Published in final edited form as:

Nature. 2019 October ; 574(7777): 273–277. doi:10.1038/s41586-019-1618-0.

Coordinated Alterations in RNA Splicing and Epigenetic Regulation Drive Leukemogenesis

Akihide Yoshimi[1,12], **Kuan-Ting Lin**[2,12], **Daniel H. Wiseman**[3,4,12], **Mohammad Alinoor Rahman**[2], **Alessandro Pastore**[1], **Bo Wang**[1], **Stanley Chun-Wei Lee**[1], **Jean-Baptiste Micol**[5], **Xiao Jing Zhang**[1], **Stephane de Botton**[5], **Virginie Penard-Lacronique**[5], **Eytan M. Stein**[6],

[*]**Corresponding Author:** Omar Abdel-Wahab, Zuckerman 701, Memorial Sloan Kettering Cancer Center, 408 E. 69th Street, New York, NY USA 10065, Phone: 646-888-3487, Fax: 646-422-0890, abdelwao@mskcc.org.

Author contributions

A.Y., K.T.L., A.R.K., and O.A.-W. designed the study. A.Y., S.C.-W.L., J.B.M., X.J.Z., H.C., R.E.M., D.I., T.R.A., K.B., F.S. and E.J.W. performed animal experiments. K.T.L. and M.A.R. performed RNA-seq analyses and minigene assays, respectively, under the supervision of A.R.K. A.P. performed DNA methylation and ChIP-seq analyses. T.R.A. and E.J.W. provided antibodies to detect Integrator components and assays for snRNA cleavage. H.D., R.K.B and F.A. performed RNA-seq analyses. D.H.W., T.C.P.S., D.P.W., S.D.B., V.P.-L., E.S., and R.L.L. provided clinical samples. D.H.W. and C.C. provided clinical correlative data for primary datasets. D.H.W. performed ChIP-seq experiments under the supervision of T.C.P.S. A.M.I. provided *Idh2*R140Q knock-in mice. R.L.L. provided *Tet2* knockout mice. A.Y., K.T.L., D.H.W., and O.A.-W. prepared the manuscript with help from all co-authors.

Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA
Akihide Yoshimi, Alessandro Pastore, Bo Wang, Stanley Chun-Wei Lee, Xiao Jing Zhang, Hana Cho, Rachel E. Miles, Daichi Inoue, Andrew M. Intlekofer, Ross L. Levine & Omar Abdel-Wahab
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
Kuan-Ting Lin, Mohammad Alinoor Rahman & Adrian R. Krainer
Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK
Daniel H. Wiseman, Tim C.P. Somervaille, Fabio Amaral & Fabrizio Simeoni
Division of Cancer Sciences, The University of Manchester, Manchester, UK
Daniel H. Wiseman & Kiran Batta
Gustave Roussy, Université Paris-Saclay, Villejuif, France
Jean-Baptiste Micol, Stephane de Botton & Virginie Penard-Lacronique
Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA
Eytan M. Stein, Ross L. Levine & Omar Abdel-Wahab
Department of Biochemistry & Molecular Biology, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA
Todd R. Albrecht & Eric J. Wagner
Manchester Cancer Research Centre Biobank, The University of Manchester, Manchester, UK
Deepti P. Wilks
Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds, UK
Catherine Cargo
Department of Biomolecular Chemistry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA
Heidi Dvinge
Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
Robert K. Bradley

Competing interests

A.M.I. has served as a consultant/advisory board member for Foundation Medicine. E.M.S. has served on advisory boards for Astellas Pharma, Daiichi Sankyo, Bayer, Novartis, Syros, Pfizer, PTC Therapeutics, AbbVie, Agios, and Celgene and has received research support from Agios, Celgene, Syros and Bayer. R.L.L. is on the Supervisory Board of Qiagen and the Scientific Advisory Board of Loxo, reports receiving commercial research grants from Celgene, Roche, and Prelude, has received honoraria from the speakers bureaus of Gilead and Lilly, has ownership interest (including stock, patents, etc.) in Qiagen and Loxo, and is a consultant/advisory board member for Novartis, Roche, Janssen, Celgene, and Incyte. A.R.K. is a founder, director, advisor, stockholder, and chair of the SAB of Stoke Therapeutics, and receives compensation from the company; A.R.K. is a paid consultant for Biogen; he is a member of the SABs of Skyhawk Therapeutics, Envisagenics BioAnalytics, and Autoimmunity Biologic Solutions, and has received compensation from these companies in the form of stock; A.R.K. is a research collaborator of Ionis Pharmaceuticals and has received royalty income from Ionis through his employer, Cold Spring Harbor Laboratory. O.A.-W. has served as a consultant for H3 Biomedicine, Foundation Medicine Inc., Merck, and Janssen; O.A.-W. has received personal speaking fees from Daiichi Sankyo. O.A.-W. has received prior research funding from H3 Biomedicine unrelated to the current manuscript. D.I., R.K.B. and O.A.-W. are inventors on a provisional patent application (patent number FHCC.P0044US.P) applied for by Fred Hutchinson Cancer Research Center on the role of reactivating BRD9 expression in cancer by modulating aberrant BRD9 splicing in *SF3B1* mutant cells.

Hana Cho[1], **Rachel E. Miles**[1], **Daichi Inoue**[1], **Todd R. Albrecht**[7], **Tim C.P. Somervaille**[3], **Kiran Batta**[4], **Fabio Amaral**[3], **Fabrizio Simeoni**[3], **Deepti P. Wilks**[8], **Catherine Cargo**[9], **Andrew M. Intlekofer**[1], **Ross L. Levine**[1,6], **Heidi Dvinge**[10], **Robert K. Bradley**[11], **Eric J. Wagner**[7], **Adrian R. Krainer**[2], **Omar Abdel-Wahab**[1,6,*]

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

[3]Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK

[4]Division of Cancer Sciences, The University of Manchester, Manchester, UK

[5]Gustave Roussy, Université Paris-Saclay, Villejuif, France

[6]Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

[7]Department of Biochemistry & Molecular Biology, The University of Texas Medical Branch at Galveston, Galveston, Texas, 77550, USA

[8]Manchester Cancer Research Centre Biobank, The University of Manchester, Manchester, UK.

[9]Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds LS29 7TF, UK

[10]Department of Biomolecular Chemistry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, 53706, USA

[11]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA

[12]Equal contribution, co-first authors

Transcription and pre-mRNA splicing are key steps in the control of gene expression and mutations in genes regulating each of these processes are common in leukemia[1,2]. Despite the frequent overlap of mutations affecting epigenetic regulation and splicing in leukemia, how these processes influence one another to promote leukemogenesis is not understood and functional evidence that mutations in RNA splicing factors initiate leukemia does not exist. Here through analyses of transcriptomes from 982 acute myeloid leukemia (AML) patients, we identified frequent overlap of mutations in *IDH2* and *SRSF2* which together promote leukemogenesis through coordinated effects on the epigenome and RNA splicing. While mutations in either *IDH2* or *SRSF2* imparted distinct splicing changes, co-expression of mutant IDH2 altered the splicing effects of mutant SRSF2 and resulted in more profound splicing changes than either mutation alone. Consistent with this, co-expression of mutant IDH2 and SRSF2 resulted in lethal myelodysplasia with proliferative features *in vivo* and enhanced self-renewal in a manner not observed with either mutation alone. *IDH2/SRSF2* double-mutant cells exhibited aberrant splicing and reduced expression of *INTS3*, a member of the Integrator complex[3], concordant with increased stalling of RNA polymerase II (RNAPII). Aberrant *INTS3* splicing contributed to leukemogenesis in concert with mutant

IDH2 and was dependent on mutant SRSF2 binding to *cis* elements in *INTS3* mRNA and increased DNA methylation of *INTS3*. These data identify a pathogenic cross talk between altered epigenetic state and splicing in a subset of leukemias, provide functional evidence that mutations in splicing factors drive myeloid malignancy development, and uncover spliceosomal changes as a novel mediator of *IDH2*-mutant leukemogenesis.

Mutations in RNA splicing factors are common in cancer and impart specific changes to splicing that are identifiable by mRNA sequencing (RNA-seq)[4–6]. Somatic mutations involving the Proline 95 residue of the spliceosome component SRSF2 are among the most recurrent in myeloid malignancies and alter SRSF2's binding to RNA in a sequence-specific manner[6,7]. We analyzed RNA-seq data from 179 AML patients from The Cancer Genome Atlas (TCGA)[1] to evaluate for spliceosomal alterations. Aberrant splicing events characteristic of *SRSF2* mutations, including *EZH2*[6,7] poison exon inclusion, were observed in 19 patients (P = 1.6e-12; Fisher's exact test; Fig. 1a, Extended Data Fig. 1a, b, and Supplementary Table 1). Although only one *SRSF2* mutant patient was reported in the TCGA AML publication[1], mutational analysis of RNA-seq data identified *SRSF2* hotspot mutations in each of these 19 patients (19/178 = 11%). Therefore, these data retrospectively identify *SRSF2* as amongst the most commonly mutated genes in the TCGA AML cohort.

Interestingly, 47% of *SRSF2* mutant patients had a co-existing *IDH2* mutation and conversely, 56% of *IDH2* mutant patients had a co-existing *SRSF2* mutation (P = 1.7e-06; Fisher's exact test; Fig. 1b, Extended Data Fig. 1c, d, and Supplementary Table 2). Similar results were seen in RNA-seq data from 498 and 263 AML patients from the Beat-AML[8] and Leucegene[9] studies, respectively (Fig. 1c, d, Extended Data Fig. 1e–j, and Supplementary Table 2). Across these datasets variant allele frequencies of *IDH2* and *SRSF2* mutations were high and significantly correlated (Extended Data Fig. 1k), suggesting their common placement as early events in AML.

Beyond these datasets, combined *IDH2* and *SRSF2* mutations were identified in 5.2 – 6.2% of 1,643 unselected consecutive AML patients in clinical practice (Supplementary Table 3). Although not statistically significant, *IDH2/SRSF2* double-mutant AML cases had the shortest overall survival across the four studied genotypes (Extended Data Fig. 2a). While *IDH2/SRSF2* double-mutant patients were mostly intermediate cytogenetic risk, their prognosis was comparable to those with adverse cytogenetic risk (Extended Data Fig. 2b). *IDH2/SRSF2* double-mutant AML patients were also significantly older than *IDH2* single-mutant or *IDH2/SRSF2* WT patients (Extended Data Fig. 2b; clinical and genetic features are summarized in Extended Data Fig. 2 and Supplementary Table 3).

Mutations in *IDH2* confer neomorphic enzymatic activity which results in the generation of 2-hydroxyglutarate (2HG)[10]. 2HG production, in turn, induces DNA hypermethylation via the competitive inhibition of αKG-dependent enzymes TET1–3. Unsupervised hierarchical clustering of DNA methylation data from the TCGA AML cohort revealed that *IDH2/SRSF2* double-mutant AML cases form a distinct cluster with higher DNA methylation than *IDH2* single-mutant AML (Extended Data Fig. 1l–o). Collectively, these data identify *IDH2/SRSF2* double-mutant leukemia as a recurrent genetically defined AML subset with a distinct epigenomic profile.

We next sought to understand the basis for co-enrichment of *IDH2* and *SRSF2* mutations. Although mutations in splicing factors are frequent in leukemias, to date there is no functional evidence that they can transform cells *in vivo*. Overexpression of *IDH2*^{R140Q} or *IDH2*^{R172K} mutants in bone marrow (BM) cells from *Vav*-cre *Srsf2*$^{P95H/+}$ or *Vav*-cre *Srsf2*$^{+/+}$ mice revealed a clear collaborative effect between mutant *IDH2* and *Srsf2* (Extended Data Fig. 3a). Four weeks post-transplantation, the peripheral blood (PB) of recipient mice transplanted with *IDH2*/*Srsf2* double-mutant cells had a substantially greater percentage of GFP^{+} cells than in an *Srsf2* WT background (Fig. 2a and Extended Data Fig. 3b, c). Moreover, these mice exhibited significant myeloid skewing, macrocytic anemia, and thrombocytopenia of greater magnitude than seen with mutant *IDH2* (Extended Data Fig. 3d–h). *IDH2*/*Srsf2* double-mutants showed no difference in plasma 2HG levels than *IDH2* single-mutants (Extended Data Fig. 3i, j). Serial replating of BM cells from leukemic mice revealed markedly enhanced clonogenicity of *IDH2*/*Srsf2* double-mutant cells compared with other genotypes, exhibiting a blastic morphology and immature immunophenotype (Extended Data Fig. 3k–m). Consistent with these *in vitro* results, mice transplanted with *IDH2*/*Srsf2* double-mutant cells developed a lethal myelodysplastic syndrome (MDS) characterized by pancytopenia, macrocytosis, myeloid dysplasia, expansion of immature BM progenitors, and splenomegaly (Fig. 2b and Extended Data Fig. 3n–w). At the same time, *IDH2/Srsf2* double-mutant cells were serially transplantable in sublethally irradiated recipients (Fig. 2c and Extended Data Fig. 3x), a feature not present in single-mutant controls. *IDH2* single-mutant controls, in contrast, developed leukocytosis, myeloid skewing without clear dysplasia, and less pronounced splenomegaly, while *Srsf2* single-mutant cells had impaired repopulation capacity. These results provide the first evidence that spliceosomal gene mutations can promote leukemogenesis *in vivo*.

We next sought to verify the effects of mutant *Idh2* and *Srsf2* using models in which both mutants were expressed from endogenous loci. *Mx1*-cre *Srsf2*$^{P95H/+}$ mice were crossed to *Idh2*$^{R140Q/+}$ mice to generate control, *Idh2*^{R140Q} single-mutant, *Srsf2*^{P95H} single-mutant, and *Idh2/Srsf2* double knock-in (DKI) mice (Extended Data Fig. 4a). As expected, 2HG levels in PB mononuclear cells were increased and 5-hydroxymethylcytosine levels in cKit^{+} BM cells were decreased from *Idh2* single-mutant and DKI primary mice compared to controls (Extended Data Fig. 4b, c). We next performed non-competitive transplantation, wherein each mutation was induced, alone or together following stable engraftment in recipients. DKI mice showed stable engraftment overtime, similar to *Idh2* single-mutant or control mice (Extended Data Fig. 4d). However, DKI mice developed a lethal MDS with proliferative features and significantly shorter survival compared to controls (Fig. 2d). In competitive transplantation, expression of mutant *Idh2*^{R140Q} rescued the impaired self-renewal capacity of *Srsf2* single-mutant cells (Fig. 2e). These observations were supported by increased hematopoietic stem/progenitor cells in DKI mice compared to *Srsf2* single-mutant or control mice in primary and serial transplantation (Extended Data Fig. 4e–i). These results confirm cooperativity between mutant IDH2 and SRSF2 in promoting leukemogenesis *in vivo*.

Given prior data identifying 2HG-mediated inhibition of TET2 as a mechanism of *IDH2* mutant leukemogenesis[11], we also evaluated if loss of *TET2* might promote transformation of *SRSF2* mutant cells. However, deletion of *Tet2* in an *Srsf2* mutant background was

insufficient to rescue the impaired self-renewal capacity of *Srsf2* single-mutant cells (Extended Data Fig. 4j–n). Similarly, restoration of TET2 function did not affect the self-renewal capacity of *Idh2/Srsf2* double-mutant cells *in vivo* (Extended Data Fig. 4o–r). These data indicated that the collaborative effects of mutant Idh2 and Srsf2 are not solely dependent on TET2. Consistent with this, combined *Tet2/Tet3* silencing partially rescued the impaired replating capacity of *Srsf2* mutant cells *in vitro* (Extended Data Fig. 4r, s) and the impaired self-renewal of *Srsf2* mutant cells *in vivo* (Extended Data Fig. 4t–v). However, since FTO and ALKBH5, which play a role in RNA processing as *N*6-methyladenosine (m^6A) RNA demethylases[12,13], are also αKG-dependent, we investigated the effects of their loss on cooperativity with mutant *Srsf2*. However, collaborative effects were not observed between loss of *Fto* or *Alkbh5* and $Srsf2^{P95H}$ (Extended Data Fig. 4w, x).

To understand the basis for cooperation between *IDH2* and *SRSF2* mutations, we next analyzed RNA-seq from the TCGA (n = 179 patients), Beat-AML (n = 498 patients), and Leucegene (n = 263 patients) cohorts in addition to two previously unpublished RNA-seq datasets targeting defined *IDH2/SRSF2* genotype combinations (n = 42 patients) and the knock-in mouse models. This revealed that *IDH2/SRSF2* double-mutant cells consistently harbor more aberrant splicing events than *SRSF2* single-mutant cells. Moreover, *IDH2* mutations alone were associated with a small but reproducible change in RNA splicing (Fig. 3a, b, Extended Data Fig. 5a–g, and Supplementary Table 4–20). In contrast, *TET2/SRSF2* co-mutant AML had fewer changes in splicing than *IDH2/SRSF2* co-mutant AML (Extended Data Fig. 5h–m and Supplementary Table 21, 22).

The majority of splicing changes associated with *SRSF2* mutations involved altered cassette exon splicing consistent with *SRSF2* mutations promoting inclusion of C-rich RNA sequences[6,7]. The sequence specificity of mutant SRSF2 on splicing was not influenced by concomitant *IDH2* mutations (Extended Data Fig. 5n–q) and a number of these events were validated by RT-PCR of primary AML samples from an independent cohort (Fig. 3c). Among the mis-splicing events in *IDH2/SRSF2* double-mutant AML was a complex event in *INTS3* involving intron retention (IR) across two contiguous introns and skipping of the intervening exon (Fig. 3b, c, Extended Data Fig. 5e–f, 5r–y, 6a–c). Aberrant *INTS3* splicing was demonstrated in isogenic and non-isogenic leukemia cells with or without *IDH2* and/or *SRSF2* mutations (Fig. 3d and Extended Data Fig. 6d–f), and *INTS3* transcripts with both IR and exon skipping resulted in nonsense-mediated decay (Extended Data Fig. 6g–j). Consistent with these observations, INTS3 protein expression was reduced in *SRSF2* mutant cells (Fig. 3d, Extended Data Fig. 6e, f, k–n, and Supplementary Table 23). Moreover, silencing of INTS3 was associated with reduced protein levels of additional Integrator subunits in *SRSF2* mutant AML compared to *SRSF2* WT AML. Consistent with these observations, steady-state protein expression levels of Integrator subunits were correlated with one another (Extended Data Fig. 6o). Overall, these data indicate that aberrant splicing and consequent loss of INTS3 was a consistent feature of *IDH2/SRSF2* double-mutant cells and associated with reduced expression of multiple Integrator subunits.

We next sought to understand how *IDH2* mutations, which impact the epigenome, might influence splicing catalysis. Splice site choice is influenced by *cis* regulatory elements engaged by RNA binding proteins as well as RNAPII elongation, which itself is regulated by

DNA cytosine methylation and histone modifications[14]. We therefore generated a controlled system to dissect the contribution of RNA binding elements and DNA methylation to *INTS3* IR. We constructed a minigene of *INTS3* spanning exons 4 and 5 and the intervening intron 4 (Extended Data Fig. 7a–c). Transfection of this minigene into leukemia cells harboring combinations of *IDH2/SRSF2* mutations revealed that *INTS3* intron 4 retention is driven by mutant SRSF2 and further enhanced in the *IDH2/SRSF2* double-mutant setting (Extended Data Fig. 7d). SRSF2 normally binds C- or G-rich motif sequences in RNA equally well to promote splicing[15]. Leukemia-associated mutations in *SRSF2* promote its avidity for C-rich sequences while reducing the ability to recognize G-rich sequences[6,7]. Interestingly, exon 4 of *INTS3* harbors the greatest number of predicted SRSF2 binding motifs over the entire *INTS3* genomic region (Extended Data Fig. 7c). We evaluated the role of putative SRSF2 motifs in regulating *INTS3* splicing by mutating all six CCNG motifs in exon 4 to G-rich sequences. In this G-rich version of the minigene, IR no longer occurred (*INTS3*-GGNG; Extended Data Fig. 7e). Conversely, when all G-rich SRSF2 motifs were converted to C-rich sequences (*INTS3*-CCNG), IR became evident (Extended Data Fig. 7f). These results confirmed the sequence-specific activity of mutant SRSF2 in *INTS3* IR and identified a role for mutant IDH2 in regulating splicing.

Given that *IDH2* mutations promote increased DNA methylation and that DNA methylation can impact splicing[14], we generated genome-wide maps of DNA cytosine methylation from AML patients across four genotypes (Supplementary Table 23). This revealed that differentially spliced events in *IDH2* single-mutant as well as *IDH2/SRSF2* double-mutant AML (compared to *IDH2/SRSF2* WT and *SRSF2* single-mutant AML) harbored significant hypermethylation of DNA. Thus regions of differential DNA hypermethylation significantly overlapped with regions of differential RNA splicing (Fig. 3e and Extended Data Fig. 7j).

The above results suggest a strong link between increased DNA methylation mediated by mutant IDH2 and altered RNA splicing by mutant SRSF2. To evaluate this further, we next examined DNA methylation levels around endogenous *INTS3* exon 4–6 by targeted bisulfite sequencing. This revealed increased DNA methylation at all CpG dinucleotides in this region in *IDH2/SRSF2* double-mutant cells compared to control or single-mutant cells (Fig. 3f and Extended Data Fig. 7k). A functional role of DNA methylation at these sites was verified by evaluating splicing in versions of the *INTS3* minigene in which each CG dinucleotide was converted to an AT to prevent cytosine methylation. In these CG to AT versions of the minigene, *IDH2* mutations no longer promoted mutant *SRSF2*-mediated IR (Extended Data Fig. 7g–i). As further confirmation of the influence of mutant *IDH2* on *INTS3* splicing, cell-permeable 2HG increased *INTS3* IR while treatment of *IDH2/SRSF2* double-mutant cells with the DNA methyltransferase inhibitor 5-aza-2’-deoxycytidine (5-AZA-CdR) inhibited *INTS3* IR (Extended Data Fig. 7l, m).

Given that changes in epigenetic state may impact splicing by influencing RNAPII stalling[14,16], we evaluated the abundance of RNAPII through ChIP-seq in isogenic *SRSF2*^{WT} and *SRSF2*^{P95H} cells as well as the primary AML patient samples. This revealed increased promoter-proximal transcriptional pausing and decreased RNAPII occupancy over gene bodies in *SRSF2* mutant cells, which was further enhanced in *IDH2/SRSF2* double-mutant cells (Fig. 4a, b, Extended Data Fig. 7n–q, and Supplementary Table 23).

Transcriptional pausing was also evident at *INTS5* and *INTS14* in *SRSF2* mutant cells (Extended Data Fig. 7r, s), which, in combination with aberrant splicing of several Integrator subunits (Supplementary Table 24), suggested impaired function of the entire Integrator complex in *SRSF2* mutant cells. Similar to DNA cytosine methylation levels, RNAPII was more abundant over differentially spliced regions between *SRSF2* single-mutant AML and $SRSF2^{WT}$ AML, and further enhanced over differentially spliced regions between *SRSF2* single-mutant and *IDH2/SRSF2* double-mutant AML (Fig. 4c and Extended Data Fig. 7t).

The above data provide further links between increased DNA cytosine methylation and RNAPII stalling with altered RNA splicing in *IDH2/SRSF2* double-mutant AML. To further evaluate this model, we performed anti-RNAPII ChIP across 4,766 bp of *INTS3* locus in isogenic leukemia cells (Fig. 3f). This revealed striking accumulation of RNAPII across this locus in *IDH2/SRSF2* double-mutant cells. Treatment with 5-AZA-CdR significantly reduced RNAPII stalling, which was coupled with decreased aberrant *INTS3* splicing (Extended Data Fig. 7k–m). These data reveal that *IDH2* and *SRSF2* mutations coordinately dysregulate splicing through alterations in RNAPII stalling in addition to aberrant sequence recognition of *cis* elements in RNA.

INTS3 encodes a component of the Integrator complex that participates in small nuclear RNA (snRNA) processing[3] in addition to RNAPII pause-release[17]. Consistent with this, *SRSF2* single-mutant cells had altered snRNA cleavage similar to those seen with direct INTS3 downregulation, which was exacerbated in *IDH2*/*SRSF2* double-mutant cells (Extended Data Fig. 8a–h). Attenuation of INTS3 expression in *SRSF2* mutant cells caused a blockade of myeloid differentiation, an effect further enhanced in an *IDH2* mutant background (Extended Data Fig. 8i–n). Importantly, direct Ints3 downregulation in the $Idh2^{R140Q/+}$ background resulted in enhanced clonogenic capacity of cells with an immature morphology and immunophenotype (Fig. 4d and Extended Data Fig. 8o–r) and promoted clonal dominance of Idh2 mutant cells (Extended Data Fig. 9a–d). Moreover, mice transplanted with $Idh2^{R140Q/+}$/anti-Ints3 shRNA treated BM cells exhibited myeloid skewing, anemia, and thrombocytopenia (Extended Data Fig. 9e–g), and developed a lethal MDS with proliferative features, phenotypes resembling those seen in *IDH2*/*Srsf2* double-mutant mice (Fig. 4e and Extended Data Fig. 9g, h).

The defects in snRNA processing in *SRSF2* single-mutant and *IDH2/SRSF2* double-mutant cells were partially rescued by INTS3 cDNA expression (Extended Data Fig. 8s–x). In addition, restoration of INTS3 expression released *SRSF2* single-mutant and *IDH2/SRSF2* double-mutant HL-60 cells from differentiation block (Extended Data Fig. 8y, z). Xenografts of *IDH2/SRSF2* double-mutant HL-60 cells demonstrated that forced expression of INTS3 induced myeloid differentiation and slowed leukemia progression *in vivo* (Extended Data Fig. 9j–s). Collectively, these data suggest that INTS3 loss due to aberrant splicing by mutant IDH2 and SRSF2 contributes to leukemogenesis.

Although INTS3 loss resulted in measurable changes in snRNA processing, the degree of snRNA mis-processing did not have a significant impact on splicing as determined by RNA-seq of $IDH2^{R140Q}$ mutant HL-60 cells with INTS3 silencing. In contrast, INTS3 depletion in these cells significantly affected transcriptional programs associated with myeloid

differentiation, multiple oncogenic signaling pathways, RNAPII elongation-linked transcription, and DNA repair (Extended Data Fig. 10a–d and Supplementary Table 25). This latter association of INTS3 loss with DNA repair is potentially consistent with previous reports[18,19].

These data uncover an important role for RNA splicing alterations in *IDH2* mutant tumorigenesis and identify perturbations in Integrator as a novel driver of transformation of *IDH2* and *SRSF2* mutant cells. However, INTS3 is not known to be recurrently affected by coding-region alterations in leukemias. We therefore evaluated *INTS3* splicing across 32 additional cancer types as well as normal blood cells to evaluate if aberrant *INTS3* splicing might be a common mechanism in AML. This revealed that while *INTS3* mis-splicing is most evident in *IDH2/SRSF2* mutant AML, *INTS3* aberrant splicing is also prevalent across other molecular subtypes of AML but not present in blood cells from healthy subjects or RNA-seq data from > 7,000 samples from other cancer types (Fig. 4f and Extended Data Fig. 10e, f). To further evaluate the effects of enforced INTS3 expression in splicing WT myeloid leukemia, we utilized *MLL-AF9/Nras*^{G12D} murine leukemia (RN2) cells. INTS3 overexpression reduced colony-forming capacity of RN2 cells (Extended Data Fig. 10g, h) and enhanced differentiation of RN2 cells, resulting in decelerated leukemia progression *in vivo* (Fig. 4g and Extended Data Fig. 10i–s).

These data highlight a role for INTS3 loss in broad genetic subtypes of AML. Further efforts to determine how Integrator loss promotes leukemogenesis, and other non-mutational mechanisms mediating *INTS3* aberrant splicing, will be critical. To this end, it is important to note that prior work has identified that both Integrator[17,20] as well as SRSF2[21] play a direct role in modulating transcriptional pause-release. The striking accumulation of RNAPII at certain mis-spliced loci here are consistent with recent data suggesting that mutant SRSF2 is defective in promoting RNAPII pause-release[22]. Identifying how aberrant splicing mediated by mutant SRSF2 is influenced by altered RNAPII pause release may therefore be enlightening.

In addition to modifying splicing in *SRSF2* mutant cells, *IDH2* mutations themselves were associated with reproducible changes in splicing in hematopoietic cells. Intriguingly, there is a strong correlation between aberrant splicing in *IDH2* and *IDH1* mutant low-grade gliomas (LGG) (P = 2.2e-16 (binominal proportion test), Extended Data Fig. 10t–w, and Supplementary Table 26–28). A significant number of splicing events dysregulated in *IDH2* mutant AML from the TCGA and Leucegene cohorts were differentially spliced in *IDH2* mutant versus *IDH1/2* WT LGG (P = 1.8e-09 and P = 1.3e-08, respectively; binominal proportion test). These data suggest that *IDH1/2* mutations impart a consistent effect on splicing regardless of tumor type. Finally, these results have important translational implications given the substantial efforts to pharmacologically inhibit mutant IDH1/2 as well as mutant splicing factors[23,24]. The frequent co-existence of *IDH2* and *SRSF2* mutations underscores the enormous therapeutic potential for modulation of splicing in the ~50% of *IDH2* mutant leukemia patients who also harbor a spliceosomal gene mutation.

METHODS

Data reporting

The number of mice in each experiment was chosen to provide 90% statistical power with a 5% error level. Otherwise, no statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Animals

All animals were housed at Memorial Sloan Kettering Cancer Center (MSK). All animal procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSK. 6–8 week female CD45.1 C57BL/6 mice were purchased from The Jackson Laboratory (Stock No: 002014). Male and female CD45.2 *Srsf2*$^{P95H/+}$ conditional knock-in mice, *Idh2*$^{R140Q/+}$ conditional knock-in mice, and *Tet2* conditional knockout mice (all on C57BL/6 background) were also analyzed and used as bone marrow donors (generation of these mice were as described[6,26,27]). For BM transplantation assays with IDH2 overexpression, *Srsf2*$^{P95H/+}$ and littermate control mice were crossed to *Vav*-cre transgenic mice[28]. CBC analysis was performed on PB collected from submandibular bleeding, using a Procyte Dx Hematology Analyzer (IDEXX Veterinary Diagnostics). For all mouse experiments, the mice were monitored closely for signs of disease or morbidity daily and were sacrificed for visible tumor formation at tumor volume > 1 cm^3, failure to thrive, weight loss > 10% total body weight, open skin lesions, bleeding, or any signs of infection. In none of the experiments were these limits exceeded.

Bone marrow (BM) transplantation assays

Freshly dissected femurs and tibias were isolated from *Mx1*-cre, *Mx1*-cre/*Idh2*$^{R140Q/+}$, *Mx1*-cre *Srsf2*$^{P95H/+}$, *Mx1*-cre *Idh2*$^{R140Q/+}$*Srsf2*$^{P95H/+}$, *Mx1*-cre *Tet2*$^{fl/fl}$, or *Mx1*-cre *Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ CD45.2$^+$ mice. BM was flushed with a 3-cc insulin syringe into cold PBS supplemented with 2% bovine serum albumin to generate single-cell suspensions. BM cells were pelleted by centrifugation at 1,500 rpm for 4 min and red blood cells (RBCs) were lysed in ammonium chloride-potassium bicarbonate lysis (ACK) buffer for 3 min on ice. After centrifugation, cells were resuspended in PBS/2% BSA, passed through a 40μm cell strainer, and counted. For competitive transplantation experiments, 0.5×10^6 BM cells from *Mx1*-cre, *Mx1*-cre *Idh2*$^{R140Q/+}$, *Mx1*-cre *Srsf2*$^{P95H/+}$, *Mx1*-cre *Idh2*$^{R140Q/+}$*Srsf2*$^{P95H/+}$, *Mx1*-cre *Tet2*$^{fl/fl}$, or *Mx1*-cre *Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ CD45.2$^+$ mice were mixed with 0.5×10^6 wild-type (WT) CD45.1$^+$ BM and transplanted via tail-vein injection into 8-week old lethally irradiated (900 cGy) CD45.1$^+$ recipient mice. The CD45.1$^+$:CD45.2$^+$ ratio was confirmed to be approximately 1:1 by flow cytometry analysis pre-transplant. To activate the conditional alleles, mice were treated with 3 doses of polyinosinic:polycytidylic acid (pIpC; 12mg/kg/day; GE Healthcare) every second day via intra-peritoneal injection. Peripheral blood chimerism was assessed every 4 weeks by flow cytometry. For noncompetitive transplantation experiments, 1×10^6 total BM cells from *Mx1*-cre, *Mx1*-cre *Idh2*$^{R140Q/+}$, *Mx1*-cre *Srsf2*$^{P95H/+}$, *Mx1*-cre *Idh2*$^{R140Q/+}$*Srsf2*$^{P95H/+}$, *Mx1*-cre *Tet2*$^{fl/fl}$, or *Mx1*-cre *Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ CD45.2$^+$ mice were injected into lethally irradiated (950 cGy) CD45.1$^+$

recipient mice. Peripheral blood chimerism was assessed as described for competitive transplantation experiments. Additionally, for each bleeding whole blood cell counts were measured on an automated blood analyzer. Animals that were lost due to pIpC toxicity were excluded from analysis.

Retroviral transduction and transplantation of primary hematopoietic cells

Vav-cre *Srsf2*$^{+/+}$ and *Vav*-cre *Srsf2*$^{P95H/+}$ mice were treated with a single dose of 5-fluoruracil (150 mg/kg) followed by BM harvest from the femurs, tibias and pelvic bones 5 days later. RBCs were removed by ACK lysis buffer, and nucleated BM cells were transduced with viral supernatants containing MSCV-*IDH2*$^{WT/R140Q/R172K}$-IRES-GFP for 2 days in RPMI/20% FCS supplemented with mouse stem cell factor (mSCF, 25 ng/mL), mouse Interleukin-3 (mIL3, 10 ng/mL) and mIL6 (10 ng/mL), followed by injection of ~0.5 × 10^6 cells per recipient mouse via tail vein injection into lethally irradiated (950 cGy) CD45.1$^+$ mice. Transplantation of primary BM cells with TET2 catalytic domain cDNA and anti-Ints3 or Tet3 shRNAs was similarly performed. For secondary transplantation experiments, 8-week old, lethally (900–950 cGy) or sub-lethally (450–700 cGy) irradiated C57/BL6 recipient mice were injected with 1 × 10^6 MDS with proliferative feature cells. *IDH2*^{WT}+*Srsf2*^{WT} and *IDH2*^{WT}+*Srsf2*^{P95H} mice were sacrificed at day 315 post-transplant to harvest BM for the serial transplantation. All cytokines were purchased from R&D Systems.

Flow cytometry analyses and antibodies

Surface-marker staining of hematopoietic cells was performed by first lysing cells with ACK lysis buffer and washing cells with ice-cold PBS. Cells were stained with antibodies in PBS/2% BSA for 30 minutes on ice. For hematopoietic stem/progenitor staining, cells were stained with the following antibodies: B220-APCCy7 (clone: RA3–6B2; purchased from BioLegend; catalog #: 103224; dilution: 1:200); B220-Bv711 (RA3–6B2; BioLegend; 103255; 1:200); CD3-PerCPCy5.5 (17A2; BioLegend; 100208; 1:200); CD3-APC (17A2; BioLegend; 100236; 1:200); CD3-APCCy7 (17A2; BioLegend; 100222; 1:200); Gr1-PECy7 (RB6–8C5; eBioscience; 25-5931-82; 1:500); CD11b-PE (M1/70; eBioscience; 12-0112-85; 1:500); CD11b-APCCy7 (M1/70; BioLegend; 101226; 1:200); CD11c-APCCy7 (N418; BioLegend; 117323; 1:200); NK1.1-APCCy7 (PK136; BioLegend; 108724; 1:200); Ter119-APCCy7 (BioLegend; 116223: 1:200); cKit-APC (2B8; BioLegend; 105812; 1:200); cKit-PerCPCy5.5 (2B8; BioLegend; 105824; 1:100); cKit-Bv605 (ACK2; BioLegend; 135120; 1:200); Sca1-PECy7 (D7; BioLegend; 108102; 1:200); CD16/CD32 (FcγRII/III)-Alexa700 (93; eBioscience; 56-0161-82; 1:200); CD34-FITC (RAM34; BD Biosciences; 553731; 1:200); CD45.1-FITC (A20; BioLegend; 110706; 1:200); CD45.1-PerCPCy5.5 (A20; BioLegend; 110728; 1:200); CD45.1-PE (A20; BioLegend; 110708; 1:200); CD45.1-APC (A20; BioLegend; 110714; 1:200); CD45.2-PE (104; eBioscience; 12-0454-82; 1:200); CD45.2-Alexa700 (104; BioLegend; 109822; 1:200); CD45.2-Bv605 (104; BioLegend; 109841; 1:200); CD48-Bv711 (HM48–1; BioLegend; 103439; 1:200); CD150 (9D1; eBioscience; 12-1501-82; 1:200). DAPI was used to exclude dead cells. For sorting human leukemia cells, cells were stained with a lineage cocktail including CD34-PerCP (8G12; BD Biosciences; 345803; 1:200); CD117-PECy7 (104D2; eBioscience; 25-1178-42; 1:200); CD33-APC (P67.6; BioLegend; 366606; 1:200); HLA-DR-FITC (L243; BioLegend;

307604; 1:200); CD13-PE (L138; BD Biosciences; 347406; 1:200); CD45-APC-H7 (2D1; BD Biosciences; 560178; 1:200). The composition of mature hematopoietic cell lineages in the BM, spleen and peripheral blood was assessed using a combination of CD11b, Gr1, B220, and CD3. For the hematopoietic stem and progenitor analysis, a combination of CD11b, CD11c, Gr1, B220, CD3, NK1.1, and Ter119 was stained as lineage-positive cells. All the FACS sorting was performed on FACS Aria, and analysis was performed on an LSRII or LSR Fortessa (BD Biosciences). For western blotting, DNA dot blot assays, and chromatin immunoprecipitation (ChIP) assays, the following antibodies were used: INTS1 (purchased from Bethyl laboratories; catalog #: A300–361A; dilution: 1:1,000), INTS2 (Abcam; ab74982; 1:1,000), INTS3 (Bethyl laboratories; A300–427A; 1:1,000, Abcam; ab70451; 1:1,000), INTS4 (Bethyl laboratories; A301–296A; 1:1,000), INTS5 (Abcam; ab74405; 1:1,000), INTS6 (Abcam; ab57069; 1:1,000), INTS7 (Bethyl laboratories; A300–271A; 1:1,000), INTS8 (Bethyl laboratories; A300–269A; 1:1,000), INTS9 (Bethyl laboratories; A300–412A; 1:1,000), INTS11 (Abcam; ab84719; 1:1,000), Flag-M2 (Sigma-Aldrich; F-1084; 1:1,000), Myc-tag (Cell Signaling; 2276S; 1:1,000), β-actin (Sigma-Aldrich; A-5441; 1:2,000), 5-Hydroxymehylcytosine (5hmC) (Active motif; 39769), RNA polymerase II CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095), RNA polymerase II CTD repeat YSPTSPS (phospho S5) (Abcam; ab5408), and UPF1 (Abcam; ab109363; 1:1,000).

Minigene assay

We constructed *INTS3*-WT minigene spanning exons 4 to 5 of human *INTS3* into pcDNA3.1(+) vector (Invitrogen) using BamHI and XhoI sites, respectively. Artificial mutations were engineered into *INTS3*-WT minigene using the QuikChange Site-Directed Mutagenesis Kit (Agilent) to generate *INTS3*-GGNG, *INTS3*-CCNG, *INTS3*-WT_CG(−) *INTS3*-GGNG_CG(−), and *INTS3*-CCNG_CG(−) minigenes, respectively, and the sequences of inserts were verified by Sanger sequencing. Plasmids (1 μg) were transfected using Lipofectamine™ LTX reagent with PLUS™ reagent (Invitrogen) including 0.2 μg of EGFP and 0.8 μg of *INTS3* minigene, per well of a 6-well plate. Total RNA was extracted 48 hrs after transfection using TRIzol® reagent (Ambion), followed by DNase I treatment (Qiagen). cDNA was synthesized with an oligo-dT primer using ImProm-II™ reverse transcriptase (Promega). Radioactive PCR was done with ^{32}P-α-dCTP, 1.25 units of AmpliTaq® (Invitrogen) and 26 cycles using primer pairs 5’-GCTTGGTACCGAGCTCGGATC-3’ (vector specific forward primer) and 5’-CAGTTCCCGTACCAACCACAC-3’ (reverse primer for *INTS3* versions of minigene), or 5’-CAGTTCCATTACCAACCACAC-3’ (reverse primer for *INTS3*_CG(−) versions of minigene). Products were run on a 5% PAGE and the bands were quantified using a Typhoon FLA 7000 (GE Healthcare). EGFP was used as a control for transfection efficiency and exogenous EGFP was amplified using a vector specific forward primer and reverse primer on EGFP. EGFP products were loaded after we ran the INTS3 products for 20–30 min. Percentages of intron 4 retention were normalized against exogenous EGFP.

Cell culture

K562 (human chronic myeloid/erythroleukemia cell line) and HL-60 (human promyelocytic leukemia cell line) leukemia cells, K052 (human multilineage leukemia cell line) leukemia

cells, TF1 (human erythroleukemia cell line) leukemia cells, *MLL-AF9/Nras*^{G12D} murine leukemia (RN2) cells[29], and Ba/F3 (murine pro-B cell line) cells were cultured in RPMI/10% FCS (Fetal Calf Serum, heat inactivated), RPMI/20% FCS, RPMI/10% FCS + human Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF, R&D Systems; 5 ng/mL), and RPMI/10% FCS + mIL3 (R&D Systems; 1 ng/mL), respectively. None of the cell lines above were listed in the data base of commonly misidentified cell lines maintained by ICLAC and NCBI Biosample.

MSCV-*IDH2*$^{WT/R140Q/R172K}$-IRES-GFP, MSCV-3xFlag-*INTS3*-puro, MSCV-IRES-3xFlag-*INTS3*-mCherry, MSCV-IRES-TET2 catalytic domain cDNA-mCherry ("TET2CD"), and empty vectors of these constructs were used for retroviral overexpression studies and pRRLSIN.cPPT.PGK-mCherry.WPRE-*SRSF2*$^{WT/P95H}$ constructs were used for lentiviral overexpression studies. TET2CD cDNA fragment with Myc tag was generated by PCR amplification using pCMVT$_N$T-TET2CD[30] as a template and inserted in the BglII restriction sites of MSCV-IRES-mCherry. Retroviral supernatants were produced by transfecting 293 GPII cells with cDNA constructs and the packaging plasmid VSV.G using XtremeGene9 (Roche) or Polyethylenimine Hydrochloride (Polysciences, Inc.). Lentiviral supernatants were produced by similarly transfecting HEK293T cells with cDNA constructs and the packaging plasmid VSV.G and psPAX2. Virus supernatants were used for transduction in the presence of polybrene (5 μg/mL). GFP$^+$mCherry$^+$ double-positive HL-60 cells and mCherry$^+$ positive K562 cells were FACS-sorted to obtain cells expressing WT/mutant IDH2 and SRSF2 in various combination. Isogenic HL-60 cells transduced with 3xFlag-tagged INTS3 or empty vector were obtained by puromycin selection (1 μg/mL). In order to let the cells fully establish epigenetic changes, they were analyzed after culture for more than 30 days.

For in vitro colony-forming assays, single-cell suspension was prepared and 15,000 cells/1.5 mL were plated in triplicates in cytokine supplemented methylcellulose medium (MethoCult™ GF M3434; StemCell Technologies), and colonies were enumerated every week. For the colony-forming assays shown in Extended Data Fig. 3k, *IDH2*^{WT}+*Srsf2*^{WT} and *IDH2*^{WT}+*Srsf2*^{P95H} mice were sacrificed at day 315 post-transplant to harvest BM as controls.

shRNA-mediated silencing

shRNAs against human INTS3 (hINTS3), mouse Ints3 (mInts3), and mouse Tet3 (mTet3) were cloned into MLS-E-Cherry and/or MLS-E-GFP vector and those against human UPF1 (hUPF1), mouse Fto (mFto), and mouse Alkbh5 (mAlkbh5) were cloned into LT3GEPIR (pRRL) Lenti-GFP-Puro-Tet-ON all-in-one vector. The antisense sequences were: hINTS3–1: TTTTCGAAACATAACCAGGTTA; hINTS3–2: TAAATATTAGGTACAGAGGCTT; mInts3–1: TTAAAAACAATTTAAAACTCGA; mInts3–2: TACAAATGCAGACTGACAGGAA; mInts3–3: TTCTTATCCTGAAAGGAGGGGA; mInts3–4: TTTAAAACTCGATTATCTTTGC; mInts3–5: TAATCTTACAAGGTCCCGGCCA; mTet3–1: TTATTAAGACCAAACCTGGCTA; mTet3–2: TTAAATGAAGTGTAGGCCATGC; mTet3–3: TTAAATGGAATTTTAAAACTAC; mTet3–4: GCCTGTTAGGCAGATTGTTCT; mTet3–5: GCTCCAACGAGAAGCTATTTG; hUPF1–1: TGGTATTACAGTAAACCACGCA; hUPF1–2:

TTGTGATTTAAACTCGTCACCA; mFto-1: TTCTAAGATATAATCCAAGGTG; mFto-2: TCTGGTTTCTGCTGTACTGGTA; mAlkbh5–1: TTGAACTGGAACTTGCAGCCGA; mAlkbh5–2: TTCATCAGCAGCATACCCACTG. mCherry$^+$ or GFP$^+$ cells with shRNAs against hINTS3, mInts3, or mTet3 were FACS-sorted.

Semi-quantitative and quantitative RT-PCR and mRNA stability assay

Total RNA was isolated using TRIzol reagent (Life Sciences) with standard RNA extraction protocol for snRNA quantification or using an RNeasy Mini or Micro kit (Qiagen) with DNase I treatment (Qiagen). For cDNA synthesis, total RNA was reverse transcribed with EcoDry kits (Random Hexamer or Oligo dT kits; Clontech), SuperScript (Invitrogen), RNA-Quant cDNA synthesis Kit (System Biosciences), or Verso cDNA Synthesis Kit (Thermo Fisher Scientific). Primers used in reverse-transcriptase polymerase chain reactions (RT-PCR) were: *INTS3* – Fwd1: TGAGTCGTGATGGCATGAAT (exon 4), Rev1: TCTTCACCAGTTCCCGTACC (exon 5; for detection of intron 4 retention), Rev2: CTGCTCTTCAGGACCCACTC (exon 7; for detection of exon 5 skipping); *NDUFAF6* – Fwd: GCCTGTGGCCATTGAACTAT, Rev: ACAATGCCTTGTGCTTTTCC; *PHF21A* – Fwd: TCCATGGCCTGGAACTTTAG, Rev: GCCAGGATGGTGTTCTTCAT; *GLYR1* – Fwd: AGGTCAGGCCCAGTTCTCTT, Rev: TCACGTCTAAGCGTCCAGTGFI*GAPDH* – Fwd: GCAAATTCCATGGCACCGTC, Rev: TCGCCCCACTTGATTTTGG.

The PCR cycling conditions (33 cycles) chosen were as follows: (1) 30 s at 95 °C (2) 30 s at 60 °C (3) 30 s at 72 °C with a final 5-min extension at 72 °C. Reaction products were analyzed on 2% agarose gels. The bands were visualized by ethidium bromide staining.

Quantitative real-time reverse transcriptase PCR (qPCR) analyses were performed on an Applied Biosystems QuantStudio 6 Flex cycler using SYBR Green Master Mix (Roche). The following primers were used: *hINTS3* – Fwd2: CTGCAGGATACCTGCCGTA (exon 4), Rev3: CTTTCCCGTTCCTGACAGAG (intron 5; for specific quantification of transcript with intron 4 retention); Fwd1: TGAGTCGTGATGGCATGAAT (exon 4), Rev4: GGCTGTAACATCTCCACCTGA (exon 4–6; for specific quantification of transcript with exon 5 skipping); Fwd3: GGGCAATGCTGAGAGAGAAG (exon 14), Rev5: TGCCTCTGCATTGTCATAGC (exon 15); *mInts3* – Fwd: GTGGCTGTTATTGACTCTGCAC, Rev: CAGGTTCCCCATCATCACAT; *mFto* – Fwd: CACTTGGCTTCCTTACCTGACCCCC, Rev: GGTATGCTGCCGGCCTCTCGG; *mAlkbh5* – Fwd: CGGCCTCAGGACATTAAGGA, Rev: TCGCGGTGCATCTAATCTTG; Total U2snRNA – Fwd: CTTCTCGGCCTTTTGGCTAAGAT, Rev: GTACTGCAATACCAGGTCGATGC; Uncleaved U2snRNA – Fwd: ACGTCCTCTATCCG +AGGACAATA, Rev: GCAGGTGCTACCGTCTCTCAC; Total U4snRNA – Fwd: GCAGTATCGTAGCCAATGAGGTCTA, Rev: CCAGTGCCGACTATATTGCAAGTC;

Uncleaved U4snRNA – Fwd: CGTAGCCAATGAGGTCTATCCG, Rev: CCTCTGTTGTTCAACTGCAAGAAA; *hGAPDH*–Fwd: GCAAATTCCATGGCACCGTC, Rev: TCGCCCCACTTGATTTTGG; *mGapdh* – Fwd: TGGAGAAACCTGCCAAGTATG, Rev: GGAGACAACCTGGTCCTCAG.

All samples, including the template controls were assayed in triplicate. The relative number of target transcripts was normalized to the housekeeping gene found in the same sample. The relative quantification of target gene expression was performed with the standard curve or comparative cycle threshold (CT) method.

mRNA stability assay was performed as previously described[6]. Briefly, anti-*UPF1* shRNA- or control shRNA lentivirus-infected K562 *SRSF2*P95H knock-in cells were generated by puromycin selection (1 μg/mL) for 7 days and shRNAs against UPF1 were expressed by doxycycline (2 μg/mL) for 2 days. GFP (shRNA)-positive cells were FACS sorted, treated with 2.5 μg/ml Actinomycin D (Life Technologies), and harvested at 0, 2, 4, 8, and 12 hrs.

Chromatin immunoprecipitation (ChIP)

Cells were crosslinked and collected. Chromatin was broken down into 200 – 1000 bp fragments using an E220 Focused-ultrasonicator. An antibody was added into the lysate and incubated overnight at 4 °C. Twenty microliters of ChIP–grade Protein A/G Dynabeeds was added into each IP tube and incubated for 2 hours. IP samples were washed and crosslinks reversed by adding proteinase K and incubating overnight at 65 °C. DNA was purified with AMPureXP beads and eluted DNA was subjected to qPCR to measure the enrichment. RNA polymerase II antibody (05–623; EMD Millipore, Billerica, MA, USA) was used in this study. Primer sequences used for ChIP-PCR were as follows: Intron 3–1 – Fwd: atacccggcccttgctatac, Rev: gcaacttccttagcctgctg; Intron 3–1 – Fwd: atacccggcccttgctatac, Rev: gcaacttccttagcctgctg; Intron 3–2 – Fwd: ctggcaggtgaaaagcagat, Rev: ggcaggggagagaaaagc; Intron 3–3 – Fwd: agcaggcttttctgcctcat, Rev: tttctttccacaggggtcct; Exon 4 – Fwd: cgggacttagctctggtgag, Rev: cctgagtacggcaggtatcc; Intron 4 – Fwd: ctctgtcaggaacgggaaag, Rev: tgtgagtttgagaagggagcta; Exon 5 – Fwd: acgggaactggtgaagagtg, Rev: ctgggctctcctcctttctt; Intron 5–1 – Fwd: ctccacccccattatctgaa, Rev: aaatgtcagggtctgttctgtg; Intron 5–2 – Fwd: tcggtgacatctgtctgagc, Rev: cagtgggctaatggtgaggt; Intron 5–3 – Fwd: aacactgatgctcctgttttga, Rev: actatgccttgccccaggt; Intron 5–4 – Fwd: gctgttgtcagccacctgta, Rev: tttggcccttgaaaatgaac; Intron 5–5 – Fwd: tgtgttaattctgccccaca, Rev: ggatgtcctgagtcctgcac; Intron 5–6 – Fwd: gtaatgggatggcagtcagg, Rev: cctgatttcaaaaggggaaa; Exon 6 – Fwd: agcaaaggtagcatccacca, Rev: cttgcctccccctctctaac; Intron 6–1 – Fwd: tttgatccagacctccttgg, Rev: gcaggggagaaaaggatacc; Intron 6–2 – Fwd: gggggtacatattgggcttt, Rev: gaaagcctcacctccaaaca; Intron 6–3-CTCF binding site–Fwd: ctcctcccaacgttcacact, Rev: atccgtgcccagagcacta; Intron 6–4 – Fwd: agggggcctttcaactctt, Rev: atggggacaggacgtatttg; Intron 6–5 – Fwd: ttccctgccttccaacag, Rev: tcccagttgctttaaaaggagt.

ChIP-seq libraries were prepared as previously described[31] and sequenced by the Integrated Genomics Operation (IGO) at MSK with 50 bp paired-end reads.

ChIP-sequencing of primary human AML samples

ChIP was performed as previously described[32] using the following antibodies: RNAPolII-Ser2P antibody - ChIP Grade (Abcam ab5095), RNAPlI-Ser5P antibody [4H8] (Abcam ab5408), and anti-HP1γ antibody, clone 42s2 (05–690 from Merck Millipore). Libraries were size selected with AMPure beads (Beckman Coulter) for 200–800 base pair size range and quantified by qPCR using a KAPA Library Quantification Kit. ChIP-seq data were

generated using the NextSeq platform from Illumina with 2 × 75 bp Hi Output (all samples pooled, and sequenced on four consecutive runs before merger of FASTQ files).

Histological analyses

Mice were sacrificed and autopsied, and dissected tissue samples were fixed in 4% paraformaldehyde, dehydrated, and embedded in paraffin. Paraffin blocks were sectioned at 4 μm and stained with hematoxylin and eosin (H&E). Images were acquired using an Axio Observer A1 microscope (Carl Zeiss) or scanned using a MIRAX Scanner (Zeiss).

Patient Samples

Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center (under MSK IRB protocol 06–107), Université Paris-Saclay (under declaration DC-200–725 and authorization AC-2013–1884), and the University of Manchester (institution project approval 12-TISO-04), and conducted in accordance with the Declaration of Helsinki protocol. Written informed consent was obtained from all participants. Manchester samples were retrieved from the Manchester Cancer Research Centre Haematological Malignancy Tissue Biobank, which receives sample donations from all consenting leukemia patients presenting to The Christie Hospital (REC Reference 07/H1003/161+5; HTA license 30004; instituted with approval of the South Manchester Research Ethics Committee). Patient samples were anonymized by the Hematologic Oncology Tissue Bank of MSK, Biobank of Gustave Roussy, and the Manchester Cancer Research Centre Haematological Malignancy Tissue Biobank.

Mutational analysis of patient samples

Genomic DNA is routinely extracted from mononuclear cell samples submitted to the Manchester Cancer Research Centre Haematological Tissue Biobank. Targeted sequencing for recurrent myeloid mutations, using either: (a) a 54 gene panel (TruSight™ Myeloid; Illumina), pooling 96 samples with 5% PhiX onto a single NextSeq high output, 2 × 151 bp sequencing run; VCF files were analyzed using Illumina's Variant Studio software; (b) a 40 gene panel (Oncomine Myeloid Research Assay; ThermoFisher), processing eight samples per Ion 530 chip on the IonTorrent platform; data analysis performed using the Ion Reporter software; (c) a 27 gene custom panel (48 × 48 Access Array; Fluidigm) sequenced by Leeds HMDS on the MiSeq platform (300v2); or (d) MSK HemePACT[33] targeting all coding regions of 585 genes known to be recurrently mutated in leukemias, lymphomas, and solid tumors. All panels provide sufficient coverage to detect minimum variant allele fraction 5% for all genes, except for the Access Array panel and *SRSF2*; all samples genotyped by this approach underwent manual Sanger sequencing of *SRSF2* exon 1 using the following primers (tagged with Fluidigm Access Array sequencing adaptors CS1/CS2): Fwd: acactgacgacatggttctacacccgtttacctgcggctc, Rev: tacggtagcagagacttggtctccttcgttcgctttcacgacaa.

Statistics and reproducibility

Statistical significance was determined by (1) unpaired two-sided Student's t-test after testing for normal distribution, (2) one-way or two-way ANOVA followed by Tukey's,

Sidak's, or Dunnett's multiple comparison test, or (3) Kruskal-Wallis tests with uncorrected Dunn's test where multiple comparisons should be adjusted (unless otherwise indicated). Data were plotted using GraphPad Prism 7 software as mean values, with error bars representing standard deviation. For categorical variables, statistical analysis was done using Fisher's exact test or Chi-square test (two-sided). Representative WB and PCR results are shown from three or more than three biologically independent experiments. Representative flow cytometry results and cytomorphology are shown from biological replicates ($n \geq 3$). *P, **P, and ***P represent *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$, respectively, unless otherwise specified.

mRNA isolation, sequencing, and analysis

RNA was extracted as shown above. Poly(A)-selected, unstranded Illumina libraries were prepared with a modified TruSeq protocol. 0.5× AMPure XP beads were added to the sample library to select for fragments < 400 bp, followed by 1× beads to select for fragments >100 bp. These fragments were then amplified with PCR (15 cycles) and separated by gel electrophoresis (2% agarose). 300-bp DNA fragments were isolated and sequenced on an Illumina HiSeq 2000 (~100M 101 bp reads per sample).

Primary samples from the Manchester Cancer Research Centre Haematological Malignancies Biobank with known *IDH2/SRSF2* mutation genotype were FACS sorted to enrich for blasts on a FACS Aria III sorter using a panel including the following antibodies (all mouse anti-human): CD34-PerCP (8G12, BD); CD117-PECy7 (104D2, eBioscience); CD33-APC (P67.6, BioLegend); HLA-DR-FITC (L243, BioLegend); CD13-PE (L138, BD); CD45-APC-H7 (2D1, BD). RNA was extracted immediately using a Qiagen Micro RNeasy kit. All RNA samples had RIN values > 8. Poly(A)-selected, strand-specific SureSelect (Agilent) mRNA libraries were prepared using 200 ng RNA according to the manufacturer's protocol. Libraries were pooled and sequenced (2 × 101 bp paired end) to > 100 million reads per sample on two HiSeq 2500 high throughput runs before retrospective merger of FASTQ files for downstream alignment and splicing analysis as described below. Transcriptional analysis was done using gene set enrichment analysis (GSEA)[34].

Publicly available RNA-sequencing data

Unprocessed RNA-sequencing (RNA-seq) reads of TCGA and Leucegene datasets (human AML patients) were downloaded from NCI's Genomic Data Commons Data Portal (GDC Legacy Archive; TCGA-LAML dataset) and NCBI's Sequence Read Archive (SRA; accession numbers SRP056295). The TCGA dataset consists of paired-end 2 × 50 bp libraries, with an average read count of 76.92 M. The Leucegene dataset consists of paired-end 2 × 100 bp libraries, with an average read count of 50.40 M per sample. The RNA-seq samples in the Leucegene dataset have 1~3 sequencing runs (~50 M each run), and only one run was used to represent each RNA-seq sample.

Genome and splice junction annotations

Human assembly hg38 (GRCh38) and Ensembl database (human release 87) were used as the reference genome and gene annotation, respectively. RNA-seq reads were aligned by using 2-pass STAR 2.5.2a[35]. Known splice junctions from the gene annotation and new

junctions identified from the alignments of the TCGA dataset were combined to create the database of alternative splicing events for splicing analysis.

Mutational analysis for the RNA-seq data

Samtools (1.3.1) were used to generate variant call format (VCF) files for 7 target genes: *IDH1*, *IDH2*, *TET2*, *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2* with mpileup parameters (-Bvu). The VCF files were further processed by our in-house scripts to filter out mutations whose VAF was lower than 15%. The filtered VCF files were used for variant effect predictor (version 89.4) to annotate the consequences of the mutations. We defined "control" patient samples as those without mutations in the 7 target genes, *IDH2* mutated samples as those with only *IDH2* mutations but no mutations in the other 6 target genes, *SRSF2* mutated samples as those with only *SRSF2* mutations but no mutations in the other 6 target genes, Double-mutant samples as those with both *IDH2* and *SRSF2* mutations but no mutations in the other 5 target genes, and "Others" as those with mutations in *IDH1, TET2, SF3B1, U2AF1*, and *ZRSR2*.

Identification and quantification of differential splicing

The inclusion ratios of alternative exons or introns were estimated by using PSI-Sigma[25]. Briefly, the new PSI index considers all isoforms in a specific gene region and can report the PSI value of individual exons in a multiple-exon-skipping or more complex splicing event. The database of splicing events was constructed based on both gene annotation and the alignments of RNA-seq reads. A new splicing event not known to the gene annotation is labeled as "Novel" and a splicing event whose reference transcript is known to induce nonsense-mediated decay is labeled as "NMD" in Supplementary Tables. The inclusion ratio of an intron retention isoform is estimated based on the median of 5 counts of intronic reads at the 1^{st}, 25^{th}, 50^{th}, 75^{th}, and 99^{th} percentiles in the intron. A splicing event is reported when both sample-size and statistical criteria are satisfied. The sample-size criterion requires a splicing event to have more than 20 supporting reads in more than 75% of the two populations in the comparison. For example, for a comparison of 130 control versus 6 *IDH2* mutant samples, a splicing event would be reported only when having more than 98 controls and 5 *IDH2* mutant samples with more than 20 supporting reads. In addition, a splicing event is reported only when it has more than 10% PSI change in the comparison and has a *P*-value lower than 0.01.

To generate Fig. 4f, RNA-seq reads were mapped and PSI values were calculated using junction-spanning reads as previously described[36,37]. All reads mapping to the INTS3 introns (chr1:153,718,433–153,722,231; hg19) were extracted from the bam files and the per-nucleotide coverage was calculated. Data from normal peripheral blood and BM mononuclear cells and $CD34^+$ cord blood cells are combined and shown as normal hematopoietic cells.

Motif enrichment and distribution

Motif analysis was done by using MEME SUITE[38]. Briefly, the sequences of alternative exons of exon-skipping events were extracted from a given strand of the reference genome. The sequences were used as the input for MEME SUITE to search for motifs. One

occurrence per sequence was set to be the expected site distribution. The width of motif was set to 5. The top 1 motif was selected based on the ranking of E-value.

Heatmap and sample clustering (differential splicing)

The heatmaps and sample clustering were done by using MORPHEUS (software.broadinstitute.org/morpheus/). The individual values in the matrix for the analysis were PSI values of a splicing event from a given RNA-seq sample. Splicing events were selected based on three criteria: (1) present in both TCGA and Leucegene datasets; (2) more than 15% PSI changes; and (3) false discovery rate smaller than 0.01. Unsupervised hierarchical clustering was based on one minus Pearson's correlation (complete linkage).

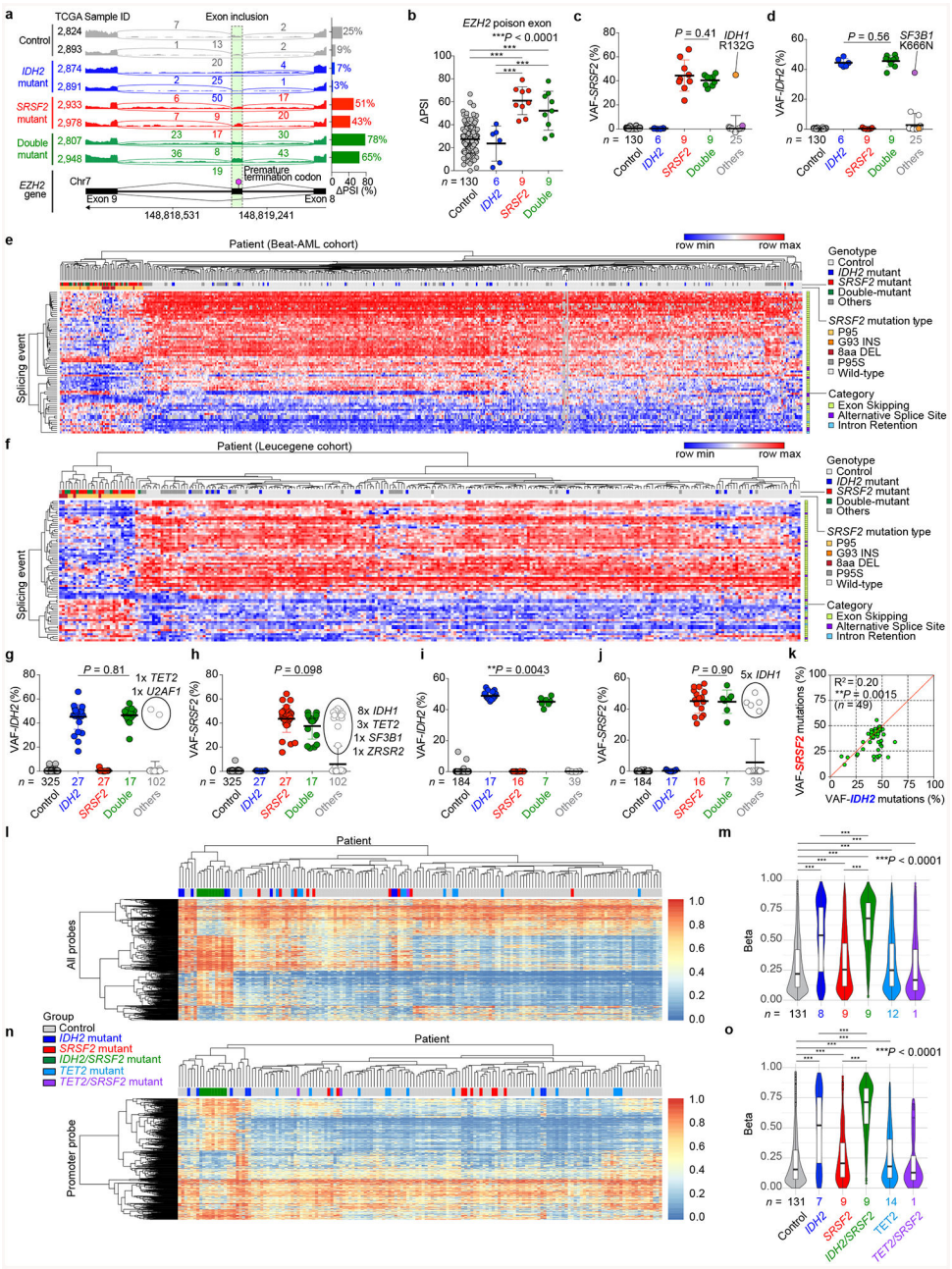
Correlation between global changes in splicing and DNA methylation

DNA methylation levels were determined by enhanced reduced representation bisulfite sequencing (eRRBS) while differentially spliced events were obtained from RNA-seq data. In Fig. 3e, Overlaps of differentially methylated regions of DNA with differential splicing was obtained by evaluating differential cytosine methylation in 500 bp segments of DNA at genomic coordinates at which differential RNA splicing were observed comparing AML with distinct *IDH2/SRSF2* genotypes shown ("WT" represents patients without mutations in *IDH1*/*IDH2*/Spliceosomal genes).

DATA ABAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. RNA-seq, ChIP-seq, and eRRBS data have been deposited in NCBI Sequence Read Archive (SRA) under accession number SRP133673. Gel source data can be found in Supplementary Fig. 1. Other data that support this study's findings are available from the authors upon reasonable request.

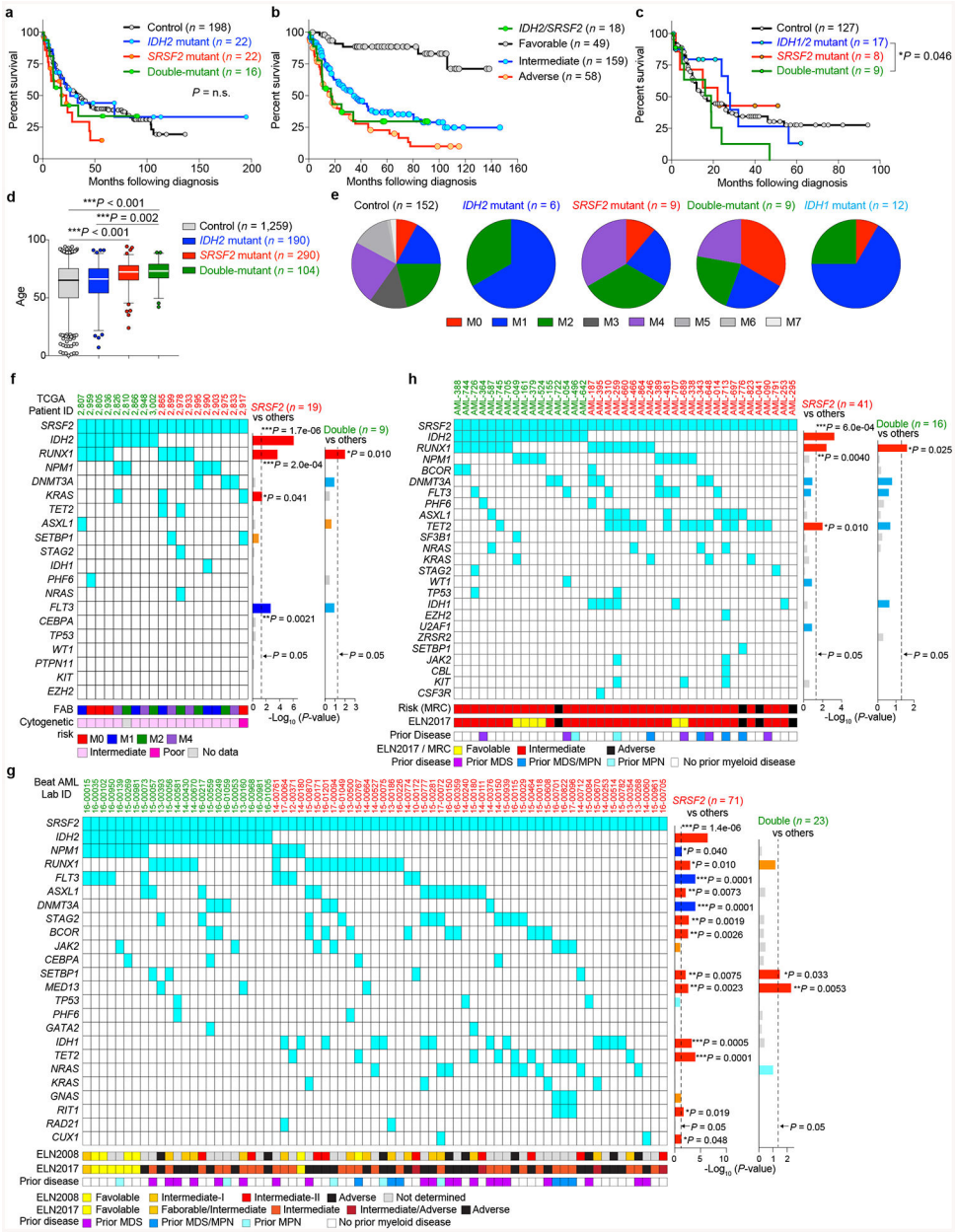
Extended Data



Extended Data Fig. 1 |. Mutant *SRSF2*-mediated splicing events in acute myeloid leukemia (AML).

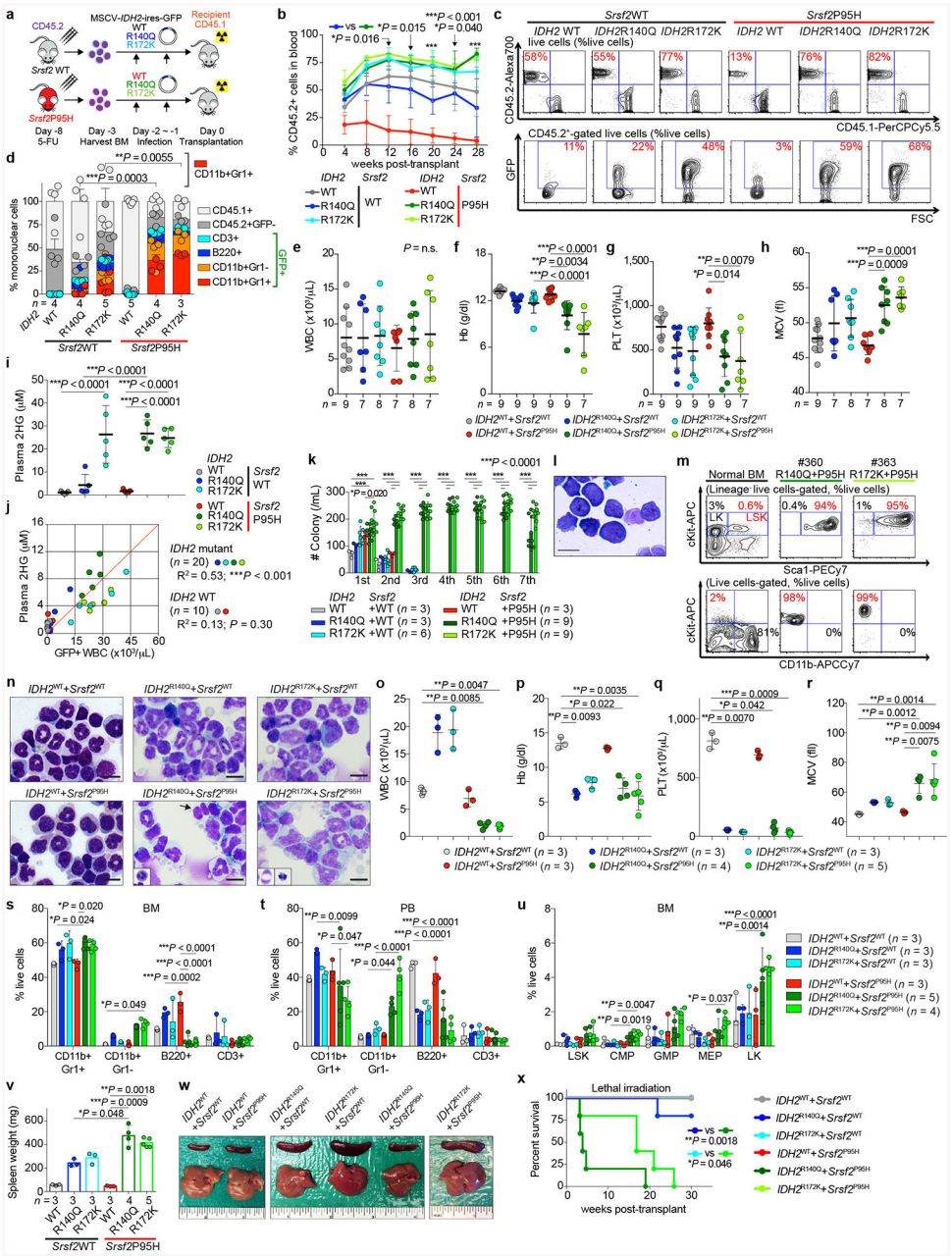
a, Representative Sashimi plots of RNA-seq data from the TCGA showing the poison exon inclusion event in *EZH2* ("Control" represents samples that are wild-type (WT) for the following 7 genes: *IDH1, IDH2, TET2, SRSF2, SF3B1, U2AF1*, and *ZRSR2*; *"IDH2* mutant" refers to patients with an *IDH2* mutation and no mutation in the other 6 genes; "*SRSF2* mutant" refers to patients with an *SRSF2* mutation and no mutation in the other 6 genes; "Double-mutant" refers to patients with an *IDH2* and *SRSF2* mutation and no mutation in the other 5 genes; "Others" refers to patients with mutations in *IDH1, TET2, SF3B1, U2AF1*, or *ZRSR2*; figure made using Integrative Genomics Viewer (IGV 2.3)[39]). **b**,

ΔPSI (Percent-Spliced-In) values of *EZH2* poison exon inclusion (the number of analyzed patients is indicated; the mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test; Note that patients classified as "Others" include one *SRSF2*^{P95L} mutant patient with coexisting *IDH1*^{R132G} mutation (TCGA ID: 2990) and one *IDH2*^{R140Q} mutant patient with an *SF3B1*^{K666N} mutation (TCGA ID: 2973), which were excluded from the analyses shown above. **c**, **d**, **g**, **h**, **i**, **j**, Variant allele frequencies (VAFs) of *SRSF2* mutations affecting the Proline 95 residue (**c**, **h**, **j**) and *IDH2* mutations affecting *IDH2* Arginine 140 or 172 (**d**, **g**, **i**) in TCGA (**c**, **d**), Beat-AML (**g**, **h**), and Leucegene (**i**, **j**) datasets (the mean ± s.d.; a two-sided Student's t-test). **e**, **f**, Heat map based on the ΔPSI of mutant *SRSF2*-specific splicing events in AML from Beat-AML (**e**) and Leucegene (**f**) cohorts. "8aa DEL" represents samples with 8 amino acid deletions in SRSF2 starting from Proline 95, which has similar effects on splicing as point mutations affecting SRSF2 P95. Detailed information of splicing events shown is available in Supplementary Table 1. **k**, Variant allele frequencies (VAFs) of *IDH2* (x-axis) and *SRSF2* mutations (y-axis) in *IDH2/SRSF2* double-mutant AML determined by RNA-seq data from the TCGA, Beat-AML, Leucegene, and our previously unpublished cohorts (Pearson correlation coefficient; *P*-value (two-tailed) was calculated by Prism7). **l**, **n**, Unsupervised hierarchical clustering of DNA methylation levels of all probes (**l**) or at the promoter probes (**n**) in the TCGA AML cohort based on *IDH2/SRSF2/TET2* genotypes. **m**, **o**, DNA methylation levels of AML samples from each genotype are quantified and visualized from **l** and **n** as violin plots (the mean represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn to 2.5 and 97.5 percentiles; one-way ANOVA with Tukey's multiple comparison test).** $P < 0.01$; *** $P < 0.001$.



Extended Data Fig. 2 |. Clinical relevance of co-existing *IDH2/SRSF2* mutations in AML. **a-c**, Kaplan-Meier survival analysis of AML patients from the Manchester/Christie Biobank dataset ((**a**): based on *IDH2/SRSF2* genotype (n = 258); (**b**): based on cytogenetic risk (n = 284)) and the TCGA (**c**) (n = 161) based on *IDH1/IDH2/SRSF2* genotypes (Log-rank (Mantel-Cox) test (two-sided)). **d**, Age at diagnosis of patients from the TCGA, Beat-AML, and Manchester/Christie Biobank cohorts combined (the mean represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn to 2.5 and 97.5 percentiles; samples below 2.5 percentile and above 97.5 percentile are shown as plots; one-way ANOVA with Tukey's multiple comparison test). **e**, Distribution of French-American-British (FAB) classification of AML patients with indicated genotypes from the TCGA cohort. **f-h**, Mutations co-existing with *IDH2/SRSF2* double-mutant and *SRSF2*

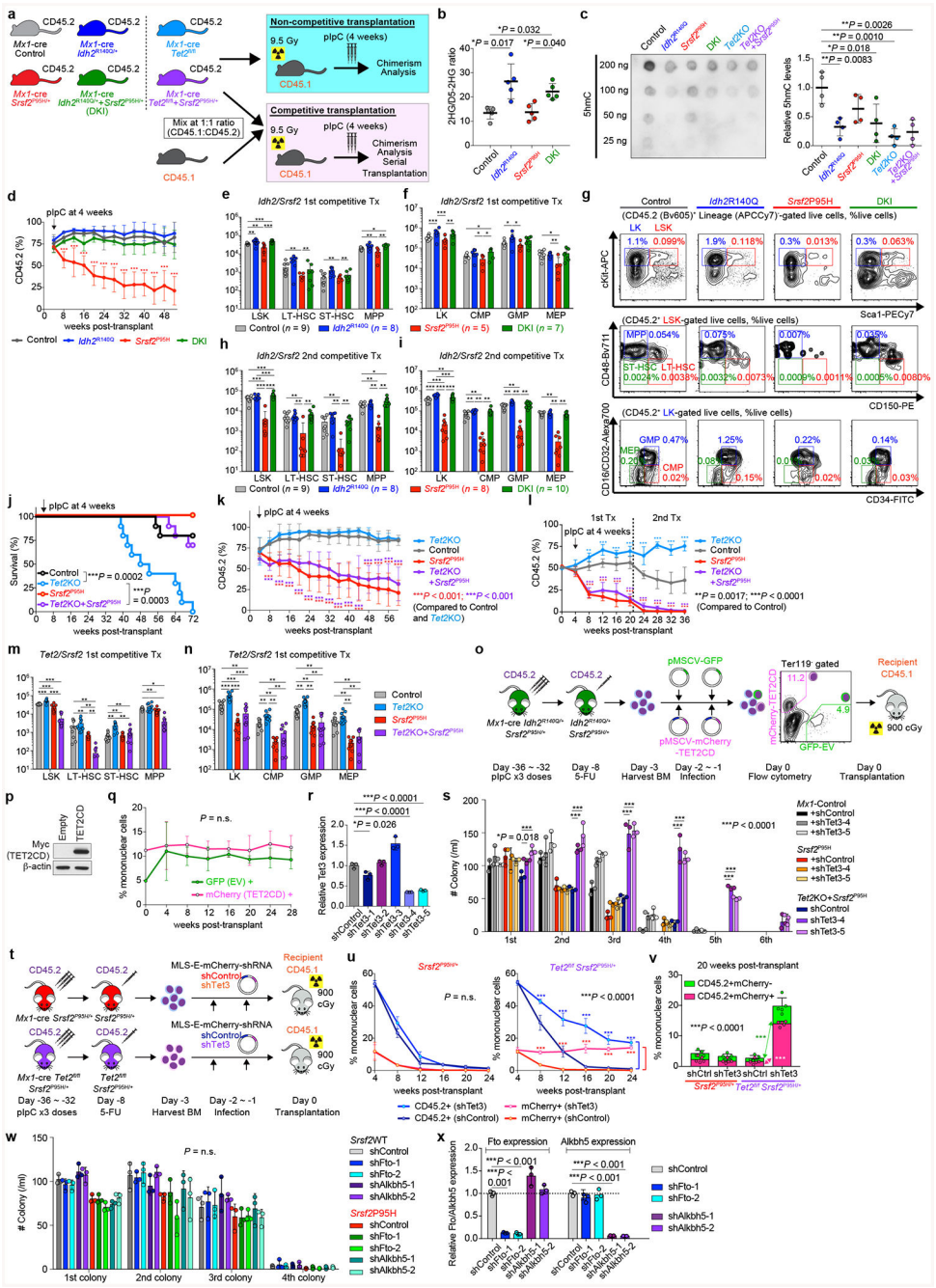
single-mutant AML from the TCGA (**f**), Beat-AML (**g**), and Manchester/Christie Biobank (**h**) cohorts are shown with FAB Classification, cytogenetic risk, prior history of myeloid disorders, and genetic risk stratification based on European LeukemiaNet (ELN) 2008 and ELN2017 guidelines (the number of patients is indicated; *P*-values on the right represent statistical significance of co-occurrence (red and orange) or mutual exclusivity (blue and light blue) of each gene mutation with *SRSF2* (including those in *IDH2/SRSF2* double-mutant AML) or co-existing *IDH2* and *SRSF2* mutations; Fisher's exact test (two-sided)). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.



Extended Data Fig. 3 |. Mutant *IDH2* cooperates with mutant *Srsf2* to generate lethal MDS with proliferative features *in vivo*.

a, Schematic of bone marrow (BM) transplantation model. **b**, **c**, Chimerism of CD45.2$^+$ cells in the peripheral blood (PB) of recipient mice over time (**b**) (n = 5 per group at 4 weeks; the mean percentage ± s.d.; two-way ANOVA with Tukey's multiple comparison test) and representative flow cytometry data showing the chimerism of CD45.2$^+$ vs CD45.1$^+$ (top) or GFP$^+$ (bottom) cells in PB at 16 weeks post-transplant (**c**) (representative results from five recipient mice; the percentages listed represent the percent of cells within live cells). **d**, Composition of PB mononuclear cells (PBMNCs) at 28 weeks post-transplant (the number of analyzed animals is indicated; the mean + s.d.; two-way ANOVA with Tukey's multiple

comparison tests statistical significances were detected in % of $CD11b^{+}Gr1^{+}$ cells in *IDH2*^{R140Q} + *Srsf2*^{WT} vs *IDH2*^{R140Q} + *Srsf2*^{P95H} and in *IDH2*^{R172K} + *Srsf2*^{WT} vs *IDH2*^{R172K} + *Srsf2*^{P95H}). **e-h**, Blood counts at 20 weeks post-transplant (WBC (**e**); Hb (**f**); PLT (**g**); MCV, mean corpuscular volume (**h**); the number of analyzed animals is indicated; the mean ± s.d.; one-way ANOVA with Tukey's multiple comparison tests. **i**, Plasma 2HG levels at 20 weeks post-transplant (2HG levels were quantified as described[40]; $n = 5$ per group were randomly selected; the mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **j**, Correlations between plasma 2HG levels and number of GFP^{+} cells in peripheral blood at 24 weeks post-transplant ($n = 5$ per group; the Pearson correlation coefficient (R^2) and *P*-values (two-tailed) were calculated using PRISM 7). **k**, Colony numbers from serial replating assays of BM cells harvested from end-stage mice from Fig. 2b are shown (the mean value ± s.d. represented by lines above the box; the number of analyzed animals is indicated; two-way ANOVA with Tukey's multiple comparison test). **l**, Giemsa staining of *IDH2*^{R140Q} + *Srsf2*^{P95H} double-mutant cells from the 6th plating (scale bar, 10 μm; original magnification × 400; representative result from nine biologically independent experiments). **m**, Immunophenotype of colony cells at the 6th plating. Normal BM cells were used as a control (the percentage listed represent the percent of cells within live cells; representative result from nine recipient mice). **n**, Cytomorphology of BM mononuclear cells (BMMNCs) from recipient mice at end-stage. BM cells from *IDH2* single-mutant and *IDH2/Srsf2* double-mutant groups have increased granulocytes. In addition, *IDH2/Srsf2* double-mutant groups had proliferation of monoblastic and monocytic cells as well as dysplastic features such as abnormally segmented neutrophils (black arrow and inset) and binucleated erythroid precursors with irregular nuclear contours (insets) (scale bar, 10 μm; original magnification × 400; representative results from three controls and nine recipients are shown; number of animals indicated in **o-r**). **o-r**, Blood counts at end-stage (WBC (**o**); Hb (**p**); PLT (**q**); MCV (**r**); the number of analyzed animals is indicated; the mean ± s.d.; Kruskal-Wallis tests with uncorrected Dunn's test). **s-u**, Results from flow cytometry analysis of BM (**s**) and PB (**t**) mature lineages as well as BM hematopoietic stem/progenitor cells (HSPC) from two tibias, two femurs, and two pelvic bones (**u**) are quantified (LSK: $Lineage^{-}Sca1^{+}cKit^{+}$; LT-HSC: long-term hematopoietic stem cell; ST-HSC: short-term HSC; MPP: multi-potent progenitor; LK: $Lineage^{-}Sca1^{-}cKit^{+}$; CMP: common myeloid progenitor; GMP: granulocyte-monocyte progenitor; MEP: megakaryocyte-erythroid progenitor; the number of analyzed animals is indicated; the mean + s.d. is represented; two-way ANOVA with Tukey's multiple comparison test). **v**, **w**, Spleen weight at end-stage (the number of analyzed animals is indicated; the mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test) (**v**) and representative photographs of spleens from recipient mice from **v** (**w**) (each photograph was taken with an inch ruler). **x**, Kaplan-Meier survival analysis of serially transplanted recipient mice that were lethally irradiated ($n = 5$ per group; Log-rank (Mantel-Cox) test (two-sided)). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

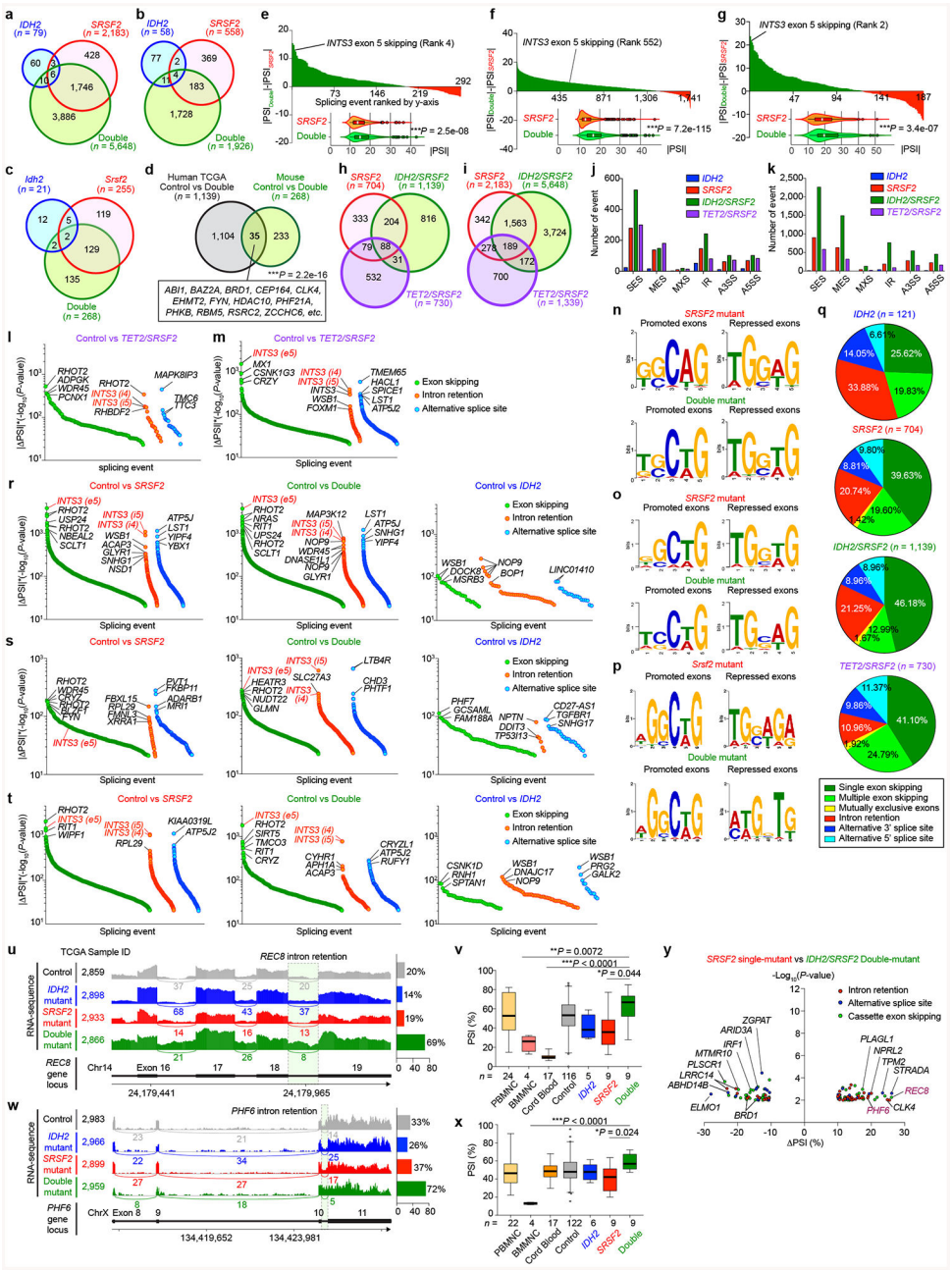


Extended Data Fig. 4 |. Collaborative effects of mutant *Idh2* and mutant *Srsf2* are not dependent on *Tet2* loss alone.

a, Schematic of competitive and non-competitive transplantation assays of CD45.2$^+$ *Mx1*-cre control, *Mx1*-cre *Idh2*$^{R140Q/+}$, *Mx1*-cre *Srsf2*$^{P95H/+}$, *Mx1*-cre *Idh2*$^{R140Q/+}$ *Srsf2*$^{P95H/+}$ mice, *Mx1*-cre *Tet2*$^{fl/fl}$, *Mx1*-cre *Tet2*$^{fl/fl}$ *Srsf2*$^{P95H/+}$ mice into CD45.1$^+$ recipient mice. **b**, 2HG levels of bulk PBMNCs from primary *Mx1*-cre mice were measured at 3 months post-pIpC (polyinosinic:polycytidylic acid) and normalized to internal standard (D-2-hydroxyglutaric-2,3,3,4,4-*d5* acid; D5–2HG) (2HG and D5–2HG levels were quantified as described[40]; $n = 5$ per group; the mean ± s.d.; one-way ANOVA with Tukey's multiple

comparison test). **c**, DNA extracted from sorted cKit$^+$ BM cells from primary *Mx1*-cre mice at 1 month post-pIpC was probed with antibodies specific for 5-hydroxymethylcytosine (5hmC) (left). Relative intensity of each dot was measured by ImageJ and divided by input DNA amount for comparison (right; $n = 4$; intensity of each dot divided by amount of input DNA was combined per genotype; representative results from three biologically independent experiments with similar results; the mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **d**, Chimerism of PB CD45.2$^+$ cells in non-competitive transplantation (pIpC was injected at 4 weeks post-transplant; the mean ± s.d.; $n = 10$ (Control and *Idh2*^{R140Q}), $n = 8$ (*Srsf2*^{P95H}), and $n = 9$ (DKI) at 0 week; two-way ANOVA with Tukey's multiple comparison test; *P*-values from comparison between *Srsf2*^{P95H} and each of other groups are shown). **e**-**i**, Absolute number of BM HSPCs from two tibias, two femurs, and two pelvic bones were measured in the primary (**e**, **f**) and serial (**h**, **i**) competitive transplant of *Idh2/Srs2* mutant cells, and representative flow cytometry of BM HSPCs from the primary competitive transplant of *Idh2/Srsf2* mutant cells from **e**, **f** (the percentage listed represents the percent of cells within live cells) (the number of analyzed animals is indicated; the mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **j**, Kaplan-Meier survival analysis of CD45.1$^+$ recipient mice transplanted non-competitively with BM cells from CD45.2$^+$ *Mx1*-cre control, *Mx1*-cre *Tet2*$^{fl/fl}$, *Mx1*-cre *Srsf2*$^{P95H/+}$, and *Mx1*-cre *Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ mice (pIpC was injected at 4 weeks post-transplant; $n = 10$ per genotype; Log-rank (Mantel-Cox) test (two-sided)). **k**, **l**, Chimerism of PB CD45.2$^+$ cells in non-competitive (**k**) ($n = 10$ (Control and *Tet2*KO), $n = 8$ (*Srsf2*^{P95H}), and $n = 5$ (*Tet2*KO + *Srsf2*^{P95H}) at 0 week) or competitive (**l**) ($n = 9$ (Control), $n = 10$ (*Tet2*KO), $n = 8$ (*Srsf2*^{P95H}), and $n = 10$ (*Tet2*KO + *Srsf2*^{P95H}) at 0 week) transplantation (pIpC was injected at 4 weeks post-transplant; percentages of CD45.2$^+$ cells at pre-transplant are also shown as data at 0 weeks in **l**; the mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **m**, **n**, Absolute number of BM HSPCs from two tibias, two femurs, and two pelvic bones were measured in the primary competitive transplant of *Tet2/Srsf2* mutant cells ($n =$ 10 per genotype; the mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **o**, Schematic of TET2 catalytic domain (CD: catalytic domain; EV: empty vector) retroviral BM transplantation model. **p**, Western blot analysis confirming the expression of Myc-tagged TET2CD in Ba/F3 cells transduced with or without TET2CD (representative images from two biologically independent experiments with similar results). **q**, Chimerism of mCherry-TET2CD$^+$ and GFP-EV$^+$ cells in PB of recipient mice over time ($n = 10$; the mean percentage ± s.d.; two-way ANOVA with Sidak's multiple comparison test). **r**, qPCR of *Tet3* in the first colony cells from **s** ($n = 3$; the mean ± s.d.; a two-sided Student's t-test). **s**, Colony numbers from serial replating assays of BM cells from *Mx1*-cre control, *Mx1*-cre *Srsf2*$^{P95H/+}$, and *Mx1*-cre *Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$ mice transduced with anti-Tet3 short-hairpin RNAs (shRNAs) ($n = 3$; the mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **t**, Schematic of anti-*Tet3* shRNA (shTet3) retroviral BM transplantation model. **u**, **v**, Chimerism of mCherry$^+$ cells in CD45.2$^+$ donor cells in PB of recipient mice over time (**u**; left panel: *Mx1*-cre *Srsf2*$^{P95H/+}$, right panel: *Mx1*-cre *Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$; $n =$ 5 per group) and at 20 weeks post-transplant (**v**) (the mean percentage ± s.d.; two-way ANOVA with Sidak's multiple comparison test). **w**, Colony numbers from serial replating assays of either *Mx1*-cre *Srsf2*$^{+/+}$ or *Srsf2*$^{P95H/+}$ BM cells transduced with an shRNA against *Fto* or *Alkbh5*. BM cells were harvested at 1 month post-pIpC ($n = 3$; the mean value

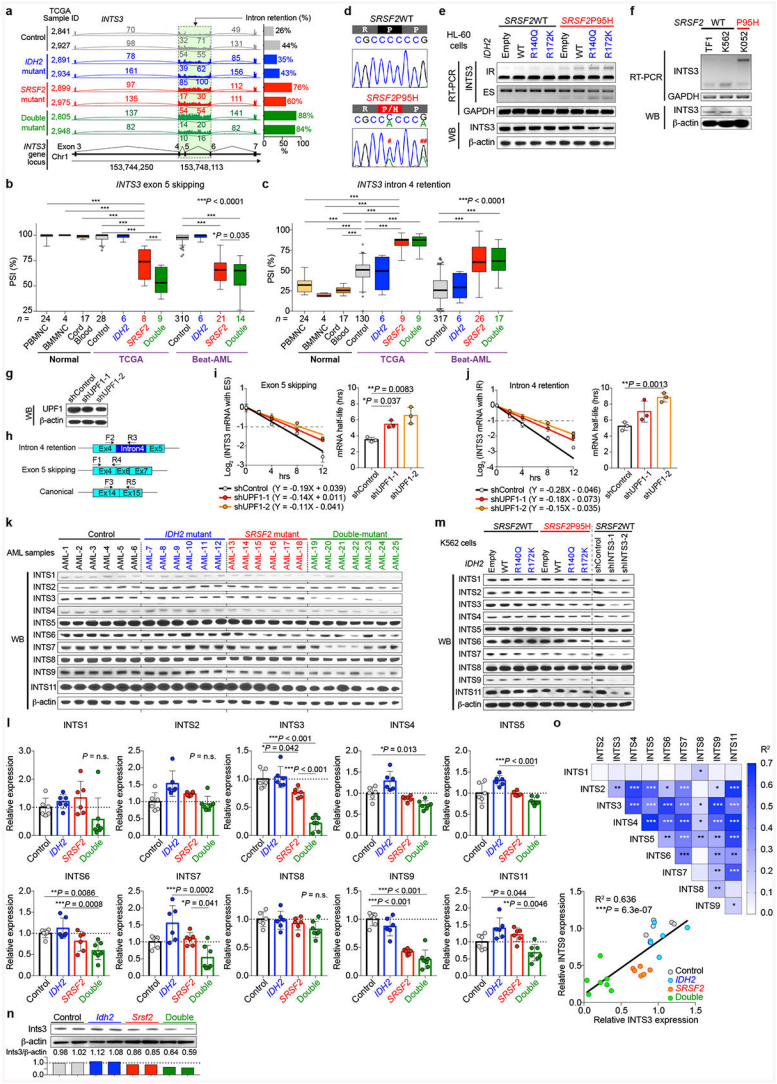
± s.d.; two-way ANOVA with Tukey's multiple comparison test). **x**, qPCR of *Fto* or *Alkbh5* in Ba/F3 cells transduced with shRNAs targeting mouse *Fto* or *Alkbh5* ($n = 3$; the mean value ± s.d.; one-way ANOVA with Tukey's multiple comparison test). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.



Extended Data Fig. 5 |. *IDH2* mutations augment the RNA splicing defects of *SRSF2* mutant leukemia.

a-c, Venn diagram showing numbers of differentially spliced events from the Beat-AML cohort (**a**), Unpublished Collaborative Cohort_2 (**b**), and murine Lin^{-}cKit^{+} bone marrow cells at 12 weeks post-pIpC (**c**) based on *IDH2/SRSF2* mutant genotypes. **d**, Venn diagram showing the numbers of overlapping alternatively spliced events between *IDH2/SRSF2* double-mutant AMLs and mouse models (***P = 2.2e-16; binominal test). **e-g**, Δ|PSI| (Δ|PSI| = |PSI|$_{Double}$ - |PSI|$_{SRSF2}$) values for each overlapping mis-spliced event in *SRSF2* single-mutant and *IDH2/SRSF2* double-mutant AML from the TCGA (**e**), Beat-AML cohort (**f**) and Unpublished Collaborative Cohort_2 (**g**) are ranked by y-axis. Spliced events shown

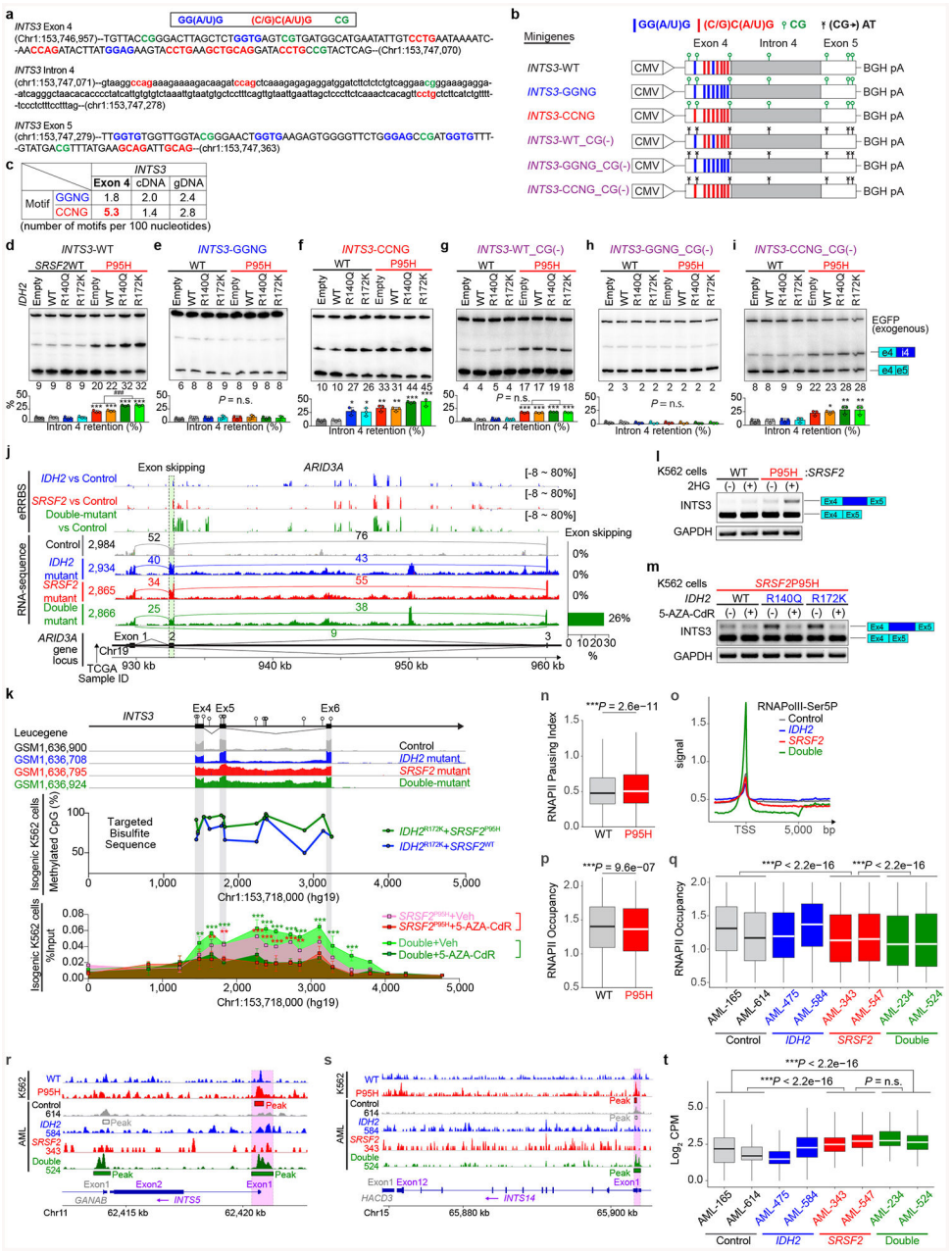
in green and red represent events that are more robust in *IDH2/SRSF2* double-mutant and *SRSF2* single-mutant AML, respectively, in terms of |PSI| values. The mean |PSI| value of each event was visualized as violin plots on the bottom (n = 292, n = 1,741, and n = 187, respectively; PSI values were calculated using PSI-Sigma; the mean value is represented by the thick white line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn down to the 2.5 and 97.5 percentiles; samples below 2.5 percentile and above 97.5 percentile are shown as plots; paired two-tailed Student t-test). **h**, **i**, Venn diagram of numbers of differentially spliced events from the TCGA (**h**) and Beat-AML (**i**) datasets based on *IDH2/TET2/SRSF2* genotypes. **j**, **k**, Absolute numbers of each class of alternative splicing event from TCGA (**j**) and Beat-AML (**k**) datasets are shown (SES; single-exon skipping, MES; multiple-exon skipping, MXS; mutually-exclusive splicing, A5SS; alternative 5' splice site, A3SS; alternative 3' splice site, IR; intron retention). **l**, **m**, Differentially spliced events (|ΔPSI| > 10% and P < 0.01 were used as thresholds) in indicated genotype from the TCGA (**l**) (n = 730 differentially spliced events) and Beat-AML (**m**) (n = 1,339 differentially spliced events) cohorts are ranked by y-axis and class of event (PSI and P-values adjusted for multiple comparisons were calculated using PSI-Sigma; e5: exon 5; i4/5: intron 4/5). **n-p**, Sequence logos of nucleotide motifs of exons preferentially promoted or repressed in splicing in *SRSF2* single-mutant (top) or *IDH2/SRSF2* double-mutant (bottom) AML from the TCGA cohort (**n**), Beat-AML cohort (**o**), and mouse models (**p**). **q**, Percentage of each class of alternative splicing event in indicated genotype from TCGA cohort is shown in pie-chart. **r-t**, Differentially spliced events (|ΔPSI| > 10% and P < 0.01 were used as thresholds) in indicated genotype from the Beat-AML (**r**) (n = 2,183, 5,648, and 79 differentially spliced events, respectively), Unpublished Collaborative Cohort_2 (**s**) (n = 558, 1,926, and 94 differentially spliced events, respectively), and Leucegene cohort (**t**) (n = 2,571, 787, and 122 differentially spliced events, respectively) are ranked by y-axis and class of event (PSI and P-values adjusted for multiple comparisons were calculated using PSI-Sigma). **u**, **w**, Representative Sashimi plots of RNA-seq data showing the intron retention events in *REC8* (**u**) and *PHF6* (**q**) from the TCGA dataset. **v**, **x**, PSI values for intron retention events in *REC8* (**v**) and *PHF6* (**x**) in normal PBMNCs (GSE58335[41]), BMMNCs (GSE61410[41]), cord blood CD34^{+} cells (GSE48846[42]), and AML samples with indicated genotypes (the median value is represented by the thick line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn down to the 2.5 and 97.5 percentiles; samples below 2.5 percentile and above 97.5 percentile are shown as plots; PSI and P-values adjusted for multiple comparisons were calculated using PSI-Sigma; one-way ANOVA with Tukey's multiple comparison test; *P < 0.05; **P < 0.01; ***P < 0.001). **y**, Volcano plots of aberrant splicing events in TCGA AML data comparing *SRSF2* single-mutant and *IDH2/SRSF2* double-mutant AML (n = 122 differentially spliced events; PSI and P-values adjusted for multiple comparisons were calculated using PSI-Sigma; |ΔPSI| > 10% and P < 0.01 were used as thresholds).



Extended Data Fig. 6 |. Aberrant *INTS3* transcripts undergo nonsense-mediated decay and impact of INTS3 loss extends to other members of the Integrator complex.

a, Representative Sashimi plots of RNA-seq data from the TCGA showing intron retention in *INTS3*. **b**, **c**, PSI values for *INTS3* exon 5 skipping (**b**) and intron 4 retention (**c**) in normal PBMNC (GSE58335[41]), BMMNC (GSE61410[41]), cord blood CD34^{+} cells (GSE48846[42]), and AML samples with indicated genotypes (the number of RNA-seq samples analyzed is indicated; PSI and *P*-values adjusted for multiple comparisons were calculated using PSI-Sigma; the mean value is represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn to 2.5 and 97.5 percentiles; samples below 2.5 percentile and above 97.5 percentile are shown as plots; one-way ANOVA with Tukey's multiple comparison test). **d**, Sanger sequencing of cDNA showing WT or mutant *SRSF2* expression in isogenic K562 knock-in cells ($^{\#}$ a nonsynonymous mutation that alters P95. $^{\#\#}$ a synonymous mutation that does not change the amino acid). **e**, RT-PCR and WB analysis of INTS3 in isogeneic HL-60 cells with various combinations of *IDH2/SRSF2* mutations (IR: intron retention; ES: exon skipping;

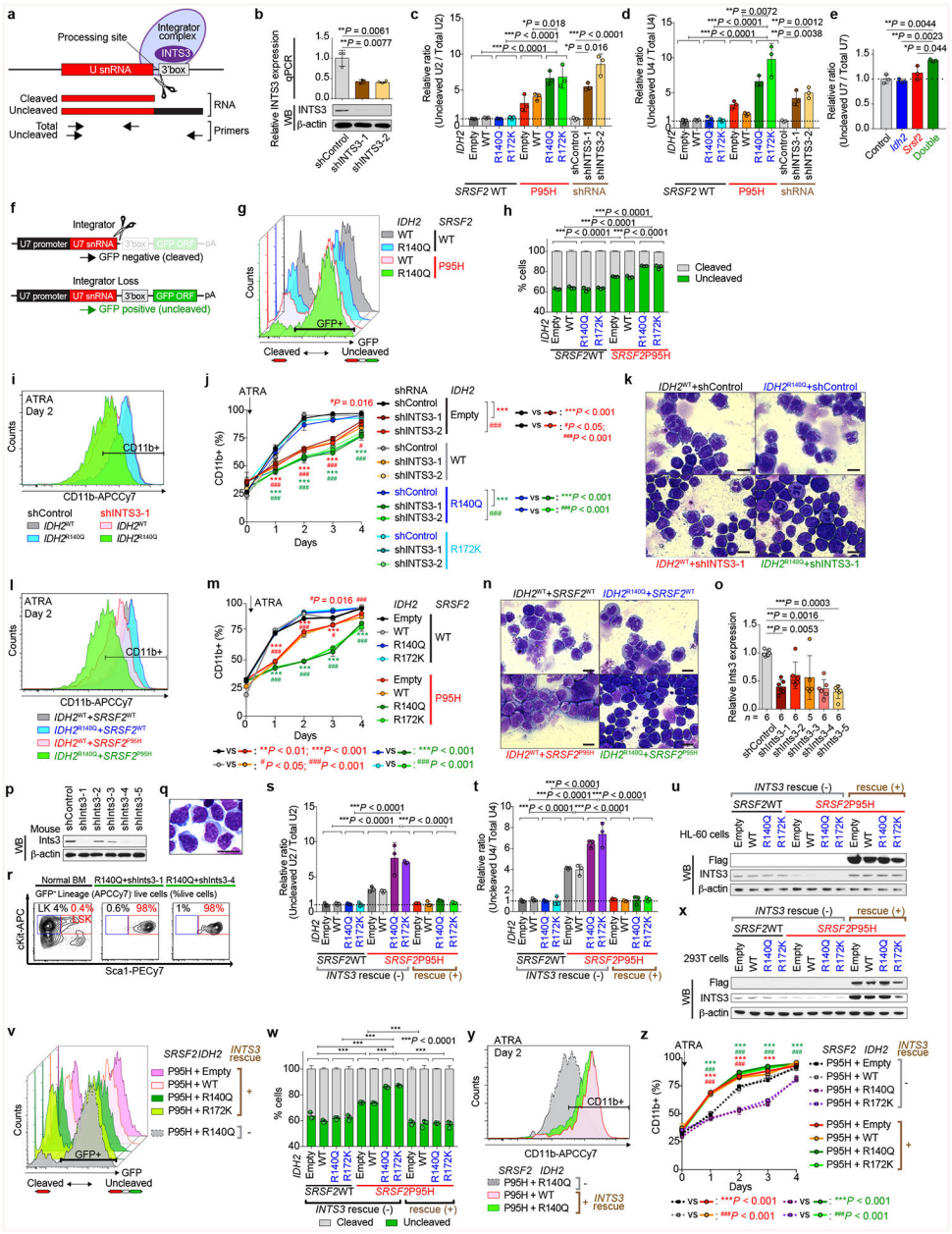
representative results from three biologically independent experiments with similar results). **f**, RT-PCR and WB of INTS3 in non-isogenic myeloid leukemia cell lines. *SRSF2* genotypes are shown together (representative results from three independent experiments with similar results). **g**, WB analysis of K562 *SRSF2*^{P95H} knock-in cells transduced with shRNAs against UPF1 (representative results from three biologically independent experiments with similar results). **h**, Primers used to specifically measure *INTS3* isoform with intron 4 retention and exon 5 skipping, and those for the normal *INTS3* isoform. **i**, **j**, Half-life of *INTS3* transcripts with exon 5 skipping (**i**) and intron 4 retention (**j**) were measured by qPCR ($n = 3$; the mean ± s.d.; a two-sided Student's t-test). **k**, **l**, WB analysis of protein lysates from AML patient samples with indicated *IDH2/SRSF2* genotypes (**k**). Expression level of each Integrator subunit was quantified using ImageJ and relative expression levels are shown in **l** where the mean expression levels of control samples were set as 1 ($n = 6$ for control, *IDH2* single-mutant, and *SRSF2* single-mutant AML, and $n = 7$ for *IDH2/SRSF2* double-mutant AML; Detailed information of the primary patient samples used for this analysis is provided in Supplementary Table 23; the mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **m**, WB analysis of protein lysates from isogenic K562 cells with indicated *IDH2/SRSF2* genotypes (left) or with INTS3 knockdown (right) (representative results from three biologically independent experiments are shown). **n**, WB analysis of murine Lin$^-$cKit$^+$ BM cells at 12 weeks post-pIpC based on *Idh2/Srsf2* mutant genotypes. (expression level of Ints3 was quantified using ImageJ and relative expression levels are shown below; $n = 2$ animals per genotype were analyzed). **o**, Correlation among indicated Integrator subunits and *P*-value were calculated in Excel(15.40) and R^2 values are visualized as a Heatmap generated by Prism 7 (top). Correlation between INTS3 and INTS9 protein expression is shown (bottom) ($n = 25$ from **k**; the Pearson correlation coefficient (R^2) and *P*-values (two-tailed) were calculated in Excel(15.40)). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.



Extended Data Fig. 7 |. DNA hypermethylation at *INTS3* enhances *INTS3* mis-splicing, which is associated with RNA polymerase II (RNAPII) stalling.

a, Sequence of human *INTS3* exon 4, intron 4 and exon 5, and schematic of INTS3 minigene constructs. GG(A/U)G motifs, (C/G)C(A/U)G motifs, and CG dinucleotides are highlighted in blue, red, and green, respectively. **b**, Schematic of *INTS3* minigene constructs. **c**, Table revealing the number of GGNG or CCNG motifs in exon 4, entire cDNA of *INTS3*, or entire genomic DNA (gDNA) of *INTS3* per 100 nucleotides. **d-i**, Radioactive RT-PCR results of INTS3 minigene assays using indicated versions of the minigene in isogenic K562 cells. Percentage of intron 4 retention were normalized against exogenous EGFP ($n = 3$; the mean percentage ± s.d.; one-way ANOVA with Tukey's multiple

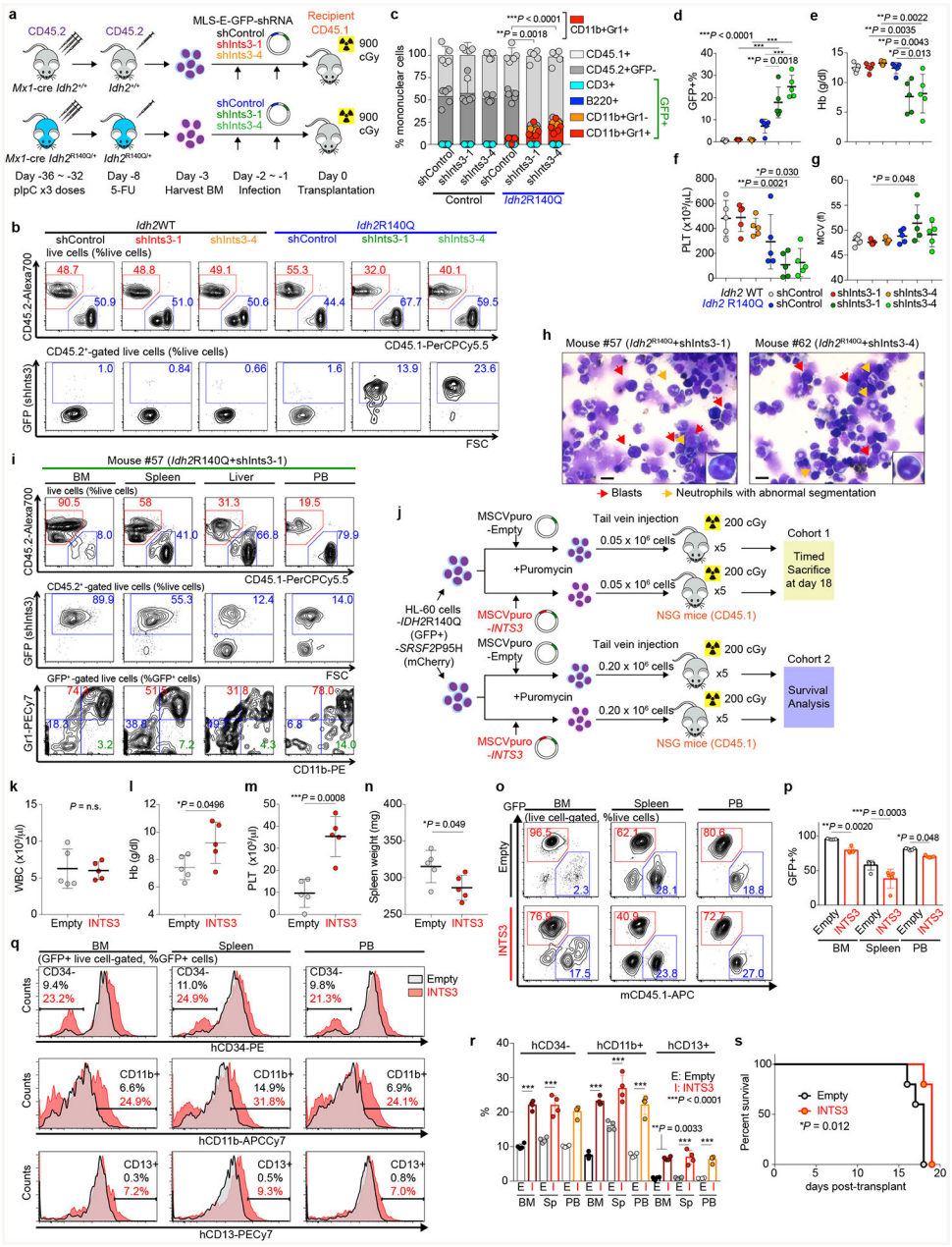
comparison test). **j**, Mean percentage of methylated CpGs at *ARID3A* in AML patient samples with indicated genotypes determined by enhanced reduced representation bisulfite sequence (eRRBS) (n = 3 patients per genotype), followed by IGV plots of RNA-seq data of *ARID3A* from the TCGA. **k**, Results of targeted bisulfite sequence (n = 1 per genotype) and RNAPII-Ser2P ChIP-walking experiments are represented as shown in Fig. 3f (n = 3; the mean percentage ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **l**, **m**, RT-PCR results detecting *INTS3* intron retention in isogenic K562 cells harboring various combinations of *IDH2* and *SRSF2* mutations that were treated with cell-permeable 2HG at 0.5 μM (**l**) or 5-AZA-CdR at 5 μM (**m**) for 8 days (representative results from three biologically independent experiments with similar results). **n**, RNAII pausing index in isogenic $SRSF2^{WT}$ or $SRSF2^{P95H}$ mutant K562 cells was calculated as previously described[20] as a ratio of normalized ChIP-Seq reads of RNAPII-Ser5P on TSSs (+/− 250 bp) over that of the corresponding bodies (+500 to +1000 from TSSs) (the median value is represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn down to the 2.5 and 97.5 percentiles; each box plot was made by analyzing ChIP-seq data from one cell line; two-sided Student's t-test). **o**, Metagene plots showing genome-wide RNAPII-Ser5P occupancy in primary AML patient samples with indicated genotypes (TSS: transcription start site; patient samples used for this analysis are described in Supplementary Table 23). **p**, **q**, RNAPII occupancy representing ChIP-Seq reads of RNAPII-Ser2P over gene bodies was calculated for isogenic K562 cells (**p**) and AML samples (**q**) (the median value is represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn down to the 2.5 and 97.5 percentiles; each box plot was made by analyzing ChIP-seq data from one cell line (**p**) or one primary AML sample (**q**); two-sided Student's t-test (**p**) and one-way ANOVA with Tukey's multiple comparison test (**q**)). **r**, **s**, Genome browser view of ChIP-seq signal for RNAPII-Ser5P at *INTS5* (**r**) and *INTS14* (**s**) in isogenic K562 cells with or without *SRSF2* mutation (n = 1) and primary AML samples with indicated genotype (results generated from n = 2 primary AML samples are shown). **t**, RNAPII abundance over the differentially spliced regions between *IDH2/SRSF2* wild-type control and *SRSF2* single-mutant AML determined by RNAPII-Ser2P ChIP-seq (y-axis: Log_2 (Counts per million); the median value is represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn down to the 2.5 and 97.5 percentiles; each box plot was made by analyzing ChIP-seq data from one primary AML sample; one-way ANOVA with Tukey's multiple comparison test). *P < 0.05; **P < 0.01; ***P < 0.001.



Extended Data Fig. 8 |. Loss of INTS3 impairs uridine-rich small nuclear RNA (snRNAs) processing and blocks myeloid differentiation.

a, Schematic of snRNA processing site and qPCR primers for detecting cleaved or uncleaved snRNA. **b**, qPCR (top; $n = 3$; the mean ± s.d.; a two-sided Student's t-test) and representative WB of INTS3 in HL-60 cells transduced with short-hairpin RNAs (shRNAs) targeting human INTS3 (bottom; representative results from three biologically independent experiments). **c**-**e**, **s**, **t**, qPCR results of U2 (**c**, **s**) and U4 (**d**, **t**) snRNAs in isogenic HL-60 cells and U7 snRNA in murine cells from Extended Data Fig. 6n (**e**). Ratio of uncleaved/ total snRNAs expression was compared ($n = 3$, the mean ratio ± s.d.; one-way ANOVA with Tukey's multiple comparison test; the largest *P*-values calculated among 2 × 2 comparisons of two components from different groups are shown. For example, *P*-values were calculated

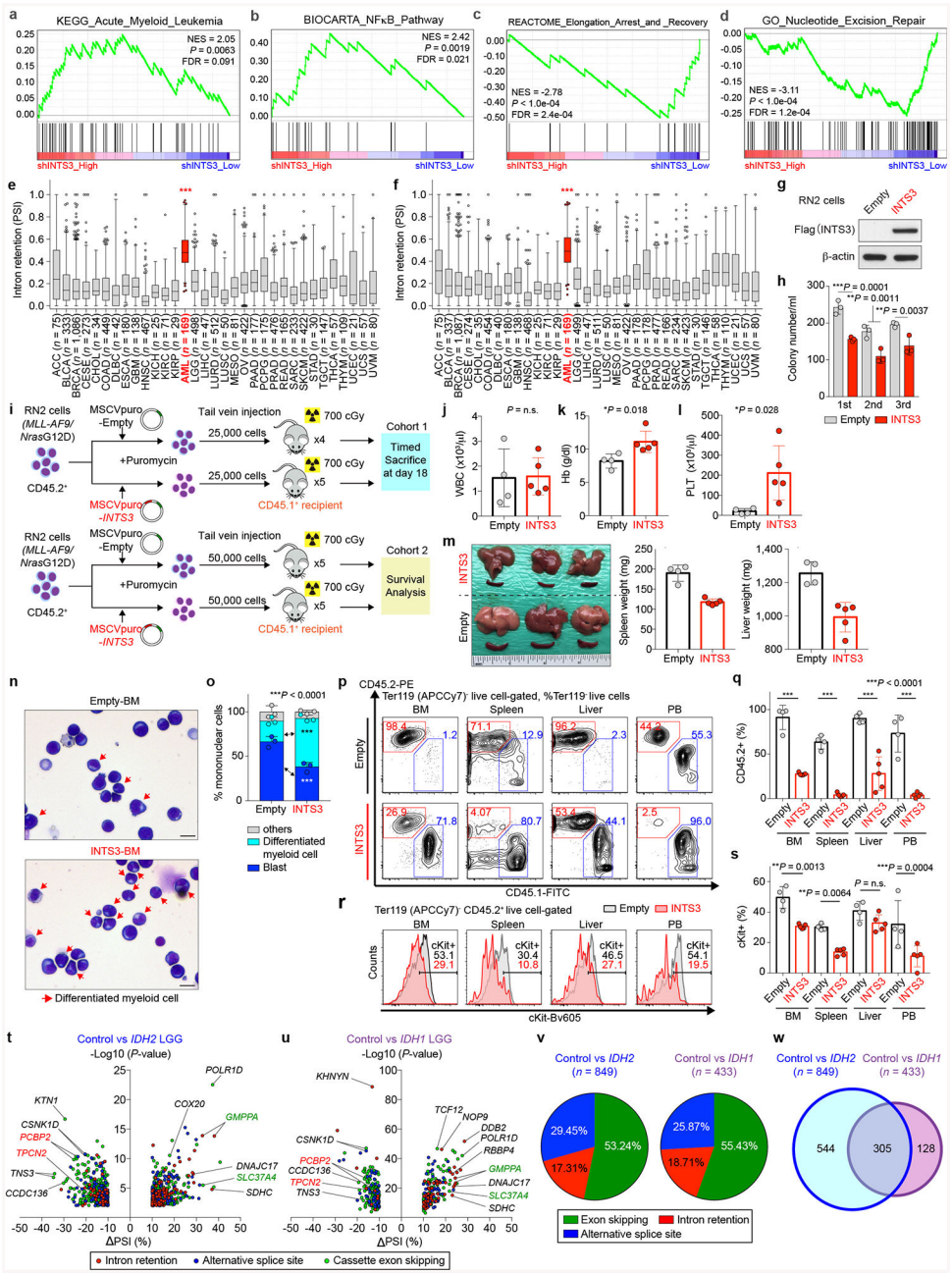
from the following four comparisons; bars 1 vs 3, 2 vs 3, 1 vs 4, 2 vs 4). **f**, Schematic of the U7 snRNA-GFP reporter. **g**, **v**, Flow cytometry analysis of 293T cells transduced with U7 snRNA-GFP reporter and *IDH2*/*SRSF2*/*INTS3* constructs as labeled on the right (representative results from three biologically independent experiments are shown). **h**, **w**, Quantification of % GFP^- and GFP^+ 293T cells ($n = 3$ biologically independent experiments, the mean percentage ± s.d.; one-way ANOVA with Tukey's multiple comparison test; *P*-values are shown as in **c**). **i**, **l**, **y**, Flow cytometry analysis of CD11b expression in isogenic HL-60 cells after ATRA treatment for two days (representative results from three biologically independent experiments are shown). **j**, **m**, **z**, Quantification of percentages of $CD11b^+$ HL-60 cells over time ($n = 3$; the mean percentage ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **k**, **n**, Cytomorphology of isogenic HL-60 cells after ATRA treatment for two days (Giemsa staining; scale bar, 10 μm; original magnification × 400; representative results from three biologically independent experiments are shown). **o**, **p**, qPCR (**o**) (the mean ± s.d.; Kruskal-Wallis tests with uncorrected Dunn's test) and WB (**p**) of Ints3 in Ba/F3 cells transduced with shRNAs targeting mouse Ints3. **q**, **r**, Representative cytomorphology (**q**) and immunophenotype (**r**) of colony cells at the 6th colony. Normal BMMNCs were used as a control (the percentage listed represent the percent of cells within live cells; representative results from three biologically independent experiments are shown). **u**, **x**, WB of proteins extracted from HL-60 cells (**u**) assayed in **s-t** and **y-z** and 293T cells (**x**) assayed in **v-w** (representative results from three biologically independent experiments). $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{\#}P < 0.05$; $^{\#\#}P < 0.01$; $^{\#\#\#}P < 0.001$.



Extended Data Fig. 9 |. Mutant Idh2 cooperates with Ints3 loss to generate a lethal myeloid neoplasm *in vivo*.

a, Schematic of anti-*Ints3* shRNA (shInts3) retroviral BM transplantation model. **b**, Flow cytometry data showing the chimerism of CD45.2$^+$ vs CD45.1$^+$ (top) or GFP$^+$ (bottom) cells in PB at 4 weeks post-transplant (the percentages listed represent the percent of cells within live cells; representative results from five recipient mice). **c**, Composition of PBMNCs at 4 weeks post-transplant (n = 5 per group; the mean + s.d.; represented by lines above the box. statistical significance was detected in % of CD11b$^+$Gr1$^+$ cells; by two-way ANOVA with Tukey's multiple comparison test. **d-g**, Chimerism of GFP$^+$ cells in PB (**d**) and blood counts of recipients at 4 weeks post-transplant (Hb (**e**); PLT (**f**); MCV (**g**); n = 5 per group; the mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **h**, Giemsa staining

of BMMNCs from moribund mice with indicated genotypes (red and yellow arrows represent blastic cells and dysplastic neutrophils, respectively; inset, representative neutrophils with abnormal segmentation; scale bar, 10 μm; original magnification × 400; representative results from five mice per genotype). **i**, Flow cytometry data of BM, spleen, liver, and PB from *Idh2*^{R140Q}+shInts3 mice (representative results from five mice). **j**, Schematic of HL-60 xenograft model where recipient mice from Cohort 1 were sacrificed at day 18 post-transplant and mice from Cohort 2 were observed for survival analysis until end-stage. **k-n**, Blood counts (WBC (**k**); Hb (**l**); PLT (**m**)) and spleen weight (**n**) of mice from Cohort 1 at day 18 post-transplant (the mean ± s.d.; $n = 5$ per group; a two-sided Student's t-test). **o**, **p**, Representative flow cytometry data of BM, spleen, and PB from the recipient mice from Cohort 1 (**o**) (the percentage represents the percent of cells within live cells) and the mean percentage of GFP$^+$ cells (**p**) ($n = 5$ per group; the mean + s.d.; two-way ANOVA with Sidak's multiple comparison test). **q**, **r**, Representative flow cytometry data of BM, spleen, and PB from Cohort 1 (**q**) (the percentage represents the percent of cells within GFP$^+$ live cells) and the mean percentage of hCD34$^-$, hCD11b$^+$, and hCD13$^+$ cells (**r**) ($n = 4$ per group; the mean + s.d.; two-way ANOVA with Sidak's multiple comparison test). **s**, Kaplan-Meier survival analysis of recipient mice from Cohort 2 ($n = 5$ per group; Log-rank (Mantel-Cox) test (two-sided)). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.



Extended Data Fig. 10 |. Gene expression and biological consequences of INTS3 loss and impact of *IDH1/2* mutations on splicing in low-grade glioma.

a-d, Gene set enrichment analysis (GSEA) based on RNA-seq data generated from isogenic *IDH2*^{R140Q} mutant HL-60 cells with or without INTS3 depletion. Representative results from gene sets associated with leukemogenesis and myeloid differentiation (**a**), oncogenic signaling pathways (**b**), RNAPII elongation-linked transcription (**c**), and DNA damage response (**d**) with statistical significance ($P < 0.01$) are shown (y-axis; Enrichment score; NES: Normalized enrichment score; FDR: False discovery rate; RNA-seq data generated from isogenic HL-60 cells in duplicate were analyzed using GSEA[34]). **e**, **f**, PSI values for *INTS3* intron 4 (**e**) and 5 (**f**) retention events across 33 cancer cell types (the same datasets

were analyzed in Fig. 4f; ACC: adrenocortical carcinoma, BLCA: bladder urothelial carcinoma, BRCA: breast invasive carcinoma, CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL: cholangiocarcinoma, DLBC: diffuse large B-cell lymphoma, ESCA: esophageal carcinoma, GBM: glioblastoma mutiforme, HNSC: head and neck squamous cell carcinoma, KICH: kidney chromophobe, KIRC: kidney renal clear cell carcinoma, KIRP: kidney renal papillary cell carcinoma, LGG: low-grade glioma, LIHC: liver hepatocellular carcinoma, LUSC: lung squamous cell carcinoma, MESO: mesothelioma, OV: ovarian serous cystadenocarcinoma, PRAD: prostate adenocarcinoma, READ: rectum adenocarcinoma, SARC: sarcoma, SKCM: skin cutaneous melanoma, STAD: stomach adenocarcinoma, TGCT: testicular germ cell tumors, THCA: thyroid carcinoma, THYM: thymoma, UCEC: uterine corpus endometrial carcinoma, UCS: uterine carcinosarcoma, UVM: uveal melanoma; the median value is represented by the line inside the box and the box expands from the 25^{th} to 75^{th} percentiles with whiskers drawn down to the 2.5 and 97.5 percentiles; samples below 2.5 percentile and above 97.5 percentile are shown as plots; one-way ANOVA with Dunnett's multiple comparison test; $^{***}P < 0.001$ represents the *P*-values from all the comparisons between AML and any of other 32 non-AML cancer type). **g**, WB analysis confirming overexpression of 3× Flag-tagged INTS3 in RN2 (*MLL-AF9/Nras*^{G12D}) leukemia cells (representative results from three biologically independent experiments). **h**, Colony numbers from serial replating assays of RN2 cells with or without INTS3 overexpression ($n = 3$; the mean + s.d. represented by lines above the box; two-way ANOVA with Sidak's multiple comparison test). **i**, Schematic of INTS3 retroviral BM transplantation models where recipient mice from Cohort 1 were sacrificed at day 18 post-transplant and mice from Cohort 2 were observed for survival analysis until end-stage. **j-l**, Blood counts (WBC (**j**); Hb (**k**); PLT (**l**)) of mice from Cohort 1 at day 18 post-transplant (the mean ± s.d.; $n = 4$ ("Empty" group); $n = 5$ ("INTS3" group) recipient mice; a two-sided Student's t-test). **m**, Representative photograph of spleens and livers from Cohort 1 with an inch scale (left), and spleen (middle) and liver weight (right) ($n = 4$ (Empty); $n = 5$ (INTS3); the mean ± s.d.; two-sided Student's t-test). **n**, **o**, Representative Giemsa staining (**n**) (red arrows represent differentiated cells; scale bar, 10 μm; original magnification × 400) and percentages of blasts, differentiated myeloid cells, and other cells in BMMNCs (**o**) from moribund mice from Cohort 2 ($n = 3$ per genotype; 100 cells per mouse were classified; the mean percentage + s.d.; two-way ANOVA with Sidak's multiple comparison test). **p**, **q**, Representative flow cytometry analysis of BM, spleen, liver, and PB (**p**) and percentages of CD45.2$^+$ cells in Ter119$^-$ live cells (**q**) in recipient from Cohort 1 ($n = 4$ (Empty); $n = 5$ (INTS3); the mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **r**, **s**, Representative flow cytometry analysis showing cKit expression in RN2 cells with or without INTS3 overexpression (**r**) and quantification of cKit$^+$ cells (**s**) from Cohort 1 ($n = 4$ (Empty); $n = 5$ (INTS3); the mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **t**, **u**, Volcano plots of aberrant splicing events in the LGG TCGA dataset based on *IDH2* (**t**) or *IDH1* (**u**) mutant genotypes. $|\Delta PSI| > 10\%$ and $P < 0.01$ were used as thresholds ($n = 849$ and $n = 433$ differentially spliced events, respectively; RNA-seq data were analyzed using PSI-Sigma). **v**, Percentage of each class of alternative splicing event in *IDH2* (left) and *IDH1* (right) mutant LGG is shown in pie-chart. **w**, Venn diagram of numbers of alternatively spliced events from the LGG TCGA dataset based on *IDH1/IDH2*

mutant genotypes. “Control” represents LGG with wild-type *IDH1* and *IDH2*. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dennis Liang Fei, Yun (Nancy) Huang, Eric Wang, Iannis Aifantis, Minal Patel, Alan S. Shih, Alex Penson, Eunhee Kim, Young Rock Chung, Benjamin H. Durham, and Hiroyoshi Kunimoto for technical support, Jeremy Wilusz for sharing his recent data on Integrator, and Brian J. Druker for sharing the Beat-AML RNA-seq data. A.Y. is supported by grants from the Aplastic Anemia and MDS International Foundation (AA&MDSIF) and the Lauri Strauss Leukemia Foundation. A.Y. is a Special Fellow of The Leukemia and Lymphoma Society. A.Y., S.C.-W.L., and D.I. are supported by the Leukemia and Lymphoma Society Special Fellow Award. A.Y. and D.I. are supported by JSPS Overseas Research Fellowships. D.H.W. is supported by a Bloodwise Clinician Scientist Fellowship (15030). D.H.W. and K.B. are supported by fellowships from The Oglesby Charitable Trust. S.C.-W.L. is supported by the NIH/NCI (K99 CA218896) and the ASH Scholar Award. T.C.P.S. is supported by Cancer Research UK grant number C5759/A20971. E.J.W. is supported by grants from the CPRIT (RP140800) and the Welch Foundation (H-1889-20150801). R.K.B. and O.A.-W. are supported by grants from NIH/NHLBI (R01 HL128239) and the Dept. of Defense Bone Marrow Failure Research Program (W81XWH-16-1-0059). A.R.K. and O.A.-W. are supported by grants from the Starr Foundation (I8-A8-075) and the Henry & Marilyn Taub Foundation. O.A.-W. is supported by grants from the Edward P. Evans Foundation, the Josie Robertson Investigator Program, the Leukemia and Lymphoma Society, and the Pershing Square Sohn Cancer Research Alliance.

REFERENCES

1. Cancer Genome Atlas Research, N. et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 368, 2059–2074 (2013). [PubMed: 23634996]
2. Papaemmanuil E et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122, 3616–3627 (2013). [PubMed: 24030381]
3. Wu Y, Albrecht TR, Baillat D, Wagner EJ & Tong L Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. Proc Natl Acad Sci U S A 114, 4394–4399 (2017). [PubMed: 28396433]
4. Darman RB et al. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3’ Splice Site Selection through Use of a Different Branch Point. Cell Rep 13, 1033–1045 (2015). [PubMed: 26565915]
5. Ilagan JO et al. U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res 25, 14–26 (2015). [PubMed: 25267526]
6. Kim E et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell 27, 617–630 (2015). [PubMed: 25965569]
7. Zhang J et al. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc Natl Acad Sci U S A 112, E4726–4734 (2015). [PubMed: 26261309]
8. Tyner JW et al. Functional genomic landscape of acute myeloid leukaemia. Nature 562, 526–531 (2018). [PubMed: 30333627]
9. Lavallee VP et al. The transcriptomic landscape and directed chemical interrogation of MLL-rearranged acute myeloid leukemias. Nat Genet 47, 1030–1037 (2015). [PubMed: 26237430]
10. Dang L et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 465, 966(2010). [PubMed: 20559394]
11. Figueroa ME et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell 18, 553–567 (2010). [PubMed: 21130701]
12. Jia G et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 7, 885–887 (2011). [PubMed: 22002720]

13. Zheng G et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 49, 18–29 (2013). [PubMed: 23177736]
14. Naftelberg S, Schor IE, Ast G & Kornblihtt AR Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu Rev Biochem 84, 165–198 (2015). [PubMed: 26034889]
15. 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 31, 162–174 (2012). [PubMed: 22002536]
16. Shukla S et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74–79 (2011). [PubMed: 21964334]
17. Gardini A et al. Integrator regulates transcriptional initiation and pause release following activation. Mol Cell 56, 128–139 (2014). [PubMed: 25201415]
18. Huang J, Gong Z, Ghosal G & Chen J SOSS complexes participate in the maintenance of genomic stability. Mol Cell 35, 384–393 (2009). [PubMed: 19683501]
19. Li Y et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. J Biol Chem 284, 23525–23531 (2009). [PubMed: 19605351]
20. Stadelmayer B et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nat Commun 5, 5531(2014). [PubMed: 25410209]
21. Ji X et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell 153, 855–868 (2013). [PubMed: 23663783]
22. Chen L et al. The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. Mol Cell 69, 412–425 e416 (2018). [PubMed: 29395063]
23. Seiler M et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nat Med 24, 497–504 (2018). [PubMed: 29457796]
24. Stein EM et al. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. Blood 130, 722–731 (2017). [PubMed: 28588020]
25. Lin KT & Krainer AR PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. Bioinformatics (2019).

Online-only references

26. Moran-Crusio K et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 20, 11–24 (2011). [PubMed: 21723200]
27. Shih AH et al. Combination Targeted Therapy to Disrupt Aberrant Oncogenic Signaling and Reverse Epigenetic Dysfunction in IDH2- and TET2-Mutant Acute Myeloid Leukemia. Cancer Discov 7, 494–505 (2017). [PubMed: 28193779]
28. Georgiades P et al. VavCre transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. Genesis 34, 251–256 (2002). [PubMed: 12434335]
29. Zuber J et al. Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. Nat Biotechnol 29, 79–83 (2011). [PubMed: 21131983]
30. Lee M et al. Engineered Split-TET2 Enzyme for Inducible Epigenetic Remodeling. J Am Chem Soc 139, 4659–4662 (2017). [PubMed: 28294608]
31. Kleppe M et al. Dual Targeting of Oncogenic Activation and Inflammatory Signaling Increases Therapeutic Efficacy in Myeloproliferative Neoplasms. Cancer Cell 33, 29–43 e27 (2018). [PubMed: 29249691]
32. Maiques-Diaz A et al. Enhancer Activation by Pharmacologic Displacement of LSD1 from GFI1 Induces Differentiation in Acute Myeloid Leukemia. Cell Rep 22, 3641–3659 (2018). [PubMed: 29590629]
33. Cheng DT et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J Mol Diagn 17, 251–264 (2015). [PubMed: 25801821]

34. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550 (2005). [PubMed: 16199517]

35. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

36. Dvinge H & Bradley RK Widespread intron retention diversifies most cancer transcriptomes. Genome Med 7, 45(2015). [PubMed: 26113877]

37. Hubert CG et al. Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. Genes Dev 27, 1032–1045 (2013). [PubMed: 23651857]

38. Bailey TL et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37, W202–208 (2009). [PubMed: 19458158]

39. Robinson JT et al. Integrative genomics viewer. Nat Biotechnol 29, 24–26 (2011). [PubMed: 21221095]

40. Intlekofer AM et al. Hypoxia Induces Production of L-2-Hydroxyglutarate. Cell Metab 22, 304–311 (2015). [PubMed: 26212717]

41. Dvinge H et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. Proc Natl Acad Sci U S A 111, 16802–16807 (2014). [PubMed: 25385641]

42. Macrae T et al. RNA-Seq reveals spliceosome and proteasome genes as most consistent transcripts in human cancer cells. PLoS One 8, e72884(2013). [PubMed: 24069164]

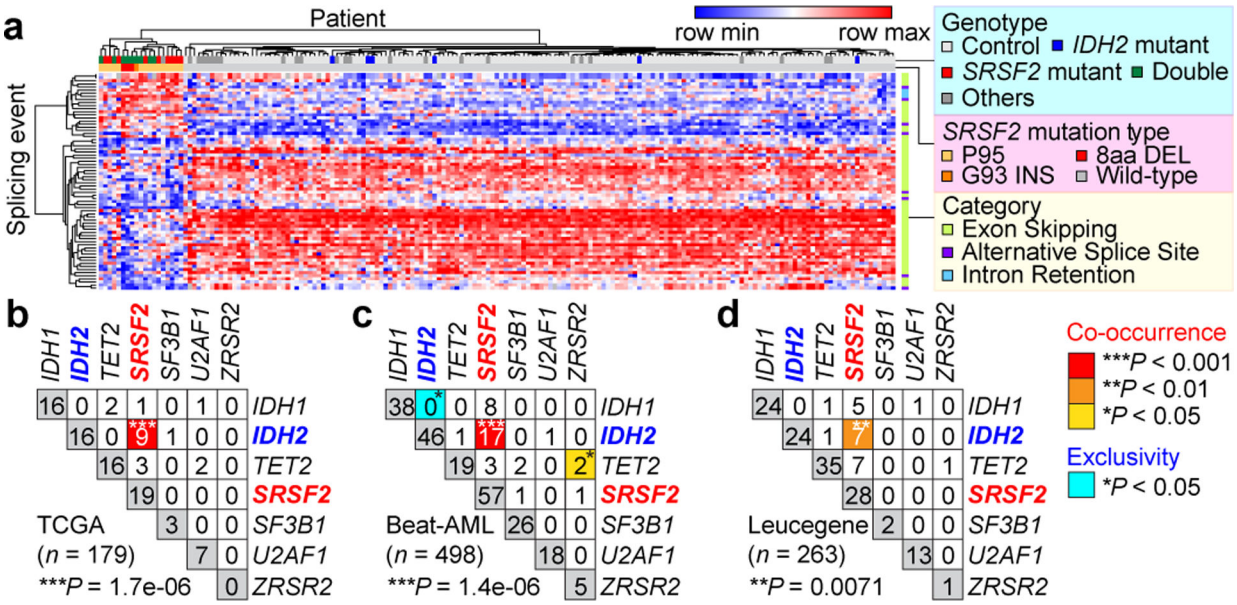


Fig. 1 |. Frequent co-existing *IDH2* and *SRSF2* mutations in acute myeloid leukemia (AML).
a, Heatmap of ΔPSI (Percent-Spliced-In) values for mutant *SRSF2*-specific splicing events in TCGA AML samples. **b-d**, Co-occurrence of mutations in *IDH1/2*, *TET2*, and RNA splicing factors in the TCGA (**b**), Beat-AML (**c**), and Leucegene (**d**) cohorts (number of patients indicated; co-occurrence or exclusivity noted by color-coding; Fisher's exact test (two-sided)).

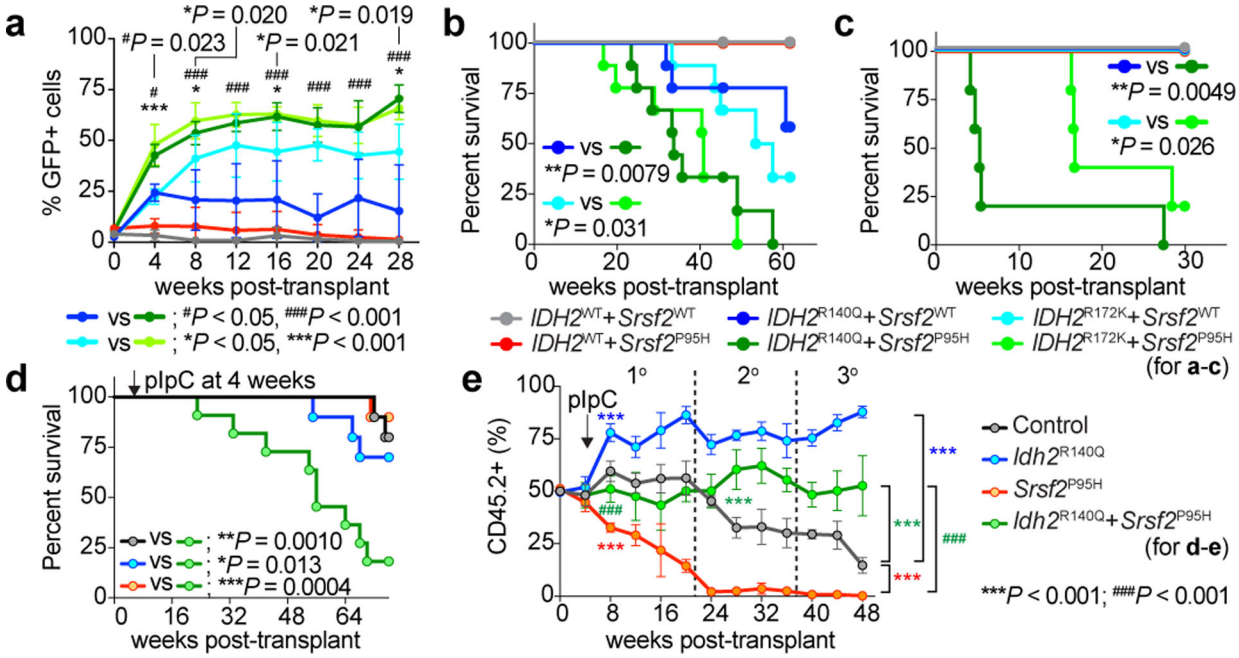


Fig. 2 |. Mutant *IDH2* cooperates with mutant *Srsf2* to promote leukemogenesis.

a, Chimerism of GFP$^+$ cells in the blood of recipients over time (n = 5 per group; data at 0 week represent transduction efficiency; the mean percentage ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **b**-**d**, Kaplan-Meier survival analysis of primary recipients (**b**) (n = 10 mice per genotype), recipients of serial transplant (**c**) (n = 5), and primary recipients transplanted non-competitively with BM cells from knock-in mice (**d**) (n = 10) (Log-rank (Mantel-Cox) test (two-sided)). **e**, Chimerism of PB CD45.2$^+$ cells in competitive transplantation (n = 10 mice per group; the mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test).

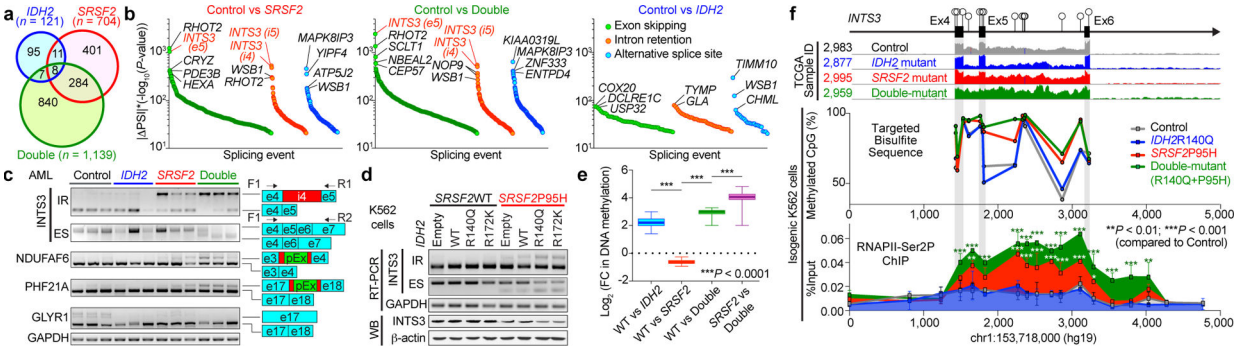


Fig. 3 |. Collaborative effects of mutant *IDH2* and *SRSF2* on aberrant splicing.

a, Venn diagram showing numbers of differentially spliced events from TCGA AML samples. **b**, Differentially spliced events (|ΔPSI| > 10% and $P < 0.01$) in indicated genotype are ranked by y-axis ((|ΔPSI × (−Log_{10}(*P*-value)) and class of event (e5: exon 5; i4/5: intron 4/5) (PSI and *P*-values adjusted for multiple comparisons were calculated using PSI-Sigma[25]). **c**, Representative RT-PCR results of aberrantly spliced transcripts in AML patient samples (pEx: exon with premature stop-codon; $n = 3$ patients per genotype; three technical replicates with similar results). **d**, RT-PCR and WB of INTS3 in isogenic K562 cells (representative images from three biologically independent experiments with similar results). **e**, Mean log_2 fold-change in DNA cytosine methylation (y-axis) at regions of genomic DNA encoding mRNA which undergo differential splicing (x-axis). DNA methylation levels were determined by eRRBS ($n = 3$ per genotype; the mean represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn down to the 2.5 and 97.5 percentiles; one-way ANOVA with Tukey's multiple comparison test; ***$P < 2.2e-16$). **f**, Diagram of the genomic locus of *INTS3* around exons 4–6 with CpG dinucleotides, representative RNA-seq from four AML patients, targeted bisulfite sequencing ($n = 1$ per genotype), and results of anti-RNAPII-Ser2P ChIP-walking experiments ($n = 3$; the mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

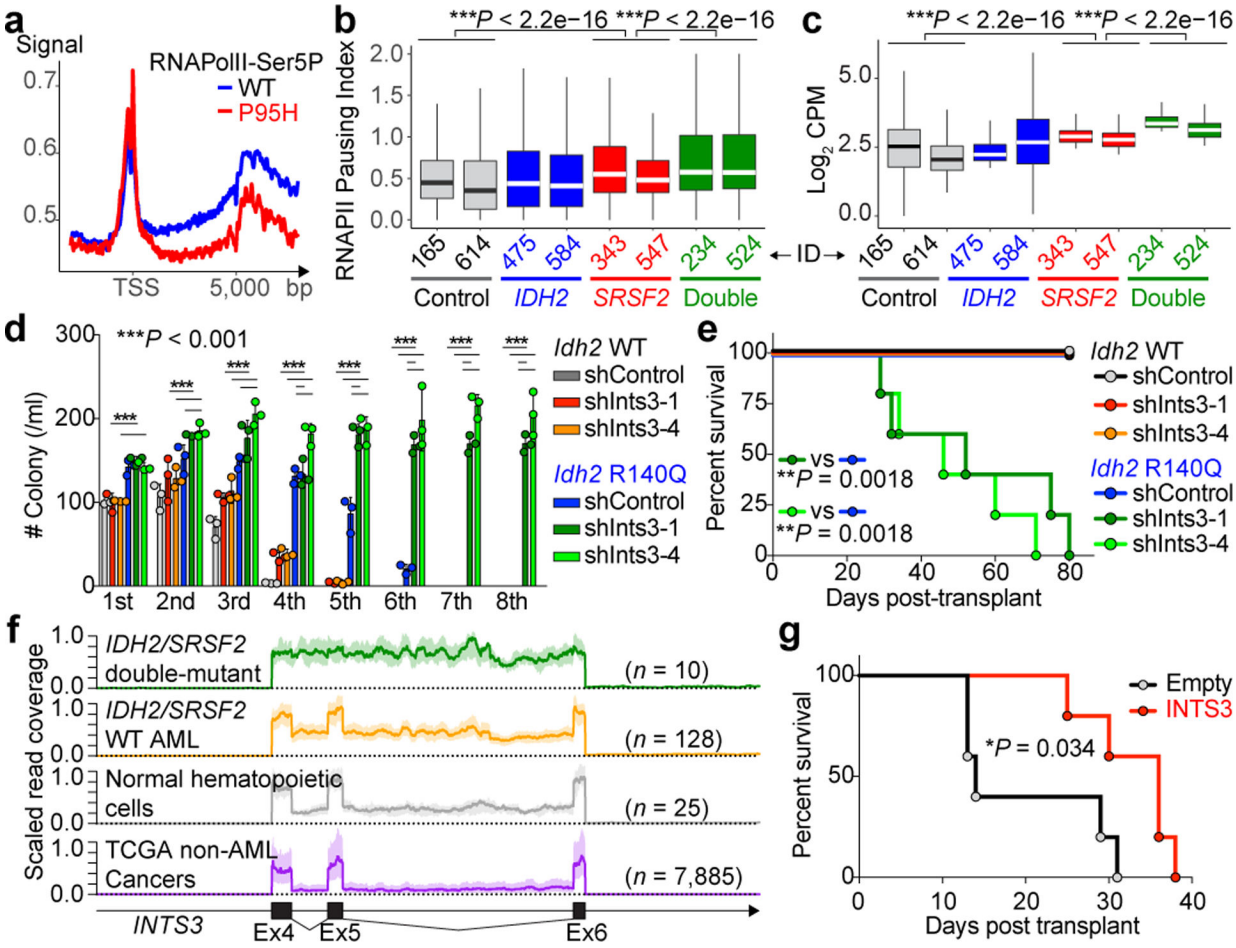


Fig. 4 |. RNAPII stalling in *IDH2/SRSF2* double-mutant AML and contribution of INTS3 loss to leukemogenesis.

a, Metagene plot of genome-wide RNAPII-Ser5P occupancy in isogenic *SRSF2*^{WT} or *SRSF2*^{P95H} mutant cells. **b**, **c**, RNAPII pausing index[20] in primary AML samples calculated as the ratio of normalized ChIP-Seq reads of RNAPII-Ser5P on TSSs (± 250 bp) over that of the corresponding bodies (+500 to +1000 from TSSs) (**b**) and RNAPII abundance over the differentially spliced regions between *SRSF2* single-mutant and *IDH2/SRSF2* double-mutant AML determined by RNAPII-Ser2P ChIP-seq (y-axis: Log_2 (Counts per million)) (**c**) (x-axis: patient ID; each box plot was generated based on ChIP-seq data from an individual primary AML sample; the mean is represented by the line inside the box and the box expands from the 25th to 75th percentiles with whiskers drawn to 2.5 and 97.5 percentiles; one-way ANOVA with Tukey's multiple comparison test). **d**, Colony numbers from serial replating assays of either *Mx1*-cre *Idh2*$^{+/+}$ or *Idh2*$^{R140Q/+}$ BM cells transduced with shRNA against Ints3 (n = 3 biologically independent experiments; the mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **e**, **g**, Kaplan-Meier survival analysis of recipients (n = 5 per group; Log-rank (Mantel-Cox) test (two-sided)). **f**, RNA-seq read coverage between exons 4–6 of *INTS3* ± 1,000 bp of *INTS3* is scaled and shown as mean

(thick line) ± s.d. (light color) (generated from TCGA datasets; sample list noted in legend for Extended Data Fig. 10e).