LETTER

https://doi.org/10.1038/s41586-019-1618-0

Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis

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], 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]*

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 leukaemia[1,2]. Despite the frequent overlap of mutations affecting epigenetic regulation and splicing in leukaemia, how these processes influence one another to promote leukaemogenesis is not understood and, to our knowledge, there is no functional evidence that mutations in RNA splicing factors initiate leukaemia. Here, through analyses of transcriptomes from 982 patients with acute myeloid leukaemia, we identified frequent overlap of mutations in *IDH2* and *SRSF2* that together promote leukaemogenesis through coordinated effects on the epigenome and RNA splicing. Whereas 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* and *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 leukaemogenesis 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 crosstalk between altered epigenetic state and splicing in a subset of leukaemias, provide functional evidence that mutations in splicing factors drive myeloid malignancy development, and identify spliceosomal changes as a mediator of *IDH2*-mutant leukaemogenesis.

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 binding to RNA in a sequence-specific manner[6,7]. We analysed RNA-seq data in The Cancer Genome Atlas (TCGA)[1] from 179 patients with acute myeloid leukaemia (AML), evaluating them for spliceosomal alterations. Aberrant splicing events characteristic of *SRSF2* mutations, including *EZH2*[6,7] poison exon inclusion, were observed in 19 patients ($P = 1.6 \times 10^{-12}$; Fisher's exact test; Fig. 1a, Extended Data Fig. 1a, b, Supplementary Table 1). Although only one patient with a mutation in *SRSF2* was reported in the TCGA AML publication[1], mutational analysis of RNA-seq data identified *SRSF2* hotspot mutations in each of these 19 patients (11% of the patients with AML). Therefore, these data retrospectively identify *SRSF2* as one of the most commonly mutated genes in the TCGA AML cohort.

Notably, 47% of patients with mutated *SRSF2* also had a mutation in *IDH2* and conversely, 56% of patients with mutated *IDH2* had a mutation in *SRSF2* ($P = 1.7 \times 10^{-6}$; Fisher's exact test; Fig. 1b, Extended Data Fig. 1c, d, Supplementary Table 2). Similar results were seen in RNA-seq data from 498 and 263 patients with AML from the Beat AML[8] and Leucegene[9] studies, respectively (Fig. 1c, d, Extended Data Fig. 1e–j, 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 patients with AML in clinical practice (Supplementary Table 3). Although not statistically significant, *IDH2* and *SRSF2* double-mutant AML cases had the shortest overall survival across the four studied genotypes (Extended Data Fig. 2a). Whereas patients with *IDH2* and *SRSF2* double mutations had mostly intermediate cytogenetic risk, their prognosis was comparable to those with adverse cytogenetic risk (Extended Data Fig. 2b). The patients with *IDH2* and *SRSF2* double mutations were also significantly older than those with mutations in *IDH2* only, or with wild-type *IDH2* and *SRSF2* (Extended Data Fig. 2d; clinical and genetic features are summarized in Extended Data Fig. 2 and Supplementary Table 3).

Mutations in *IDH2* confer neomorphic enzymatic activity that results in the generation of 2-hydroxyglutarate (2HG)[10], which in turn induces DNA hypermethylation via the competitive inhibition of the α-ketoglutarate-dependent enzymes TET1–TET3. Unsupervised hierarchical clustering of DNA methylation data from the TCGA AML cohort revealed that *IDH2* and *SRSF2* double-mutant AML cases form a distinct cluster with higher DNA methylation than *IDH2* single-mutant AML cases (Extended Data Fig. 1l-o). Collectively, these data identify *IDH2* and *SRSF2* double-mutant leukaemia 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 frequently found in leukaemias, there is no functional evidence that they can transform cells in vivo. Overexpression of human $IDH2^{R140Q}$ or $IDH2^{R172K}$ 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 after transplantation, the peripheral blood of recipient mice transplanted with *IDH2* and

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 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, USA. [7]Department of Biochemistry & Molecular Biology, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA. [8]Manchester Cancer Research Centre Biobank, The University of Manchester, Manchester, UK. [9]Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds, UK. [10]Department of Biomolecular Chemistry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA. [11]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [12]These authors contributed equally: Akihide Yoshimi, Kuan-Ting Lin, Daniel H. Wiseman. *e-mail: abdelwao@mskcc.org

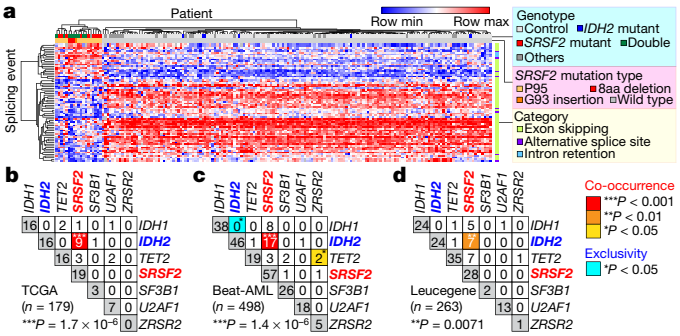


Fig. 1 | Frequently co-occurring *IDH2* and *SRSF2* mutations in AML. **a**, Heat map of per cent-spliced-in values for mutant *SRSF2*-specific splicing events in TCGA AML samples. **b**–**d**, Co-occurrence of mutations in *IDH1*, *IDH2*, *TET2* and RNA-splicing factors in the TCGA (**b**), Beat AML (**c**) and Leucegene (**d**) cohorts. Number of patients indicated; co-occurrence or exclusivity noted by colour coding; two-sided Fisher's exact test. Double refers to *SRSF2* and *IDH2* double mutant throughout.

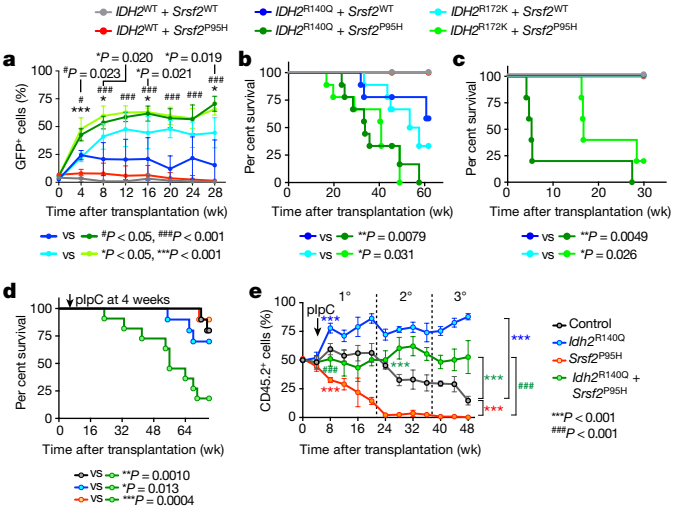


Fig. 2 | Mutant *IDH2* cooperates with mutant *Srsf2* to promote leukaemogenesis. **a**, Chimerism of GFP$^+$ cells in the blood of recipients transplanted with BM cells with indicated genotypes over time ($n = 5$ per group; data at 0 week represent transduction efficiency; mean percentage ± s.d.; two-way analysis of variance (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 peripheral blood CD45.2$^+$ cells in competitive transplantation. $n = 10$ mice per group; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test.

Srsf2 double-mutant cells had a substantially higher percentage of GFP$^+$ cells than in an *Srsf2* wild-type background (Fig. 2a, Extended Data Fig. 3b, c). Moreover, these mice exhibited significant myeloid skewing, macrocytic anaemia and thrombocytopaenia of greater magnitude than seen with mutant *IDH2* (Extended Data Fig. 3d–h). *IDH2* and *Srsf2* double mutants showed no difference in plasma 2HG levels from *IDH2* single mutants (Extended Data Fig. 3i, j). Serial replating of BM cells from leukaemic mice revealed markedly enhanced clonogenicity of *IDH2* and *Srsf2* double-mutant cells compared with other genotypes; the *IDH2* and *Srsf2* cells exhibited a blastic morphology and immature immunophenotype (Extended Data Fig. 3k–m). Consistent with these in vitro results, mice transplanted with *IDH2* and *Srsf2* double-mutant cells developed a lethal myelodysplastic syndrome (MDS) characterized by pancytopaenia, macrocytosis, myeloid dysplasia, expansion of immature BM progenitors and splenomegaly (Fig. 2b, Extended Data Fig. 3n–w). The *IDH2* and *Srsf2* double-mutant cells were also serially transplantable in sublethally irradiated recipients (Fig. 2c, Extended Data Fig. 3x). By contrast, *IDH2* single-mutant controls developed leukocytosis, myeloid skewing without clear dysplasia and less pronounced splenomegaly, whereas *Srsf2* single-mutant cells exhibited impaired repopulation capacity. These results provide evidence that spliceosomal gene mutations can promote leukaemogenesis 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 with *Idh2*$^{R140Q/+}$ mice to generate control, *Idh2*^{R140Q} single-mutant, *Srsf2*^{P95H} single-mutant and *Idh2* and *Srsf2* double knock-in (DKI) mice (Extended Data Fig. 4a). As expected, 2HG levels in peripheral blood mononuclear cells were increased and 5-hydroxymethylcytosine levels in KIT$^+$ BM cells were decreased in *Idh2* single-mutant and DKI primary mice compared with controls (Extended Data Fig. 4b, c). We next performed non-competitive transplantation, in which each mutation was induced alone or together following stable engraftment in recipients. DKI mice showed stable engraftment over time, 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 an increase in haematopoietic stem–progenitor cells in DKI mice compared with *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 leukaemogenesis in vivo.

On the basis of data identifying 2HG-mediated inhibition of TET2 as a mechanism of *IDH2* mutant leukaemogenesis[11], we also evaluated whether 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* and *Srsf2* double-mutant cells in vivo (Extended Data Fig. 4o–q). These data indicated that the collaborative effects of mutant *Idh2* and *Srsf2* are not solely dependent on TET2. Consistent with this, combined silencing of *Tet2* and *Tet3* 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). Because FTO and ALKBH5—which have roles in RNA processing as *N*6-methyladenosine (m^{6}A) RNA demethylases[12,13]—are also dependent on α-ketoglutarate, 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 analysed RNA-seq data from the TCGA ($n = 179$ patients), Beat AML ($n = 498$ patients) and Leucegene ($n = 263$ patients) cohorts as well as two previously unpublished RNA-seq datasets targeting defined *IDH2* and *SRSF2* genotype combinations ($n = 42$ patients) and the knock-in mouse models. This revealed that cells with mutations in both *IDH2* and *SRSF2* consistently contained more aberrant splicing events than cells with mutations in *SRSF2* only. Moreover, *IDH2* mutations were associated with a small but reproducible change in RNA splicing (Fig. 3a, b, Extended Data Fig. 5a–g, Supplementary Tables 4–20). By contrast, AML cases in which both *TET2* and *SRSF2* were mutated had fewer changes in splicing than those in which *IDH2* and *SRSF2* were mutated (Extended Data Fig. 5h–m, Supplementary Tables 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 PCR with reverse transcription

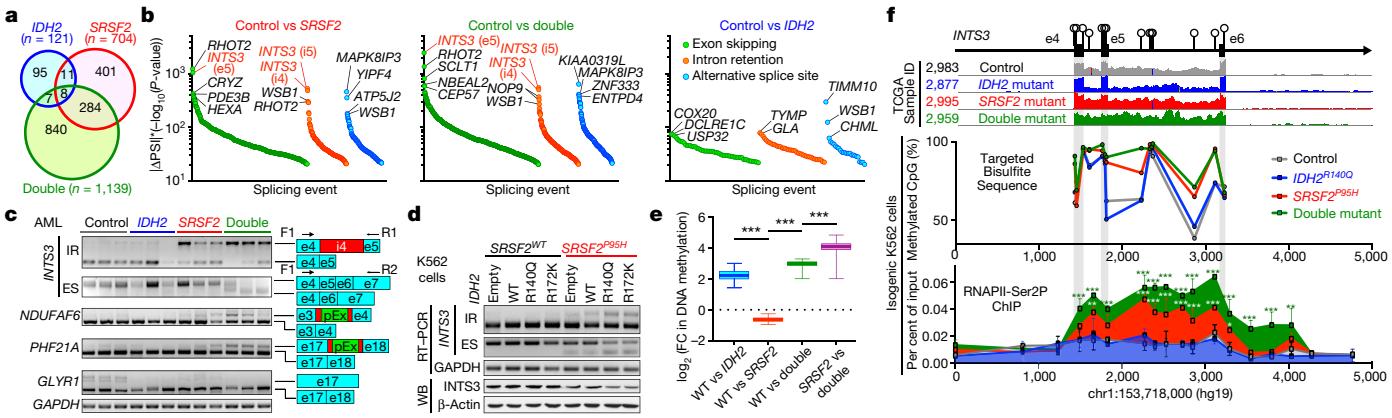


Fig. 3 | Collaborative effects of mutant *IDH2* and *SRSF2* on aberrant splicing. a, Venn diagram showing numbers of differentially spliced events from TCGA AML samples. **b**, Number of differentially spliced events ($|\Delta PSI| > 10\%$ and $P < 0.01$) in indicated genotypes are ranked by $((|\Delta PSI| \times |(-\log(P))$ according to class of event (e5, exon 5; i4 and i5, intron 4 and 5, respectively). PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma[25]. **c**, Representative RT–PCR results of aberrantly spliced transcripts in samples from patients with AML (pEx: exon with premature stop codon; $n = 3$ patients per genotype; three technical replicates with similar results). ES, exon skipping; IR, intron retention; F1, R1 and R2 represent primers used for RT–PCR. **d**, RT–PCR and western blots in isogenic K562 human leukaemia cells (representative images from three biologically independent experiments with similar results). **e**, Mean fold change (expressed in $\log_2$) in DNA cytosine methylation (*y* axis) at regions of genomic DNA encoding mRNA that undergo differential splicing (*x* axis). DNA methylation levels were determined by enhanced reduced representation bisulfite sequencing (eRRBS). $n = 3$ per genotype; the line represents mean, box edges represent the 25th and 75th percentiles and whiskers show 2.5th and 97.5th percentiles; one-way ANOVA with Tukey's multiple comparison test; $^{***}P < 2.2 \times 10^{-16}$. **f**, Genomic locus of *INTS3* around exons 4–6 with CpG dinucleotides indicated (top), representative RNA-seq from four patients with AML (top; $n = 1$ per genotype), results of targeted bisulfite sequencing (middle; $n = 1$ per genotype) and results of ChIP-walking experiments targeting RNAPII phosphorylated on Ser2 (Ser2P) (bottom; $n = 3$; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test compared with control.). Double mutant, $IDH2^{R140Q}SRSF2^{P95H}$. $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

(RT–PCR) of primary AML samples from an independent cohort (Fig. 3c). Among the mis-splicing events in AML with mutations in both *IDH2* and *SRSF2* was a complex event in *INTS3* involving intron retention across two contiguous introns and skipping of the intervening exon (Fig. 3b, c, Extended Data Figs. 5e–g, r–y, 6a–c). Aberrant *INTS3* splicing was demonstrated in isogenic and non-isogenic leukaemia cells with or without *IDH2* and/or *SRSF2* mutations (Fig. 3d, Extended Data Fig. 6d–f), and *INTS3* transcripts with both intron retention 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, Supplementary Table 23). Moreover, silencing of INTS3 was associated with reduced protein levels of additional integrator subunits in *SRSF2* mutant AML compared to *SRSF2* wild-type 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* and *SRSF2* double-mutant cells and was associated with reduced expression of multiple integrator subunits.

We next sought to understand how *IDH2* mutations, which affect 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 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* intron retention. 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 leukaemia cells containing combinations of *IDH2* and *SRSF2* mutations revealed that retention of *INTS3* intron 4 is driven by mutant SRSF2 and further enhanced in the *IDH2* and *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]. Leukaemia-associated mutations in *SRSF2* promote its avidity for C-rich sequences while reducing the ability to recognize G-rich sequences[6,7]. Of note, exon 4 of *INTS3* contains the highest 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, intron retention no longer occurred (*INTS3*-GGNG) (Extended Data Fig. 7e). Conversely, when all G-rich SRSF2 motifs were converted to C-rich sequences (*INTS3*-CCNG), intron retention became evident (Extended Data Fig. 7f). These results confirmed the sequence-specific activity of mutant SRSF2 in *INTS3* intron retention and identified a role for mutant IDH2 in regulating splicing.

Because *IDH2* mutations promote increased DNA methylation and DNA methylation can affect splicing[14], we generated genome-wide maps of DNA cytosine methylation from patients with AML across four genotypes (Supplementary Table 23). This revealed that differentially spliced events in *IDH2* single-mutant as well as *IDH2* and *SRSF2* double-mutant AML (compared to *IDH2* and *SRSF2* wild-type and *SRSF2* single-mutant AML) contained significant hypermethylation of DNA. Thus regions of differential DNA hypermethylation significantly overlapped with regions of differential RNA splicing (Fig. 3e, 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* exons 4–6 by targeted bisulfite sequencing. This revealed increased DNA methylation at all CpG dinucleotides in this region in *IDH2* and *SRSF2* double-mutant cells compared to control or single-mutant cells (Fig. 3f, 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 intron retention (Extended Data Fig. 7g–i). As further confirmation of the influence of mutant *IDH2* on *INTS3* splicing, cell-permeable 2HG increased *INTS3* intron retention whereas treatment of *IDH2* and *SRSF2* double-mutant cells with the DNA methyltransferase inhibitor 5-aza-2′-deoxycytidine (5-AZA-CdR) inhibited *INTS3* intron retention (Extended Data Fig. 7l, m).

Given that changes in epigenetic state may affect splicing by influencing RNAPII stalling[14,16], we evaluated the abundance of RNAPII

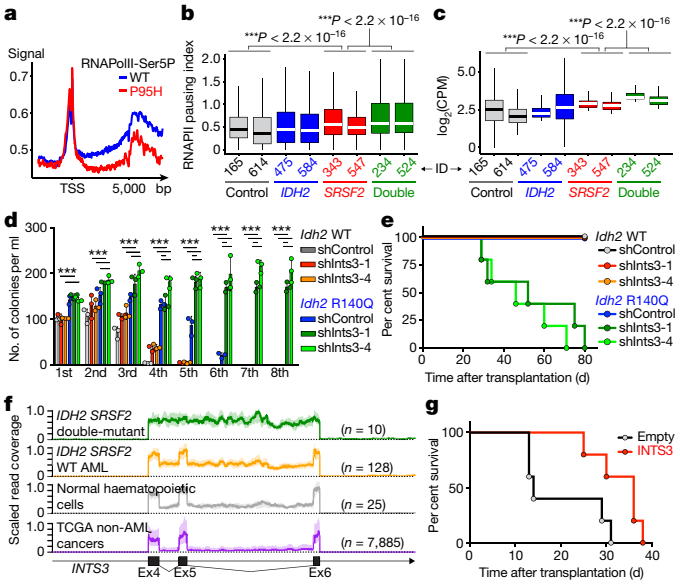


Fig. 4 | RNAPII stalling in *IDH2* and *SRSF2* double-mutant AML and contribution of INTS3 loss to leukaemogenesis. a, Metagene plot of genome-wide RNAPII Ser5 phosphorylation (Ser5P) in isogenic $SRSF2^{WT}$ or $SRSF2^{P95H}$ mutant K562 cells. **b**, **c**, RNAPII pausing index[20] in primary AML samples calculated as the ratio of normalized ChIP–seq reads of RNAPII Ser5P on transcription start sites (TSSs) (± 250 bp) over that of the corresponding bodies (+500 bp to +1000 bp from TSSs) (**b**) and RNAPII abundance over the differentially spliced regions between *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant AML determined by RNAPII Ser2P ChIP–seq (*y* axis shows $\log_2$(counts per million (CPM))) (**c**). *x* axis shows patient ID. Box plots were generated from ChIP–seq data from an individual primary AML sample; the line represents mean, box edges represent 25th and 75th percentiles and whiskers show 2.5th and 97.5th 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 targeting *Ints3* (shInts3-1 or shInts3-4) or control shRNA (shControl). $n = 3$ biologically independent experiments; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test. $^{***}P < 0.004$. **e**, **g**, Kaplan–Meier survival analysis of recipients. $n = 5$ per group; log-rank (Mantel–Cox) test (two-sided). In **e**: shInts3-1 versus shControl, $P = 0.0018$; shInts3-4 versus shControl, $P = 0.0018$. In **g**, $P = 0.034$ (INTS3 versus empty vector control). **f**, RNA-seq read coverage between exons 4–6 of *INTS3* and 1,000 bp either side of *INTS3* is scaled and shown as mean (line) ± s.d. (shaded region) (generated from TCGA datasets; sample list shown in legend for Extended Data Fig. 10e).

using chromatin precipitation with DNA sequencing (ChIP–seq) in isogenic $SRSF2^{WT}$ and $SRSF2^{P95H}$ cells as well as the primary samples from patients with AML. This revealed increased promoter-proximal transcriptional pausing and decreased RNAPII occupancy over gene bodies in *SRSF2* mutant cells, which was further enhanced in *IDH2* and *SRSF2* double-mutant cells (Fig. 4a, b, Extended Data Fig. 7n–q, 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 in *SRSF2* single-mutant AML than in *SRSF2* wild-type AML, and further enhanced over differentially spliced regions in *IDH2* and *SRSF2* double-mutant AML compared with those in *SRSF2* single-mutant AML (Fig. 4c, Extended Data Fig. 7t).

The above data provide further links between increased DNA cytosine methylation and RNAPII stalling with altered RNA splicing in *IDH2* and *SRSF2* double-mutant AML. To further evaluate this model, we performed ChIP for RNAPII across 4,766 bp of the *INTS3* locus in isogenic leukaemia cells (Fig. 3f). This revealed substantial accumulation of RNAPII across this locus in *IDH2* and *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* and *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). Notably, direct *Ints3* downregulation in the $Idh2^{R140Q/+}$ background resulted in enhanced clonogenic capacity of cells with an immature morphology and immunophenotype (Fig. 4d, Extended Data Fig. 8o–r) and promoted clonal dominance of *Idh2* mutant cells (Extended Data Fig. 9a–d). Moreover, mice transplanted with $Idh2^{R140Q/+}$ BM cells treated with short hairpin RNA (shRNA) targeting *Ints3* exhibited myeloid skewing, anaemia and thrombocytopaenia (Extended Data Fig. 9e–g) and developed a lethal MDS with proliferative features—phenotypes resembling those seen in *IDH2* and *Srsf2* double-mutant mice (Fig. 4e, Extended Data Fig. 9h, i).

The defects in snRNA processing in *SRSF2* single-mutant and *IDH2* and *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* and *SRSF2* double-mutant HL-60 cells from differentiation block (Extended Data Fig. 8y, z). Xenografts of *IDH2* and *SRSF2* double-mutant HL-60 cells demonstrated that forced expression of INTS3 induced myeloid differentiation and slowed leukaemia 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 leukaemogenesis.

Although loss of INTS3 resulted in measurable changes in snRNA processing, the degree of snRNA mis-processing did not have a substantial effect on splicing as determined by RNA-seq of $IDH2^{R140Q}$ mutant HL-60 cells with INTS3 silencing. By contrast, INTS3 depletion in these cells significantly affected transcriptional programs associated with myeloid differentiation, multiple oncogenic signalling pathways, RNAPII elongation-linked transcription and DNA repair (Extended Data Fig. 10a–d, Supplementary Table 25). This latter association of INTS3 loss with DNA repair is potentially consistent with previous reports that sensor of single-stranded DNA complexes containing INTS3 participate in DNA damage response[18,19].

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

These data highlight a role for loss of INTS3 in broad genetic subtypes of AML. Further efforts to determine how integrator loss

promotes leukaemogenesis and other non-mutational mechanisms mediating *INTS3* aberrant splicing will be critical for understanding and targeting leukaemias with integrator loss. Previous studies have identified that integrator[17,20] and SRSF2[21] have direct roles in modulating transcriptional pause–release. The accumulation of RNAPII at certain mis-spliced loci in this study is consistent with recent data that suggest 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 informative.

In addition to modifying splicing in *SRSF2* mutant cells, *IDH2* mutations were associated with reproducible changes in splicing in haematopoietic cells. There is a strong correlation between aberrant splicing in *IDH2* and *IDH1* mutant low-grade gliomas ($P = 2.2 \times 10^{-16}$ (binominal proportion test); Extended Data Fig. 10t–w, Supplementary Tables 26–28). A significant number of splicing events that were dysregulated in *IDH2* mutant AML from the TCGA and Leucegene cohorts were differentially spliced in *IDH2* mutants versus *IDH1* and *IDH2* wild-type low-grade gliomas ($P = 1.8 \times 10^{-9}$ and $P = 1.3 \times 10^{-8}$, respectively; binominal proportion test). These data suggest that *IDH1* and *IDH2* mutations impart a consistent effect on splicing regardless of tumour type. Finally, these results have important translational implications given the substantial efforts to pharmacologically inhibit mutant IDH1 and IDH2 as well as mutant splicing factors[23,24]. The frequent coexistence of *IDH2* and *SRSF2* mutations underscores the enormous therapeutic potential for modulation of splicing in the approximately 50% of patients with *IDH2* mutant leukaemia who also have a spliceosomal gene mutation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1618-0.

Received: 24 May 2018; Accepted: 28 August 2019;
Published online 2 October 2019.

1. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
2. Papaemmanuil, E. et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627, quiz 3699 (2013).
3. Wu, Y., Albrecht, T. R., Baillat, D., Wagner, E. J. & Tong, L. Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc. Natl Acad. Sci. USA* **114**, 4394–4399 (2017).
4. Darman, R. B. 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).
5. Ilagan, J. O. et al. *U2AF1* mutations alter splice site recognition in hematological malignancies. *Genome Res.* **25**, 14–26 (2015).
6. Kim, E. et al. *SRSF2* mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
7. Zhang, J. et al. Disease-associated mutation in *SRSF2* misregulates splicing by altering RNA-binding affinities. *Proc. Natl Acad. Sci. USA* **112**, E4726–E4734 (2015).
8. Tyner, J. W. et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
9. Lavallée, V. P. et al. The transcriptomic landscape and directed chemical interrogation of *MLL*-rearranged acute myeloid leukemias. *Nat. Genet.* **47**, 1030–1037 (2015).
10. Dang, L. et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
11. Figueroa, M. E. 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).
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).
13. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
14. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84**, 165–198 (2015).
15. Daubner, G. M., Cléry, A., Jayne, S., Stevenin, J. & Allain, F. H. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.* **31**, 162–174 (2012).
16. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
17. Gardini, A. et al. Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56**, 128–139 (2014).
18. Huang, J., Gong, Z., Ghosal, G. & Chen, J. SOSS complexes participate in the maintenance of genomic stability. *Mol. Cell* **35**, 384–393 (2009).
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).
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).
21. Ji, X. et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
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 (2018).
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).
24. Stein, E. M. et al. Enasidenib in mutant *IDH2* relapsed or refractory acute myeloid leukemia. *Blood* **130**, 722–731 (2017).

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.

Mice. All mice were housed at Memorial Sloan Kettering Cancer Center (MSK). All mouse 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. Six- to eight-week-old 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 analysed 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 peripheral blood 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 euthanized for visible tumour formation at tumour 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.

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-cm^3 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 r.p.m. for 4 min and red blood cells (RBCs) were lysed in ammonium chloride-potassium bicarbonate lysis (ACK) buffer (150 mM NH_4Cl + 10 mM $KHCO_3$ + 0.1 mM EDTA; Thermo Fisher Scientific) for 3 min on ice. After centrifugation, cells were resuspended in PBS/2% BSA, passed through a 40-μm 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; 12 mg per kg (body weight) per day; GE Healthcare) every second day via intraperitoneal injection. Peripheral blood chimerism was assessed every four 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 analyser. Mice that were lost owing to pIpC toxicity were excluded from analysis.

Retroviral transduction and transplantation of primary haematopoietic cells. *Vav-cre Srsf2*$^{+/+}$ and *Vav-cre Srsf2*$^{P95H/+}$ mice were treated with a single dose of 5-fluoruracil (150 mg kg^{-1}) followed by BM collection 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^{-1}), mouse interleukin-3 (mIL3, 10 ng ml^{-1}) and mIL6 (10 ng ml^{-1}), followed by injection of about 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 unsorted 1 × 10^6 BM cells from the primary transplantation. *IDH2*^{WT} + *Srsf2*^{WT} and *IDH2*^{WT} + *Srsf2*^{P95H} mice were euthanized at day 315 post-transplant to collect BM for the serial transplantation. All cytokines were purchased from R&D Systems.

Flow cytometry analyses and antibodies. Surface-marker staining of haematopoietic 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 min on ice. For haematopoietic stem/progenitor staining, cells were stained with the following antibodies: B220-APCCy7 (clone: RA3-6B2; purchased from BioLegend; catalogue no.: 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); KIT-APC (2B8; BioLegend; 105812; 1:200); KIT-PerCPCy5.5 (2B8; BioLegend; 105824; 1:100); KIT-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 leukaemia 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 haematopoietic cell lineages in the BM, spleen and peripheral blood was assessed using a combination of CD11b, Gr1, B220, and CD3. For the haematopoietic stem and progenitor analysis, a combination of CD11b, CD11c, Gr1, B220, CD3, NK1.1 and Ter119 was stained as lineage-positive cells. Fluorescence-activated cell sorting (FACS) was performed on a FACS Aria, and analysis was performed on an LSRII or LSR Fortessa (BD Biosciences). For western blotting, DNA dot blot assays, and ChIP assays, the following antibodies were used: INTS1 (purchased from Bethyl laboratories; catalogue no.: 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), RNAPII CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095), RNAPII 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 h 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/erythroleukaemia cell line) and HL-60 (human promyelocytic leukaemia cell line) leukaemia cells, K052 (human multilineage leukaemia cell line) leukaemia cells, TF1 (human erythroleukaemia cell line) leukaemia cells, *MLL-AF9/Nras*^{G12D} murine leukaemia (RN2) cells[29], and Ba/F3 (murine pro-B cell line) cells were cultured in RPMI/10% fetal calf serum (FCS, heat inactivated), RPMI/20% FCS, RPMI/10% FCS + human granulocyte-macrophage colony-stimulating factor (GM-CSF, R&D Systems; 5 ng ml^{-1}), and RPMI/10% FCS + mIL3 (R&D Systems; 1 ng ml^{-1}), respectively. None of the cell lines above were listed in the database of commonly misidentified cell lines maintained by ICLAC and NCBI Biosample.

MSCV-*IDH2*$^{WT/R140Q/R172K}$-IRES-GFP, MSCV-3×Flag-*INTS3*-puro, MSCV-IRES-3×Flag-*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). 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^{-1}). GFP$^+$mCherry$^+$ double-positive HL-60 cells and mCherry$^+$ K562 cells were FACS-sorted to obtain cells expressing wild-type or mutant IDH2 and SRSF2 in various combinations. Isogenic HL-60 cells transduced with 3×Flag-tagged INTS3 or empty vector were obtained by puromycin selection (1 μg ml^{-1}). To let the cells fully establish epigenetic changes, they were analysed after culture for more than 30 days.

For in vitro colony-forming assays, a single-cell suspension was prepared and 15,000 cells per 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 euthanized at day 315 post-transplant to collect 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* forward1: TGAGTCGTGATGGCATGAAT (exon 4), reverse1: TCTTCACCAGTTCCCGTACC (exon 5; for detection of intron 4 retention), reverse2: CTGCTCTTCAGGACCCACTC (exon 7; for detection of exon 5 skipping); *NDUFAF6* forward: GCCTGTGGCCATTGAACTAT, reverse: ACAATGCCTTGTGCTTTTCC; *PHF21A* forward: TCCATGGCCTGGAACTTTAG, reverse: GCCAGGATGGTGTTCTTCAT; *GLYR1* forward: AGGTCAGGCCCAGTTCTCTT, reverse: TCACGTCTAAGCGTCCAGTG; *GAPDH* forward: GCAAATTCCATGGCACCGTC, reverse: 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 analysed 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*: forward2: CTGCAGGATACCTGCCGTA (exon 4), reverse3: CTTTCCCGTTCCTGACAGAG (intron 5; for specific quantification of transcript with intron 4 retention); forward1: TGAGTCGTGATGGCATGAAT (exon 4), reverse4: GGCTGTAACATCTCCACCTGA (exon 4–6; for specific quantification of transcript with exon 5 skipping); forward3: GGGCAATGCTGAGAGAGAAG (exon 14), reverse5: TGCCTCTGCATTGTCATAGC (exon 15); *mInts3*: forward: GTGGCTGTTATTGACTCTGCAC, reverse: CAGGTTCCCCATCATCACAT; *mFto*: forward: CACTTGGCTTCCTTACCTGACCCCC, reverse: GGTATGCTGCCGGCCTCTCGG; *mAlkbh5*: forward: CGGCCTCAGGACATTAAGGA, reverse: TCGCGGTGCATCTAATCTTG; Total U2snRNA: forward: CTTCTCGGCCTTTTGGCTAAGAT, reverse: GTACTGCAATACCAGGTCGATGC; uncleaved U2snRNA: forward: ACGTCCTCTATCCG+AGGACAATA, reverse: GCAGGTGCTACCGTCTCTCAC; total U4snRNA: forward: GCAGTATCGTAGCCAATGAGGTCTA, reverse: CCAGTGCCGACTATATTGCAAGTC.

Uncleaved U4snRNA: forward: CGTAGCCAATGAGGTCTATCCG, reverse: CCTCTGTTGTTCAACTGCAAGAAA; *hGAPDH*: forward: GCAAATTCCATGGCACCGTC, reverse: TCGCCCCACTTGATTTTGG; *mGapdh*: forward: TGGAGAAACCTGCCAAGTATG, reverse: 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 (C_T) method.

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

ChIP assays. Cells were crosslinked and collected. Chromatin was broken down into 200–1,000-bp fragments using an E220 Focused-ultrasonicator. An antibody was added into the lysate and incubated overnight at 4 °C. Twenty microlitres of ChIP–grade Protein A/G Dynabeeds was added into each IP tube and incubated for 2 h. 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. RNAPII antibody (05-623; EMD Millipore) was used in this study. Primer sequences used for ChIP-PCR were as follows: Intron 3-1 forward: atacccggcccttgctatac, reverse: gcaacttccttagcctgctg; Intron 3-1 forward: atacccggcccttgctatac, reverse: gcaacttcctta gcctgctg; Intron 3-2 forward: ctggcaggtgaaaagcagat, reverse: ggcaggggagagaaaagc; Intron 3-3 forward: agcaggcttttctgcctcat, reverse: tttctttccacaggggtcct; Exon 4 forward: cgggacttagctctggtgag, reverse: cctgagtacggcaggtatcc; Intron 4 forward: ctct gtcaggaacgggaaag, reverse: tgtgagtttgagaagggagcta; Exon 5 forward: acgggaactggtgaa gagtg, reverse: ctgggctctcctcctttctt; Intron 5-1 forward: ctccaccccattatctgaa, reverse: aaatgtcagggtctgttctgtg; Intron 5-2 forward: tcggtgacatctgtctgagc, reverse: cagtgggctaa tggtgaggt; Intron 5-3 forward: aacactgatgctcctgttttga, reverse: actatgccttgccccaggt; Intron 5-4 forward: gctgttgtcagccacctgta, reverse: tttggcccttgaaaatgaac; Intron 5-5 forward: tgtgttaattctgccccaca, reverse: ggatgtcctgagtcctgcac; Intron 5-6 forward: gtaatgggatggcagtcagg, reverse: cctgatttcaaaaggggaaa; Exon 6 forward: agcaaaggtagc atccacca, reverse: cttgcctccccctctctaac; Intron 6-1 forward: tttgatccagacctccttgg, reverse: gcaggggagaaaaggatacc; Intron 6-2 forward: gggggtacatattgggcttt, reverse: gaaagcctcacctccaaaca; Intron 6-3-CTCF binding site forward: ctcctcccaacgttcacact, reverse: atccgtgcccagagcacta; Intron 6-4 forward: agggggcctttcaactctt, reverse: atggggacaggacgtatttg; Intron 6-5 forward: ttccctgccttccaacag, reverse: tcccagttgctt taaaaggagt.

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–seq of primary human AML samples. ChIP was performed as previously described[32] using the following antibodies: RNAPolII-Ser2P antibody ChIP Grade (Abcam ab5095), RNAPII-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-bp 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 euthanized 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 haematoxylin 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 patients with leukaemia 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; variant call format (VCF) files were analysed using Illumina's Variant Studio software; (b) a 40 gene panel (Oncomine Myeloid Research Assay; ThermoFisher), processing 8 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 leukaemias, lymphomas, and solid tumours. 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): forward: acactgacgacatggttctacacccgtttacctgcggctc, reverse: tacggtagc agagacttggtctccttcgttcgctttcacgacaa.

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 χ^2-test (two-sided). Representative western blot 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 ($n \geq 3$). $^*P < 0.05$, $^{**}P < 0.01$ 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). DNA fragments 300 bp in length were isolated and sequenced on an Illumina HiSeq 2000 (about 100 × 10^6 101-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-seq data. Unprocessed RNA-seq reads of TCGA and Leucegene datasets (patients with AML) were downloaded from NCI's Genomic Data Commons Data Portal (GDC Legacy Archive; TCGA-LAML dataset) and NCBI's Sequence Read Archive (SRA; accession number 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 (about 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) was used to generate VCF files for seven 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 with a variant allele frequency (VAF) lower than 15%. The filtered VCF files were used for variant effect predictor (v.89.4) to annotate the consequences of the mutations. We defined control patient samples as those without mutations in the seven target genes, *IDH2* mutated samples as those with only *IDH2* mutations but no mutations in the other six target genes, *SRSF2* mutated samples as those with only *SRSF2* mutations but no mutations in the other six target genes, Double-mutant samples as those with both *IDH2* and *SRSF2* mutations but no mutations in the other five 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]. In brief, 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 labelled as 'novel' and a splicing event with a reference transcript that is known to induce nonsense-mediated decay is labelled 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 1st, 25th, 50th, 75th and 99th 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 2 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 haematopoietic cells.

Motif enrichment and distribution. Motif analysis was done by using MEME SUITE[38]. In brief, 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 on the basis of the ranking of *E*-value.

Heat map and sample clustering (differential splicing). The heat maps and sample clustering were done by using MORPHEUS (https://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 eRRBS and 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).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

RNA-seq, ChIP–seq and eRRBS data have been deposited in the NCBI Sequence Read Archive under accession number SRP133673. Gel source data are shown in Supplementary Fig. 1. Other data that support the findings of this study are available from the authors upon reasonable request.

25. Lin, K. T. & Krainer, A. R. PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics* btz438 (2019).
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).
27. Shih, A. H. 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).
28. Georgiades, P. et al. *VavCre* transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis* **34**, 251–256 (2002).
29. Zuber, J. et al. Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nat. Biotechnol.* **29**, 79–83 (2011).
30. Lee, M. et al. Engineered split-TET2 enzyme for inducible epigenetic remodeling. *J. Am. Chem. Soc.* **139**, 4659–4662 (2017).
31. Kleppe, M. et al. Dual targeting of oncogenic activation and inflammatory signaling increases therapeutic efficacy in myeloproliferative neoplasms. *Cancer Cell* **33**, 29–43 (2018).
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).
33. Cheng, D. T. 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).

34. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
35. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
36. Dvinge, H. & Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45 (2015).
37. Hubert, C. G. et al. Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* **27**, 1032–1045 (2013).
38. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
39. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
40. Intlekofer, A. M. et al. Hypoxia induces production of l-2-hydroxyglutarate. *Cell Metab.* **22**, 304–311 (2015).
41. Dvinge, H. et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc. Natl Acad. Sci. USA* **111**, 16802–16807 (2014).
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).

Acknowledgements We thank D. L. Fei, Y. Huang, E. Wang, I. Aifantis, M. Patel, A. S. Shih, A. Penson, E. Kim, Y. R. Chung, B. H. Durham and H. Kunimoto for technical support, J. Wilusz for sharing recent data on integrator and B. 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 Department 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.

Author contributions A.Y., K.-T.L., A.R.K. and O.A.-W. designed the study. A.Y., B.W., 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 mouse 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.M.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.

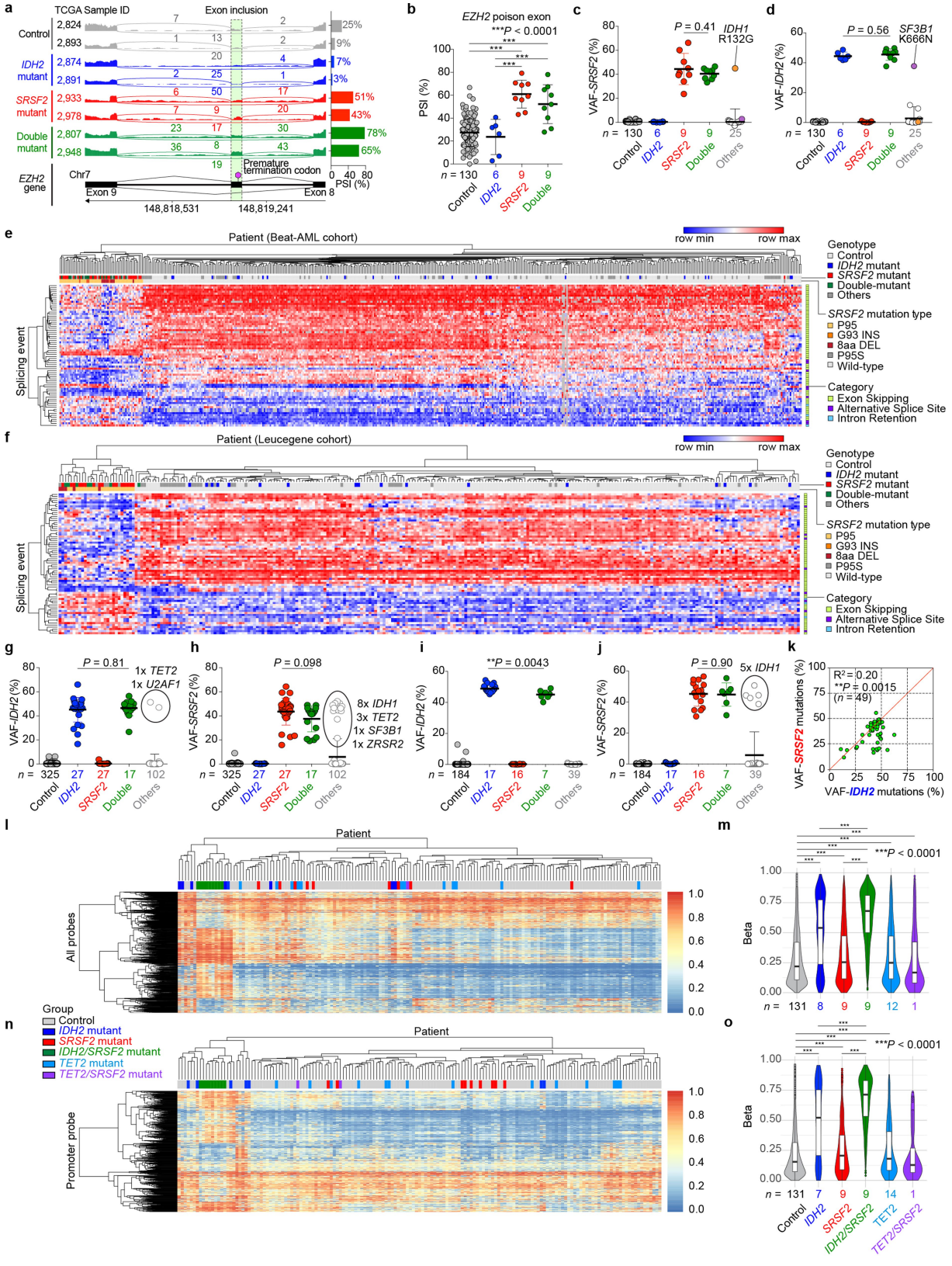
Competing interests A.M.I. has served as a consultant and advisory board member for Foundation Medicine. E.M.S. has served on advisory boards for Astellas Pharma, Daiichi Sankyo, Bayer, Novartis, Syros, Pfizer, PTC Therapeutics, AbbVie, Agios and Celgene and has received research support from Agios, 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 and patents) in Qiagen and Loxo, and is a consultant and/or advisory board member for Novartis, Roche, Janssen, Celgene and Incyte. A.R.K. is a founder, director, advisor, stockholder and chair of the Scientific Advisory Board 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, Merck and Janssen. O.A.-W. has received personal speaking fees from Daiichi Sankyo. O.A.-W. has received previous 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.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1618-0.

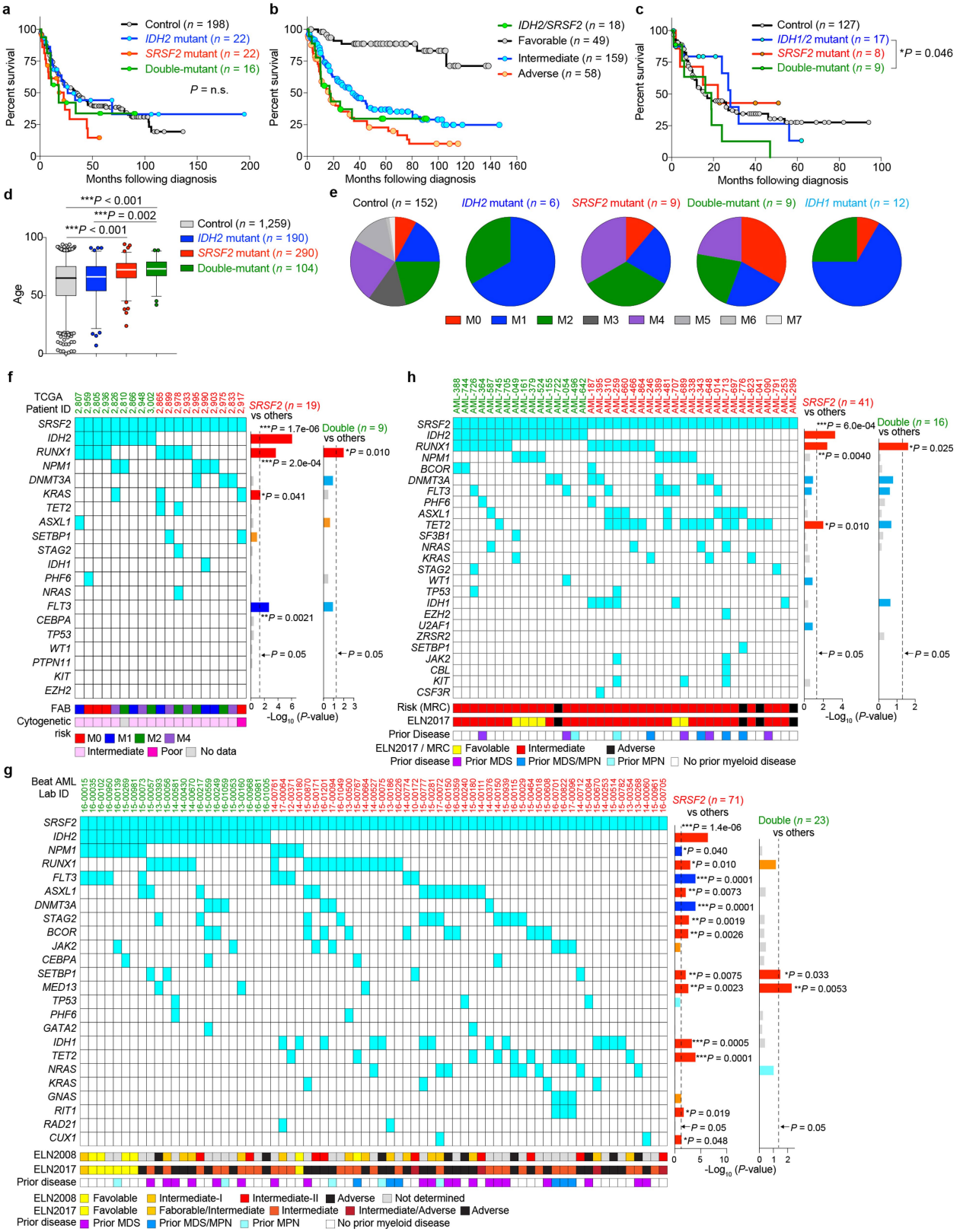
Correspondence and requests for materials should be addressed to O.A.-W.

Peer review information *Nature* thanks Rotem Karni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



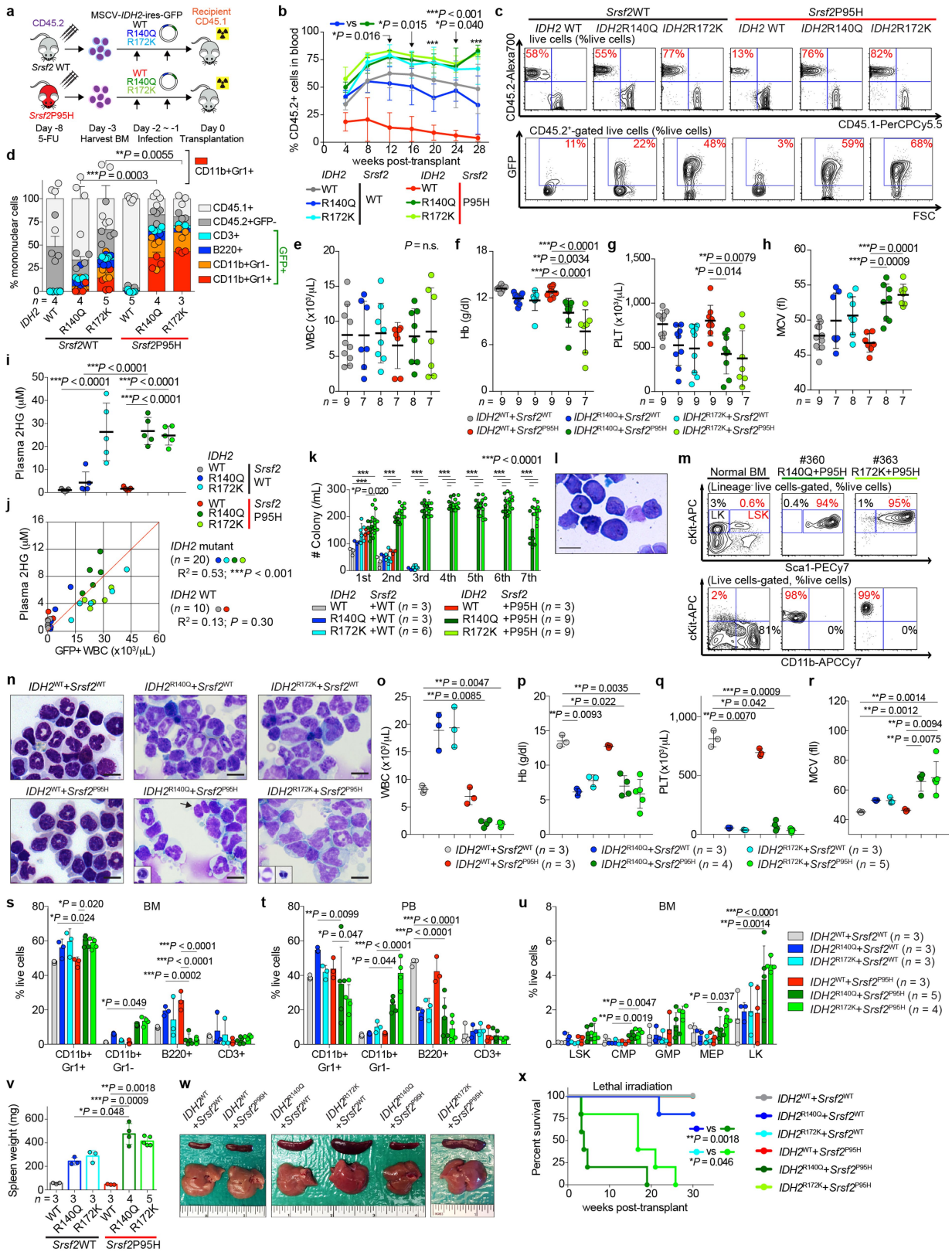
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Mutant *SRSF2*-mediated splicing events in acute myeloid leukaemia (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 seven genes: *IDH1*, *IDH2*, *TET2*, *SRSF2*, *SF3B1*, *U2AF1*, and *ZRSR2*; '*IDH2* mutant' refers to patients with an *IDH2* mutation and no mutation in the other six genes; '*SRSF2* mutant' refers to patients with an *SRSF2* mutation and no mutation in the other six genes; 'double-mutant' refers to patients with an *IDH2* and *SRSF2* mutation and no mutation in the other five 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**, PSI values of *EZH2* poison exon inclusion (the number of analysed patients is indicated; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). Note that patients classified as 'others' include one patient with an $SRSF2^{P95L}$ mutation with a coexisting $IDH1^{R132G}$ mutation (TCGA ID: 2990) and one patient with an $IDH2^{R140Q}$ mutation also having an $SF3B1^{K666N}$ mutation (TCGA ID: 2973), which were excluded from the analyses shown above. **c**, **d**, **g**–**j**, 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 (mean ± s.d.; 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**, VAFs of *IDH2* (*x* axis) and *SRSF2* mutations (*y* axis) in *IDH2* and *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* and *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 line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; one-way ANOVA with Tukey's multiple comparison test). $^{**}P < 0.01$, $^{***}P < 0.001$.



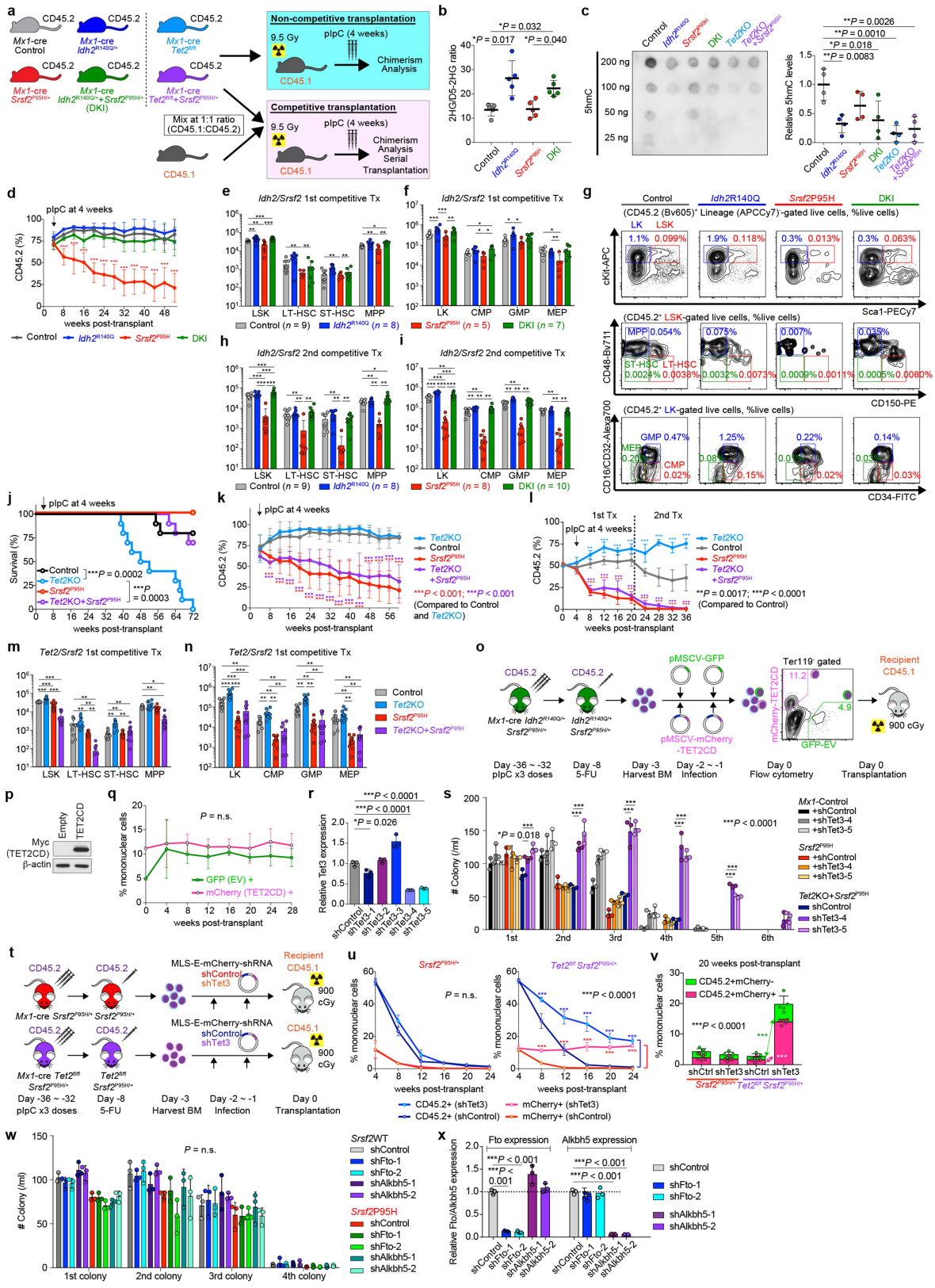
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Clinical relevance of coexisting *IDH2* and *SRSF2* mutations in AML. a–c, Kaplan–Meier survival analysis of patients with AML from the Manchester/Christie Biobank dataset (**a**: based on *IDH2* and *SRSF2* genotype ($n = 258$); **b**: based on cytogenetic risk ($n = 284$)) and the TCGA (**c**; $n = 161$; based on *IDH1*, *IDH2* and *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 line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; one-way ANOVA with Tukey's multiple comparison test). **e**, Distribution of French–American–British (FAB) classification of patients with AML with the indicated genotypes from the TCGA cohort. **f–h**, Mutations coexisting with *IDH2* and *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 LeukaemiaNet (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* and *SRSF2* double-mutant AML) or coexisting *IDH2* and *SRSF2* mutations; Fisher's exact test (two-sided)). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.



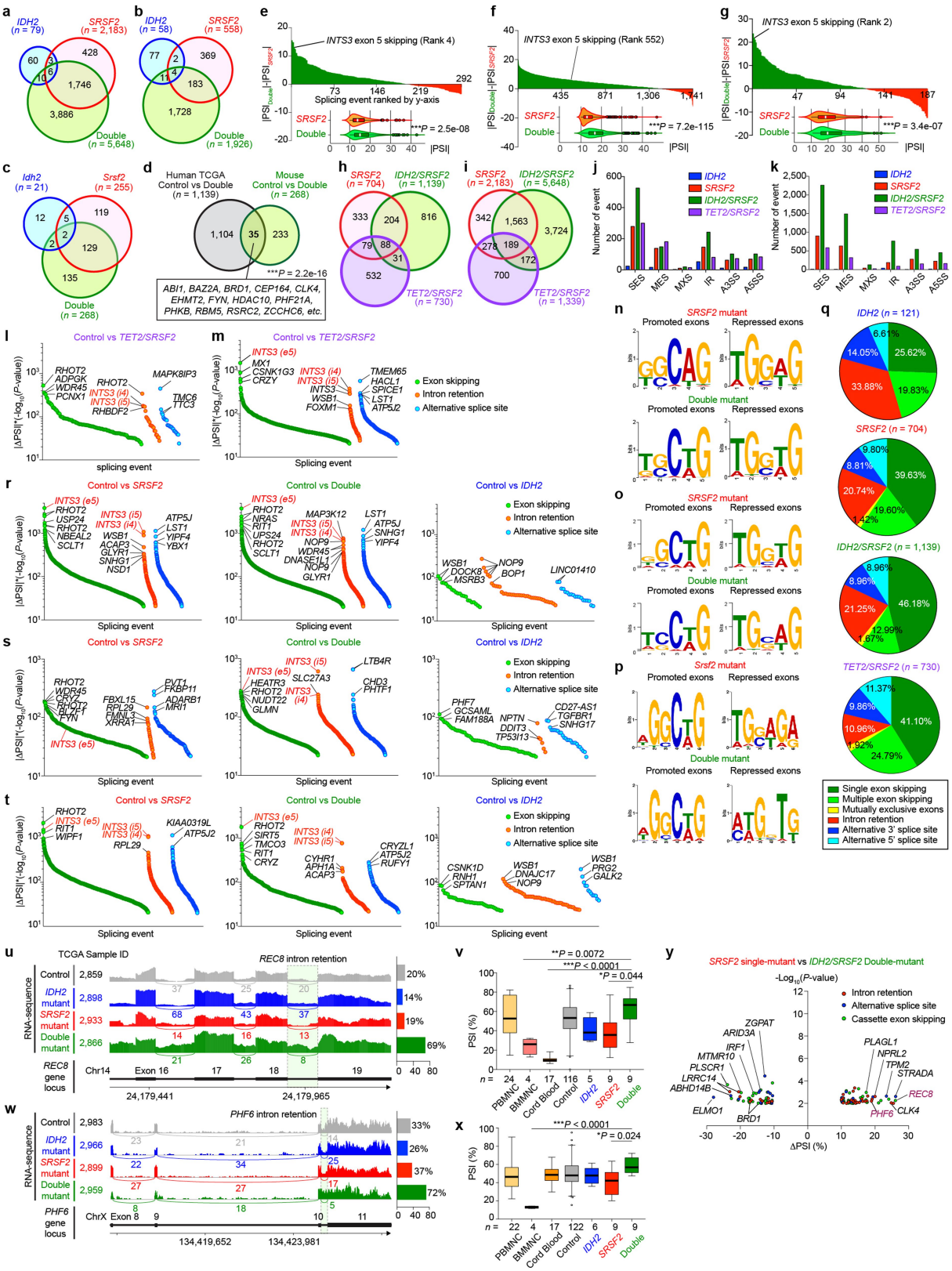
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Mutant *IDH2* cooperates with mutant *Srsf2* to generate lethal MDS with proliferative features in vivo. a, Schematic of BM transplantation model. **b**, **c**, Chimerism of CD45.2$^+$ cells in the peripheral blood of recipient mice over time (**b**) ($n = 5$ per group at 4 weeks; mean percentage ± s.d.; two-way ANOVA with Tukey's multiple comparison test) and representative flow cytometry data showing the chimerism of CD45.2$^+$ versus CD45.1$^+$ (top) or GFP$^+$ (bottom) cells in peripheral blood 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 peripheral blood mononuclear cells (PBMNCs) at 28 weeks post-transplant (the number of analysed mice is indicated; mean + s.d.; two-way ANOVA with Tukey's multiple comparison tests; statistical significances were detected in percentage of CD11b$^+$Gr1$^+$ cells in *IDH2*^{R140Q} + *Srsf2*^{WT} versus *IDH2*^{R140Q} + *Srsf2*^{P95H} and in *IDH2*^{R172K} + *Srsf2*^{WT} versus *IDH2*^{R172K} + *Srsf2*^{P95H}). **e**–**h**, Blood counts at 20 weeks post-transplant (white blood cells (WBC) (**e**); haemoglobin (Hb) (**f**); platelets (PLT) (**g**); mean corpuscular volume (MCV) (**h**); the number of analysed mice is indicated; 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 previously described[40]; $n = 5$ per group were randomly selected; 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 collected from end-stage mice from Fig. 2b are shown (mean value ± s.d. represented by lines above the box; the number of analysed mice is indicated; two-way ANOVA with Tukey's multiple comparison test). **l**, Giemsa staining of *IDH2*^{R140Q} *Srsf2*^{P95H} double-mutant cells from the sixth plating (scale bar, 10 μm; original magnification, ×400; representative result from 9 biologically independent experiments). **m**, Immunophenotype of colony cells at the sixth 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* and *Srsf2* double-mutant groups have increased granulocytes. In addition, *IDH2* and *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 3 controls and 9 recipients are shown; number of mice indicated in **o**–**r**). **o**–**r**, Blood counts at end-stage (WBC (**o**); Hb (**p**); PLT (**q**); MCV (**r**); the number of analysed mice is indicated; mean ± s.d.; Kruskal–Wallis tests with uncorrected Dunn's test). **s**–**u**, Results from flow cytometry analysis of BM (**s**) and peripheral blood (**t**) mature lineages as well as BM haematopoietic stem/progenitor cells (HSPC) from two tibias, two femurs, and two pelvic bones (**u**) are quantified (LSK, Lineage$^-$SCA1$^+$KIT$^+$; LT-HSC, long-term haematopoietic stem cell (HSC); ST-HSC, short-term HSC; MPP, multi-potent progenitor; LK: Lineage$^-$SCA1$^-$KIT$^+$; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythroid progenitor; the number of analysed mice is indicated; mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **v**, **w**, Spleen weight at end stage (**v**; the number of analysed mice is indicated; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test) 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)). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.



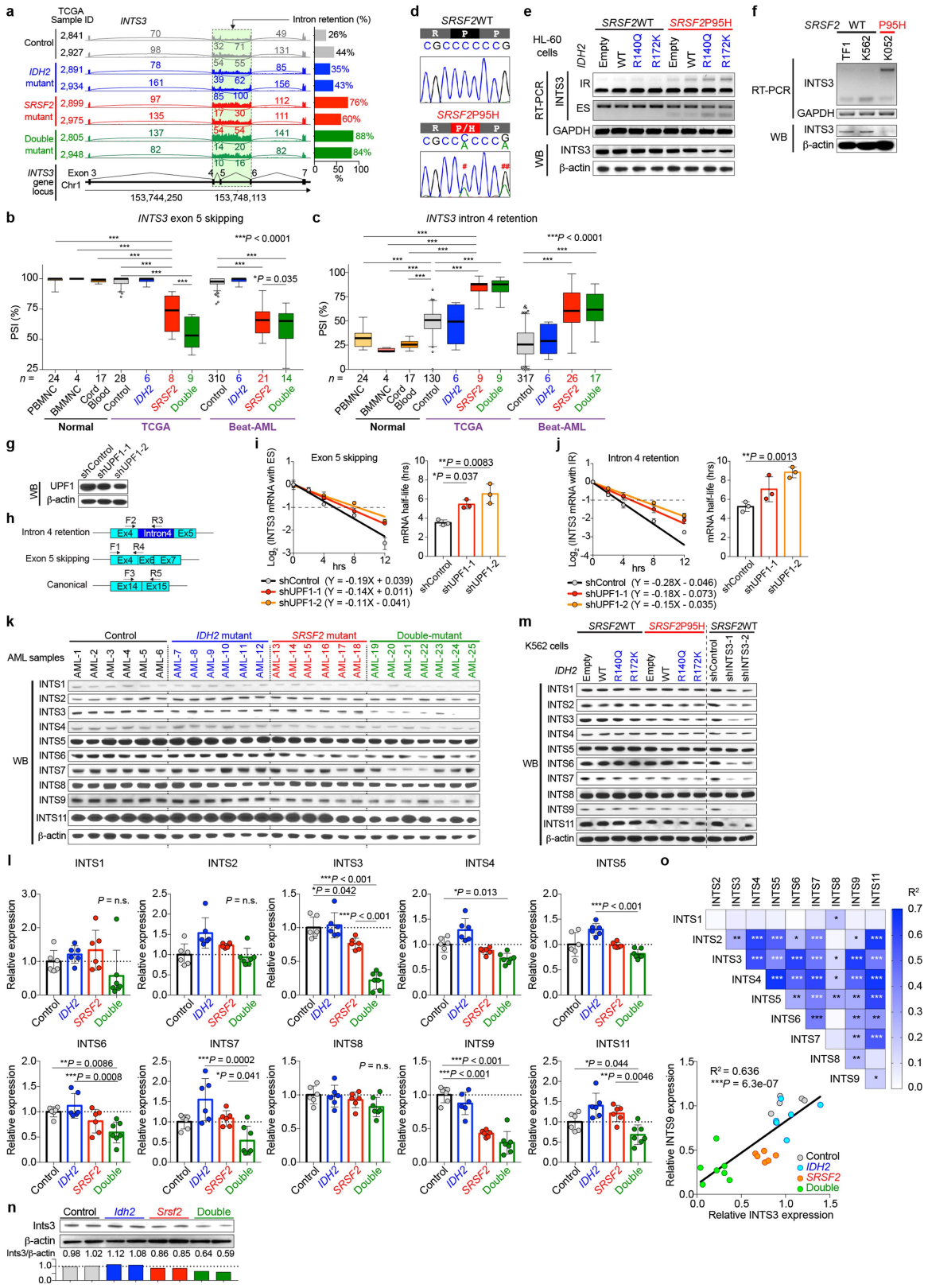
Extended Data Fig. 4 | See next page for caption.

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 three 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; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **c**, DNA extracted from sorted KIT$^+$ BM cells from primary *Mx1-cre* mice at one 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 3 biologically independent experiments with similar results; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **d**, Chimerism of peripheral blood CD45.2$^+$ cells in non-competitive transplantation (pIpC was injected at 4 weeks post-transplant; 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* and *Srs2* mutant cells, and representative flow cytometry of BM HSPCs from the primary competitive transplant of *Idh2* and *Srsf2* mutant cells from **e**, **f** (the percentage listed represents the percent of cells within live cells) (the number of analysed mice is indicated; 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 peripheral blood CD45.2$^+$ cells in non-competitive (**k**) ($n = 10$ (control and *Tet2* knockout (*Tet2*KO)), $n = 8$ (*Srsf2*^{P95H}), and $n = 5$ (*Tet2*KO + *Srsf2*^{P95H}) at 0 weeks) or competitive (**l**) ($n = 9$ (control), $n = 10$ (*Tet2*KO), $n = 8$ (*Srsf2*^{P95H}), and $n = 10$ (*Tet2*KO + *Srsf2*^{P95H}) at 0 weeks) 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**; 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* and *Srsf2* mutant cells ($n = 10$ per genotype; 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 TET2 CD in Ba/F3 cells transduced with or without TET2 CD (representative images from two biologically independent experiments with similar results). **q**, Chimerism of mCherry–TET2 CD$^+$ and GFP–EV$^+$ cells in peripheral blood of recipient mice over time ($n = 10$; mean percentage ± s.d.; two-way ANOVA with Sidak's multiple comparison test). **r**, qPCR of *Tet3* in the first colony cells from **s** ($n = 3$; 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 shRNAs targeting *Tet3* (shTet3) ($n = 3$; mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **t**, Schematic of shTet3 retroviral BM transplantation model. **u**, **v**, Chimerism of mCherry$^+$ cells in CD45.2$^+$ donor cells in peripheral blood of recipient mice over time (**u**; left, *Mx1-cre Srsf2*$^{P95H/+}$; right, *Mx1-cre Tet2*$^{fl/fl}$*Srsf2*$^{P95H/+}$; $n = 5$ per group) and at 20 weeks post-transplant (**v**) (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 collected at one month post-pIpC ($n = 3$; 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* ($n = 3$; mean value ± s.d.; one-way ANOVA with Tukey's multiple comparison test). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.



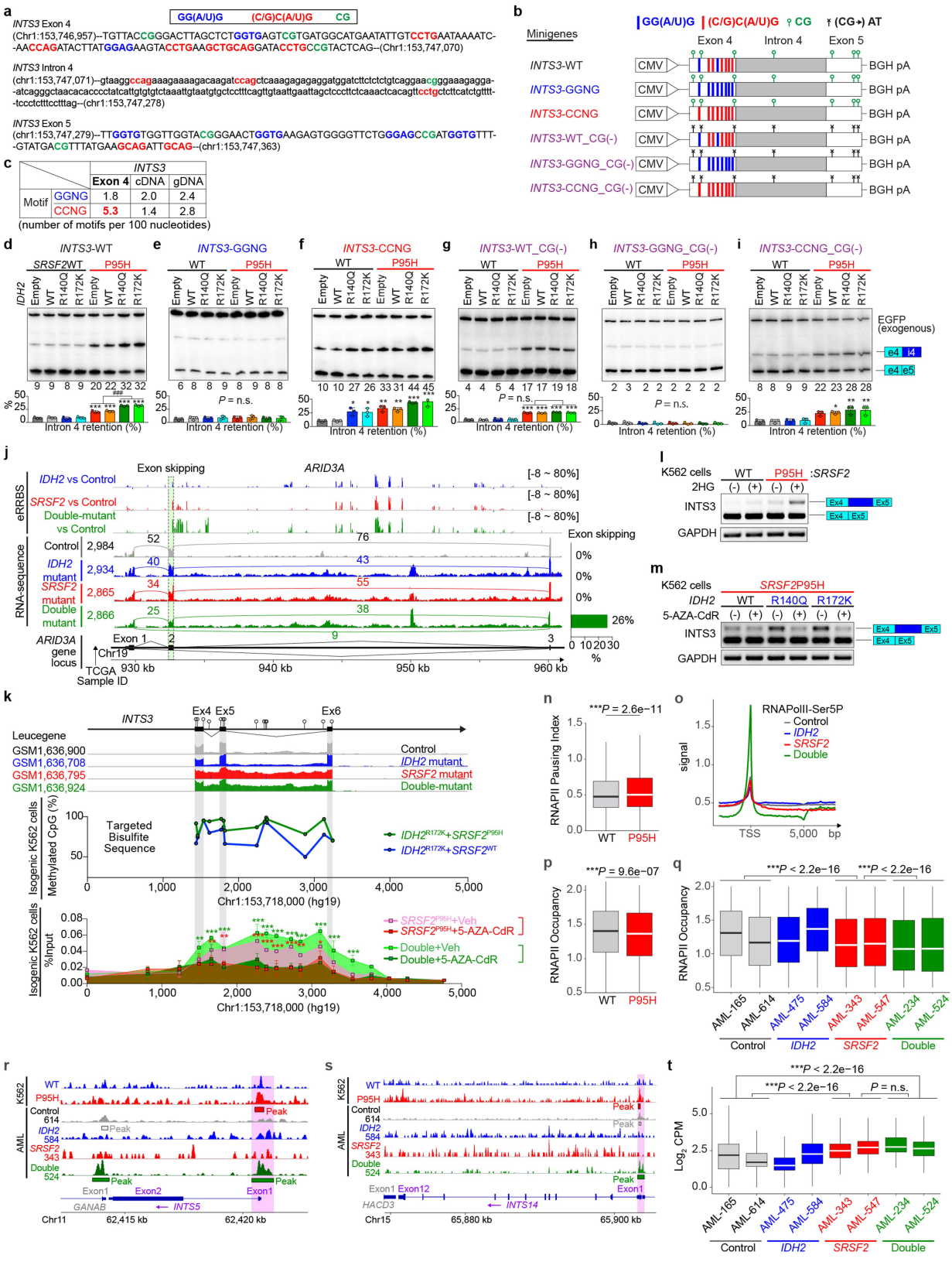
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | *IDH2* mutations augment the RNA splicing defects of *SRSF2* mutant leukaemia. a–c, Venn diagram showing numbers of differentially spliced events from the Beat AML cohort (**a**), unpublished collaborative cohort 2 (**b**) and mouse $Lin^{-}KIT^{+}$ bone marrow cells at 12 weeks post-pIpC (**c**) based on *IDH2* and *SRSF2* mutant genotypes. **d**, Venn diagram showing the numbers of overlapping alternatively spliced events between *IDH2* and *SRSF2* double-mutant AMLs and mouse models ($^{***}P = 2.2 \times 10^{-16}$; binominal test). **e–g**, $\Delta|\text{PSI}|$ ($\Delta|\text{PSI}| = |\text{PSI}|_{\text{Double}} - |\text{PSI}|_{SRSF2}$) values for each overlapping mis-spliced event in *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant AML from the TCGA (**e**), Beat AML cohort (**f**) and unpublished collaborative cohort 2 (**g**) are plotted along the *y* axis. Spliced events shown in green and red represent events that are more robust in *IDH2* and *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 line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; 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* and *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. **l**, **m**, Differentially spliced events ($|\Delta\text{PSI}| > 10\%$ 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). **n–p**, Sequence logos of nucleotide motifs of exons preferentially promoted or repressed in splicing in *SRSF2* single-mutant (top) or *IDH2* and *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 a pie-chart. **r–t**, Differentially spliced events ($|\Delta\text{PSI}| > 10\%$ 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 line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; 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; $|\Delta\text{PSI}| > 10\%$ and $P < 0.01$ were used as thresholds).



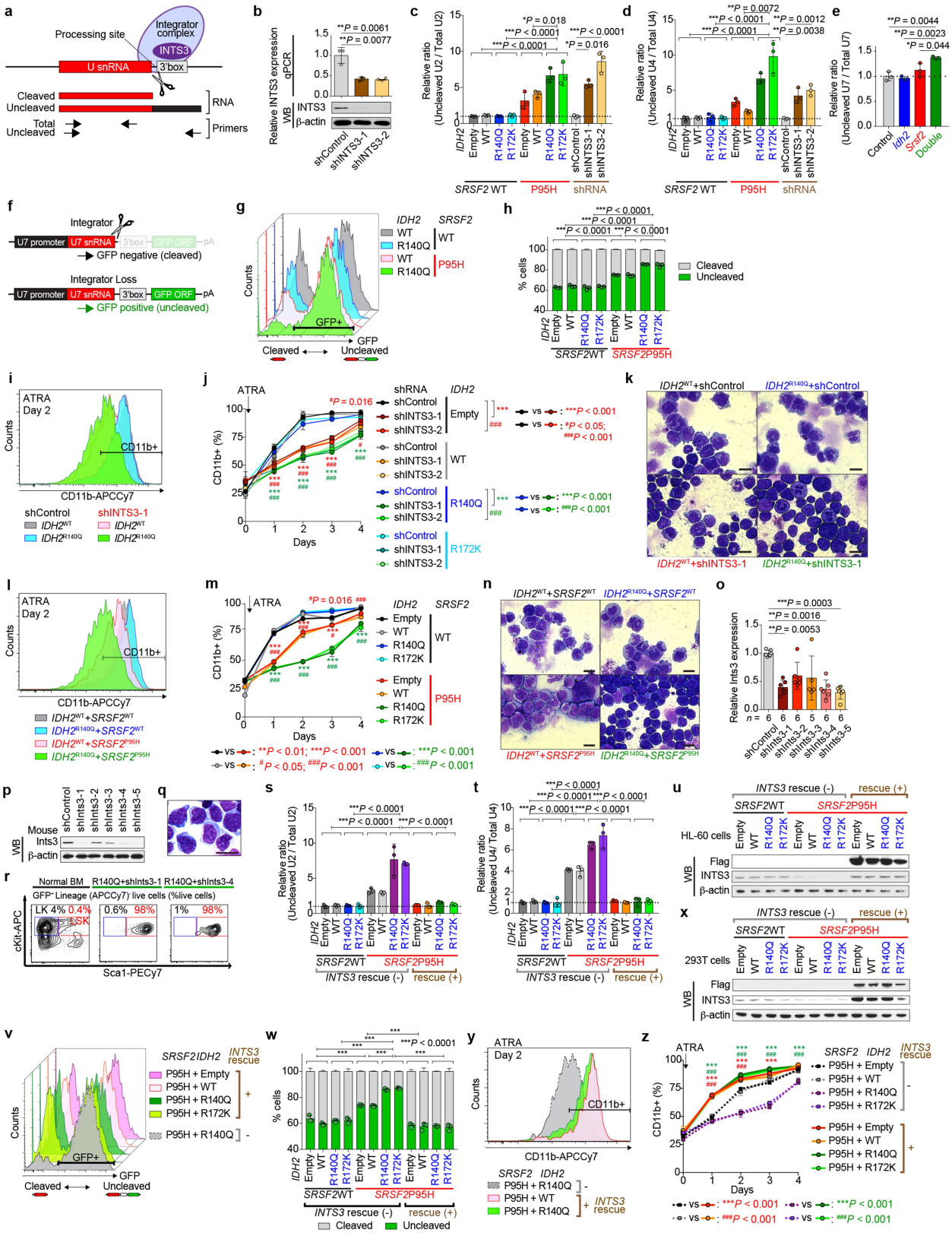
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Aberrant *INTS3* transcripts undergo nonsense-mediated decay, and effect 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 analysed is indicated; PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma; the line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; one-way ANOVA with Tukey's multiple comparison test). **d**, Sanger sequencing of cDNA showing wild-type 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 western blot analysis of INTS3 in isogeneic HL-60 cells with various combinations of *IDH2* and *SRSF2* mutations. IR: intron retention; ES: exon skipping. Representative results from three biologically independent experiments with similar results. **f**, RT–PCR and western blot of INTS3 in non-isogenic myeloid leukaemia cell lines. *SRSF2* genotypes are shown together (representative results from three independent experiments with similar results). **g**, Western blot 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 ($n = 3$; mean ± s.d.; a two-sided Student's *t*-test). **k**, **l**, Western blot analysis of protein lysates of samples from patients with AML with the indicated *IDH2* and *SRSF2* genotypes (**k**). Expression level of each integrator subunit was quantified using ImageJ and relative expression levels are shown in **l**, in which 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* and *SRSF2* double-mutant AML; detailed information of the primary patient samples used for this analysis is provided in Supplementary Table 23; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **m**, Western blot analysis of protein lysates from isogenic K562 cells with indicated *IDH2* and *SRSF2* genotypes (left) or with INTS3 knockdown (right) (representative results from three biologically independent experiments are shown). **n**, Western blot analysis of murine Lin^-KIT^+ BM cells at 12 weeks post-pIpC based on *Idh2* and *Srsf2* mutant genotypes. Expression level of INTS3 was quantified using ImageJ and relative expression levels are shown below; $n = 2$ mice per genotype were analysed. **o**, Correlation among indicated Integrator subunits and *P* value were calculated in Excel and R^2 values are visualized as a heat map 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). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.



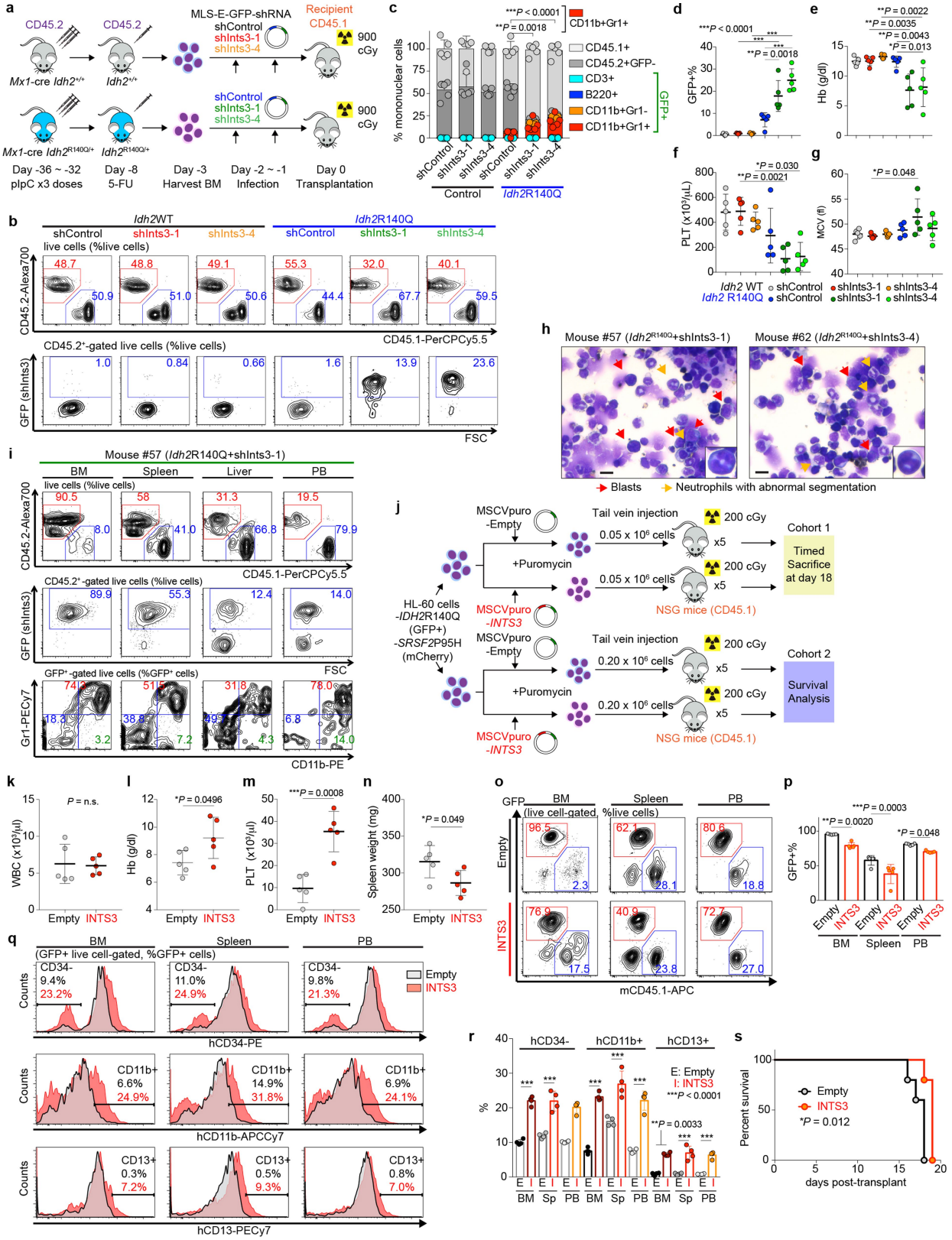
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | DNA hypermethylation at *INTS3* enhances *INTS3* mis-splicing, which is associated with 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 ($n = 3$; mean percentage ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **j**, Mean percentage of methylated CpGs at *ARID3A* in samples from patients with AML with indicated genotypes determined by eRRBS ($n = 3$ patients per genotype), followed by IGV plots of RNA-seq data of *ARID3A* from the TCGA. **k**, Results of eRRBS ($n = 1$ per genotype) and RNAPII-Ser2P ChIP-walking experiments are represented as shown in Fig. 3f ($n = 3$; 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 containing 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 +1,000 from TSSs) (the line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; each box plot was made by analysing ChIP–seq data from one cell line; two-sided Student's *t*-test). **o**, Metagene plots showing genome-wide RNAPII-Ser5P occupancy in primary samples from patients with AML with indicated genotypes (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 line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; each box plot was made by analysing 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* and *SRSF2* wild-type control and *SRSF2* single-mutant AML determined by RNAPII-Ser2P ChIP–seq (*y* axis, $\log_2$(counts per million); the line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; each box plot was made by analysing ChIP–seq data from one primary AML sample; one-way ANOVA with Tukey's multiple comparison test). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.



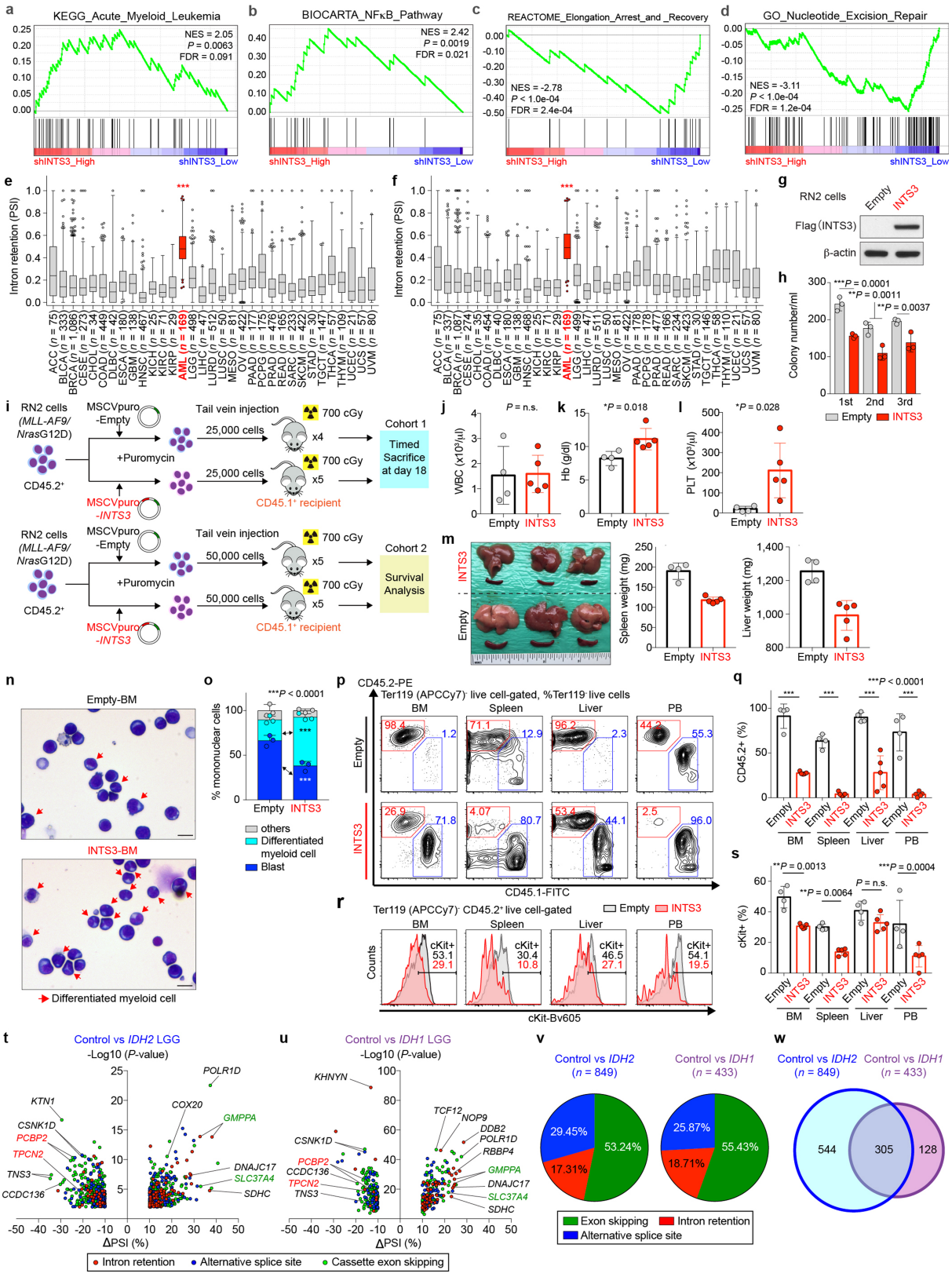
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Loss of INTS3 impairs uridine-rich small nuclear RNA processing and blocks myeloid differentiation. **a**, Schematic of snRNA processing site and qPCR primers for detecting cleaved or uncleaved snRNA. **b**, qPCR (top; $n = 3$; mean ± s.d.; a two-sided Student's *t*-test) and representative western blot of INTS3 in HL-60 cells transduced with 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 ($n = 3$, 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 versus 3, 2 versus 3, 1 versus 4, 2 versus 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* and *INTS3* constructs as labelled on the right (representative results from three biologically independent experiments are shown). **h**, **w**, Quantification of per cent GFP^- and GFP^+ 293T cells ($n = 3$ biologically independent experiments, 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 ($n = 3$; 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 of *Ints3* (**o**) (mean ± s.d.; Kruskal–Wallis tests with uncorrected Dunn's test) and western blot of INTS3 (**p**) in Ba/F3 cells transduced with shRNAs targeting mouse *Ints3*. **q**, **r**, Representative cytomorphology (**q**) and immunophenotype (**r**) of colony cells at the sixth 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**, Western blot of proteins extracted from HL-60 cells (**u**) assayed in **s**-**t** and **y**-**z** and 293T cells (**x**) assayed in **v** and **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 | See next page for caption.

Extended Data Fig. 9 | Mutant Idh2 cooperates with Ints3 loss to generate a lethal myeloid neoplasm in vivo. a, Schematic of shRNA targeting *Ints3* (shInts3) retroviral BM transplantation model. **b**, Flow cytometry data showing the chimerism of $CD45.2^+$ versus $CD45.1^+$ (top) or GFP^+ (bottom) cells in peripheral blood at four 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 four weeks post-transplant ($n = 5$ per group; mean + s.d.; represented by lines above the box. statistical significance was detected in percentage of $CD11b^+Gr1^+$ cells; by two-way ANOVA with Tukey's multiple comparison test). **d**–**g**, Chimerism of GFP^+ cells in peripheral blood (**d**) and blood counts of recipients at four weeks post-transplant (Hb (**e**); PLT (**f**); MCV (**g**); $n = 5$ per group; 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 peripheral blood from $Idh2^{R140Q}$ mice treated with shInts3 (representative results from five mice). **j**, Schematic of HL-60 xenograft model in which recipient mice from cohort 1 were euthanized 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 (mean ± s.d.; $n = 5$ per group; a two-sided Student's *t*-test). **o**, **p**, Representative flow cytometry data of BM, spleen, and peripheral blood 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; mean + s.d.; two-way ANOVA with Sidak's multiple comparison test). **q**, **r**, Representative flow cytometry data of BM, spleen and peripheral blood 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; 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 | See next page for caption.

Extended Data Fig. 10 | Gene expression and biological consequences of INTS3 loss, and effect of *IDH1* and *IDH2* mutations on splicing in low-grade glioma. a–**d**, 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 leukaemogenesis and myeloid differentiation (**a**), oncogenic signalling 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 analysed 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 analysed 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, oesophageal carcinoma; GBM, glioblastoma multiforme; 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 tumours; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma. The line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; 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**, Western blot analysis confirming overexpression of 3× Flag-tagged INTS3 in RN2 (*MLL-AF9* $Nras^{G12D}$) leukaemia cells (representative results from three biologically independent experiments). **h**, Colony numbers from serial replating assays of RN2 cells with or without INTS3 overexpression ($n = 3$; 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 in which recipient mice from cohort 1 were euthanized 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 (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); 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; mean percentage + s.d.; two-way ANOVA with Sidak's multiple comparison test). **p**, **q**, Representative flow cytometry analysis of BM, spleen, liver, and peripheral blood (**p**) and percentages of $CD45.2^+$ cells in $Ter119^-$ live cells (**q**) in recipient from cohort 1 ($n = 4$ (empty); $n = 5$ (INTS3); mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **r**, **s**, Representative flow cytometry analysis showing KIT expression in RN2 cells with or without INTS3 overexpression (**r**) and quantification of KIT^+ cells (**s**) from cohort 1 ($n = 4$ (Empty); $n = 5$ (INTS3); 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 analysed 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* and *IDH2* mutant genotypes. 'Control' represents LGG with wild-type *IDH1* and *IDH2*. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

nature research

Corresponding author(s): Omar Abdel-Wahab

Last updated by author(s): Jul 17, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No code was used for data collection.
Data analysis	The inclusion ratios of alternative exons or introns were estimated by using PSI-Sigma. GraphPad Prism 7 software was used to analyze the data and to make figures. RNA-seq reads were aligned by using 2-pass STAR 2.5.2a. Samtools (1.3.1) were used to generate variant call format (VCF) files for 6 target genes: IDH1, IDH2, 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. Motif analysis was done by using MEME SUITE. The heatmaps and sample clustering were done by using MORPHEUS (software.broadinstitute.org/morpheus/). VCF files from the TruSightTM Myeloid 54 gene panel from Illumina were analyzed using Illumina's Variant Studio software while those from a 40 gene panel (Oncomine Myeloid Research Assay; ThermoFisher), processing eight samples per Ion 530 chip on the IonTorrent platform were analyzed using the Ion Reporter software. ChIP-seq reads were mapped to the genome by calling Bowtie v1.0.048 with the arguments '-v 2 -k 1 -m 1 --best --strata'. Peaks were called using MACS2 v2.1.1.2016030952 against input control libraries with P < 1e-5 and subsequently filtered to remove peaks contained within ENCODE blacklisted regions and the mitochondrial genome. Subsequent data analysis was performed with Bioconductor in the R programming environment. Consensus peaks between samples were called using the soGGI package v1.14.0. Peaks were annotated using the ChIPseeker package v1.18.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The RNA sequencing data have been deposited in NCBI Sequencing Read Archive (SRA) under accession number SRP133673.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the in vivo experiments, the number of mice in each experiment was chosen to provide 90% statistical power with a 5% error level. For the RNA-seq experiments, maximal sample sizes available were used to obtain statistical power as much as possible to detect significant splicing alterations.
Data exclusions	No data were excluded from the analyses.
Replication	The experiments were repeated at a minimum of 3 times for all the in vitro experiments. All attempts at replication were successful.
Randomization	Animals were assigned to experimental group based on genotype and there was no drug treatment groups, therefore randomization was not utilized.
Blinding	For survival and blood count analyses of mice, actual measurements were carried out by a member of the lab who did not have knowledge of which alleles were expected to alter survival or blood count parameters. All other experiments were not blinded and it was not necessary to be as they were less subjective.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Flow cytometry 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); 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).
Western blotting, DNA dot blot assays, and ChIP: 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).

Validation

All antibodies were validated by the supplier for human samples, and were checked in the lab by Western blotting on cell lysate and by comparing to the manufacturer's or in-house results.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

K562, HL-60, TF1, and HEK293T cells were obtained from the American Type Culture Collection (ATCC) . K052 cells were obtained from JCRB Cell Bank. Ba/F3 cells were obtained from DSMZ. 293 GPII cells were purchased from Clontech. The isogenic K562 cell lines with or without SRSF2 P95H were generated at Horizon Discovery. MLL-AF9/NrasG12D murine leukemia (RN2) cells were obtained from Dr. Iannis Aifantis (NYU School of Medicine).

Authentication

An aliquot of each cell lines were authenticated using ATCC/JCRB/DSMZ DNA fingerprinting.

Mycoplasma contamination

All cell lines are frequently tested for mycoplasma contamination. Cell lines used in this study were verified to be mycoplasma negative before undertaking any experiments with them.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

6-8 week female CD45.1 C57BL/6 mice were purchased from The Jackson Laboratory (Stock No: 002014) . Male and female CD45.2 Srsf2P95H/+ conditional knock-in mice, Idh2R140Q/+ conditional knock-in mice, and Tet2 conditional knockout mice (all on C57BL/6 background) were also analyzed and used as bone marrow donors.

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve samples collected from the field.

Ethics oversight

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 MSKCC.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics

The covariate-relevant population characteristics of the human research participants of Memorial Sloan Kettering Cancer Center (MSKCC), Université Paris-Saclay, and the University of Manchester are provided below: Samples were obtained from acute myeloid leukemia (AML) patients treated at MSKCC, Université Paris-Saclay, and the University of Manchester. All samples were viably frozen and used to extract DNA, RNA, and protein. All subjects with AML were eligible for inclusion regardless of age, sex, or race.

Recruitment

All the participants were recruited without knowing their genotypes. Samples were genotyped and classified based on IDH2/SRSF2 genotypes. Samples that had mutations in IDH1, SF3B1, U2AF1, or ZRSR2 were excluded from RNA-seq and targeted RNA and protein analyses. Then RNA-seq was performed to analyze the splicing alterations. Therefore, there was no self-selection bias and it is unlikely that bias, if any, impacted the splicing analysis.

Ethics oversight

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.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	The ChIP-seq data (and RNA-seq data and eRRBS data) have been deposited in NCBI Sequence Read Archive (SRA) under accession number SRP133673.
Files in database submission	Sample_WT-WT-614, Sample_WT-WT-247, Sample_RQ-WT-584, Sample_RQ-WT-475, Sample_WT-PH-547, Sample_WT-PH-343, Sample_RQ-PH-524, Sample_RQ-PH-475, k562_WT_input, k562_WT_si_PolII-Ser2, k562_WT_si_PolII-Ser5, k562_p95_input, k562_p95H_si_PolII-Ser2, k562_p95H_si_PolII-Ser5
Genome browser session (e.g. UCSC)	The ChIP-seq data (and RNA-seq data and eRRBS data) have been deposited in NCBI Sequence Read Archive (SRA) under accession number SRP133673.

Methodology

Replicates	Two primary AML patient samples per IDH2/SRSF2 genotype (IDH2/SRSF2 WT/WT, Mutant/WT, WT/Mutant. and Mutant/Mutant) were used for the ChIP-seq experiments.
Sequencing depth	An average of 75 million paired reads was generated per sample (125 bp single-end).
Antibodies	Antibodies used for ChIP were as follows: RNA polymerase II CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095) RNA polymerase II CTD repeat YSPTSPS (phospho S5) (Abcam; ab5408)
Peak calling parameters	Narrow peaks were called using the callpeak function from MACS2 v2.1.1.20160309 against matched input samples, using default parameters and a P-value cutoff of 1e-5, according to the ENCODE Histone ChIP-seq Data Standards and Processing Pipeline (https://www.encodeproject.org/chip-seq/histone/).
Data quality	For all samples, a P-value cutoff of 1e-5 against input was used. All peaks were called at a q-value of < 0.017. For each sample, the number of peaks with a fold-change > 5, and the average total number of peaks called is 19,200.
Software	ChIP-seq reads were mapped to the genome by calling Bowtie v1.0.048 with the arguments '-v 2 -k 1 -m 1 --best --strata'. Peaks were called using MACS2 v2.1.1.2016030952 against input control libraries with P < 1e-5 and subsequently filtered to remove peaks contained within ENCODE blacklisted regions and the mitochondrial genome. Subsequent data analysis was performed with Bioconductor in the R programming environment. Consensus peaks between samples were called using the soGGI package v1.14.0. Peaks were annotated using the ChIPseeker package v1.18.0.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	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.
Instrument	All the FACS sorting was performed on FACS Aria, and analysis was performed on an LSRII or LSR Fortessa (BD Biosciences).
Software	FlowJo Ver.9 was used for analysis of flow cytometry data.
Cell population abundance	To check the purity of GFP and/or mCherry positivity in post-sort samples, the sorted samples were analyzed for GFP and/or mCherry by FACS Aria (BD Biosciences), and samples with >99% purity were used for analyses.
Gating strategy	The FSC/SSC gates of the starting cell population was set in order to include all the lineages of mouse hematopoietic cells such as granulocytes, monocytes, and lymphocytes. Then doublet cells were excluded by SSC-H vs SSC-W and FSC-H vs FSC-W gating. The boundaries between “positive” and “negative” staining cell population were defined by using unstained and single color-stained controls that were prepared by staining the whole BM mononuclear cells from B6 mice at 8-12 weeks with antibodies against mouse CD11b. “Positive” staining cell population was defined as CD11b+ population and “negative” staining cell population was defined by unstained control. The boundary for each fluorescence was set between these “positive” and “negative” staining cell populations.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.