factor mutant MDS. A, Sashimi plot showing the *AKAP8* SE event in *SRSF2* mutant MDS patient samples which results in the inclusion of an exon containing a premature termination codon (PTC). **B**, AKAP8 downregulation in CD34$^+$ cells of *SRSF2* mutant MDS patients (RNA-seq data). **C-D**, PCR band isoform ratio quantification (**C**) and semi-quantitative band visualization using an Agilent 2100 bioanalyzer (**D**) of the aberrant splicing event in *AKAP8* measured in CD34$^+$

cells of *SRSF2* mutant MDS patient samples, SF-WT MDS patient samples and healthy controls. **E**, Sashimi plot showing the *SEPT2* A3SS event in *SF3B1* mutant MDS patient samples which results in a frameshift and generation of a PTC. **F**, *SEPT2* downregulation in $CD34^+$ cells of *SF3B1* mutant MDS patients (RNA-seq data). **G-H**, PCR band isoform ratio quantification (**G**) and semi-quantitative band visualization using an Agilent 2100 bioanalyzer (**H**) of the aberrant splicing event in *SEPT2* measured in $CD34^+$ cells of *SF3B1* mutant MDS patient samples, SF-WT MDS patient samples and healthy controls. **I-J**, PCR band ratio quantification plots for aberrant splicing events identified in *METTL17* (**I**) and *PFKM* (**J**) in *SRSF2* mutant MDS, measured in $CD34^+$ cells of *SRSF2* mutant MDS patient samples, SF-WT MDS patient samples and healthy controls. All PCR band ratio quantification plots are generated with a minimum of 5 samples per group and data are represented as mean ± SEM. For PCR band ratio quantification plots (**C, G, I and J**), p-values were obtained from a one-way ANOVA with Tukey's post-test (for normally distributed data with groups of equal variance), or a non-parametric Kruskal-Wallis with Dunn's post-test. For *AKAP8* and *SEPT2* expression plots (**B and F**) statistical significance was obtained from DESeq2. *$p<0.05$, **$p< 0.01$, ***$p< 0.001$.

Figure S1


A
ID
Subtype
RNA Splicing
SF3B1
SRSF2
U2AF1
ZRSR2
DNA Methylation
TET2
DNMT3A
IDH1
IDH2
Chromatin modifiers
ASXL1
EZH2
Pathway Signaling
RUNX1
JAK2
TP53
Cohesin
STAG2
RA
RARS
RAEB1
RAEB2
CMML

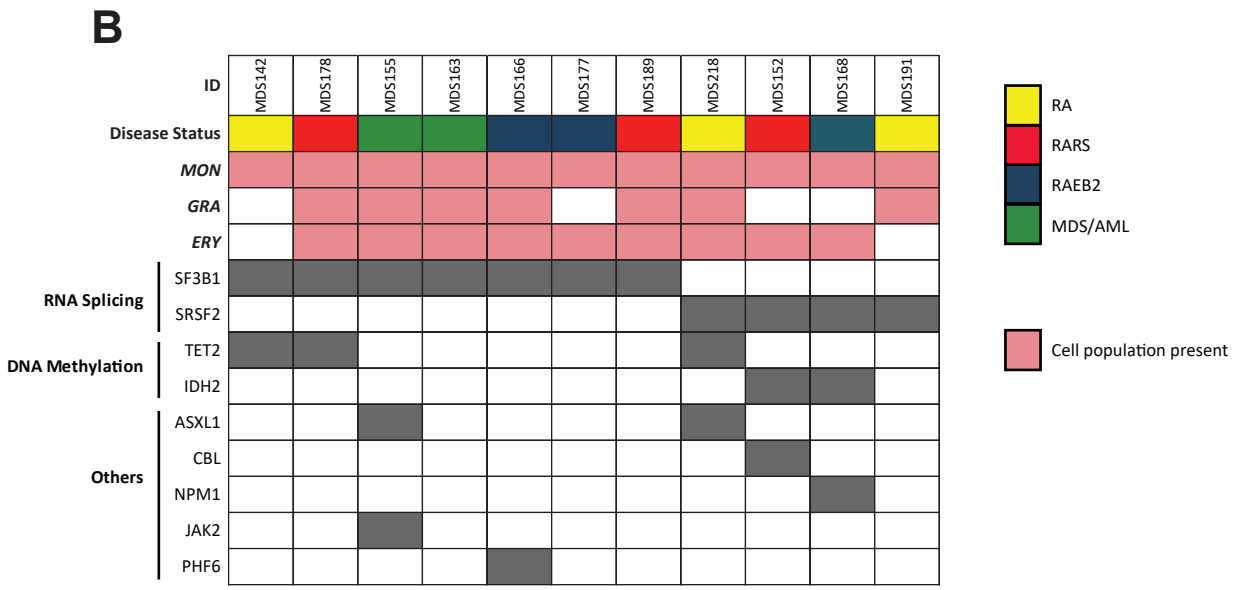

B
ID
MDS142
MDS178
MDS155
MDS163
MDS166
MDS177
MDS189
MDS218
MDS152
MDS168
MDS191
Disease Status
MON
GRA
ERY
RNA Splicing
SF3B1
SRSF2
DNA Methylation
TET2
IDH2
Others
ASXL1
CBL
NPM1
JAK2
PHF6
RA
RARS
RAEB2
MDS/AML
Cell population present

Figure S2

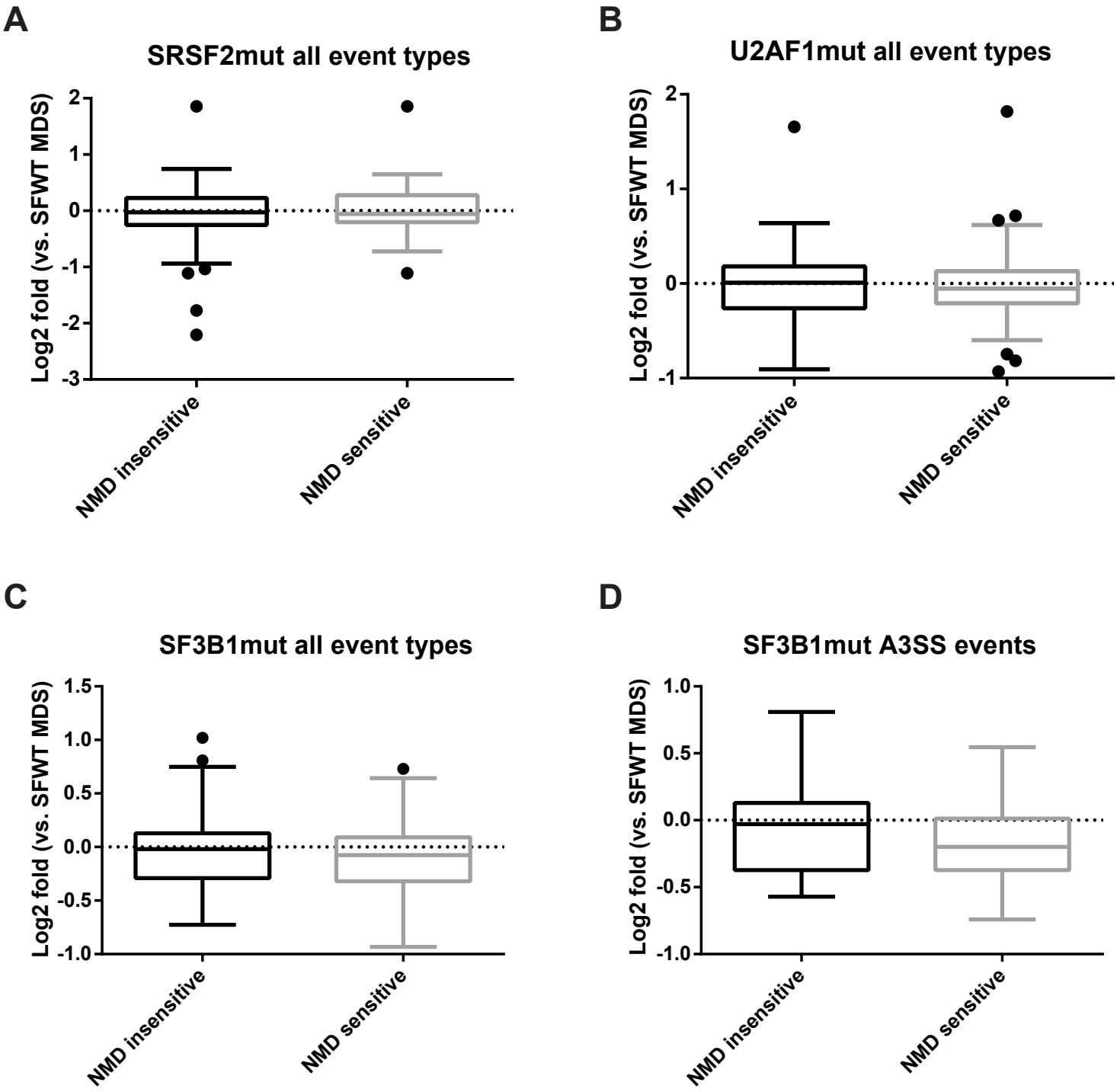
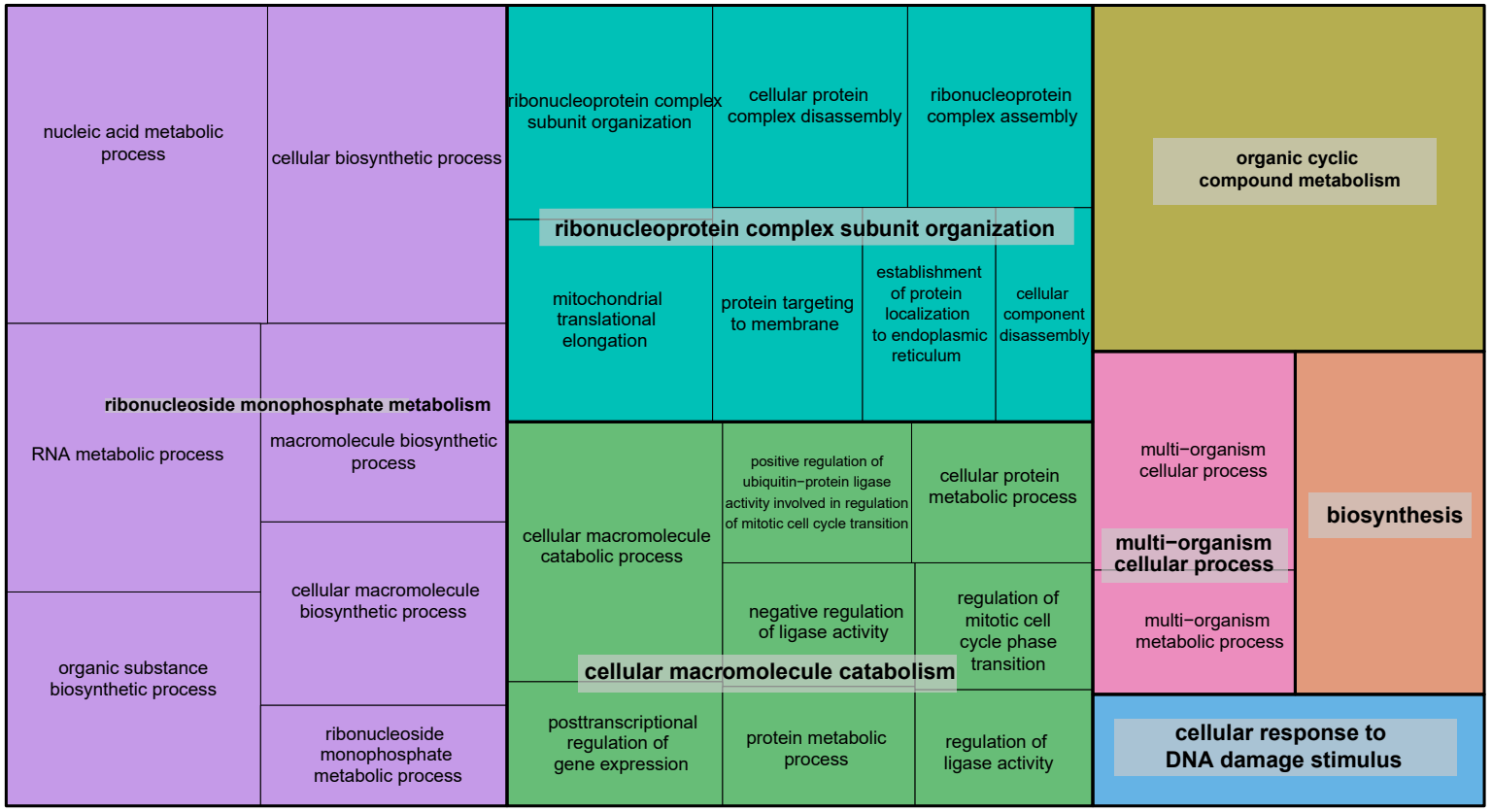

A
SRSF2mut all event types
Log2 fold (vs. SFWT MDS)
2
1
0
-1
-2
-3
NMD insensitive
NMD sensitive
B
U2AF1mut all event types
Log2 fold (vs. SFWT MDS)
2
1
0
-1
NMD insensitive
NMD sensitive
C
SF3B1mut all event types
Log2 fold (vs. SFWT MDS)
1.5
1.0
0.5
0.0
-0.5
-1.0
NMD insensitive
NMD sensitive
D
SF3B1mut A3SS events
Log2 fold (vs. SFWT MDS)
1.0
0.5
0.0
-0.5
-1.0
NMD insensitive
NMD sensitive

Figure S3

A


nucleic acid metabolic process
cellular biosynthetic process
ribonucleoprotein complex subunit organization
cellular protein complex disassembly
ribonucleoprotein complex assembly
organic cyclic compound metabolism
ribonucleoprotein complex subunit organization
mitochondrial translational elongation
protein targeting to membrane
establishment of protein localization to endoplasmic reticulum
cellular component disassembly
ribonucleoside monophosphate metabolism
RNA metabolic process
macromolecule biosynthetic process
positive regulation of ubiquitin−protein ligase activity involved in regulation of mitotic cell cycle transition
cellular protein metabolic process
multi−organism cellular process
biosynthesis
cellular macromolecule catabolic process
multi−organism cellular process
cellular macromolecule biosynthetic process
negative regulation of ligase activity
regulation of mitotic cell cycle phase transition
multi−organism metabolic process
organic substance biosynthetic process
cellular macromolecule catabolism
posttranscriptional regulation of gene expression
protein metabolic process
regulation of ligase activity
ribonucleoside monophosphate metabolic process
cellular response to DNA damage stimulus

B

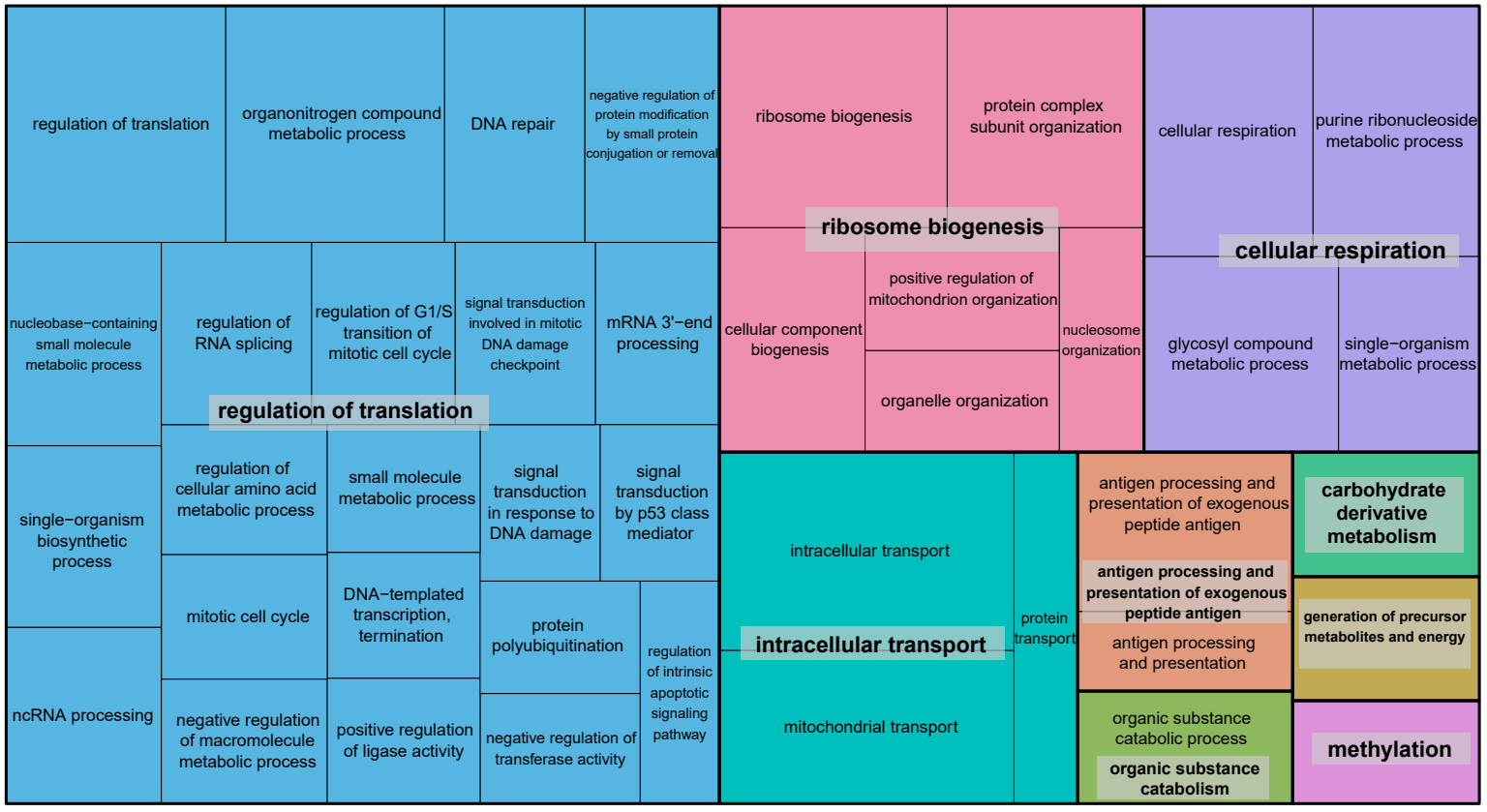

regulation of translation
organonitrogen compound metabolic process
DNA repair
negative regulation of protein modification by small protein conjugation or removal
ribosome biogenesis
protein complex subunit organization
cellular respiration
purine ribonucleoside metabolic process
nucleobase−containing small molecule metabolic process
regulation of RNA splicing
regulation of G1/S transition of mitotic cell cycle
signal transduction involved in mitotic DNA damage checkpoint
mRNA 3'−end processing
ribosome biogenesis
positive regulation of mitochondrion organization
cellular component biogenesis
nucleosome organization
cellular respiration
organelle organization
glycosyl compound metabolic process
single−organism metabolic process
regulation of translation
single−organism biosynthetic process
regulation of cellular amino acid metabolic process
small molecule metabolic process
signal transduction in response to DNA damage
signal transduction by p53 class mediator
intracellular transport
antigen processing and presentation of exogenous peptide antigen
carbohydrate derivative metabolism
mitotic cell cycle
DNA−templated transcription, termination
protein polyubiquitination
regulation of intrinsic apoptotic signaling pathway
protein transport
antigen processing and presentation of exogenous peptide antigen
generation of precursor metabolites and energy
intracellular transport
antigen processing and presentation
ncRNA processing
negative regulation of macromolecule metabolic process
positive regulation of ligase activity
negative regulation of transferase activity
mitochondrial transport
organic substance catabolic process
methylation
organic substance catabolism

Figure S4

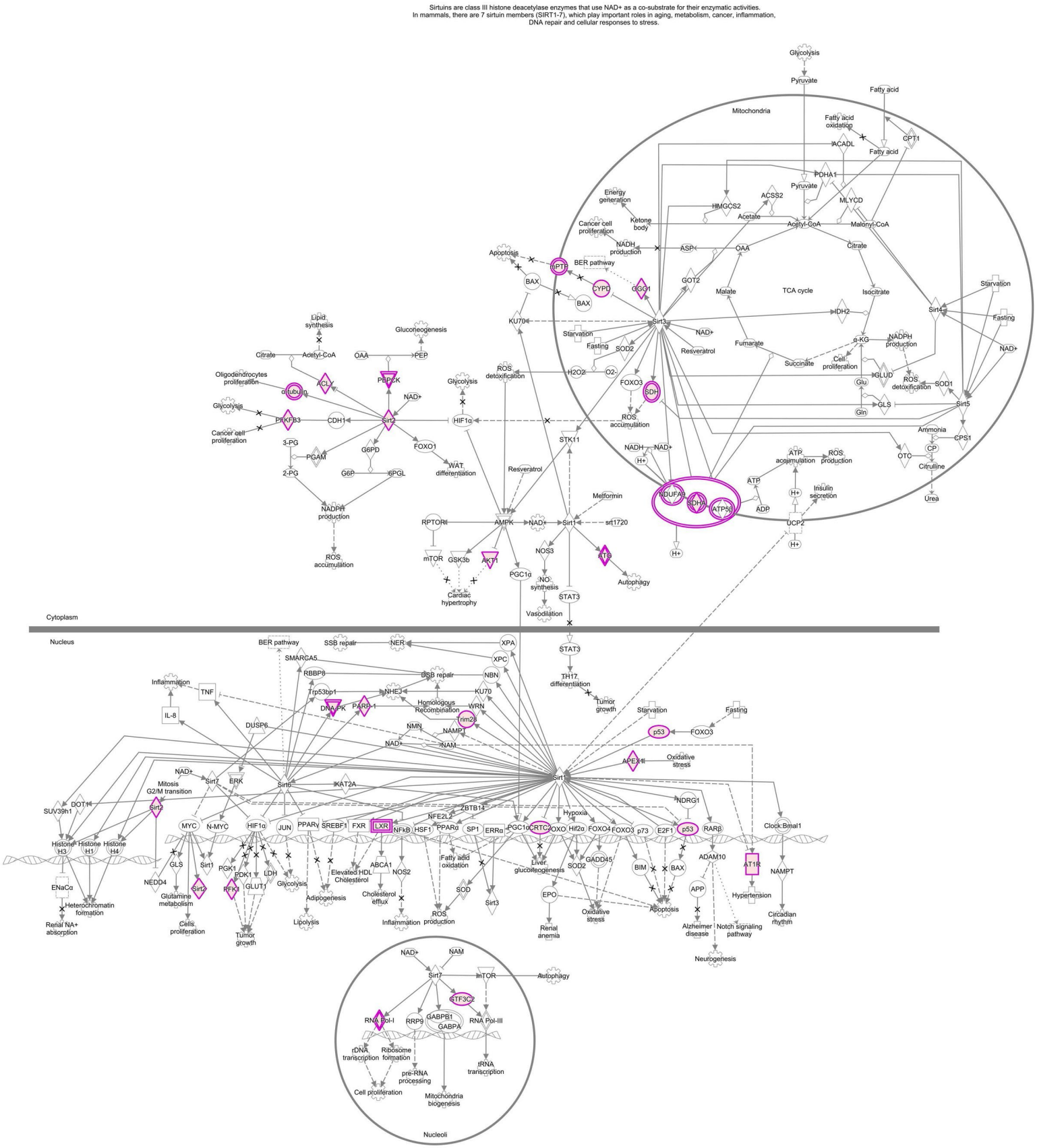
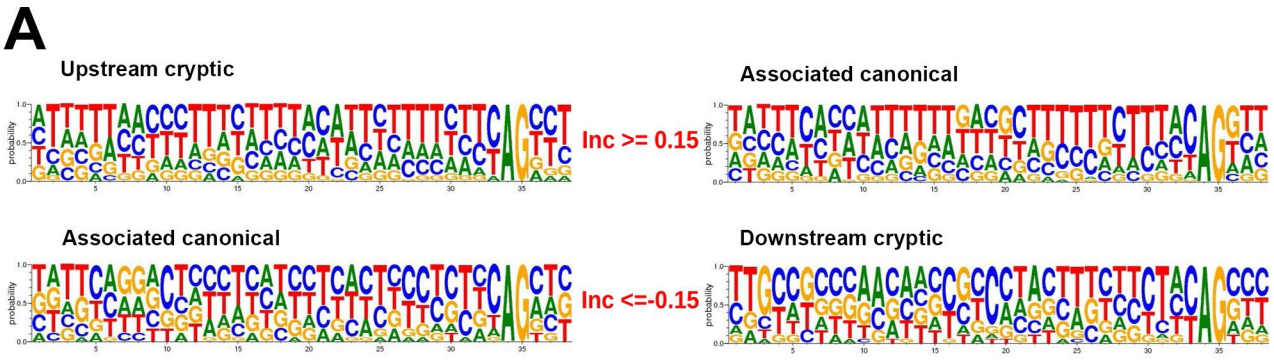
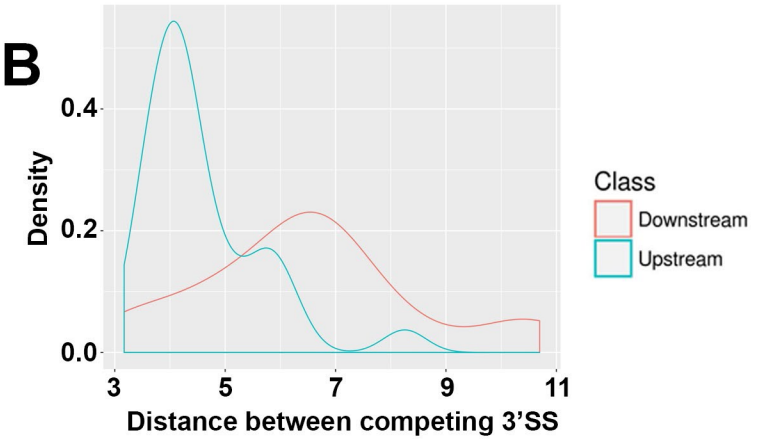
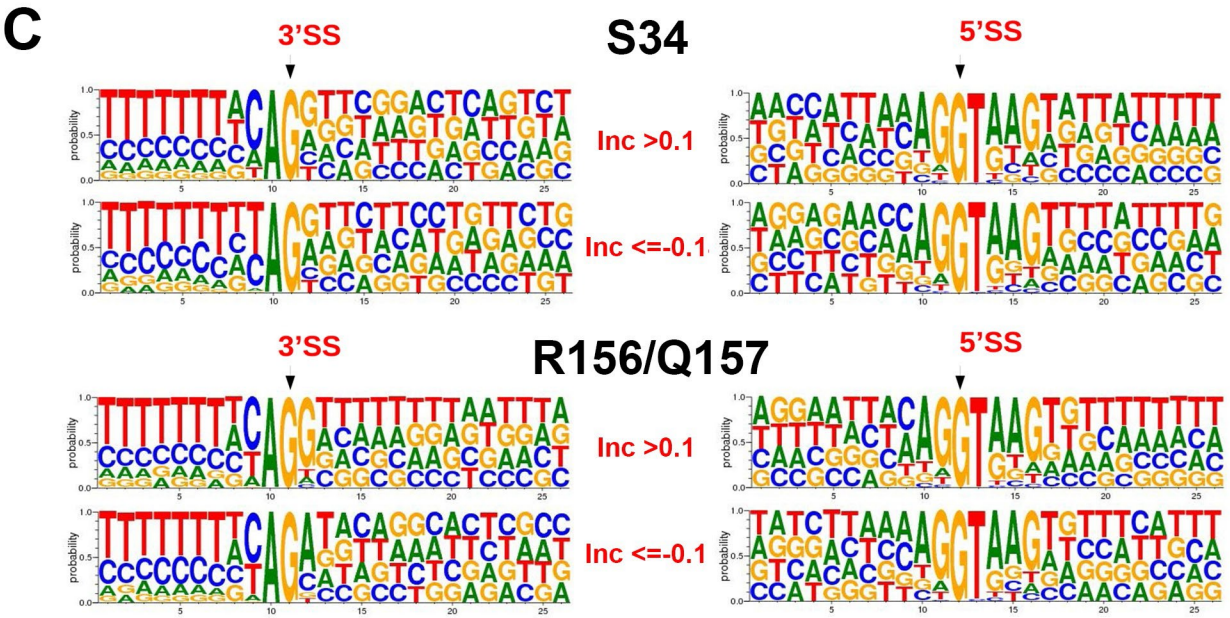
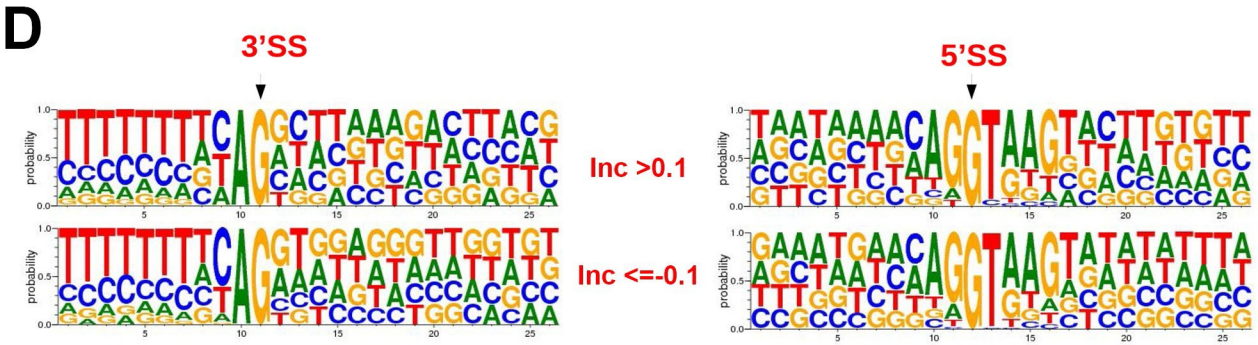


Figure S5


A
Upstream cryptic
Associated canonical
Inc >= 0.15
Associated canonical
Downstream cryptic
Inc <=-0.15


B
Density
0.4
0.2
0.0
3
5
7
9
11
Distance between competing 3'SS
Class
Downstream
Upstream


C
3'SS
S34
5'SS
Inc >0.1
Inc <=-0.1
3'SS
R156/Q157
5'SS
Inc >0.1
Inc <=-0.1


D
3'SS
5'SS
Inc >0.1
Inc <=-0.1

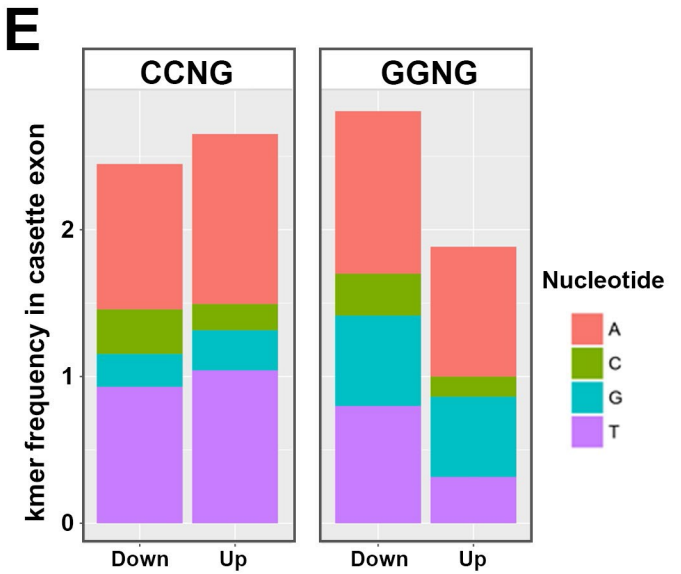

E
CCNG
GGNG
kmer frequency in casette exon
2
1
0
Down
Up
Down
Up
Nucleotide
A
C
G
T

Figure S6

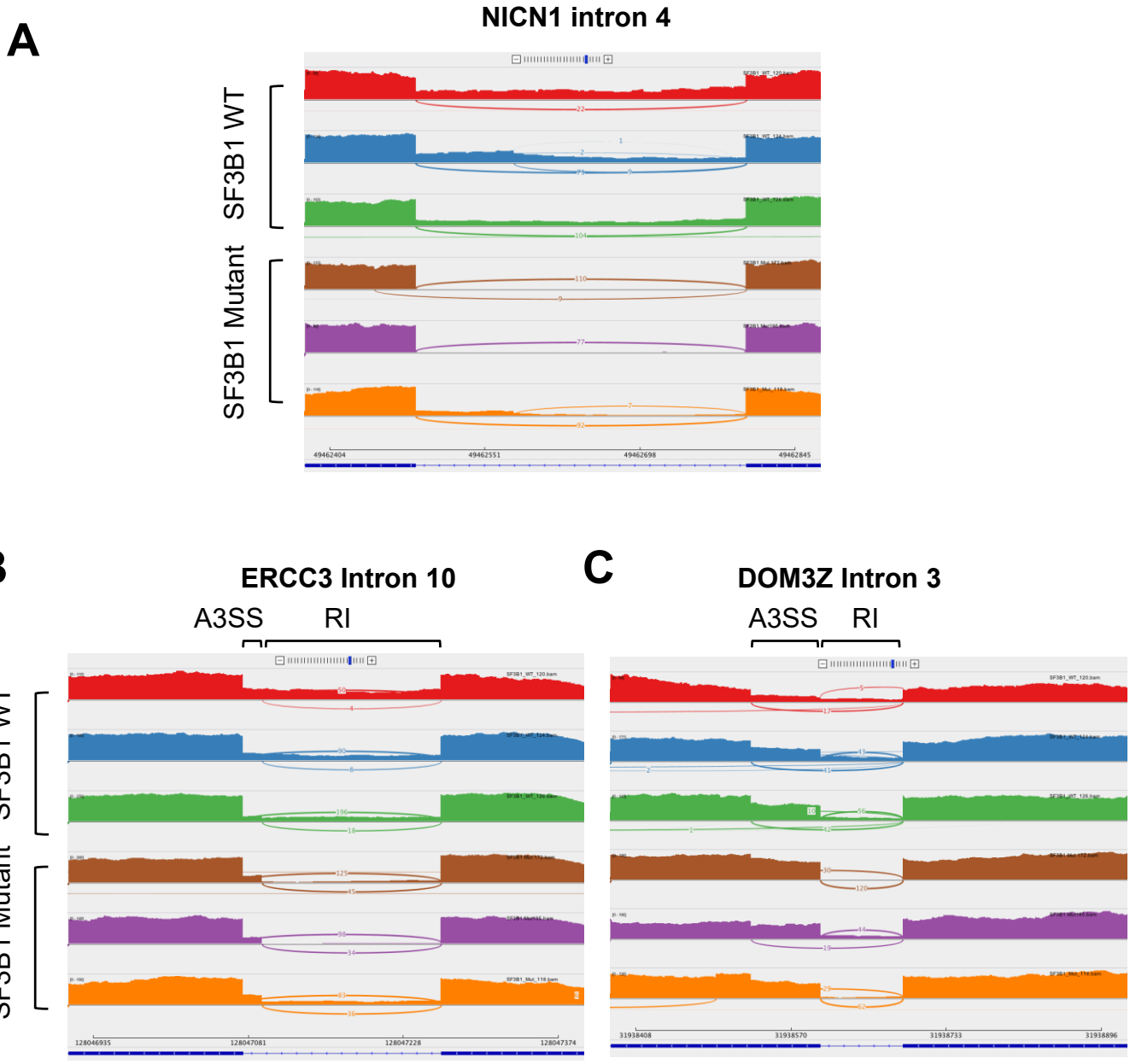

A
NICN1 intron 4
SF3B1 WT
SF3B1 Mutant
49462404
49462551
49462698
49462845
B
ERCC3 Intron 10
A3SS
RI
SF3B1 WT
SF3B1 Mutant
128046935
128047081
128047228
128047374
C
DOM3Z Intron 3
A3SS
RI
31938408
31938570
31938733
31938896

Figure S7

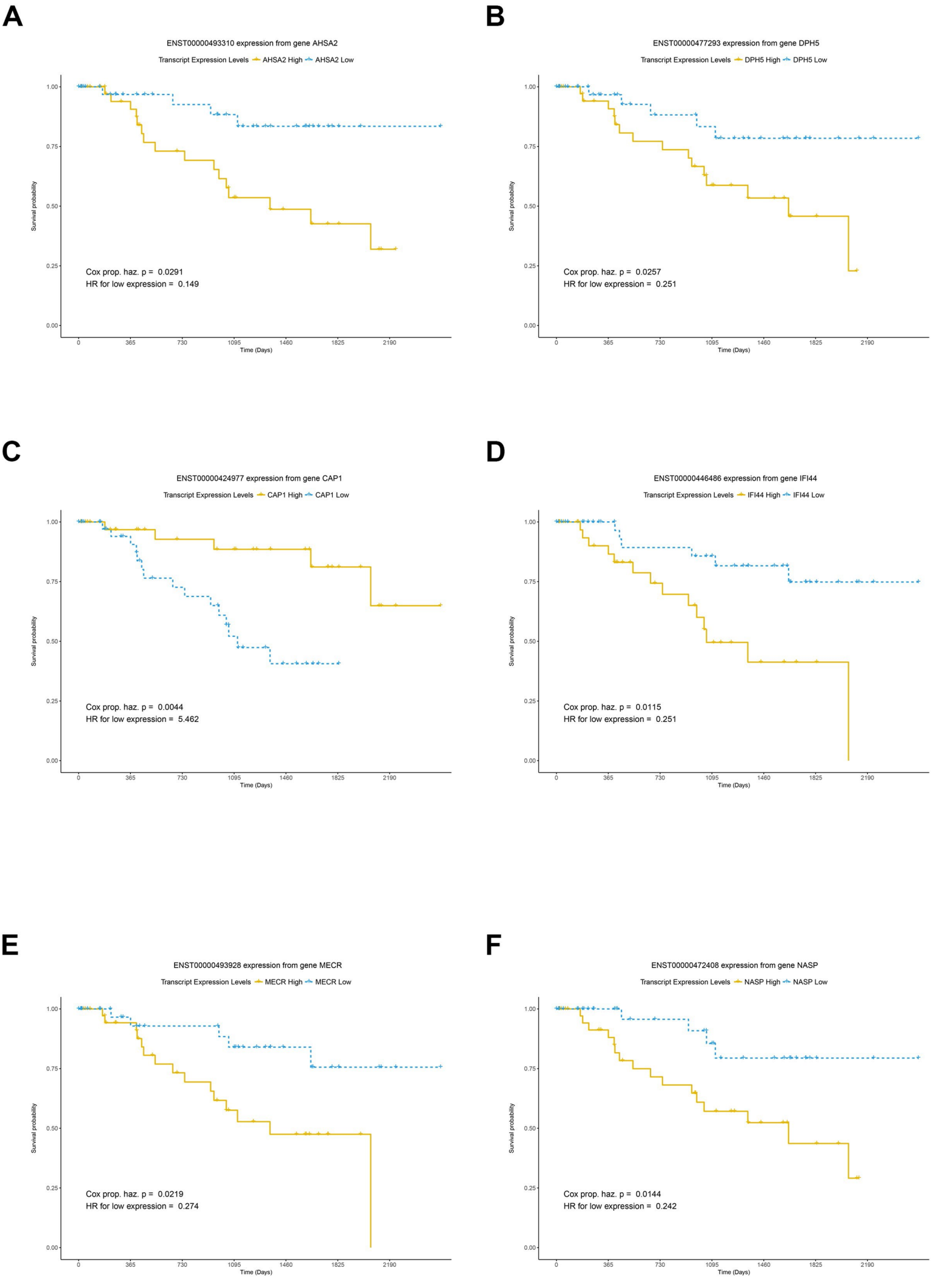

A
ENST00000493310 expression from gene AHSA2
Transcript Expression Levels AHSA2 High AHSA2 Low
Survival probability
Cox prop. haz. p = 0.0291
HR for low expression = 0.149
Time (Days)
B
ENST00000477293 expression from gene DPH5
Transcript Expression Levels DPH5 High DPH5 Low
Survival probability
Cox prop. haz. p = 0.0257
HR for low expression = 0.251
Time (Days)
C
ENST00000424977 expression from gene CAP1
Transcript Expression Levels CAP1 High CAP1 Low
Survival probability
Cox prop. haz. p = 0.0044
HR for low expression = 5.462
Time (Days)
D
ENST00000446486 expression from gene IFI44
Transcript Expression Levels IFI44 High IFI44 Low
Survival probability
Cox prop. haz. p = 0.0115
HR for low expression = 0.251
Time (Days)
E
ENST00000493928 expression from gene MECR
Transcript Expression Levels MECR High MECR Low
Survival probability
Cox prop. haz. p = 0.0219
HR for low expression = 0.274
Time (Days)
F
ENST00000472408 expression from gene NASP
Transcript Expression Levels NASP High NASP Low
Survival probability
Cox prop. haz. p = 0.0144
HR for low expression = 0.242
Time (Days)

Figure S8

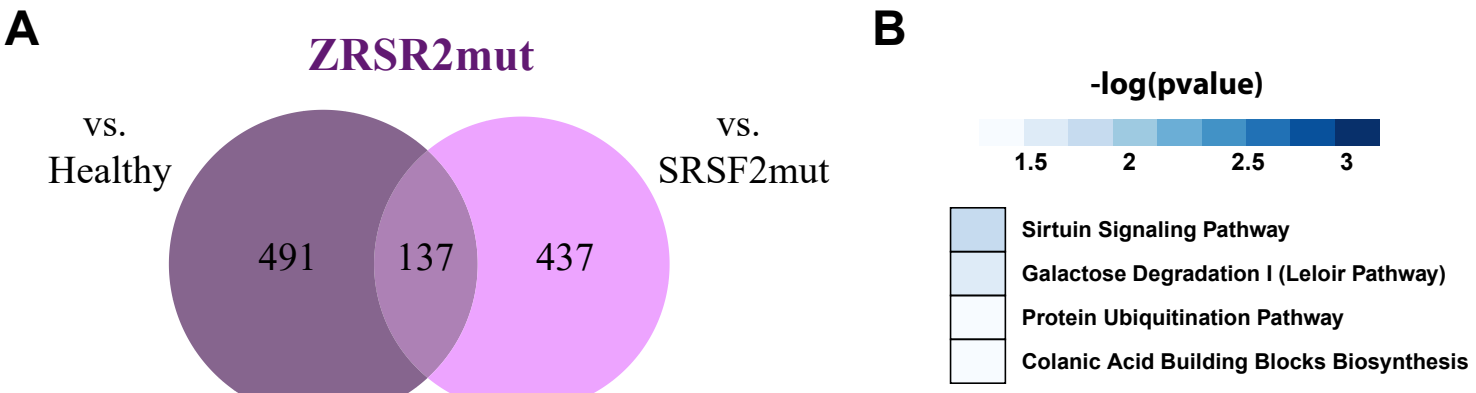

A
ZRSR2mut
vs.
Healthy
vs.
SRSF2mut
491
137
437
B
-log(pvalue)
1.5
2
2.5
3
Sirtuin Signaling Pathway
Galactose Degradation I (Leloir Pathway)
Protein Ubiquitination Pathway
Colanic Acid Building Blocks Biosynthesis

Figure S9

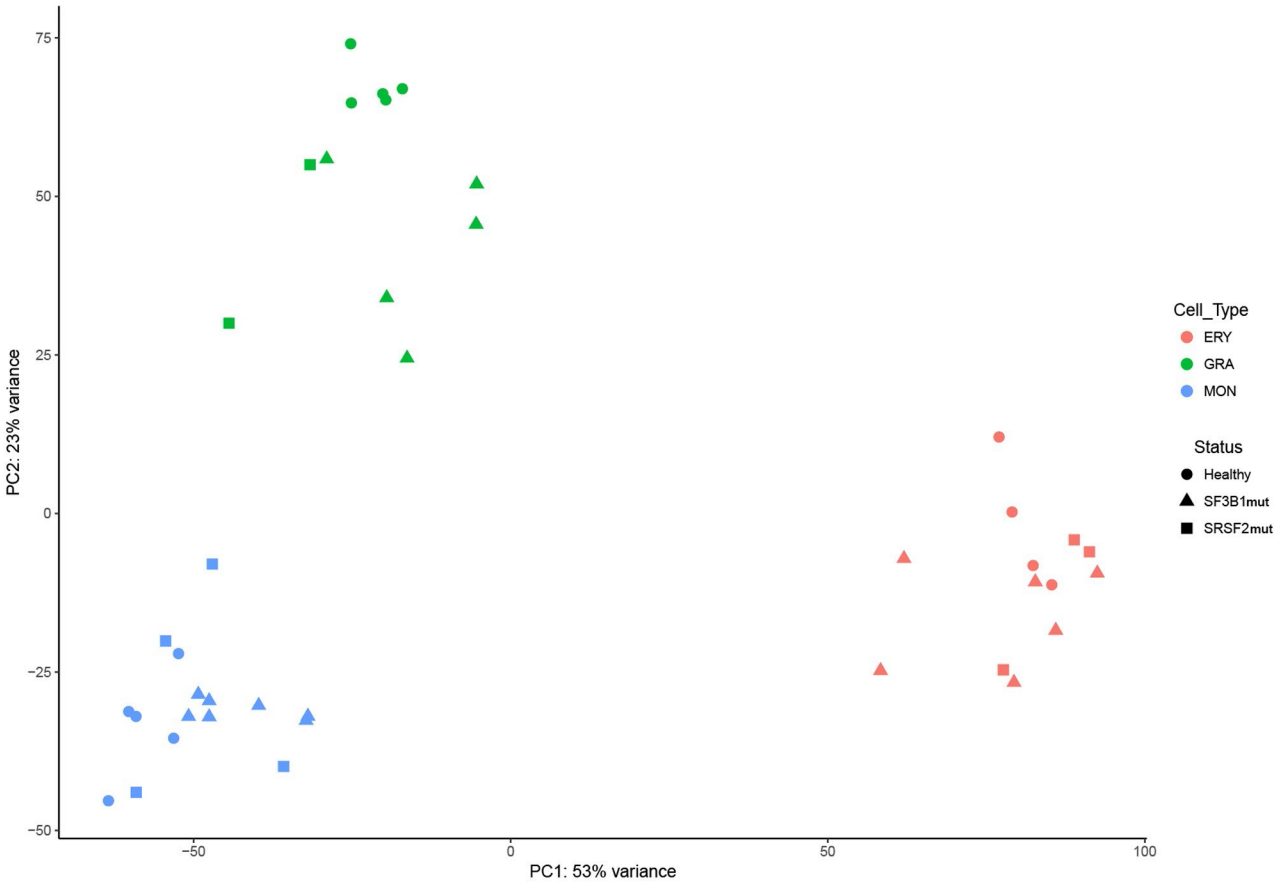
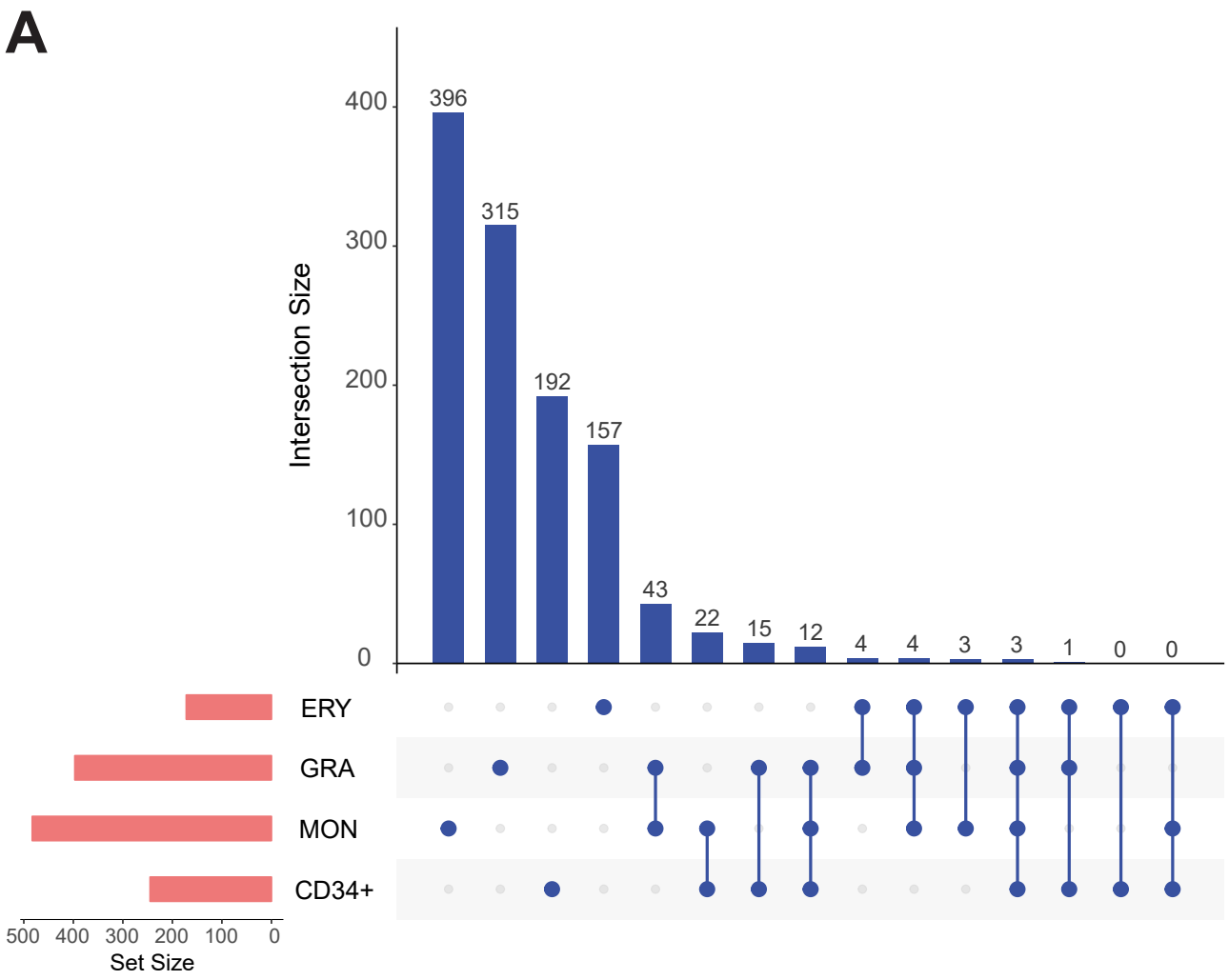

75
50
25
0
−25
−50
PC2: 23% variance
−50
0
50
100
PC1: 53% variance
Cell_Type
ERY
GRA
MON
Status
Healthy
SF3B1mut
SRSF2mut

Figure S10


A
Intersection Size
400
300
200
100
0
396
315
192
157
43
22
15
12
4
4
3
3
1
0
0
ERY
GRA
MON
CD34+
500 400 300 200 100 0
Set Size

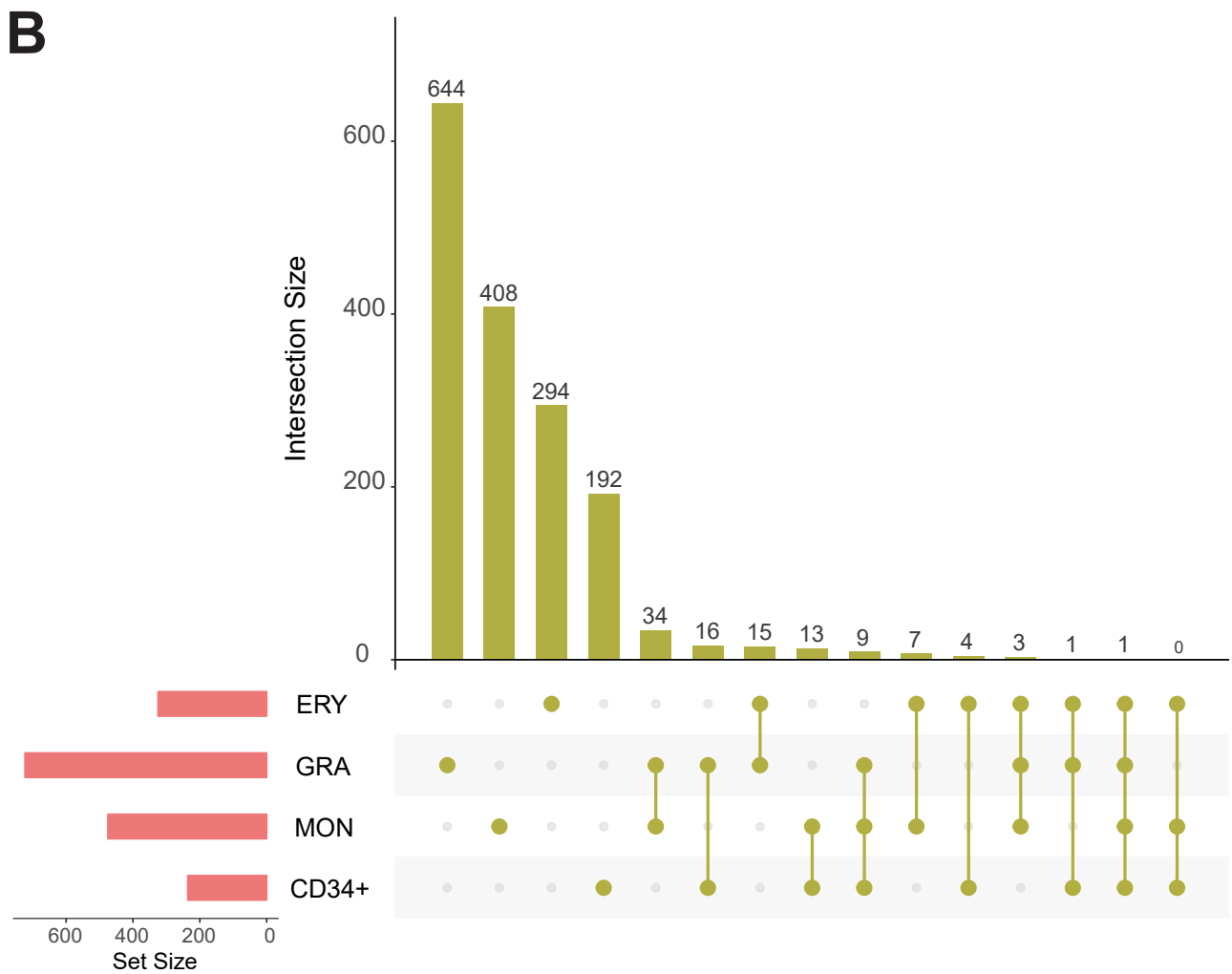
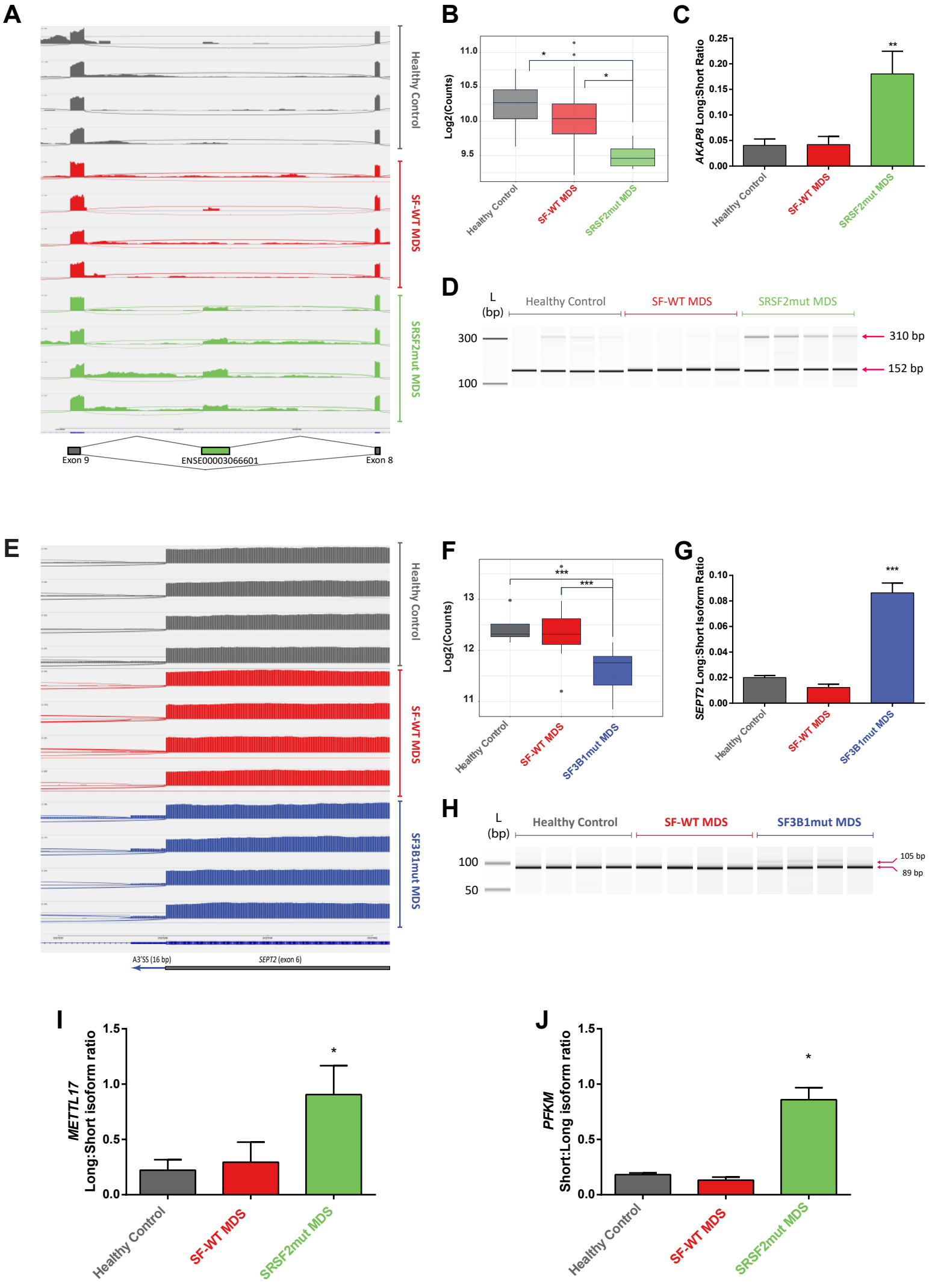

B
Intersection Size
600
400
200
0
644
408
294
192
34
16
15
13
9
7
4
3
1
1
0
ERY
GRA
MON
CD34+
600 400 200 0
Set Size

Figure S11


A
Healthy Control
SF-WT MDS
SRSF2mut MDS
Exon 9
ENSE00003066601
Exon 8
B
Log2(Counts)
11.0
10.5
10.0
9.5
*
*
Healthy Control
SF-WT MDS
SRSF2mut MDS
C
AKAP8 Long:Short Ratio
0.25
0.20
0.15
0.10
0.05
0.00
**
Healthy Control
SF-WT MDS
SRSF2mut MDS
D
L
(bp)
Healthy Control
SF-WT MDS
SRSF2mut MDS
300
100
310 bp
152 bp
E
Healthy Control
SF-WT MDS
SF3B1mut MDS
A3'SS (16 bp)
SEPT2 (exon 6)
F
Log2(Counts)
13
12
11
***
***
Healthy Control
SF-WT MDS
SF3B1mut MDS
G
SEPT2 Long:Short Isoform Ratio
0.10
0.08
0.06
0.04
0.02
0.00
***
Healthy Control
SF-WT MDS
SF3B1mut MDS
H
L
(bp)
Healthy Control
SF-WT MDS
SF3B1mut MDS
100
50
105 bp
89 bp
I
METTL17
Long:Short isoform ratio
1.5
1.0
0.5
0.0
*
Healthy Control
SF-WT MDS
SRSF2mut MDS
J
PFKM
Short:Long isoform ratio
1.5
1.0
0.5
0.0
*
Healthy Control
SF-WT MDS
SRSF2mut MDS