SUPPLEMENTARY METHODS

Single cell DNA sequencing analysis

Out of the 58 dual RNA splicing factor mutant patients included in our bulk sequencing analysis, 30 were treated at either MSKCC or Mayo Clinic and therefore potentially available for single cell analysis. From these 30 samples, only 15 samples were available as viably frozen cells and, ultimately, only 11 out of the 15 samples had viability >90% when thawed.

Pooled single cell DNA libraries were sequenced on Illumina MiSeq with paired-end multiplex runs (2 x 150 bp). Raw sequencing reads in FASTQ files were processed using the Tapestri Pipeline, which uses Bluebee's High Performance Genomics Platform. The pipeline trims adapter information, aligns sequences to the reference genome (hg19), assigns sequence reads to individual cell barcodes, calls genotypes using GATK, and generates annotated loom files. Resulting data that met the following filtering criteria were included for downstream analysis (**Supplementary Table 5**): read depth > 10, genotype quality (i.e., difference between the likelihoods of the two most likely genotypes on a scale of 0-99) > 30, alternate allele frequency (i.e., number of reads that support each of the reported alleles) > 20, variants genotyped in > 10% of cells, variants mutated in >1% of cells, and cells with >50% of genotypes present. VAFs were determined by aggregated reads from filtered cells, i.e. number of reads with variant of interest / number of total reads. The co-occurrence or mutual exclusivity of the dual splicing factor alterations was calculated at the same time as the rate of allele dropout (ADO) using the Bayesian statistical package STAN[1]. Since ADO of the relevant splicing factors can strongly influence the observed degree of co-occurrence, it was modeled under the assumption that splicing factor alterations were diploid and either wild-type or heterozygous in all cells and also that mutant and wild-type alleles are equally likely to be amplified. The estimated median ADO rate was 22% overall (IQR: 14% - 26%) but 73.9% (ICR: 68 - 77%) at the locus encoding SRSF2 P95. We attempted to address this by jointly estimating the co-occurrence of *SF3B1, SRSF2,* and *U2AF1* mutations and ADO at both loci under the assumption that hotspot mutations in these genes were diploid and either wild-type or heterozygous in all cells.

Subclones from each sample were identified using Tapestri Insights and annotated NGS data were subsequently analyzed in R. Oncoprints were generated where each column represents an individual cell and each row represents a mutation. Cells were clustered by subclone and then arranged by size of subclones from large to small. Pairwise associations between mutations were evaluated by Fisher's exact test with correction for multiple testing

using the Benjamini-Hochberg method. The order of mutations in each sample was inferred based on the principle of maximum parsimony and depicted on fishplots in accordance with the proportions of the subclones.

U2AF1 allele-specific sequencing

Cryopreserved BM MNCs from a patient with co-occurring U2AF1^{S34F} and U2AF1^{Q157R} mutations by bulk and single cell DNA sequencing were thawed and lysed in TRI Reagent®. Total RNA was extracted using the Direct-zol^{TM} RNA MiniPrep. Next, RT-PCR was performed using the QIAGEN One-Step RT-PCR Kit with the following primers: forward, 5'-GGCACCGAGAAAGACAAAGT-3'; reverse, 5'-AGCTCTCTGGAAATGGGCTT-3'. The PCR product was confirmed by gel electrophoresis and purified using the QIAquick PCR Purification Kit. Next, the U2AF1 PCR product was cloned into the pGEM®-T vector at a 3:1 insert to vector molar ratio. The ligated product was transformed into NEB® Turbo Competent E. coli (High Efficiency). Transformed bacteria were then plated onto LB agar plates with 100 µg/ml ampicillin, 0.1 mM IPTG, and 80 µg/ml X-Gal and incubated for overnight hours at 37°C. The resulting white colonies generally contain inserts whereas the blue colonies contain empty vectors. White colonies were thus selected for Sanger sequencing using the T7 primer which binds upstream of the insert on the vector backbone. U2AF1^{S34F} results from a C-to-T mutation that changes the codon from TCT to TTT. U2AF1^{Q157R} results from an A-to-G mutation that changes the codon from CAG to CGG.

Genome annotations and RNA-seq read mapping

Genome annotations for RNA-seq read mapping to the NCBI GRCh37 / UCSC hg19 human genome assembly were created as previously described[2]. In brief, genome annotations from Ensembl release 71 and the UCSC knownGene track were merged with isoform annotations from the MISO v2.0 annotation to create a merged gene and isoform annotation[3-5]. RNA-seq reads were mapped to this merged annotation using RSEM v1.2.4, modified to call Bowtie v1.0.0 using the option '-v 2'[6,7]. RNA-seq reads that were not aligned to the transcriptome by RSEM were then aligned to a database of possible splice junctions, consisting of all possible combinations of 5' and 3' splice sites for each gene, with TopHat v2.0.8b[8]. Mapped reads from each procedure were merged together and used as input for subsequent analyses of isoform expression.

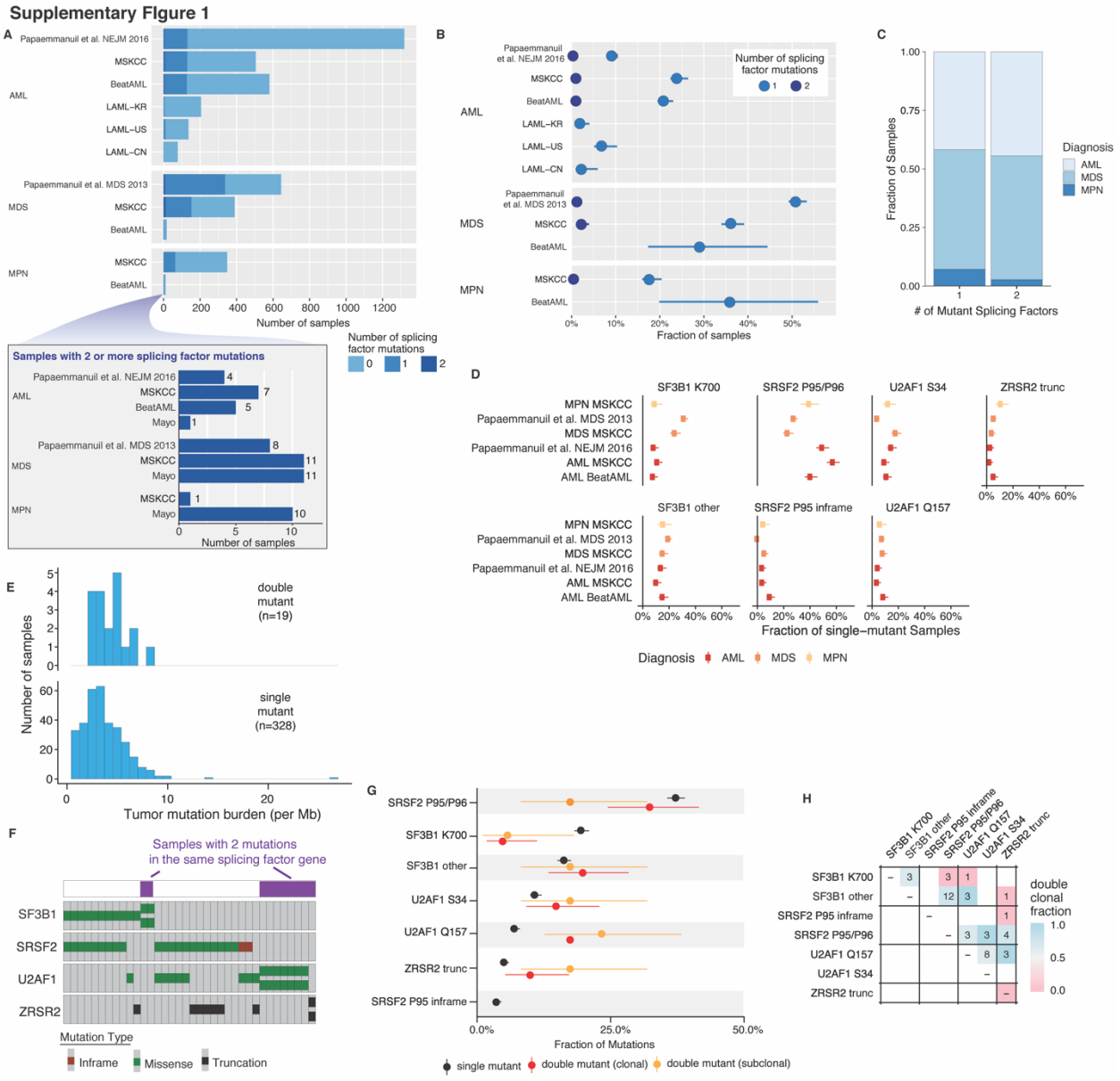
Differential RNA splicing analysis

Isoform ratios for annotated splicing events were calculated using MISO v2.0[5]. Differential splicing was identified by comparing each mutant patient sample to the median of the $SF3B1^{WT}$ controls. The most robust events for each SF3B1 mutant group were identified by restricting the analysis to events with at least 20 informative reads per sample, an associated Bayes factor of ≥5 (computed using Wagenmakers' framework)[9], and absolute change in isoform ratio of ≥10%. The top splicing events were defined as events with the greatest absolute change in isoform usage that occurred with significance in all samples from each mutation group.

Isothermal titration calorimetry (ITC)

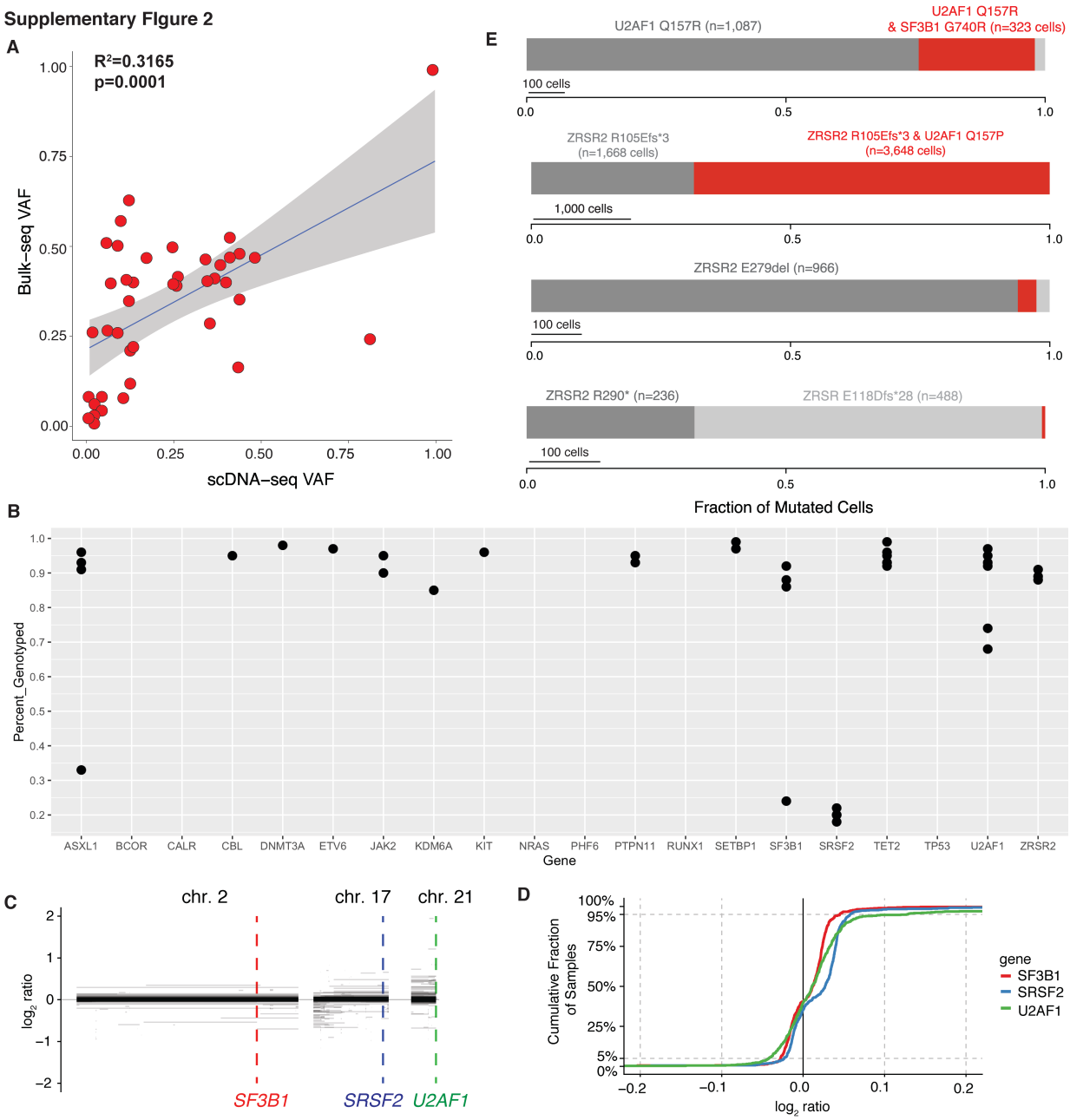
SRSF2 peptides were purified with the same protocol as previously described[10] and ITC was performed as previously described.[11] Briefly, cDNAs of SRSF2 WT and P95A, P95H, P95HL, and P95R mutants encoding amino acids 1-101 of SRSF2 (which encompasses the RNA recognition motif (RRM) of SRSF2 were cloned into the pet28a plasmid using the restriction sites *Bam*H1/*Xho*1. SRSF2 proteins were overexpressed at 37°C for 3 h in *Escherichia coli BL21 (DE3) codon plus* cells in minimal M9 medium (1 g l^{-1} ^{15}N-NH_4Cl, 2 g l^{-1} ^{13}C-glucose) using 0.1 mM IPTG. Protein was then purified by two successive nickel affinity chromatography steps and dialysed against an NMR buffer (50 mM L-Glu, 50 mM L-Arg and 20 mM Na_2HPO_4/NaH_2PO_4 at pH 5.5). A last purification step by size exclusion chromatography with a Superdex75 column (GE Healthcare) was necessary to remove residual RNases in the solution. The protein could be concentrated to over 2 mM with a 10-kDa molecular mass cutoff membrane.

To perform ITC, 5'-uCCAGu-3' RNA oligonucleotide ligand was utilized. This RNA was purchased from Dharmacon, deprotected according to the manufacturer's protocol, purified by butanol extraction, lyophilized and resuspended in NMR buffer. ITC measurements were conducted either on a MicroCal VP-ITC or MicroCal iTC200 instrument, which were calibrated according to the manufacturer's protocol. Concentrations of RNA and protein were calculated based on their optical density absorbance at 260 nm or 280 nm, respectively. The sample cell was loaded with either 1.4ml (VP-ITC) of 10 μM RNA and the syringe with 200 μM of protein, or 0.25ml (iTC200) of 10 μM protein and the syringe with 200μM RNA. Measurements were done at 37°C in the final buffer using either 35 injections of 6 μl protein (VP-ITC) or 21 injections of 2 μl RNA (iTC200). Data was integrated and normalized using the Origin 7.0 software according to a 1:1 RNA:protein ratio binding model.



Supplementary Figure 1. Characteristics of study cohort and genetic alterations in splicing factors analyzed. (A) Histogram of patients with acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), and myeloproliferative neoplasms (MPNs) with zero, one, or two mutations in RNA splicing factors studied here. Inset shows exact numbers of patients with two mutant splicing factors. **(B)** Percentage of samples within each study that contain one or two mutations in RNA splicing factors. **(C)** Fraction of samples with diagnosis of AML, MDS, or MPN based on presence of one versus two RNA splicing factor mutations. **(D)** Percentage of samples amongst single mutants that contained mutations at the $SF3B1^{K700}$, $SRSF2^{P95/P96}$, $U2AF1^{S34}$, or $U2AF1^{Q157}$ residues, other residues of SF3B1 ("SF3B1 other"), inframe deletions or insertions around the $SRSF2^{P95}$ residue ("SRSF2 P95 inframe"), and ZRSR2 truncating mutations ("ZRSR2 trunc"). **(E)** Histogram of tumor mutational burden in patients with two or more (top; "double mutant") mutant splicing factors versus a single mutant splicing factor (bottom). **(F)** Oncoprint of patients with two splicing factor mutations with annotation to describe patients with >1 mutation within the same gene. **(G)** Percentage of patients with mutations in

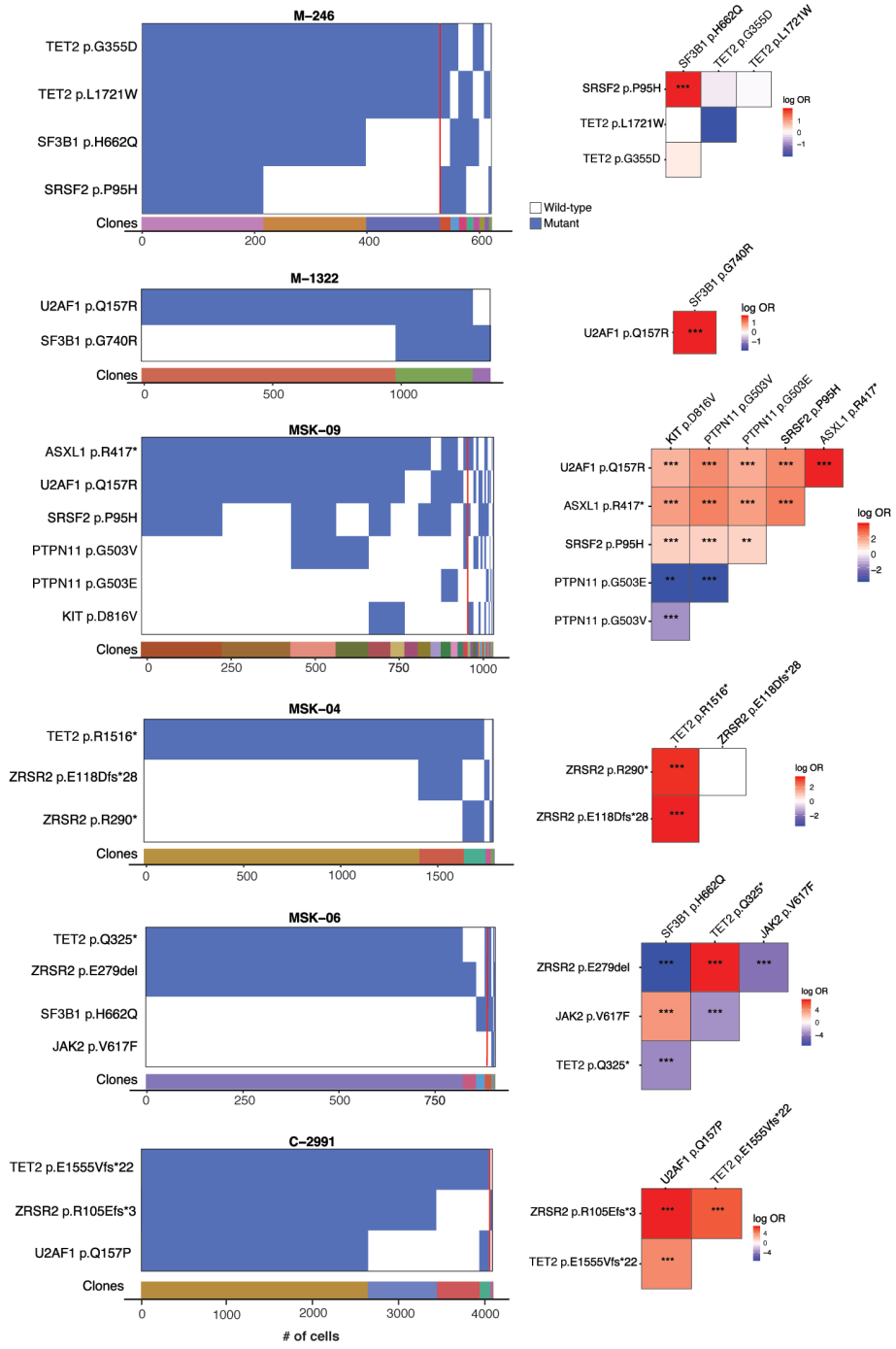
SRSF2, SF3B1, U2AF1, and ZRSR2 as a single mutation (in black) or dual mutation where ≥1 mutation was subclonal (orange) or both were clonal (red). Error bars show one standard deviation, based on a binomial distribution. **(H)** Plot describing number of patients with co-existing mutant alleles in splicing factors restricted to double-mutant patients where both mutations were clonal. The expected number was based on the fraction of samples with exactly two mutations under the assumption of no mutual exclusivity and using a Poisson distribution.



Supplementary Figure 2. Single cell genomic analysis of myeloid neoplasm patients bearing two mutations in RNA splicing factors. (A) Correlation of the variant allele fraction (VAF) from bulk sequencing and single cell DNA sequencing. The X-axis shows the VAF from the single-cell genotype data (scDNA-seq VAF). The Y-axis shows the VAF from the bulk next-generation sequencing (bulk VAF). Each dot represents a detected variant. The linear trendline was added to best fit the distribution of the dots. The shaded area around the trendline represents the 95% confidence intervals. **(B)** Genotyping efficiencies from scDNA-seq for each gene sequenced. Each dot represents a detected variant. **(C)** Copy number profiles for the 600 samples in the cohort for which they could be reliably determined (all derived using MSK IMPACT data). **(D)** Cumulative distribution of tumor/normal log-ratio at each splicing factor locus. Fewer than 5% of samples had a $\log_2$ratio more extreme of ±0.1 and fewer than 3% had a $\log_2$ratio of more extreme of ±0.2. **(E)** Fraction of mutated cells with one or two mutations in

RNA splicing factors within each unique dual splicing factor mutant patient. Red bar denotes fraction of individual cells where two mutations were identified within the same cell.

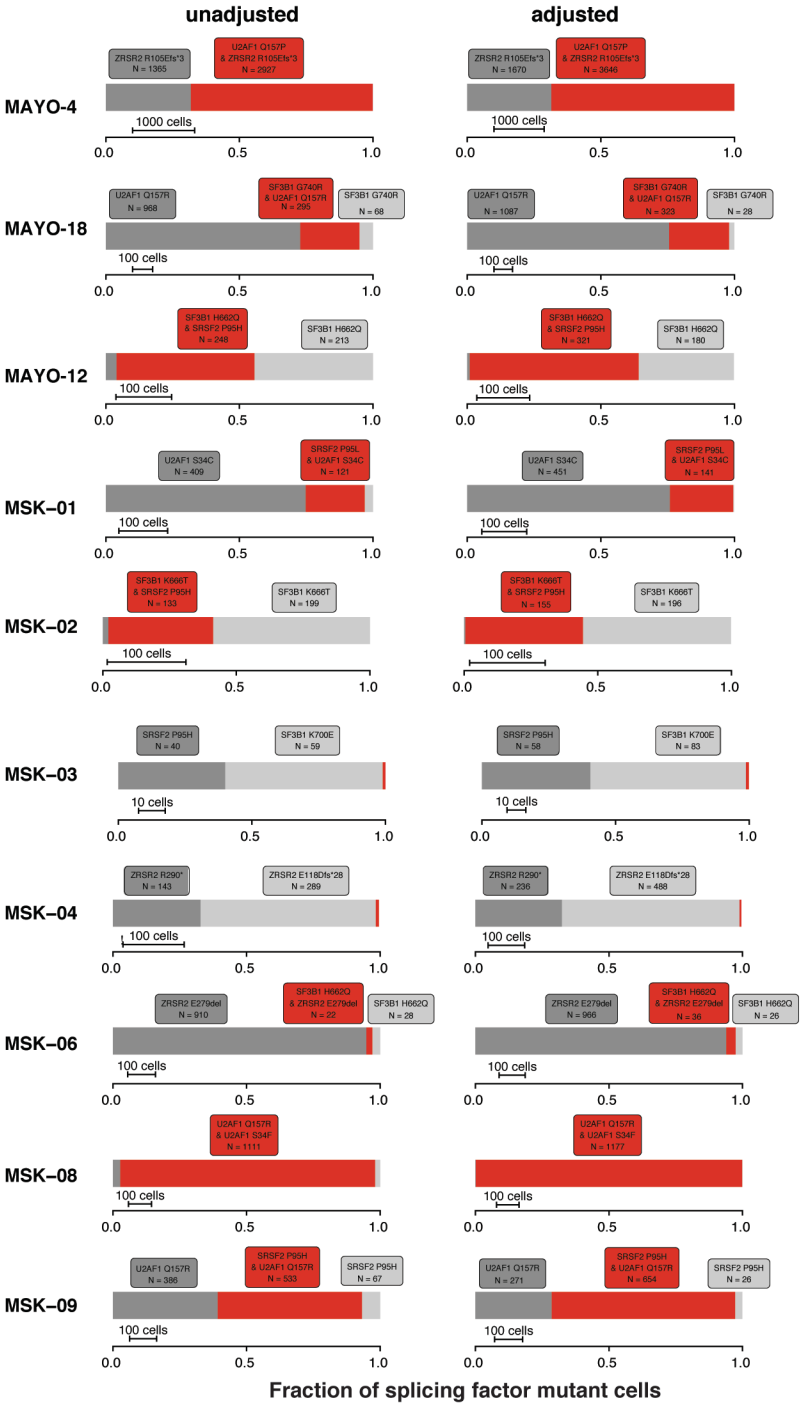
Supplementary Figure 3



Supplementary Figure 3. Oncoprints of mutations in individual cells of dual splicing factor mutant myeloid neoplasm patients. (A) Cellular-level co-occurrence of mutations in select double-splicing factor mutant patient cases (clinical diagnosis and sample ID are listed above each oncoprint). Oncoprint (left) shows the genotype of each sequenced cell for each variant, with clustering based on the genotypes of driver mutations. Each column represents a cell at the indicated scale. Mutant and wild-type cells are indicated in blue and white, respectively. The subclones located to the right of the red line comprised <1% of the total sequence cells, since such small subclones can represent false positive or negative genotypes

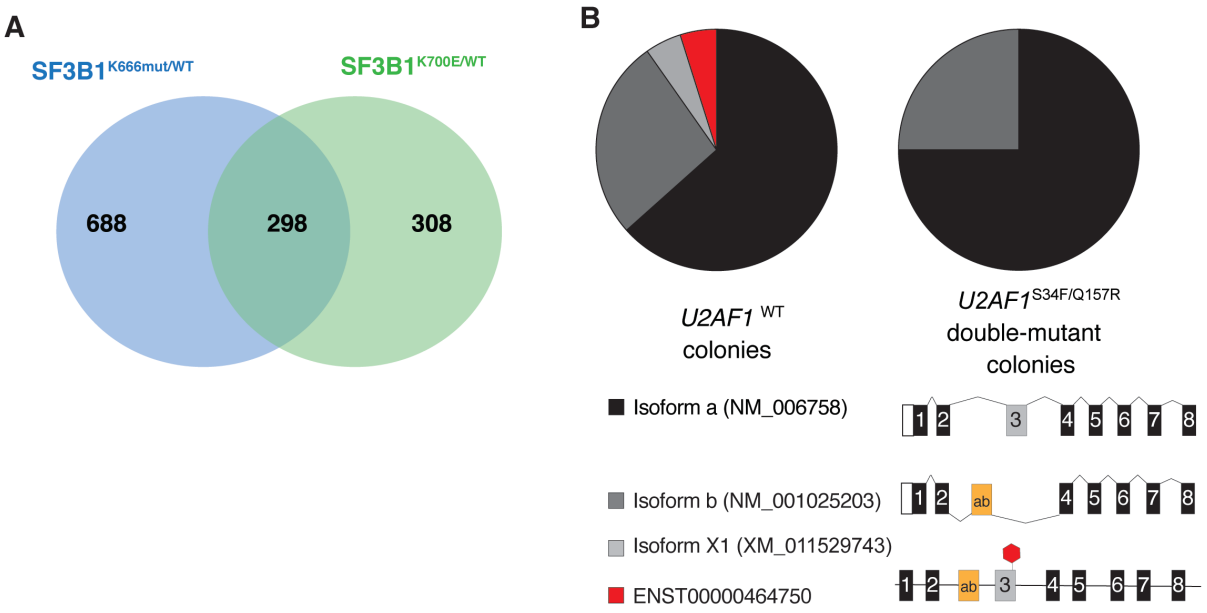
as a result of allele-drop out or multiplets. The figures on the right show the pairwise association of mutations. The color and size of each panel represent the degree of the logarithmic odds ratio (log OR). The vertical bar is a key indicating the association of the colors with the log OR. Co-occurrence and mutual exclusivity are indicated by red and blue, respectively. The statistical significance of the associations based on the false discovery rate (FDR) is indicated by the asterisks (*FDR < 0.1, **FDR < 0.05, ***FDR < 0.001).

Supplementary Figure 4



Supplementary Figure 4. Clonal structure of RNA splicing factor mutations in double mutant patients before and after adjustments for allele dropout (ADO). ADO was adjusted under the assumptions that i) mutant and wild-type alleles are equally likely to be amplified and ii) mutations are consistently heterozygous in diploid cells. The left panel shows unadjusted estimates of mutant subclones within each patient. The right panel shows adjusted estimates in corresponding patients. Red bar denotes fraction of cells with both splicing factor mutations. Gray bars denote fraction of cells with one of the two splicing factor mutations.

Supplementary Figure 5



Supplementary Figure 5. Allele-specific differences in splicing based on SF3B1 mutational hotspot and *U2AF1* isoform usage in *U2AF1*^{WT} versus *U2AF1*$^{S34F/Q157R}$ double mutant cells. (A) Venn diagram illustrating the overlap between differentially spliced events in SF3B1^{K700E} and SF3B1^{K666} mutant versus WT samples. **(B)** Pie-chart distribution of *U2AF1* mRNA isoforms in U2AF1^{WT} and U2AF1$^{S34F/Q157}$ double-mutant clones. Abbreviations: CMML: chronic myelomonocytic leukemia; PMF: primary myelofibrosis; t-MDS: therapy-related myelodysplastic syndrome.

Supplemental References

1. Carpenter B, Gelman A, Hoffman MD, et al. Stan: A Probabilistic Programming Language. *2017*. 2017;76(1):32.
2. Dvinge H, Ries RE, Ilagan JO, Stirewalt DL, Meshinchi S, Bradley RK. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc Natl Acad Sci U S A*. 2014;111(47):16802-16807.
3. Flicek P, Ahmed I, Amode MR, et al. Ensembl 2013. *Nucleic Acids Res*. 2013;41(Database issue):D48-55.
4. Meyer LR, Zweig AS, Hinrichs AS, et al. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res*. 2013;41(Database issue):D64-69.
5. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*. 2010;7(12):1009-1015.
6. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
7. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
8. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105-1111.
9. Wagenmakers EJ, Lodewyckx T, Kuriyal H, Grasman R. Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. *Cogn Psychol*. 2010;60(3):158-189.
10. Daubner GM, Clery A, Jayne S, Stevenin J, Allain FH. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J*. 2012;31(1):162-174.
11. Kim E, Ilagan JO, Liang Y, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer cell*. 2015;27(5):617-630.

Supplementary Table 1. All splicing factor mutations included in analysis.

Splicing Factor	Mutation (count)
SF3B1	K700E (188), K666N (51), K666R (16), K666T (12), K666Q (7), K666E (2), K666M (2), H662Q (26), H662D (3), H662Y (1), E622D (20), E622V (2), R625L (9), R625C (8), R625G (2), N626D (1), G740E (2), G740R (2), V701F (2), I704N (2), Y623C (1), T663I (1), G742D (1), X833_splice (3), X742_splice (1)
SRSF2	P95H (185), P95L (99), P95R (67), P95T (9), P95A (6), P95_R102del (23), R94dup (13), P96L (12), P96Rfs*136 (1), S101Rfs*21 (11)
U2AF1	S34F (97), S34Y (16), S34C (1), Q157P (54), Q157R (32)
ZRSR2	R290* (6), R295* (5), R126* (3), W340* (3), K29Efs*26 (2), W75* (2), W291* (2), K16Ifs*40 (1), Y18* (1), R30Vfs*8 (1), E33Lfs*23 (1), R36* (1), E46* (1), E49* (1), I53Mfs*6 (1), L71* (1), X104_splice (1), R105Efs*3 (1), Q110* (1), E118Dfs*28 (1), E123Gfs*42 (1), E133* (1), W153* (1), Q154* (1), N155Tfs*10 (1), R169* (1), C181* (1), S188* (1), K190Nfs*48 (1), Q213* (1), Q235* (1), L237* (1), Y240* (1), D242Efs*13 (1), V253Sfs*36 (1), F256Lfs*32 (1), X258_splice (1), Q275* (1), X276_splice (1), S285Ffs*4 (1), F287Lfs*18 (1), E279del (1), C312Vfs*? (1), X313_splice (1), Q319* (1), N343Ifs*? (1), L348* (1), E365Gfs*? (1), Y373* (1), E394Nfs*? (1), R467Vfs*? (1)

Supplementary Table 2. All patient samples with double splicing factor mutations.

Study	Sample	Diagnosis	Category	Mutation 1	VAF 1	Mutation 2	VAF 2	Cytogenetics
Mayo	MAYO-2	CMML	MDS/MPN	ZRSR2 p.R290*	0.67	ZRSR2 p.N343Ifs*?	0.20	45,X,-Y[20]
Mayo	MAYO-4	CMML	MDS/MPN	ZRSR2 p.R105Efs*3	0.92	U2AF1 p.Q157P	0.47	46,XY[20]
Mayo	MAYO-5	CMML	MDS/MPN	SF3B1 p.K700E	0.23	SRSF2 p.P95H	N/A	46,XX[20]
Mayo	MAYO-7	CMML	MDS/MPN	SF3B1 p.K700E	0.32	SRSF2 p.P95L	N/A	46,XX[20]
Mayo	MAYO-8	CMML	MDS/MPN	SF3B1 p.Y623C	N/A	SRSF2 p.P95H	N/A	46,XY[20]
Mayo	MAYO-9	CMML	MDS/MPN	SRSF2 p.P95L	N/A	ZRSR2 p.E46*	N/A	46,XY[20]
Mayo	MAYO-10	ET	MPN	SF3B1 p.K666N	0.43	SRSF2 p.P95H	0.11	46,XY[20]
Mayo	MAYO-12	PMF	MPN	SRSF2 p.P95H	0.50	SF3B1 p.H662Q	0.48	46,XX[30]
Mayo	MAYO-13	PMF	MPN	U2AF1 p.Q157R	0.15	U2AF1 p.S34F	0.14	46,XX,del(20)(q11.2q13.3)[13]/46,XX[7]
Mayo	MAYO-16	PMF	MPN	U2AF1 p.Q157R	0.43	U2AF1 p.S34Y	0.42	45,X,-X[20]
Mayo	MAYO-18	PMF	MPN	U2AF1 p.Q157R	0.39	SF3B1 p.G740R	0.12	46,XY dup(1q),del(13q),t(7;12)[17]/46,XY[3]
Mayo	MAYO-20	PMF	MPN	ZRSR2 p.W153*	0.42	ZRSR2 p.Y18*	0.35	46,XY,del(12)(p11.2p13)[2]/46,XY,del(13)(q12q22)[1]/46,XY[17]
Mayo	MAYO-21	PMF	MPN	ZRSR2 p.W291*	0.40	ZRSR2 p.R290*	0.07	46,XY[20]
Mayo	MAYO-22	PMF	MPN	ZRSR2 p.F256Lfs*32	0.10	SRSF2 p.P95_R102del	0.05	46,XY[30]
Mayo	MAYO-23	PMF	MPN	ZRSR2 p.L348*	0.77	U2AF1 p.Q157R	0.47	46,XY[20]
Mayo	MAYO-25	PMF	MPN	SF3B1 p.T663I	0.27	U2AF1 p.Q157R	0.24	46,XY[20]
Mayo	MAYO-27	RARS	MDS	SF3B1 p.K700E	0.33	SRSF2 p.P95H	0.17	45,X,-Y [17]/46,XY[3]
Mayo	MAYO-28	RARS	MDS	SF3B1 p.H662D	0.31	SF3B1 p.K700E	0.08	46,XX [20]
Mayo	MAYO-29	RARS	MDS	SRSF2 p.P95L	0.38	SF3B1 p.K666Q	0.38	46,XY[20]
Mayo	MAYO-30	RARS	MDS	U2AF1 p.Q157R	0.41	SF3B1 p.K666R	0.38	46,XY[20]
Mayo	MAYO-32	RARS_T	MDS	SF3B1 p.G740E	0.50	SRSF2 p.P95H	0.45	46,XX[20]
Mayo	MAYO-33	AML	AML	SRSF2 p.P95H	0.45	U2AF1 p.S34Y	0.44	48,XY,+8,+21[20]
MSKCC	MSK-01	CMML-2	MDS/MPN	SRSF2 p.P95L	0.54	U2AF1 p.S34C	0.47	47,XY,+8[20]
MSKCC	MSK-02	MDS/MPN-U with >15% RS	MDS/MPN	SF3B1 p.K666T	0.45	SRSF2 p.P95H	0.50	46,XY[20]
MSKCC	MSK-03	CMML-1	MDS/MPN	SF3B1 p.K700E	0.27	SRSF2 p.P95H	0.06	46,XX[20]

MSKCC	MSK-04	MDS-MLD	MDS	ZRSR2 p.R290*	0.13	ZRSR2 p.E118Dfs*28	0.08	46,XY[20]
MSKCC	MSK-06	CMML-1	MDS/MPN	ZRSR2 p.E279del	0.24	SF3B1 p.H662Q	0.26	45,X,-Y[17]/46,XY[3]
MSKCC	MSK-07	AML, NOS	AML	U2AF1 p.S34F	0.22	SRSF2 p.P96Rfs*136	0.07	46,XY[20]
MSKCC	MSK-08	MDS-EB1	MDS	U2AF1 p.S34F	0.40	U2AF1 p.Q157R	0.26	46,XY[20]
MSKCC	MSK-09	Therapy-related myeloid neoplasm	AML	U2AF1 p.Q157R	0.62	SRSF2 p.P95H	0.29	44,XY,del(5)(q22q35),-7,add(13)(p11.2),del(16)(q?13q24),-18[9]/44,idem,dup(6)(p21.1p21.3),-17,+mar[5]/46,XY,del(12)(q15q24.1),del(17)(p11.2p13),del(17)(p11.2p13),add(17)(p11.2),del(20)(q11.2q13.3)[4]/46,XY[2]
MSKCC	P-0010290-T10-TB3	AML	AML	SRSF2 p.P95H	0.53	SF3B1 p.E622V	0.09	46,XY,add(2)(q37),inv(3)(q21q26),inc[15]/46,XY[1]
MSKCC	P-0010874-T05-TB3	MDS	MDS	SRSF2 p.P96L	0.03	SF3B1 p.E622V	0.03	46,XY[20]
MSKCC	P-0017145-T05-IH3	AML-MRC	AML	U2AF1 p.Q157R	0.44	U2AF1 p.S34F	0.44	46,XY[20]
MSKCC	P-0017792-T01-TB3	MDS	MDS	SF3B1 p.R625C	0.54	SF3B1 p.K700E	0.48	46,XX[20]
MSKCC	P-0020445-T01-TB3	MDS	MDS	SF3B1 p.K700E	0.54	SF3B1 p.K666N	0.04	46,XY[20]
MSKCC	P-0022634-T01-TB3	MDS	MDS	U2AF1 p.S34F	0.42	U2AF1 p.Q157R	0.05	N/A
MSKCC	P-0023073-T01-TB3	MDS	MDS	SRSF2 p.P95H	0.66	U2AF1 p.Q157P	0.04	46,XX,del(5)(q13q31)[2]/46,X,del(X)(q22q28),del(2)(q33q37)[1]/47,XX,+15[1]/46,XX[26]
MSKCC	P-0026703-T03-IH3	MDS	MDS	ZRSR2 p.K29Efs*26	0.87	SRSF2 p.P95L	0.45	46,XY[20]
MSKCC	P-0027895-T04-IH3	MDS-EB-1	MDS	SRSF2 p.P95H	0.43	SF3B1 p.H662Q	0.42	46,XY[20]
MSKCC	P-0028132-T04-IH3	MDS	MDS	ZRSR2 p.L71*	0.86	U2AF1 p.Q157P	0.26	N/A
MSKCC	P-0029855-T03-IH3	AML	AML	U2AF1 p.S34F	0.38	U2AF1 p.Q157R	0.03	46,XY[20]
MSKCC	P-0032923-T02-IH3	AML	AML	SRSF2 p.P95H	0.15	U2AF1 p.Q157R	0.14	47,XY,t(4;6)(q31;q27),+13[1]/90~92,idemx2,inc[2]/46,XY[11]
MSKCC	P-0035091-T03-IH3	MPN	MPN	SF3B1 p.K666N	0.49	SRSF2 p.P95H	0.17	N/A

MSKCC	P-0036187-T01-IH3	MDS-RS-SLD	MDS	SF3B1 p.E622D	0.30	SRSF2 p.P95H	0.03	N/A
MSKCC	E-H-112901-T1-1-D1-1	MPN	MPN	SRSF2 p.P95A	N/A	SF3B1 p.K666N	N/A	N/A
BeatAML	13-00160	AML-MRC	AML	ZRSR2 p.R467Vfs*?	0.83	SRSF2 p.P95H	0.50	N/A
BeatAML	14-00761	AML-MRC	AML	SRSF2 p.P95L	0.70	SF3B1 p.K666N	0.48	N/A
BeatAML	14-00774	AML with minimal differentiation	AML	ZRSR2 p.I53Mfs*6	0.93	SRSF2 p.P95H	0.44	N/A
BeatAML	16-00548	AML-MRC	AML	SF3B1 p.H662Q	0.33	ZRSR2 p.W75*	0.12	N/A
BeatAML	16-01223	AML, NOS	AML	U2AF1 p.Q157R	0.49	U2AF1 p.S34F	0.43	N/A
Papaemmanuil_2013	PD6072a	RAEB	MDS	SRSF2 p.P95R	N/A	ZRSR2 p.E394Nfs*?	N/A	N/A
Papaemmanuil_2013	PD6099a	RCMD	MDS	SRSF2 p.P95R	N/A	SRSF2 p.S101Rfs*21	N/A	N/A
Papaemmanuil_2013	PD6106a	RAEB	MDS	SRSF2 p.P95R	N/A	SRSF2 p.S101Rfs*21	N/A	N/A
Papaemmanuil_2013	PD6283a	CMML	MDS/MPN	SRSF2 p.P95T	N/A	U2AF1 p.Q157R	N/A	N/A
Papaemmanuil_2013	PD6504a	RARS	MDS	SF3B1 p.K666N	N/A	SRSF2 p.P95R	N/A	N/A
Papaemmanuil_2013	PD6830a	MDS	MDS	SF3B1 p.K666Q	N/A	SRSF2 p.P95T	N/A	N/A
Papaemmanuil_2013	PD7364a	RAEB	MDS	SRSF2 p.P95R	N/A	ZRSR2 p.V253Sfs*36	N/A	N/A
Papaemmanuil_2013	PD7385a	RAEB	MDS	U2AF1 p.Q157R	N/A	U2AF1 p.S34F	N/A	N/A
Papaemmanuil_2016	PD11136a	AML	AML	SF3B1 p.K666N	0.55	SRSF2 p.P95H	0.36	N/A
Papaemmanuil_2016	PD7722a	AML	AML	U2AF1 p.S34Y	0.40	SRSF2 p.P96L	0.32	N/A
Papaemmanuil_2016	PD8084a	AML	AML	U2AF1 p.S34F	0.45	U2AF1 p.Q157R	0.42	N/A
Papaemmanuil_2016	PD8258a	AML	AML	U2AF1 p.Q157R	0.41	SF3B1 p.K700E	0.06	N/A

Supplementary Table 3. Patient samples selected for single cell DNA sequencing.

Sample	Age	Sex	Diagnosis	Category	Mutation 1	Mutation 2	Additional Mutations
MSK-01	76	M	CMML-2	MDS/MPN	U2AF1 p.S34C	SRSF2 p.P95L	TET2 p.G1361D; ASXL1 p.L823*; ETV6 p.X155_splice
MSK-02	65	M	MDS/MPN-U with >15% RS	MDS/MPN	SF3B1 p.K666T	SRSF2 p.P95H	ASXL1 p.E635Rfs*15; SETBP1 p.G870S; CBL p.C396R
MSK-03	74	F	CMML-1	MDS/MPN	SF3B1 p.K700E	SRSF2 p.P95H	TET2 p.P1725Sfs*4
MSK-04	68	M	MDS-MLD	MDS	ZRSR2 p.R290*	ZRSR2 p.E118Dfs*28	TET2 p.R1516*
MSK-06	89	M	CMML-1	MDS/MPN	ZRSR2 p.E279del	SF3B1 p.H662Q	TET2 p.Q325*
MSK-07	54	M	AML, NOS	AML	U2AF1 p.S34F	SRSF2 p.P96Rfs*148	TET2 p.N275Ifs*18
MSK-08	71	M	MDS-EB1	MDS	U2AF1 p.S34F	U2AF1 p.Q157R	KDM6A p.G1077E; DNMT3A p.C351Afs*56; SETBP1 p.D868N
MSK-09	77	M	Therapy-related myeloid neoplasm	AML	U2AF1 p.Q157R	SRSF2 p.P95H	ASXL1 p.R417*; PTPN11 p.G503V; PTPN11 p.G503E; KIT p.D816V
MAYO-4	70	M	CMML	MDS/MPN	ZRSR2 p.R105Efs*3	U2AF1 p.Q157P	TET2 p.E1555Vfs*22
MAYO-12	72	F	PMF	MPN	SF3B1 p.H662Q	SRSF2 p.P95H	TET2 p.G355D TET2 p.L1721W
MAYO-18	73	M	PMF	MPN	SF3B1 p.G740R	U2AF1 p.Q157R	N/A

* scDNA-seq VAF: aggregated VAF of filtered cells across the sample based on read count

Supplementary Table 4. Custom sequencing panel for single cell DNA sequencing.

AmpID	Variants	Protein	Chr	Start*	End*	Sequencing Primers
1	DNMT3A:c.1051delT_1	C351Afs*56	chr2	25467416	25467632	Fwd: AGTTGTTGTTTCCGCACATGA Rev: GGTCATGTCTTCAGGGCTTAGG
2	DNMT3A:c.1051delT_2	C351Afs*56	chr2	25469960	25470217	Fwd: TTGTTGTACGTGGCCTGGTG Rev: CTGCACTCAGGTACTTCCCTA
3	SF3B1:c.2218G>A; 2098A>G	G740R; K700E	chr2	198266685	198266925	Fwd: TCAAAAGGTAATTGGTGGATTTACCTT Rev: GAGAATCTGGATGATATTGTGTAACTTAGG
4	SF3B1:c.1997A>C; 1986C>A; 1865A>T	K666T, H662Q; E622V	chr2	198267324	198267549	Fwd: GCACAGCCCATAAGAATAGCTAT Rev: CTGCTGGTCTGGCTACTATGAT
5	KIT:c.2447A>T_1; 2447A>T_2	D816V	chr4	55599271	55599486	Fwd: CCTTACTCATGGTCGGATCACAA Rev:TTTGACTGCTAAAATGTGTGATATCCC
6	TET2:c.822delC_1; 822delC_2	N275Ifs*18	chr4	106155803	106156016	Fwd: CAGATTGTGTTTCCATTGCGGT Rev: TTACTGGCATTATCAGCATCATCAG
7	TET2:c.973C>T	Q325*	chr4	106156035	106156262	Fwd: TACCTGTTCCTTTCAGAAACCAGAA Rev: GTGAACACTGAGCTTTGCTTGA
8	TET2:c.2671delC	Q891Nfs*30	chr4	106157747	106157972	Fwd: TTCAAGAACAGGAGCAGAAGTCA Rev: TGTAAGAGATGCCACCTTAGAGC
9	TET2:c.4082G>A	G1361D	chr4	106190626	106190861	Fwd: GTGTCATTCCATTTTGTTTCTGGAT Rev: TGAGCACAGAAGTCCAAACATG
10	TET2:c.4393C>T	R1465*	chr4	106193748	106193960	Fwd: CGAGAATTTGGAGGAAAACCTGAG Rev: AGCTTTCTTGGCTTCTAGTTTCC
11	TET2:c.4546C>T	R1516*	chr4	106196120	106196359	Fwd: ACATTTAAGTATCCTCACTAGCCTTC Rev: TCCAGAAGCAGAATAAGAGTTGAC
12	TET2:c.5172dupT	P1725Sfs*4	chr4	106196692	106196913	Fwd: CCATACACTTTACCAGCCAAGG Rev: ATGTTTGGATTGCTCAGATTGGG
13	JAK2:c.1849G>T_1	V617F	chr9	5073699	5073902	Fwd: CTTTCTTTGAAGCAGCAAGTATGATGA Rev: CTCCTGTTAAATTATAGTTTACACTGACAC
14	JAK2:c.1849G>T_2	V617F	chr9	5122856	5123080	Fwd: GCCCTTTTAATCATAGAAGCCTCAG Rev: CATACAGAACCACTCCAAAGCTC
15	CBL:c.1186T>C	C396R	chr11	119148903	119149119	Fwd: GGCTCCACATTCCAACTATGTAAA Rev: TTTAAACCTAACCCCAAAAGCCAG
16	ETV6:c.463+2T>C	X155_splice	chr12	12006467	12006694	Fwd: GGTCATACTGCATCAGAACCATG Rev: TGTACATCCATCTCACTGCTTCG
17	PTPN11:c.1508G>A; 1508G>T	G503E; G503V	chr12	112926841	112927080	Fwd: TATTGACGTTCCCAAAACCATCC Rev: ACAGTTGTCTATCAGAGCCTGT
18	TP53:c.989T>C_1	L330P	chr17	7572907	7573129	Fwd: GGAACAAGAAGTGGAGAATGTCAG Rev: TCTTGATTTGAATTCCCGTTGTCC
19	TP53:c.989T>C_2	L330P	chr17	7576793	7577021	Fwd: GGCATTTTGAGTGTTAGACTGGAA Rev: GAGGTAAGCAAGCAGGACAAG

20	SRSF2:c.284_287delinsT; 287delC; 284C>A; 284C>T	P95_P96delinsL; P96Rfs*148; P95H; P95L	chr17	74732876	74733145	Fwd: TTTACCTGCGGCTCCGGC Rev: TCTTCGAGAAGTACGGGCGC
21	SETBP1:c.2602G>A; 2608G>A	D868N; G870S	chr18	42531848	42532089	Fwd: TAAAGGAAATCACGCTGTCCCC Rev: CAGAAAGTTGTCCACAATGAGATGC
22	ASXL1:c.1249C>T	R417*	chr20	31021138	31021366	Fwd: GGCTGAAATCAAAAGTGGCTTGT Rev: CTTAGCCTCCCCATCCTTGTA
23	ASXL1:c.1900_1922delAGAGAGG CGGCCACCACTGCCAT; 1924G>T	E635Rfs*15; G642*	chr20	31022365	31022634	Fwd: TTAAAGCCCGTGCTCTGCA Rev: TATGCTCCCCATTTAGAGGATAAGG
24	ASXL1:c.2172delG	R725Efs*19	chr20	31022644	31022858	Fwd: GAACTGCCATGTCCAGAGCT Rev: CAACTGAGGCTGCTCCACTA
25	ASXL1:c.2415dupC_1; 2423delC_1; 2415dupC_2; 2423delC_2; 2468delT	T806Hfs*16; P808Lfs*10; L823*	chr20	31022872	31023111	Fwd: GAACTGAATGTGAGTCTGGCAC Rev: ATTCGTCATCAAATGCTCTGTTCTG
26	ASXL1:c.3637_3640delCTCC	L1213Ifs*3	chr20	31024105	31024320	Fwd: CTGGAGCACCCCAAAAGAATTG Rev: AGGTAGTAAGATCTCCTGGGCT
27	U2AF1:c.470A>C; 470A>G	Q157P; Q157R	chr21	44514753	44514997	Fwd: CTCCTCACTCACCCCATCTCATA Rev: ACCTGAGTGTGTATATCTCTCTCTG
28	U2AF1:c.101C>A; 101C>G; 101C>T	S34Y; S34C; S34F	chr21	44524417	44524634	Fwd: AAACAAACCTGGCTAAACGTCG Rev: AAAGTCTGTGTCATGTTTCTGTGAG
29	ZRSR2:c.83dupA	K29Efs*26	chrX	15809034	15809244	Fwd: GTGTATATGTCTTAATCTTCCAGCCAC Rev: TTCCTCCGACACAATATTTGACTTC
30	ZRSR2:c.313del_1; 353_354insCTTT_1	R105Efs*3; E118Dfs*28	chrX	15819440	15819677	Fwd: CTCCCAGGTTCACACGATTCT Rev: AGCTTAGAGACATTAAGAAATTTCGGC
31	ZRSR2:c.313del_2; 353_354insCTTT_2	R105Efs*3; E118Dfs*28	chrX	15822125	15822364	Fwd: TGATCTTATTCAGTTTGCCAGTGT Rev: ATTCTTCAAACTCATCACACAGCAC
32	ZRSR2:c.835_837delGAA; 868C>T	E279del; R290*	chrX	15838231	15838457	Fwd: GGTCCTACAATAAATTAATAACGGGGT Rev: CATCACTTTGTCTTTTACCACAAATCG
33	BCOR:c.1888G>A	E630K	chrX	39932523	39932739	Fwd: AAACAGACTGCCATTGGGTAAC Rev: CAACCCAGAACCGAGTTTCAA
34	KDM6A:c.3230G>A_1	G1077E	chrX	44938422	44938671	Fwd: GGAAGCTAACAATGAACATATGGTAGA Rev: AATGTGTCCTTTCAAAACTCCAAAG
35	KDM6A:c.3230G>A_2	G1077E	chrX	44941925	44942126	Fwd: AAGGCATGTTTCTAATACTGTGTCTC Rev: ATGAAACAGAAGAGGAAATTGTAGAGA
36	SMC1A:c.3152G>A_1; 3152G>A_2	R1051Q	chrX	53409459	53409679	Fwd: AGATCTCATCAATGTTGGTAGCCA Rev: CATTCCTCTCTCTCAGTGGGAT

* Reference genome: hg19

Supplementary Table 5. Single cell DNA sequencing metrics.

Sample	Cells	Reads	Reads per cell	Reads per amplicon per cell
MSK-01	4197	62000000	12204	339
MSK-02	3578	63000000	14147	393
MSK-03	6701	80000000	10090	280
MSK-04	8585	68000000	6764	188
MSK-06	1442	43000000	24101	669
MSK-07	5288	45000000	7056	196
MSK-08	9542	29000000	2526	70
MSK-09	10300	48000000	3933	109
MAYO-4	10168	55000000	4681	130
MAYO-12	4015	42000000	8326	231
MAYO-18	14690	86000000	4934	137

Supplementary Table 6. Characteristics of patients and samples used to evaluate effects of SF3B1 K700E versus K666 mutations on RNA splicing by RNA-sequencing†*.

Sample ID	SF3B1 Status	Additional genetic alterations	Cytogenetics	Disease	Source	Blast %	Age	Sex	WBC
SLMSK_001	WT	KIT D816V, CEBPA H219fs*99, NUP98-NSD1 fusion	inv(5q)	AML	PB	28	23	M	2.4
SLMSK_002	WT	EZH2 Y181*, RUNX1 P261fs, TET2 C1633X	Normal	MDS/MPN	PB	7	62	F	84
SLMSK_003	WT	RUNX1 R201*, BCORL1 R945*	Normal	AML	PB	0	63	M	6.3
SLMSK_004	WT			AML	BM	68			
SLMSK_005	WT	DNMT3A R326C, NRAS G13D	Normal	AML	BM	20	46	M	3
SLMSK_006	WT	CEBPA F6fs*12, CREBBP S1108*, PICALM-MLLT10 fusion	t(10;11)(p12;q21)	AML	PB	55	35	M	15.2
SLMSK_007	WT	No mutations on Thunderbolt panel	Normal	AML	PB	5	71	M	1.7
SLMSK_008	WT	KIT D816V	t(8;21)	AML	BM	72	35	F	9.3
SLMSK_011	p.K700E	None	Normal	AML	PB	20	75	F	6.3
SLMSK_012	p.K700E			AML	BM				
SLMSK_013	p.K700E			AML	BM				
SLMSK_015	p.K700E	IDH1 R132H, JAK2 V617F	Complex	AML with MF	PB	36	59	F	2.5
SLMSK_016	p.K700E	TP53 Y243C	Normal	MDS	PB	0	88	F	6.7
SLMSK_017	p.K700E			MDS	BM	1			
SLMSK_020	p.K700E	None	Normal	RARS	BM	0	65	M	3.8
SLMSK_022	p.K700E	KRAS G12A	Monosomy 7	MDS>AML	BM	46	78	F	6.8
SLMSK_023	p.K700E	TET2 C1464X, V430fs	Normal	MDS	BM	0	70	M	9.6
SLMSK_024	p.K700E			AML	PB	73			
SLMSK_026	p.K666N			AML	PB	20			
SLMSK_027	p.K666M	ASXL1 Q925X, RUNX1 A142fs	Normal	MDS	PB	2	83	F	1.9
SLMSK_028	p.K666N	None	Normal	MDS/MPN	BM	7	61	F	3.6
SLMSK_030	p.K666N	None	t(12;17)	AML	BM	60	63	M	4
SLMSK_032	p.K666N			AML	BM				
SLMSK_033	p.K666E			AML	BM				

†Blank cells indicate information not known.

*Abbreviations: AML: acute myeloid leukemia; BM: bone marrow; MDS: myelodysplastic syndrome; MF: myelofibrosis; MPN: myeloproliferative neoplasm; PB: peripheral blood; RARS: refractory anemia with ring sideroblasts.