



Minor intron retention drives clonal hematopoietic disorders and diverse cancer predisposition

Daichi Inoue^{1,2,12}, Jacob T. Polaski^{3,12}, Justin Taylor^{4,12}, Pau Castel⁵, Sisi Chen², Susumu Kobayashi^{1,6}, Simon J. Hogg², Yasutaka Hayashi¹, Jose Mario Bello Pineda^{3,7,8}, Ettaib El Marabti², Caroline Erickson², Katherine Knorr², Miki Fukumoto¹, Hiromi Yamazaki¹, Atsushi Tanaka^{1,9}, Chie Fukui¹, Sydney X. Lu², Benjamin H. Durham¹⁰, Bo Liu², Eric Wang², Sanjoy Mehta¹⁰, Daniel Zakheim¹⁰, Ralph Garippa¹⁰, Alex Penson², Guo-Liang Chew¹¹, Frank McCormick⁵, Robert K. Bradley^{3,7} ✉ and Omar Abdel-Wahab² ✉

Most eukaryotes harbor two distinct pre-mRNA splicing machineries: the major spliceosome, which removes >99% of introns, and the minor spliceosome, which removes rare, evolutionarily conserved introns. Although hypothesized to serve important regulatory functions, physiologic roles of the minor spliceosome are not well understood. For example, the minor spliceosome component ZRSR2 is subject to recurrent, leukemia-associated mutations, yet functional connections among minor introns, hematopoiesis and cancers are unclear. Here, we identify that impaired minor intron excision via ZRSR2 loss enhances hematopoietic stem cell self-renewal. CRISPR screens mimicking nonsense-mediated decay of minor intron-containing mRNA species converged on LZTR1, a regulator of RAS-related GTPases. LZTR1 minor intron retention was also discovered in the RASopathy Noonan syndrome, due to intronic mutations disrupting splicing and diverse solid tumors. These data uncover minor intron recognition as a regulator of hematopoiesis, noncoding mutations within minor introns as potential cancer drivers and links among ZRSR2 mutations, LZTR1 regulation and leukemias.

Myelodysplastic syndromes (MDS) are clonal blood disorders characterized by impaired hematopoiesis, risk of transformation to acute myeloid leukemia (AML) and a paucity of effective treatments. More than 50% of patients with MDS carry a mutation affecting an RNA splicing factor^{1–3}. Additionally, splicing factor mutations are common to all forms of myeloid malignancies, including AML and myeloproliferative neoplasms. Splicing factor mutations in leukemia are concentrated in four genes (*SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2*). *SF3B1*, *SRSF2* and *U2AF1* are subject to heterozygous, change-of-function^{4–6} missense mutations affecting specific residues^{1–3}. By contrast, the X chromosome-encoded *ZRSR2* is enriched in nonsense and frameshift mutations in male patients, consistent with loss of function^{1–3}. For example, across 2,302 sequentially sequenced patients with myeloid neoplasm, 100% of patients with myeloid neoplasm (40 of 40) and somatic mutations in *ZRSR2* were male, and there were no females with *ZRSR2* mutations ($P < 0.00001$, Fisher's exact test; Extended Data Fig. 1a,b). By contrast, 57% of patients with wild-type (WT) *ZRSR2* and myeloid neoplasms were male (1,285 of 2,262 total patients). Moreover, *ZRSR2* (also known as U2AF35-related protein (Urp)) is the only protein of the four frequently mutated splicing factors that primarily functions in the minor spliceosome^{7,8}. While most introns are spliced by the major spliceosome ('U2-type introns'), a

small subset (<1%) of introns have distinct splice sites and branch-points that are recognized by a separate ribonucleoprotein complex, the minor spliceosome^{9–12}. Although minor ('U12-type') introns are present in only 700–800 genes in humans, their sequences and positions are highly evolutionarily conserved, more so than their U2-type counterparts¹³.

The unusually high conservation of minor introns across eukaryotes suggests that they play key regulatory roles¹³. In some cases, molecular roles were elucidated. For example, minor introns are less efficiently excised from pre-mRNA than are major introns. It was therefore postulated that minor introns serve as 'molecular switches' to regulate the expression of their host genes, in which the rate of removal of a single minor intron within a gene regulates expression of the entire host mRNA^{14,15}. However, relatively few specific functional roles have been elucidated for the minor spliceosome in regulating biological phenotypes.

Results

Zrsr2 loss promotes hematopoietic stem cell self-renewal. Given the recurrent nature of *ZRSR2* mutations in leukemias, we hypothesized that minor intron splicing might be particularly important in the hematopoietic system. The relative rarity and exquisite conservation of minor introns offered a unique opportunity to

¹Department of Hematology–Oncology, Institute of Biomedical Research and Innovation, Foundation for Biomedical Research and Innovation at Kobe, Kobe, Japan. ²Human Oncology and Pathogenesis Program, Memorial Sloan Ketterer Absolute numbers of live mature hematopoietic cells, New York, NY, USA. ³Public Health Sciences and Basic Sciences Divisions, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁴Sylvester Comprehensive Cancer Center at the University of Miami Miller School of Medicine, Miami, FL, USA. ⁵Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA. ⁶Division of Cellular Therapy, The Institute of Medical Science, the University of Tokyo, Tokyo, Japan. ⁷Department of Genome Sciences, University of Washington, Seattle, WA, USA. ⁸Medical Scientist Training Program, University of Washington, Seattle, WA, USA. ⁹Department of Immunology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan. ¹⁰Gene Editing & Screening Facility, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹¹Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore. ¹²These authors contributed equally: Daichi Inoue, Jacob T. Polaski, Justin Taylor. ✉e-mail: rbradley@fredhutch.org; abdelwao@mskcc.org

simultaneously investigate splicing factor mutations in malignant hematopoietic stem cell (HSC) disorders as well as to identify potential tissue-specific roles of the minor spliceosome. We therefore set out to understand the role of loss-of-function mutations in *ZRSR2* as observed in myeloid HSC disorders by generating a mouse model permitting time- and tissue-specific deletion of *Zrsr2* (Extended Data Fig. 1c,d; also encoded on the X chromosome in mice). Conditional Cre-mediated excision of exon 4 of *Zrsr2*, in a manner that results in an early frameshift, efficiently downregulated *Zrsr2* mRNA in long-term HSCs (LT-HSCs; lineage-negative CD150⁺CD48⁻c-Kit⁺Sca-1⁺ cells) and *Zrsr2* protein in the spleen (Fig. 1a and Extended Data Fig. 1e–h). This was accomplished by the generation of *Mx1-Cre;Zrsr2^{fl/y}* and *Mx1-Cre;Zrsr2^{fl/fl}* mice, as *Mx1-Cre* is a well-established system allowing for conditional, time-controlled and efficient deletion of genes in postnatal hematopoietic cells^{16–18}. Exon 4 was chosen for deletion because deletion of this exon causes a frameshift when skipped and this exon is present in all sequences coding for annotated *Zrsr2* isoforms and is highly conserved across species.

Prior work from several animal models of global deletion of core minor spliceosome components (including *Rnpc3* in zebrafish and mice^{19,20} and small nuclear (sn)RNA species in *Drosophila melanogaster*²¹) identified that the minor spliceosome is required for development and survival. However, tissue-specific deletion of a minor spliceosome component has not been performed. In contrast to the loss of viability that results from pan-tissue deletion of core minor spliceosome components *Rnpc3* (refs. ^{19,20}) and the U11 snRNA²², we found that hematopoietic-specific *Zrsr2* deletion in 6-week-old male and female mice enhanced proliferation as well as clonogenic capacity of *Zrsr2*-null HSCs in vitro (Fig. 1b,c and Extended Data Fig. 2a,b).

We therefore evaluated the in vivo self-renewal capacity of *Zrsr2*-null HSCs by performing bone marrow (BM) competitive transplantation assays, in which *Zrsr2* was deleted following stable reconstitution of hematopoiesis with equal numbers of CD45.2⁺ *Zrsr2*-floxed and CD45.1⁺ WT hematopoietic cells in CD45.1 recipient mice (Extended Data Fig. 2c; in these experiments, *Zrsr2* was deleted in recipient mice 4 weeks following transplantation via polyinosinic-polycytidylic acid (pIpC) administration to recipients). These assays revealed strikingly enhanced self-renewal of *Zrsr2*-deficient male and female hematopoietic cells, both in primary and secondary transplantation (Fig. 1d; such an effect was not observed in females with heterozygous deletion of *Zrsr2* (Extended Data Fig. 2d,e)). This was associated with increased numbers of *Zrsr2*-null mature B cells and myeloid cells in the blood (Extended Data Fig. 2f) as well as LT-HSCs and lineage-negative Sca-1⁺c-Kit⁺ (LSK) cells in the BM 16 weeks following primary and secondary transplantation (Fig. 1e and Extended Data Fig. 2g,h). Similar effects were observed in primary, non-transplanted *Mx1-Cre;Zrsr2*-knockout (KO) mice, in which deletion of *Zrsr2* increased numbers of LT- and short-term HSCs as well as downstream progenitor populations (Extended Data Fig. 3a; deletion was induced at 6 weeks of age). Interestingly, *Zrsr2* deletion was also associated with increased total BM mononuclear cells as well as LT-HSCs in the active phase of the cell cycle and undergoing apoptosis (Fig. 1f and Extended Data Fig. 3b,c). Given that these phenotypes are key features of human MDS, we also performed detailed morphological assessments of hematopoietic tissues from primary *Zrsr2*-KO mice, which revealed modest morphologic evidence of dysplasia (Extended Data Fig. 3d,e). Overall, *Mx1-Cre;Zrsr2*-KO mice had numerically hastened death compared to littermate controls, but this did not reach statistical significance, and there were no significant differences in blood counts based on genotype (Extended Data Fig. 3f,g). Other than increased numbers of mature B220⁺ cells in the blood and BM, numbers of B cell and T cell subsets in the BM, spleens and thymi of *Zrsr2*-KO mice were unperurbed (Extended Data Fig. 3h–m).

The above data, revealing enhanced self-renewal of *Zrsr2*-null HSCs, stand in stark contrast to recent work evaluating the effects of hotspot mutations in *SF3B1*, *SRSF2* and *U2AF1*, all of which revealed a perplexing impairment in self-renewal when these mutations were induced in mice using similar transplantation methods^{4,23,24}. We therefore repeated the above competitive transplantation assays using 6-week-old *Mx1-Cre;Sf3b1^{K700E/+}* and *Mx1-Cre;Srsf2^{P95H/+}* mice alongside *Mx1-Cre;Zrsr2^{fl/y}* mice (Extended Data Fig. 4a). We additionally included *Mx1-Cre;Tet2^{fl/fl}* mice, given the well-described effects of *Tet2* loss on increasing self-renewal and numbers of HSCs in this system¹⁶. Similar to the effects of *Tet2* deletion, *Zrsr2* loss again resulted in increased competitive advantage in vivo (Fig. 1g). This advantage was observed in the numbers and percentages of BM hematopoietic stem and progenitor cells (HSPCs) (Fig. 1h and Extended Data Fig. 4b–d). Moreover, the effect of *Zrsr2* loss was strikingly distinct from the effects of inducing leukemia-associated mutations in *Sf3b1* and *Srsf2*, which were associated with near-complete loss of hematopoiesis (Fig. 1g and Extended Data Fig. 4b). Interestingly, deletion of *Zrsr2* in the context of mutant *Sf3b1* rescued the impaired clonogenic capacity of *Sf3b1^{K700E/+}* hematopoietic precursor cells (Extended Data Fig. 4e,f). These data are quite distinct from prior reports of the lethal phenotype of combined *Sf3b1^{K700E}* and *Srsf2^{P95H}* mutations²⁵ and may explain the occasional co-occurrence of *SF3B1* and *ZRSR2* mutations in patients with myeloid neoplasm²⁶.

Global impairment in minor retention with ZRSR2 loss. We next sought to understand the mechanistic basis by which ZRSR2 loss causes aberrant HSC self-renewal and MDS. We performed high-coverage RNA-seq on BM samples from patients with MDS and diverse ZRSR2 mutations ($n=8$) and patients with MDS lacking any spliceosomal gene mutation ($n=10$) and quantified transcriptome-wide splicing patterns (Extended Data Fig. 5a and Supplementary Tables 1 and 2). ZRSR2-mutant samples were characterized by widespread, dysfunctional recognition of U12-type introns, with over one-third of U12-type introns exhibiting significantly increased retention (Fig. 2a,b). All ZRSR2-mutant samples exhibited U12-type intron retention, indicating that the diverse ZRSR2-mutant lesions represented in our cohort converge on loss of function (Extended Data Fig. 5b–d). Aberrant intron retention was specific to the minor spliceosome: we observed no transcriptome-wide association between ZRSR2 mutations and levels of U2-type intron retention (Extended Data Fig. 5e,f). We confirmed the robustness of our results by reanalyzing data from a published cohort of MDS samples (Extended Data Fig. 5g and Supplementary Table 3)⁷. ZRSR2-mutant samples in this cohort were similarly characterized by transcriptome-wide U12-type intron retention (Fig. 2c,d and Extended Data Fig. 5h). However, in contrast to our cohort, U2-type introns were also frequently retained in most ZRSR2-mutant samples (Extended Data Fig. 5i,j). As the two patient cohorts represented different ZRSR2 mutational spectra, exhibited both similarities and differences in their splicing programs and were orthogonally collected, we hypothesized that convergently mis-spliced genes might be particularly important for disease pathogenesis. We observed a striking overlap in aberrantly retained U12-type introns, with ~94% of all U12-type introns that were retained in our cohort also being retained in the previously published cohort⁷ (Fig. 2e). By contrast, very few (~4%) aberrantly retained U2-type introns were shared between cohorts (Fig. 2f).

Closer inspection of genes of potential disease relevance confirmed the robustness and specificity of U12-type intron retention in ZRSR2-mutant samples. Minor spliceosome-dependent genes such as *ATG3*, *TRAIIP* and *PARP1* exhibited striking retention of a single U12-type intron, with adjacent U2-type introns normally spliced (Fig. 2g and Extended Data Fig. 5k). In each case, the U12-type intron was specifically retained in ZRSR2-mutant

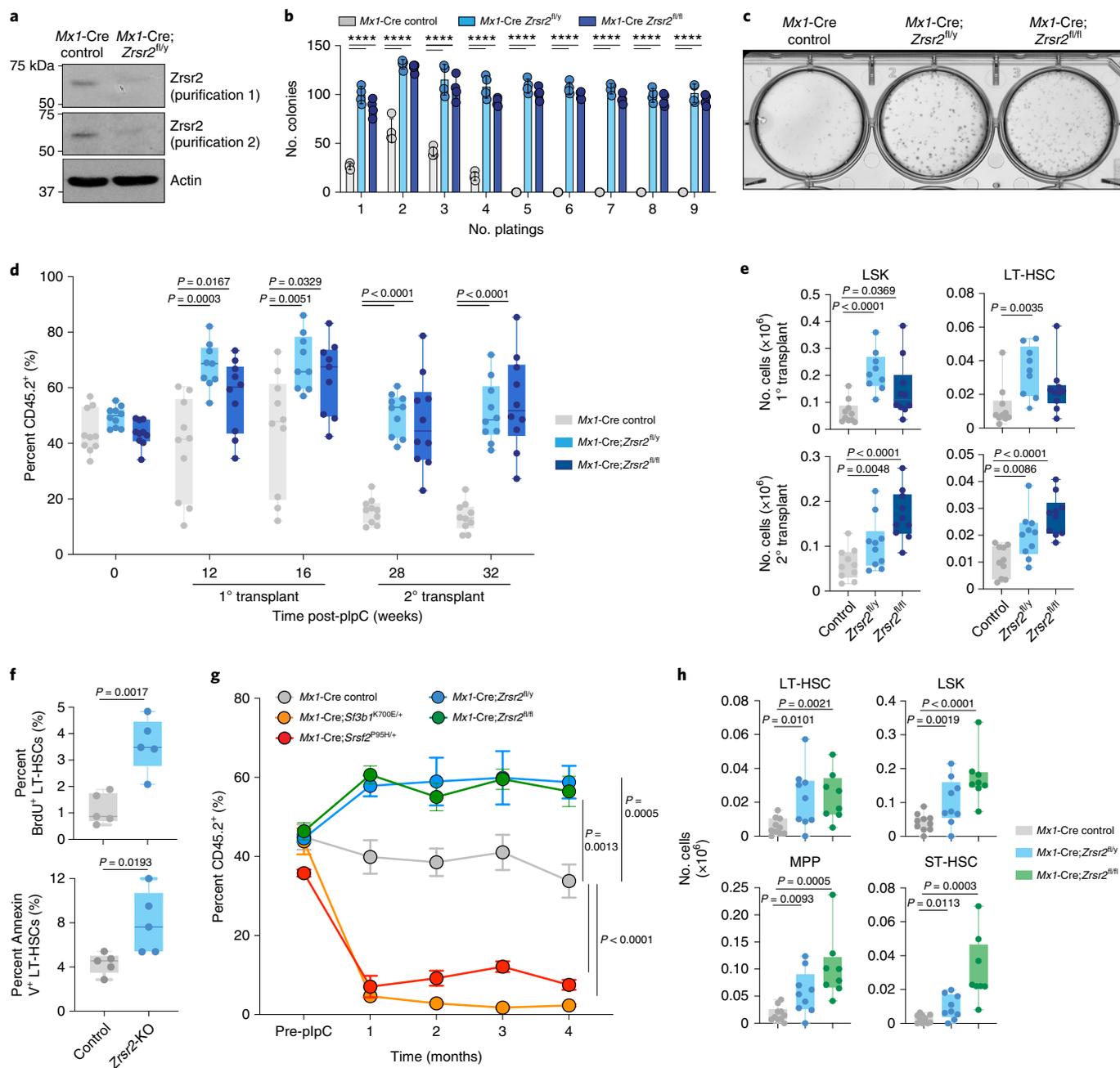


Fig. 1 | *Zrsr2* loss increases HSC self-renewal. **a**, Anti-*Zrsr2* western blot in the spleens of 6-week-old *Mx1-Cre* control and *Mx1-Cre;Zrsr2^{fl/y}* mice. The experiment was repeated three times with similar results. **b**, Number of colonies in methylcellulose colony-forming-unit assays from LT-HSCs from male and female *Zrsr2*-KO mice and controls. Mean values \pm s.d. are shown. $n = 4$ biologically independent experiments. P values were calculated by two-sided t -tests, **** $P < 0.0001$. **c**, Representative photograph of the initial methylcellulose plating from **b**. **d**, Box-and-whisker plots of the percentage of peripheral blood CD45.2⁺ cells in competitive transplantation assays after plpC administration using CD45.2⁺ *Mx1-Cre* control, *Mx1-Cre Zrsr2^{fl/y}* and *Mx1-Cre Zrsr2^{fl/fl}* mice. For box-and-whisker plots throughout, bars indicate medians, box edges indicate first and third quartile values, and whisker edges indicate minimum and maximum values. **e**, Box-and-whisker plots of absolute numbers of LSK cells and LT-HSCs in the BM of recipient mice following 16 weeks of primary (top) and secondary (bottom) transplantation from **d**. **f**, Box-and-whisker plots of percentages of bromodeoxyuridine (BrdU)⁺ and Annexin V⁺ LT-HSCs in the BM of primary 6-week-old *Zrsr2*-KO and control mice. **g**, Percentage of peripheral blood CD45.2⁺ cells in competitive transplantation assays before and after plpC administration using CD45.2⁺ *Mx1-Cre* control ($n = 10$), *Sf3b1^{K700E/+}* ($n = 9$), *Srsf2^{P95H/+}* ($n = 8$), *Zrsr2^{fl/y}* ($n = 10$) and *Tet2^{fl/fl}* ($n = 10$) mice. Error bars represent mean values \pm s.e.m. P values were calculated by two-sided t -tests, using the values at 4 months. **h**, Absolute numbers of CD45.2⁺ LSK cells, LT-HSCs, multipotent progenitors and short-term (ST)-HSCs in the BM of recipient mice following 16 weeks of competitive transplantation from **g**. P values were calculated relative to the control group by two-sided t -tests and are indicated in the figures.

MDS, but not in MDS without spliceosomal mutations, relative to normal BM in both cohorts (Fig. 2h and Extended Data Fig. 5l). To further evaluate the specificity of minor intron splicing across

leukemia-associated spliceosomal gene mutations, we quantified the number of retained U12-type introns across patients with bona fide deleterious mutations in *ZRSR2* versus known hotspot

mutations in *SF3B1*, *SRSF2* and *U2AF1* in the 427 patients with AML from the Beat AML study²⁷. This analysis revealed a striking enrichment for U12-type intron retention in patients with mutated *ZRSR2* but no such enrichment for patients with *SF3B1*, *SRSF2* or *U2AF1* mutations (Fig. 2i). U12-type intron retention was generally associated with downregulated mRNA expression of U12-type intron-containing genes (Extended Data Fig. 5m).

Consistent with the effect of *ZRSR2* mutations in MDS on splicing, hematopoietic precursors (lineage-negative c-Kit⁺ cells) from mice deficient for *Zrsr2* exhibited global increases in U12-type intron retention without significant changes in U2-type splicing (Fig. 2a,j,k; RNA-seq analysis was performed 2 months following pIpC administration to 6-week-old *Mx1-Cre;Zrsr2^{fl/y}* and *Mx1-Cre* control mice). Moreover, a number of overlapping U12-type introns were similarly retained in the BM of *Zrsr2*-KO mice but not in control mice, consistent with the high conservation of U12-type introns (Fig. 2a,l and Supplementary Table 4). These comparisons strongly suggest that U12-type intron retention is the central molecular phenotype of cells with *ZRSR2* mutations and that U2-type intron retention is a sporadic occurrence that only characterizes a subset of cases of *ZRSR2*-mutant MDS.

Mapping direct RNA binding targets of *ZRSR2*. As the above data identify a strong functional link between *ZRSR2* and minor introns, we next sought to understand the mechanistic basis for this relationship by identifying direct binding targets of *ZRSR2* in RNA. We therefore performed anti-*ZRSR2* eCLIP-seq²⁸ (enhanced UV cross-linking immunoprecipitation followed by next-generation sequencing) in human myeloid leukemia (K562) cells (Extended Data Fig. 5n,o). This revealed that ~80% of *ZRSR2*-binding sites mapped to exons, with a highly significant enrichment for *ZRSR2* binding to minor intron-containing genes, the minor introns of which were responsive versus unresponsive to *ZRSR2* mutations (Fig. 3a,b and Extended Data Fig. 5p; P value $< 2.2 \times 10^{-16}$ with an odds ratio in the range (95% confidence interval) of 2.1–2.5). *ZRSR2* binding was specifically enriched in minor introns, consistent with our analyses of the effects of *ZRSR2* loss on minor intron retention (Fig. 3c). Finally, *ZRSR2*-bound mRNA was enriched for mRNA encoding RNA regulatory proteins as well as for that corresponding to genes with known involvement in leukemia and protein processing and translation (Fig. 3d). Overall, these analyses demonstrate that minor intron-containing genes with splicing regulated by *ZRSR2* are direct binding targets of *ZRSR2*.

Both RNA-seq and eCLIP-seq analyses above identified that approximately only one-third of U12-type intron-containing genes are sensitive to loss of *ZRSR2*. To understand the specificity of *ZRSR2* for regulating the splicing of minor introns and why only a

portion of minor introns are regulated by *ZRSR2*, we next evaluated the sequence features of introns that were retained upon *ZRSR2* loss.

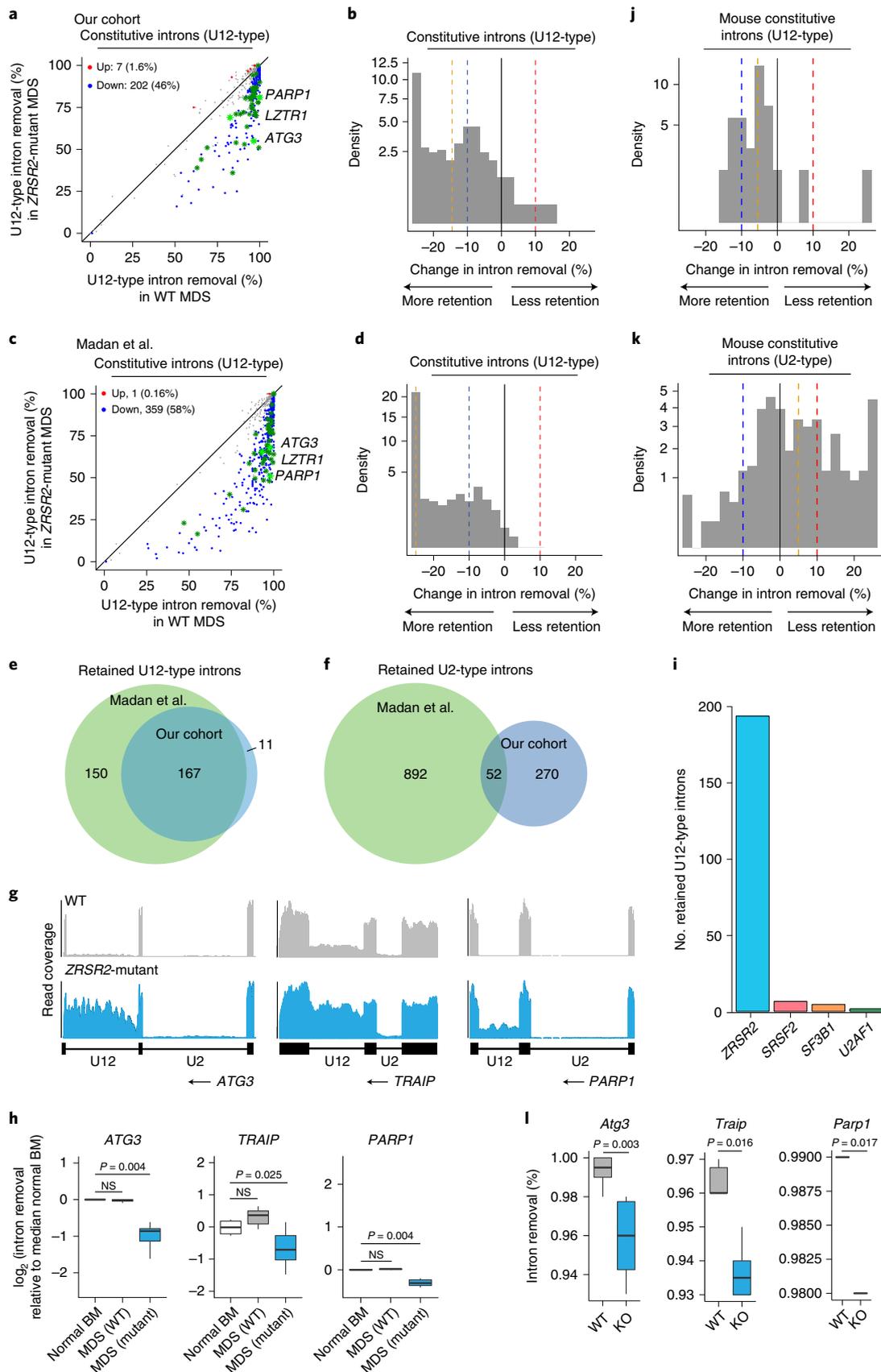
Characteristics of *ZRSR2*-regulated introns. We previously reported that, while branchpoints within U2-type introns are highly constrained in their location, branchpoints within U12-type introns exhibit a bimodal distribution, such that half of U12-type introns have branchpoints similar in location to U2-type branchpoints, while half of U12-type branchpoints occur in closer proximity (within 20 nucleotides) to the 3' splice site (3'ss)²⁹ (Fig. 3e). To test whether this bimodality was relevant to *ZRSR2* responsiveness, we augmented our previously published branchpoint annotation by querying available RNA-seq data from cohorts within the Cancer Genome Atlas (TCGA) to search for lariat-derived reads that span the 5' splice site-branchpoint junction within minor introns. Such reads are extremely rare due to typically rapid lariat degradation (hence the need for an extremely large-scale analysis) but allow for inference of branchpoint location with nucleotide-level resolution. Using this large U12-type branchpoint annotation, we discovered that introns that responded to *ZRSR2* loss had branchpoints that were significantly more proximal to the 3'ss than did unresponsive introns (two-sided Kolmogorov–Smirnov test, $P < 2.2 \times 10^{-16}$; Fig. 3f). By contrast, unresponsive U12-type introns exhibited no such spatially restricted enrichment, suggesting that branchpoint location influences U12-type intron susceptibility to retention in the absence of *ZRSR2* (ref. 29). We therefore examined the branchpoint more closely as a potential determinant of response to *ZRSR2* loss. This revealed that *ZRSR2*-responsive introns preferred adenosine nucleotides as branchpoints (Fig. 3g; $P = 1.5 \times 10^{-5}$ by two-sided binomial proportion test), had more branchpoints per intron than *ZRSR2*-unresponsive introns (Fig. 3h; $P = 0.03$ by two-sided t -test) and had branchpoints that more closely matched the U12 snRNA consensus sequence (Fig. 3i; $P = 5.2 \times 10^{-16}$ by Wilcoxon rank-sum test). *ZRSR2*-responsive minor introns additionally had less-defined polypyrimidine tracts and a reduced preference for G at the +1 position than unresponsive introns (Fig. 3j,k and Extended Data Fig. 5q). Overall, these data demonstrate that U12-type introns fall into two classes: those that are resistant to *ZRSR2* loss and those that respond strongly to *ZRSR2* loss. Responsive introns are typically characterized by a 3'ss-proximal, adenosine branchpoint that is surrounded by nucleotides that closely resemble the U12 snRNA consensus, as well as by having a weak or absent polypyrimidine tract.

Positive-enrichment screen of *ZRSR2*-regulated events. One major challenge with understanding how RNA splicing factor mutations cause disease is determining whether the observed phenotypes are causally linked to one or many aberrant splicing changes. Our transcriptomic analyses revealed that only a subset of

Fig. 2 | Widespread minor intron retention with *ZRSR2* loss. **a**, Genome-wide quantification of differential splicing of U12-type introns from our patient cohort with key overlapping mis-spliced mRNA species in mouse hematopoietic precursors (green). Each point corresponds to a single intron and illustrates the percentage of mRNA in which the intron is spliced out. Blue or red dots indicate introns that exhibit significantly increased or decreased retention in *ZRSR2*-mutant versus -WT cells, defined as an absolute change in retention of $\geq 10\%$ or absolute log fold change of ≥ 2 with associated $P \leq 0.05$. Green asterisks, minor introns differentially retained in both patients and mice; light green asterisks, minor introns in genes of particular interest (*Parp1*, *Lztr1* and *Atg3*). P values were calculated by two-sided Mann–Whitney U tests without adjustment for multiple comparisons. **b**, Distribution of U12-type intron retention in *ZRSR2*-mutant samples ($n = 8$ mutant and $n = 10$ WT samples). Blue and red dashed lines, thresholds of -10% and 10% for differential retention; gold line, median change in intron retention. **c**, As for **a** but for the Madan et al.⁷ cohort. **d**, As for **b** but for the Madan et al.⁷ cohort. **e**, Overlap of U12-type intron retention events between the two patient cohorts ($n = 8$ mutant and $n = 4$ WT samples). **f**, As for **e** but for U2-type introns. **g**, RNA-seq coverage plots of U12-type intron retention but not U2-type intron retention in our patient cohort. Plots were averaged over all samples with the indicated genotypes. **h**, Box plots quantifying splicing efficiencies of introns illustrated in **g** relative to normal marrow (median of four normal samples from Madan et al.⁷). P values were calculated by two-sided Mann–Whitney U tests. NS, not significant. **i**, Numbers of retained U12-type introns in samples bearing any of the four spliceosomal gene mutations from the Beat AML cohort. Distributions of U12-type (**j**) and U2-type (**k**) intron retention in *Zrsr2*-KO relative to WT mouse precursors. Blue and red dashed lines indicate thresholds of -10% and 10% used for differential retention; gold line indicates the median change in intron retention. **l**, Box plots illustrating intron splicing efficiencies in mouse lineage-negative c-Kit⁺ cells. For **h,l**, the middle line, hinges, notches and whiskers indicate medians, 25th and 75th percentiles, 95% confidence intervals and the most extreme data points within 1.5 \times interquartile range from the hinge. P values were calculated by two-sided Mann–Whitney U tests.

minor spliceosome-dependent genes are recurrently and robustly mis-spliced, suggesting that not all U12-type introns are equally important for disease pathogenesis.

We therefore systematically evaluated the potential role of each U12-type intron retention event in cell transformation with a functional-genomics screen. We mimicked the effects of nonsense-mediated



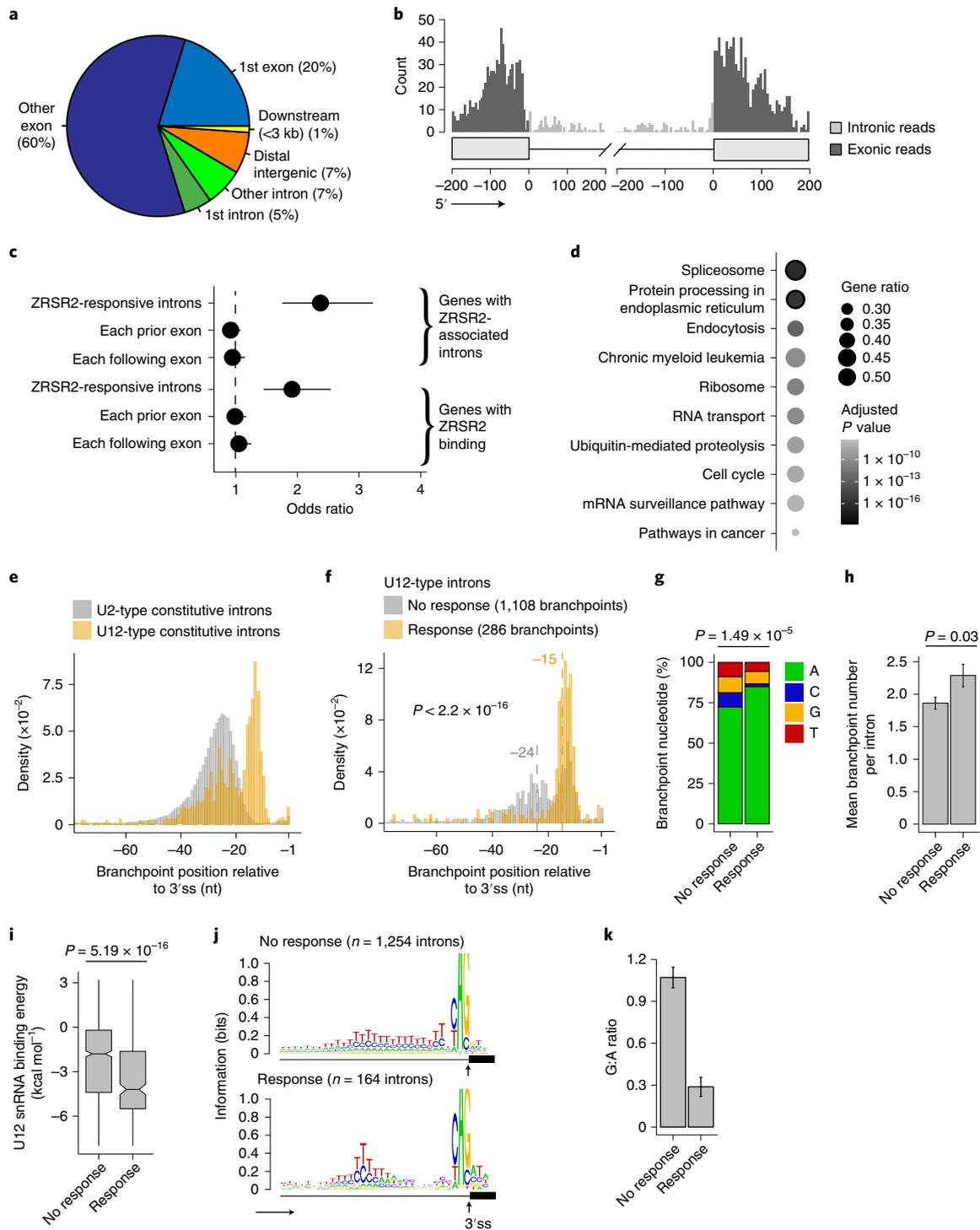


Fig. 3 | ZRSR2 RNA binding targets and features of ZRSR2-responsive introns. **a**, Genomic distribution of ZRSR2 eCLIP-seq peaks. **b**, Metaplot of ZRSR2 eCLIP sequencing reads at ZRSR2-regulated minor introns. **c**, Fisher's exact test analysis evaluating the enrichment of ZRSR2 within responsive introns and each flanking exon by eCLIP-seq in genes with ZRSR2-responsive introns or those with ZRSR2 binding. **d**, Gene ontology analysis of ZRSR2-bound genes by eCLIP-seq. **e**, Histogram of the locations of branchpoints relative to the 3'ss in U2- versus U12-type constitutive introns. nt, nucleotide. **f**, Histogram of the locations of branchpoints relative to the 3'ss for ZRSR2-unresponsive versus -responsive minor introns (P value was estimated by a two-sided Kolmogorov–Smirnov test). **g**, Branchpoint nucleotide preference for ZRSR2-unresponsive versus -responsive minor introns (P value was estimated with a two-sided binomial proportion test for a difference in the fraction of adenine branchpoints). **h**, Mean number of branchpoints within ZRSR2-unresponsive versus -responsive minor introns (P value was estimated by a two-sided t -test). Error bars represent ± 1 s.e.m. ($\text{s.d.} \times (\sqrt{n})^{-1}$). **i**, U12 snRNA binding energy for branchpoint motifs in ZRSR2-unresponsive versus -responsive minor introns (P value was estimated by a two-sided Mann–Whitney U test). **j**, Sequence logo plots of the 3'ss of ZRSR2-responsive versus -unresponsive introns (upward arrow indicates site of the 3'ss). As shown, ZRSR2-responsive introns have weaker or less-defined polypyrimidine tracts. **k**, G:A ratio at the +1 position (relative to the 3'ss). Error bars represent ± 1 s.d. estimated by bootstrapping (10,000 iterations). P value = 0 by two-sided Mann–Whitney U test.

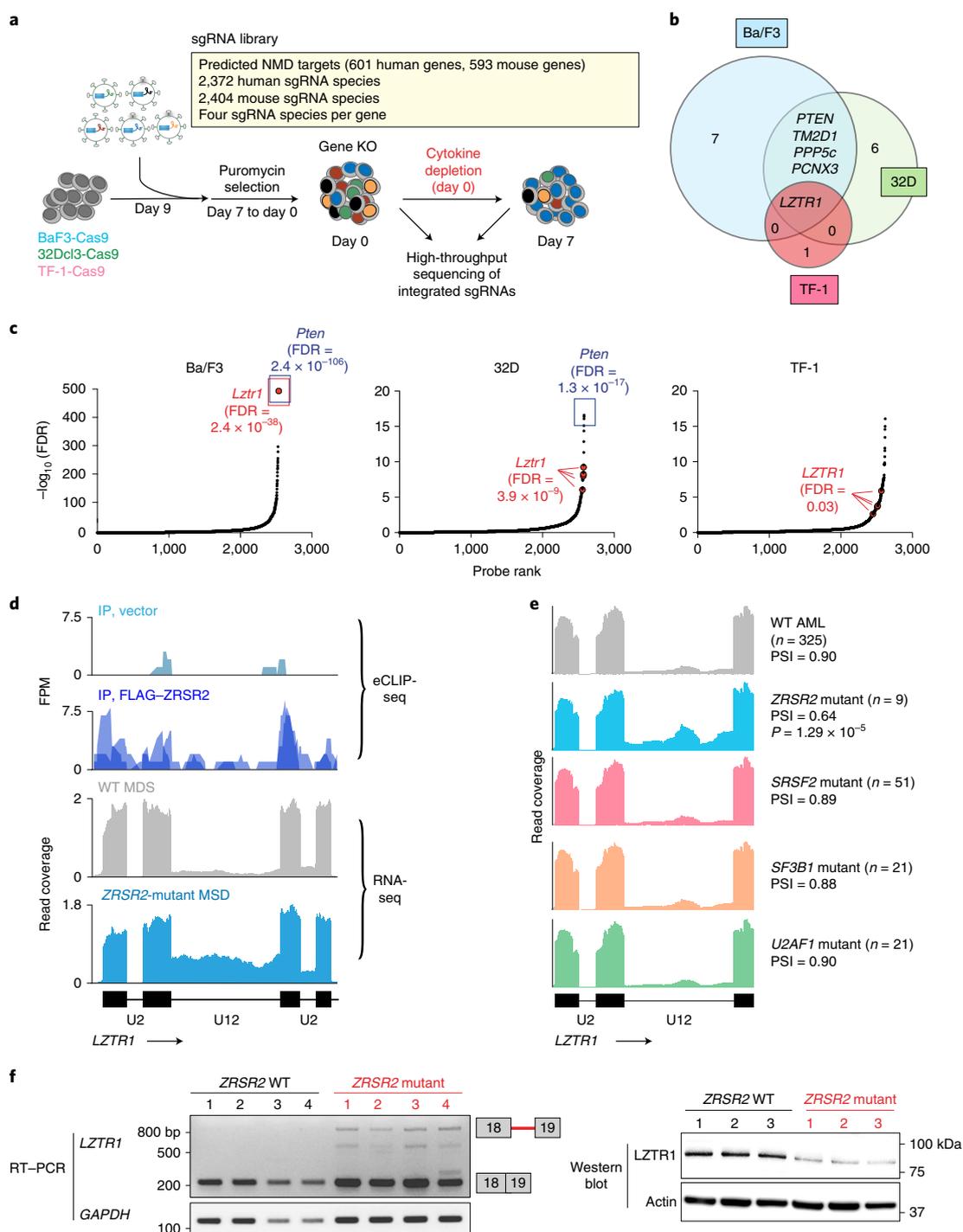


Fig. 4 | Impaired LZTR1 minor intron excision confers competitive advantage. **a**, Schematic of a positive-enrichment custom CRISPR-Cas9 pooled lentiviral screen to identify functionally important ZRSR2-regulated minor intron splicing events. **b**, Venn diagram of genes targeted by sgRNA species in the screen in **a** across Ba/F3, 32D and TF-1 cells. Numbers of genes identified in the screen in each segment of the Venn diagram are displayed. **c**, Rank plot for the $-\log_{10}$ (false discovery rate (FDR)) associated with each sgRNA in the screen from **a**. sgRNA species targeting the positive control (*Pten*) and *Lztr1* are highlighted. For the probe-level (per-sgRNA) analysis, we fitted a negative binomial generalized log-linear model and performed a likelihood-ratio test. False discovery rates were computed using the Benjamini-Hochberg method. **d**, Top, eCLIP-seq for FLAG immunoprecipitation (IP) from cells expressing FLAG empty vector or FLAG-ZRSR2 over the region of the minor intron of *LZTR1*. FPM, fragments per million mapped fragments. Biological duplicate immunoprecipitation data are overlaid. Bottom, RNA-seq coverage over the same locus in primary human MDS samples, with WT MDS data colored gray (top) and ZRSR2-mutant MDS data colored blue (bottom). **e**, RNA-seq coverage plots across the U12-type intron of *LZTR1* for patients from Beat AML with the indicated genotypes. Each coverage plot represents an average across all samples with the indicated genotype, following normalization to the total number of reads mapped to the coding genes for each sample. PSI, fraction of spliced mRNA. **f**, Qualitative reverse transcriptase (RT)-PCR gel for *LZTR1* minor intron excision (left) and *LZTR1* protein levels (right) in representative samples from patients with MDS and WT or mutant ZRSR2 ($n = 4$ distinct patient samples or genotypes for RT-PCR and $n = 3$ distinct patient samples or genotypes for the western blot).

decay (NMD) caused by loss of minor intron splicing and subsequent open reading frame disruption in *ZRSR2*-mutant cells via CRISPR–Cas9-mediated gene KO. In this assay, the protein-coding

region of each of the 601 genes, for which corresponding mRNA was identified as differentially spliced in *ZRSR2*-mutant MDS patient samples versus spliceosomal wild-type MDS patient samples

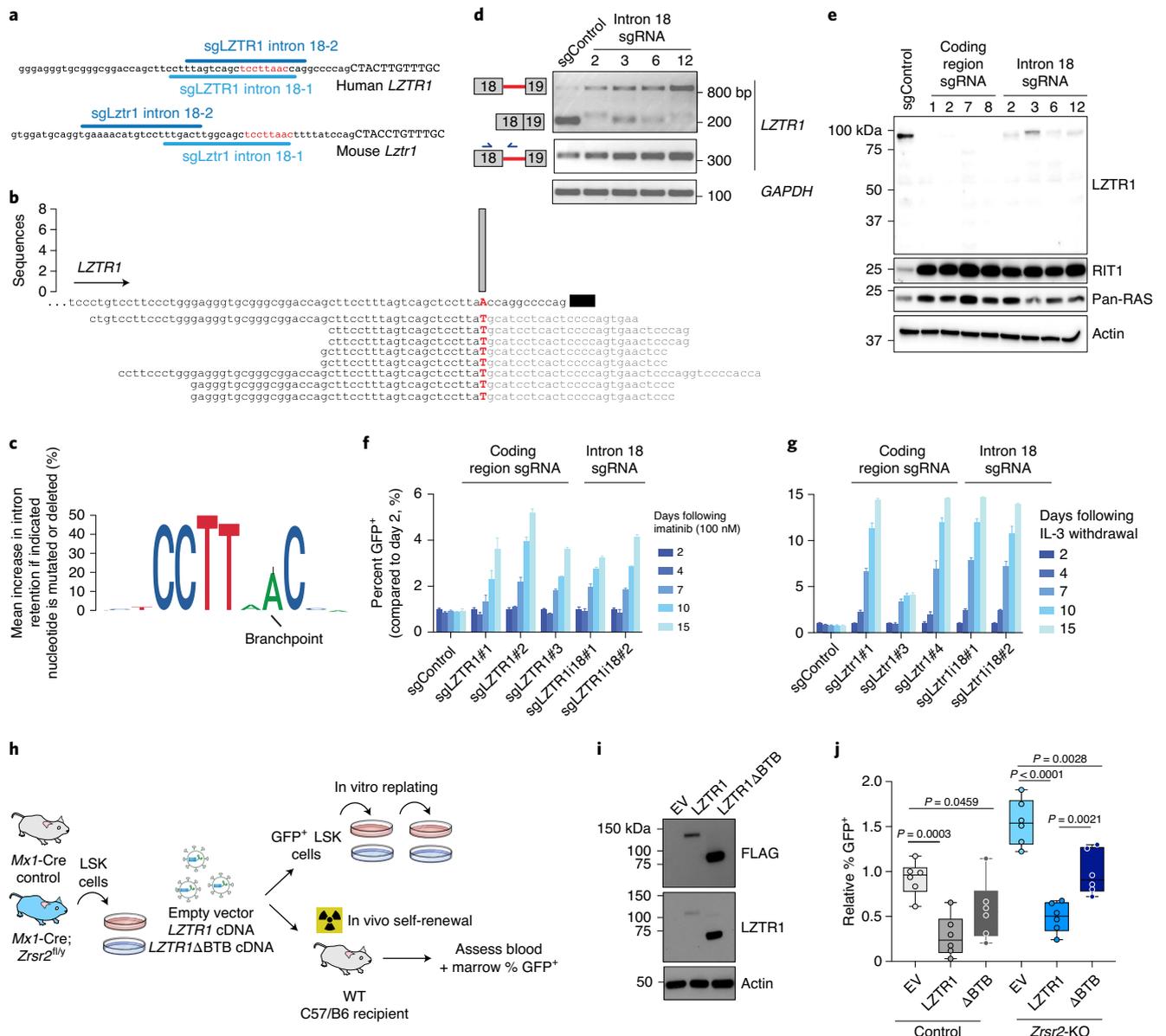


Fig. 5 | Impaired *LZTR1* minor intron excision promotes clonal advantage and effects of *LZTR1* re-expression on *Zrsr2*-null hematopoietic cells.

a, Schematic of the minor intron branchpoint binding region in *LZTR1* intron 18 with illustration of intronic sgRNA binding sequences (the U12 conserved sequence is in red). **b**, Branchpoint within the minor intron of *LZTR1* based on intron lariat-derived RNA-seq reads. The bar represents the number of supporting high-confidence reads, defined as those with a single identifying mismatch at the branchpoint characteristic of traversal of the 2'-5' linkage. **c**, Logo plot representation of minigene experiments summarizing mean increases in intron retention if the indicated nucleotide is mutated or deleted (%). The height of each nucleotide indicates its requirement for normal excision of the minor intron of *LZTR1*. **d**, RT-PCR for *LZTR1* intron 18 excision in K562 AML single-cell clones treated with intron 18-targeting sgRNA species. The experiment was repeated three times with similar results. **e**, Western blot for full-length *LZTR1* using antibodies specific to the N terminus as well as K-, N- and H-RAS ('pan-RAS') in K562 single-cell clones treated with sgRNA species targeting the protein-coding versus intronic sequence of *LZTR1*. The experiment was repeated three times with similar results. **f** and **g**, Percentages of GFP⁺ K562 cells from **d** mixed with equal proportions of unlabeled cells sensitive to imatinib (**f**) and Ba/F3 cells treated with sgRNA species targeting the protein-coding region of *LZTR1* following interleukin (IL)-3 withdrawal (**g**) (median percent relative to day 2 is plotted). **h**, Schema of the *LZTR1* cDNA experiment. Lineage-negative cells from *Mx1-Cre Zrsr2* control and *Zrsr2^{fl/y}* mice expressing empty vector, *LZTR1* cDNA or *LZTR1* cDNA lacking the sequence for BTB domains (' Δ BTB'). DAPI-GFP⁺LSK⁺ cells were then sorted and tested for replating capacity or transplanted into mice. **i**, Western blot for FLAG and *LZTR1* in cells expressing N-terminal FLAG-tagged empty vector (EV), *LZTR1* and Δ BTB *LZTR1*. The experiment was repeated twice with similar results. **j**, Relative percent of GFP⁺ blood cells from mice receiving lineage-negative cells expressing *Mx1-Cre Zrsr2* control or *Zrsr2^{fl/y}* constructs from **h**. Mean values \pm s.e.m. are shown. Bars indicate medians, box edges indicate first and third quartile values, and whisker edges indicate minimum and maximum values. *P* values were calculated relative to the control group by two-sided *t*-tests and are indicated in the figures.

and predicted to result in NMD (Supplementary Tables 2 and 3), was targeted by four single guide (sg)RNA species. This was performed as a positive-enrichment CRISPR screen using pools of lentiviral sgRNA species in cytokine-dependent mouse (32D, Ba/F3) and human (TF-1) hematopoietic cell lines stably expressing Cas9 (Fig. 4a and Supplementary Tables 5–9; similar to an approach we recently used to model mutant splicing factor (SF)3B1-induced aberrant splicing events³⁰). Following stable infection with the sgRNA library, cytokines were depleted, and sgRNA representation was evaluated before and 7 d after cytokine removal.

This screen revealed several minor intron-containing genes, the downregulation of which conferred cytokine independence to one or more cell lines (Fig. 4b). Strikingly, only one gene was significantly enriched in all three cell lines (Fig. 4c), *LZTR1*, which encodes a cullin-3 adaptor regulating ubiquitin-mediated suppression of RAS-related GTPases^{31–33} and is subject to loss-of-function mutations in glioblastoma³⁴, schwannomatosis³⁵ and the RASopathy known as Noonan syndrome³⁶ (of note, anti-*PTEN* sgRNA species were included here as a positive sgRNA control; although *PTEN* has a minor intron, its splicing did not consistently differ between *ZRSR2*-WT and -mutant cells). Inspection of transcriptomic data from primary patient samples confirmed the link between aberrant splicing of *LZTR1* and MDS pathogenesis, as we identified retention of a U12-type intron in *LZTR1* (intron 18) in samples harboring mutations in *ZRSR2* (which was specific to *ZRSR2*-mutant MDS compared to BM samples from patients with WT *ZRSR2* and MDS and unaffected individuals; Fig. 4d). Importantly, this same region is a direct binding target of *ZRSR2*, as revealed by eCLIP-seq in human myeloid leukemia cells (Fig. 4d). Moreover, this mis-splicing of *LZTR1* was specific to patients with *ZRSR2* mutations and AML versus those with other spliceosomal gene mutations (Fig. 4e). Consistent with intron retention causing NMD-inducing reading frame disruption, minor intron retention in *LZTR1* correlated with reduced *LZTR1* mRNA and *LZTR1* protein levels in MDS and AML primary patient samples as well as in mouse hematopoietic precursors (Fig. 4f, Extended Data Fig. 6a–f and Supplementary Table 1). Moreover, blocking NMD by knockdown of UPP1, a core component of the NMD machinery or by pharmacologic inhibition of NMD in *ZRSR2*-KO cells increased the mRNA stability and expression of the minor intron-retained forms of *LZTR1* and an additional *ZRSR2*-regulated intron in *CHD4* (Extended Data Fig. 6g–j). These data formally confirm that these U12-type intron-containing isoforms are NMD substrates.

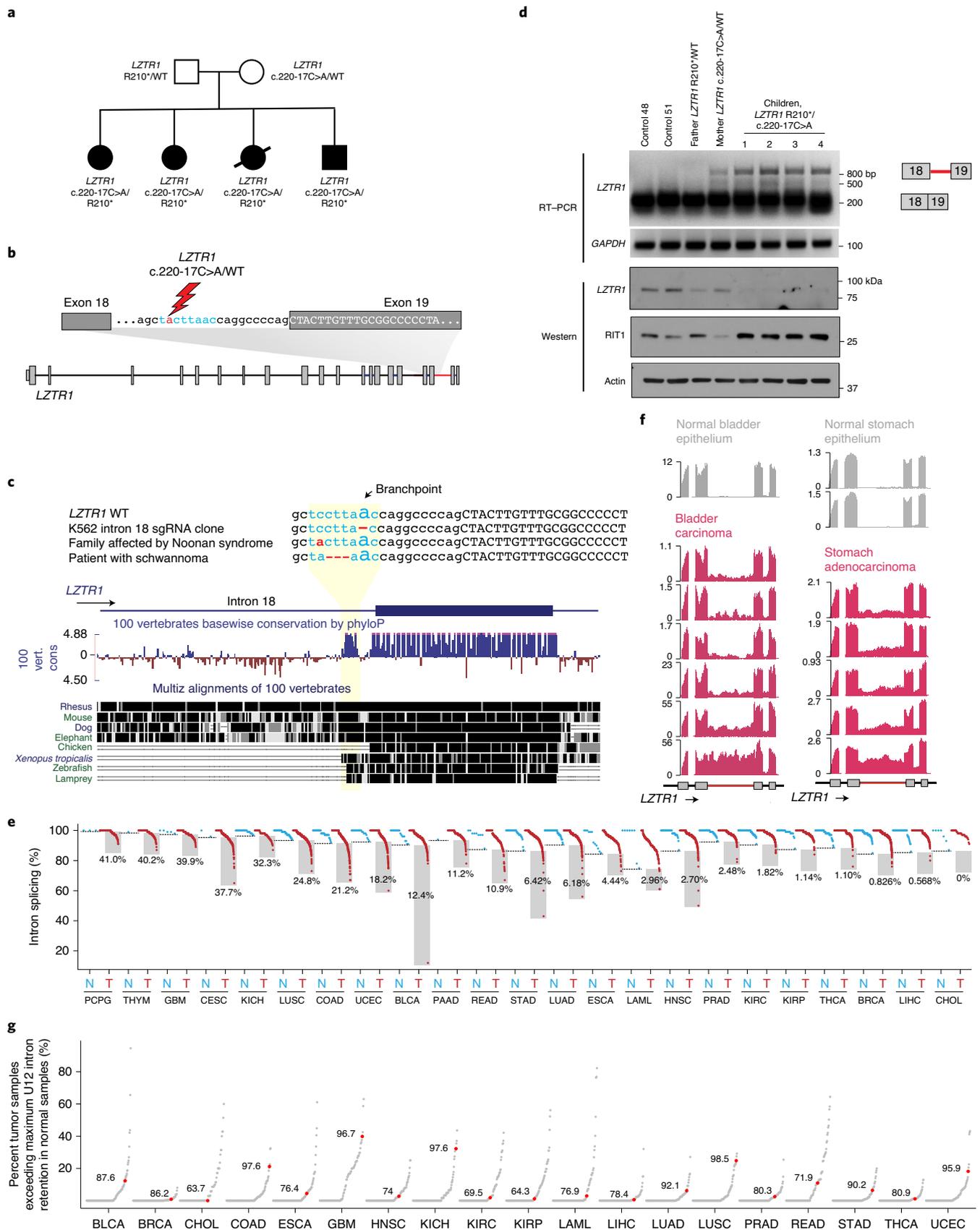
***LZTR1* intron retention is transforming.** Our functional screens revealed that *LZTR1* KO was associated with a uniquely robust competitive advantage, but CRISPR-based gene KO is an imperfect model of *LZTR1* loss due to impaired removal of its minor intron. To address this, we took advantage of a protospacer adjacent motif site deep within intron 18 of *LZTR1*, which is located adjacent to a highly conserved sequence matching the consensus motif for the minor intron branchpoint (Fig. 5a). Although recent work from our group identified that most introns use multiple branchpoints²⁹, we analyzed intron lariat-derived reads from RNA-seq data and identified that this highly conserved sequence within intron 18 of *LZTR1* contained only a single branchpoint nucleotide (Fig. 5b) located proximal to the 3' splice site, consistent with our genomic analysis of minor introns that were vulnerable to *ZRSR2* loss (Fig. 3e–f). To functionally evaluate the requirement for this putative branchpoint and the surrounding U12 consensus sequence for splicing of *LZTR1*, we generated a minigene of *LZTR1*'s minor intron and flanking exons from human cells and performed extensive mutagenesis reactions. This verified that *LZTR1* is spliced less efficiently in *ZRSR2*-null cells than in WT cells and identified the conserved U12 intronic sequence as required for *ZRSR2*-dependent excision of the minor intron (Fig. 5c and Extended Data Fig. 6k–m).

Delivering sgRNA species targeting the intronic U12 consensus sequence of *LZTR1* (which is ten nucleotides from the closest protein-coding region of *LZTR1*) induced robust *LZTR1* intron 18 retention in polyclonal mouse and human cells (Fig. 5d). We therefore generated multiple clones of *LZTR1* minor intron-mutant human K562 and Ba/F3 cells (Extended Data Fig. 6n,o) and compared the effects of impaired *LZTR1* minor intron excision with deleterious mutations in the protein-coding sequence of *LZTR1*. This revealed that both direct disruption of the protein-coding sequence of *LZTR1* and induction of minor intron retention virtually abolished *LZTR1* protein expression, without generating a truncated *LZTR1* protein product (Fig. 5e and Extended Data Fig. 6p). Similarly, inducing mutations within either the protein-coding region of *LZTR1* or its minor intron resulted in dramatic accumulation of RAS proteins, including RIT1, a RAS GTPase recently identified as an endogenous substrate of *LZTR1* (ref. ³²), and its corresponding gene is known to undergo activating mutations in RASopathies and a variety of cancers^{37,38}. Consistent with these convergent effects of protein-coding mutations and minor intron retention in *LZTR1*, both perturbations conferred cytokine independence to Ba/F3 cells

Fig. 6 | Aberrant *LZTR1* minor intron retention in Noonan syndrome, schwannomatosis and diverse cancers. **a**, Pedigree of a family affected by Noonan syndrome in which the mother and all four offspring had a mutation within the conserved U12 branchpoint sequence in *LZTR1* intron 18. All four children were diagnosed with Noonan syndrome, and the proband (black circle) developed AML. **b**, Schematic of the location of the intronic c.220-17C>A mutation. Intronic nucleotides in blue represent the conserved U12 sequence. **c**, Sequence conservation (cons) of the *LZTR1* minor intron 3' conserved sequence and surrounding intronic and exonic sequences (as estimated by phyloP³⁹). The location of the *LZTR1* intron 18 branchpoint as well as mutations in a K562 single-cell clone; the family affected by Noonan syndrome from a and a patient with schwannoma³⁵ bearing mutations in this region are shown. Conservation and repetitive element annotation from the University of California, Santa Cruz (UCSC) Genome Browser⁴⁰. Vert., vertebrate. **d**, Qualitative RT-PCR gel for *LZTR1* intron 18 excision (top) and *LZTR1* and RIT1 protein levels (bottom) in immortalized skin fibroblasts from the individuals in a as well as two healthy control individuals. The experiment was repeated three times with similar results. **e**, Splicing of the minor intron of *LZTR1* across peritumoral normal (N) and tumor (T) samples in cancers from the indicated TCGA cohorts with available matched normal samples. Normal samples for LAML comparison were from ref. ⁴¹. The horizontal dashed line in each normal column represents the maximum intron retention observed in normal control samples for each cancer type. Shaded boxes indicate cancer samples exceeding maximum intron retention in normal samples. The percentage next to each box indicates per cancer type percentage of cancer samples with high intron retention. **f**, Representative RNA-seq coverage plots for the minor intron of *LZTR1* in bladder and stomach carcinoma and peritumoral normal samples. **g**, Degree of minor (U12-type) intron retention across normal and tumor samples in cancers from TCGA. Each point corresponds to a single U12-type intron and indicates the percentage of all tumor samples in which retention of that intron exceeded the maximum corresponding retention of that intron observed in normal samples. Red dots indicate the U12-type intron of *LZTR1*; numbers indicate its percentile rank compared to all other U12-type introns. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; ESCA, esophageal 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; LAML, acute myeloid leukemia; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma.

as well as BCR-ABL inhibitor resistance to K562 cells (Fig. 5f,g; a phenotype originally used to identify LZTR1 as a regulator of mitogen-activated protein kinase signaling³¹). In fact, mutagenesis

of the exact branchpoint nucleotide within intron 18 of LZTR1 rendered K562 cells resistant to ATP-dependent or -independent ABL kinase inhibitors (imatinib or rebastinib, respectively (Fig. 5f and



Extended Data Fig. 6q)). Finally, we evaluated the impact of mutagenesis of *Lztr1*'s minor intron on the clonogenic capacity of mouse HSPCs. This was performed by generating HSC-specific conditional Cas9 knockin mice by crossing HSC-*Scl-Cre^{ERT}* *Zrsr2*-WT or *Zrsr2^{nl/y}* mice with *Rosa26-Lox-STOP-Lox-Cas9-EGFP* knockin mice. Infection of BM cells from these mice with sgRNA species targeting the conserved U12 sequence of *Lztr1*'s minor intron strongly promoted replating capacity of *Zrsr2*-WT HSPCs but not that of *Zrsr2*-mutant HSPCs (Extended Data Fig. 7a,b). Conversely, restoring expression of LZTR1 in *Zrsr2*-KO HSPCs strongly impaired their replating capacity as well as their self-renewal in vivo (which was not seen with expression of a version of LZTR1 lacking the Broad-complex, Tramtrack and Bric-abrac (BTB) domains required for cullin-3 interaction and LZTR1 function; Fig. 5h–j and Extended Data Fig. 7c,d). These data further confirm the impact of impaired *Lztr1* minor intron excision on clonogenic capacity. Of note, although RIT1 was clearly upregulated upon LZTR1 downregulation, the effects of LZTR1 downregulation were not solely dependent on RIT1 (Extended Data Fig. 7e,f). This latter point likely reflects the upregulation of multiple RAS proteins upon LZTR1 loss and underscores the need for future efforts to systematically identify LZTR1-regulated substrates in hematopoietic cells.

Impaired LZTR1 splicing in Noonan syndrome and cancer. Our finding that mutagenesis of deep intronic sequences within the minor intron of LZTR1 transformed cells, combined with the fact that LZTR1 is recurrently affected by protein-coding mutations in a variety of cancers^{34,35}, led us to search for aberrant LZTR1 minor intron excision beyond the context of ZRSR2-mutant MDS. We first studied cancer predisposition syndromes. Interestingly, in one reported family with autosomal recessive Noonan syndrome in which one child died of AML³⁶, the mother and all four children carried an intronic mutation within the branchpoint-containing region that we identified within the minor intron of LZTR1 (c.2220-17C>A; Fig. 6a,b). This same sequence is also mutated in schwannomatosis³⁵ (Fig. 6c). We established immortalized fibroblasts from each family member and unrelated controls and identified clear LZTR1 minor intron retention with consequently impaired LZTR1 protein expression and RIT1 accumulation in individuals bearing the LZTR1 minor intron mutation (Fig. 6d and Extended Data Fig. 8a–d).

We next interrogated LZTR1 minor intron splicing across the diverse cancer types profiled by TCGA. The minor intron of LZTR1 was efficiently excised in all normal samples. However, a notable subset of tumors in almost all profiled cancer types exhibited significantly increased retention that was specific to the minor intron of LZTR1. The extent of LZTR1 intron retention varied between samples and across cancer types, with a total of 11.1% of all profiled cancer samples exhibiting LZTR1 minor intron retention exceeding that observed in any peritumoral control normal tissue (Fig. 6e). In some cases, this intron retention was comparable to that observed in ZRSR2-mutant MDS, even though ZRSR2 and other minor spliceosomal factors are not recurrently mutated in those cancers (Fig. 6f). Moreover, we performed systematic analysis of U12-type as well as U2-type intron retention across all alternative introns across cancers in TCGA datasets and evaluated the degree of LZTR1 U12-type intron retention in this context. This analysis revealed that the minor intron of LZTR1 is among the most frequently retained introns across cancer types (top 10% of retained minor introns across all cancers; Fig. 6g and Extended Data Fig. 9a).

Discussion

Since the discovery of a second, independent spliceosome in most metazoans over 20 years ago¹⁰, many questions regarding the role of minor introns in cellular physiology and disease have been enigmatic. Here, we uncover a heretofore unrecognized role of minor intron excision in regulating HSC self-renewal, a molecular link

among ZRSR2 mutations and aberrant LZTR1 splicing and expression and frequent LZTR1 U12-type intron retention in diverse cancers and cancer predisposition syndromes.

Interestingly, the most common somatic mutation in LZTR1 reported to date affects a splice site³³. Given the prevalence of LZTR1 deep intronic mutations in cancer predisposition syndromes, it is reasonable to hypothesize that the minor intron of LZTR1 may be similarly subject to somatic mutations not detected by whole-exome sequencing. Consistent with this concept, LZTR1 intron retention is not limited to the U12-type intron of LZTR1; several U2-type LZTR1 introns are also commonly retained in tumors relative to normal tissues (Extended Data Fig. 9b,c).

Overall, these analyses indicate that LZTR1 is frequently dysregulated via perturbed minor intron splicing, much more so than revealed by studying protein-coding mutations alone. Given our finding of frequent post-transcriptional disruption of LZTR1 in the absence of protein-coding mutations, our data motivate study of other cancer-associated minor intron-containing genes that may be dysregulated via similar and as-yet-undetected aberrant splicing.

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/s41588-021-00828-9>.

Received: 28 May 2020; Accepted: 24 February 2021;

Published online: 12 April 2021

References

- Yoshida, K. et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
- Papaemmanuil, E. et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627 (2013).
- Haferlach, T. et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**, 241–247 (2014).
- Kim, E. et al. SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
- Ilgan, J. O. et al. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res.* **25**, 14–26 (2014).
- 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).
- Madan, V. et al. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat. Commun.* **6**, 6042 (2015).
- Gault, C. M. et al. Aberrant splicing in maize *rough endosperm3* reveals a conserved role for U12 splicing in eukaryotic multicellular development. *Proc. Natl Acad. Sci. USA* **114**, E2195–E2204 (2017).
- Tarn, W. Y. & Steitz, J. A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84**, 801–811 (1996).
- Hall, S. L. & Padgett, R. A. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239**, 357–365 (1994).
- Tarn, W. Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**, 1824–1832 (1996).
- Hall, S. L. & Padgett, R. A. Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271**, 1716–1718 (1996).
- Burge, C. B., Padgett, R. A. & Sharp, P. A. Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2**, 773–785 (1998).
- Patel, A. A., McCarthy, M. & Steitz, J. A. The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J.* **21**, 3804–3815 (2002).
- Younis, I. et al. Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *eLife* **2**, e00780 (2013).
- Moran-Crusio, K. et al. *Tet2* loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
- Challen, G. A. et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2011).
- Li, Q. et al. Hematopoiesis and leukemogenesis in mice expressing oncogenic NrasG12D from the endogenous locus. *Blood* **117**, 2022–2032 (2011).

19. Markmiller, S. et al. Minor class splicing shapes the zebrafish transcriptome during development. *Proc. Natl Acad. Sci. USA* **111**, 3062–3067 (2014).
20. Doggett, K. et al. Early developmental arrest and impaired gastrointestinal homeostasis in U12-dependent splicing-defective *Rnpc3*-deficient mice. *RNA* **24**, 1856–1870 (2018).
21. Otake, L. R., Scamborova, P., Hashimoto, C. & Steitz, J. A. The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in *Drosophila*. *Mol. Cell* **9**, 439–446 (2002).
22. Baumgartner, M. et al. Minor spliceosome inactivation causes microcephaly, owing to cell cycle defects and death of self-amplifying radial glial cells. *Development* **145**, dev166322 (2018).
23. Obeng, E. A. et al. Physiologic expression of Sf3b1^{K700E} causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. *Cancer Cell* **30**, 404–417 (2016).
24. Shirai, C. L. et al. Mutant U2AF1 expression alters hematopoiesis and pre-mRNA splicing in vivo. *Cancer Cell* **27**, 631–43 (2015).
25. Lee, S. C. et al. Synthetic lethal and convergent biological effects of cancer-associated spliceosomal gene mutations. *Cancer Cell* **34**, 225–241 (2018).
26. Taylor, J. et al. Single-cell genomics reveals the genetic and molecular bases for escape from mutational epistasis in myeloid neoplasms. *Blood* **136**, 1477–1486 (2020).
27. Tyner, J. W. et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
28. Van Nostrand, E. L. et al. Robust, cost-effective profiling of RNA binding protein targets with single-end enhanced crosslinking and immunoprecipitation (seCLIP). *Methods Mol. Biol.* **1648**, 177–200 (2017).
29. Pineda, J. M. B. & Bradley, R. K. Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev.* **32**, 577–591 (2018).
30. Inoue, D. et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432–436 (2019).
31. Bigenzahn, J. W. et al. LZTR1 is a regulator of RAS ubiquitination and signaling. *Science* **362**, 1171–1177 (2018).
32. Castel, P. et al. RIT1 oncoproteins escape LZTR1-mediated proteolysis. *Science* **363**, 1226–1230 (2019).
33. Steklov, M. et al. Mutations in *LZTR1* drive human disease by dysregulating RAS ubiquitination. *Science* **362**, 1177–1182 (2018).
34. Frattini, V. et al. The integrated landscape of driver genomic alterations in glioblastoma. *Nat. Genet.* **45**, 1141–1149 (2013).
35. Piotrowski, A. et al. Germline loss-of-function mutations in *LZTR1* predispose to an inherited disorder of multiple schwannomas. *Nat. Genet.* **46**, 182–187 (2014).
36. Johnston, J. J. et al. Autosomal recessive Noonan syndrome associated with biallelic *LZTR1* variants. *Genet. Med.* **20**, 1175–1185 (2018).
37. Berger, A. H. et al. Oncogenic *RIT1* mutations in lung adenocarcinoma. *Oncogene* **33**, 4418–4423 (2014).
38. Aoki, Y. et al. Gain-of-function mutations in *RIT1* cause Noonan syndrome, a RAS/MAPK pathway syndrome. *Am. J. Hum. Genet.* **93**, 173–180 (2013).
39. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
40. Rosenbloom, K. R. et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015).
41. Pellagatti, A. et al. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood* **132**, 1225–1240 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Patient samples. Studies were approved by the institutional review boards of Memorial Sloan Kettering Cancer Center (MSKCC) and Fred Hutchinson Cancer Research Center and were conducted in accordance with the Declaration of Helsinki protocol. Informed consents were obtained from all human participants. Next-generation sequencing was performed on DNA extracted from BM mononuclear cells and matched normal samples from fingernails. Patient samples were sequenced with the MSK-IMPACT targeted sequencing panel, and somatic mutations (substitutions and small insertions and deletions), gene-level focal copy number alterations and structural rearrangements were detected with a clinically validated pipeline as previously described^{42,43}.

Animals. All animals were housed at MSKCC using a 12-h light–12-h dark cycle and with ambient temperature maintained at 72 °F ± 2 °F (~21.5 °C ± 1 °C) with 30–70% humidity. 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. All mouse experiments were performed in accordance with a protocol approved by the MSKCC Institutional Animal Care and Use Committees (11-12-029).

Generation of *Zrsr2*-conditional knockout mice. Please see the Supplementary Note for full details.

Genetically engineered mice other than *Zrsr2*-conditional knockout mice. *Sf3b1*^{K700E}, *Srsf2*^{P95H}, *HSC-Scl-Cre*^{ERT}, *Rosa26-Lox-STOP-Lox-Cas9-EGFP*, *Tet2* floxed and *Mx1-Cre* mice were obtained from Jackson Laboratory and were previously generated and described^{4,16,23}, with the exception of *HSC-Scl-Cre*^{ERT} mice⁴⁴ (which were obtained from J.R. Göthert, University Hospital, West German Cancer Center (WTZ), University Duisburg-Essen).

Bone marrow transplantation. Please see full details in the Supplementary Note.

Colony-forming assays. LT-HSCs (lineage-negative CD150⁺CD48⁻c-Kit⁺Sca-1⁺DAPI⁻) were sorted by flow cytometry from the BM of *Mx1-Cre* WT, *Mx1-Cre;Zrsr2*^{fl/+}, *Mx1-Cre;Zrsr2*^{fl/fl} mice and seeded at a density of 100 cells per replicate into cytokine-supplemented methylcellulose medium (MethoCult M3434, Stemcell Technologies). Colonies propagated in culture were scored at day 7. The remaining cells were resuspended and counted, and a portion was taken for replating (100 cells per replicate).

LZTR1 cDNA expression experiments. Please see full details in the Supplementary Note.

Development of a custom anti-mouse *Zrsr2* antibody. The rabbit polyclonal antibody against *Zrsr2* was generated by rabbit injections (YenZym, Cys-C-Ahx-PEQEEPPQESQSQPQPQSDP-amide; cysteine (C) was assigned for single-point, site-directed conjugation to the carrier protein; Ahx was added as a linker or spacer), followed by affinity purification using standard protocols.

Antibodies, flow cytometry and western blot analysis. All antibodies for flow cytometry were purchased from BD Pharmingen, eBioscience or BioLegend. BM mononuclear cells were stained with a lineage cocktail composed of antibodies targeting CD3, CD4, CD8, B220, CD19, NK1.1, Gr-1, CD11b, Ter119 and IL-7R α . Cells were also stained with antibodies against c-Kit, Sca-1, CD150 and CD48. Cell populations were analyzed using an LSRFortessa (Becton Dickinson) and sorted with a FACSAria II instrument (Becton Dickinson). We used the following antibodies: anti-B220-APC-Cy7 (clone RA3-6B2, BioLegend, 103224, 1:200 dilution), anti-B220-PerCP-Cy5.5 (RA3-6B2, eBioscience, 45-0452-82, 1:200), anti-CD3-PE-Cy7 (17A2, BioLegend, 100220, 1:200), anti-CD3-APC-Cy7 (17A2, BioLegend, 100222, 1:200), anti-Gr-1-APC (RB6-8C5, eBioscience, 25-5931-82, 1:500), anti-CD11b-FITC (M1/70, BioLegend, 101206, 1:200), anti-CD11b-APC-Cy7 (M1/70, BioLegend, 101226, 1:200), anti-NK1.1-APC-Cy7 (PK136, BioLegend, 108724, 1:200), anti-Ter119-APC-Cy7 (Ter119, BioLegend, 116223, 1:200), anti-c-Kit-APC (2B8, BioLegend, 105812, 1:100), anti-c-Kit-PerCP-Cy5.5 (2B8, BioLegend, 105824, 1:100), anti-c-Kit-BV605 (ACK2, BioLegend, 135120, 1:100), anti-Sca-1-PE-Cy7 (D7, BioLegend, 108102, 1:100), anti-CD45.1-FITC (A20, BioLegend, 110706, 1:200), anti-CD45.1-PerCP-Cy5.5 (A20, BioLegend, 110728, 1:200), anti-CD45.1-BV711 (A20, BioLegend, 110739, 1:200), anti-CD45.1-APC (A20, BioLegend, 110714, 1:200), anti-CD45.2-PE (104, eBioscience, 12-0454-82, 1:200), anti-CD45.2-Alexa 700 (104, BioLegend, 109822, 1:200), anti-CD45.2-BV605 (104, BioLegend, 109841, 1:200), anti-CD48-PerCP-Cy5.5 (HM48-1, BioLegend, 103422, 1:100), anti-CD150-PE (9D1, eBioscience, 12-1501-82, 1:100), anti-CD127 (IL-7R α)-APC-Cy7 (A7R34, BioLegend, 135040, 1:200), anti-CD8-APC-Cy7 (53-6.7, BioLegend, 100714, 1:200), anti-CD19-APC-Cy7 (6D5, BioLegend, 115529, 1:200), anti-IgM-PE (II/41, eBioscience, 12-5790-82, 1:200), anti-IgD-FITC (11-26c.2a, BioLegend, 405704, 1:200), anti-CD135-APC (A2F10, BioLegend, 135310, 1:200), anti-CD4-APC-Cy7 (GK1.5, BioLegend, 100413, 1:200), anti-CD25-BV711 (PC61.5, BioLegend, 102049, 1:200), anti-CD44-APC (IM7, BioLegend, 103012,

1:200), anti-D43 (eBioR2/60, eBioscience, 11-0431-85, 1:200), anti-CD24-BV605 (M1/69, BD Biosciences, 563060, 1:200), anti-CD21/35-PE (4E3, eBioscience, 12-0212-82, 1:200), anti-CD93-APC (AA4.1, eBioscience, 17-5892-82, 1:200), anti-CD23-e450 (B3B4, BD Biosciences, 48-0232-82, 1:200), anti-CD16/CD32 (Fc γ R1II/III)-Alexa 700 (93, eBioscience, 56-0161-82, 1:100), anti-CD34-FITC (RAM34, BD Biosciences, 553731, 1:50), anti-CD34-PerCP (8G12, BD Biosciences, 345803, 1:50), anti-CD117-PE-Cy7 (104D2, eBioscience, 25-1178-42, 1:100), anti-CD45-APC-H7 (2D1, BD Biosciences, 560178, 1:200) and anti-Ly-51-PE (BP-1, BD Bioscience, 553735, 1:200).

The composition of immature and mature hematopoietic cell lineages in the BM, spleen and peripheral blood was assessed using a combination of antibodies against B220 (RA3-6B2), CD19 (1D3), CD3 (17A2), CD4 (GK1.5), CD8 α (53-6.7), CD11b (M1/70), CD25 (PC61.5), CD44 (IM7), Gr-1, IgM (II/41), IgD (11-26c.2a), CD43 (S11), CD24 (M1/69), Ly-51 (BP-1), CD21/35 (4E3), CD93 (AA4.1), CD23 (B3B4), c-Kit (2B8), Sca-1 (D7), CD127 (A7R34) and CD135 (A2F10).

The following antibodies were used for western blotting: anti-*Zrsr2* (custom, YenZym, 1:1,000), anti-LZTR1 (sc-390166 or sc-390166 X, Santa Cruz Biotechnology, 1:1,000), anti-RIT1 (ab53720, Abcam, 1:1,000), anti-pan-RAS antibody (Thermo Fisher Scientific, MA1-012X, 1:1,000), anti-FLAG (F-1084, Sigma-Aldrich, 1:1,000), anti-UPF1 (ab109363, Abcam, 1:1,000) and anti-actin (A-5441, Sigma-Aldrich, 1:4,000).

Cell cycle and apoptosis analyses. Please see full details in the Supplementary Note.

Histological and peripheral blood analysis. Please see full details in the Supplementary Note.

Cell lines and tissue culture. HEK293T cells were obtained from ATCC and grown in DMEM with 10% FCS. Ba/F3 and 32Dcl3 cells were grown in RPMI with 10% FCS and 1 ng ml⁻¹ murine IL-3 (PeproTech, 213-13) unless noted otherwise. TF-1 cells were grown in RPMI with 10% FCS with 2 ng ml⁻¹ recombinant human GM-CSF (R&D Systems, 215-GM) unless noted otherwise. K562 cells were cultured in RPMI with 10% FCS. Human fibroblast cells were from F. McCormick's laboratory and were cultured in DMEM with 20% FCS. All cell culture media included penicillin (100 U ml⁻¹) and streptomycin (100 μ g ml⁻¹).

Transformation of primary human fibroblasts. Please see full details in the Supplementary Note.

In vitro competition assay. Please see full details in the Supplementary Note.

CRISPR screening for ZRSR2-regulated U12-type introns. Ba/F3, 32D and TF-1 cells were transduced with LentiCas9-Blast (Addgene, 52962) and single-cell sorted into 96-well plates. Among these clones, we used a single clone with strong Cas9 expression. sgRNA libraries of NMD targets in LZTR1-mutated cells were amplified and packaged as lentiviruses. Functional viral titers were obtained by measuring puromycin (2 μ g ml⁻¹) resistance after transduction as previously published⁴⁵. A titer resulting in 30% of cells surviving puromycin selection was calculated. For the NMD library screen, which included 2,600 sgRNA species, triplicate transductions with 8.7 \times 10⁶ infected cells each were performed for a coverage of approximately 1,000 \times representation. The library included four sgRNA species against each target gene (602 human genes for TF-1-Cas9 and 594 mouse genes for Ba/F3-Cas9 and 32D-Cas9), 100 control sgRNA species and positive control sgRNA species against *Pten* for Ba/F3-Cas9 and 32D-Cas9 cells and *Nfi* for TF-1-Cas9 cells. Cells were transduced with lentivirus carrying the sgRNA library produced by 293FT cells, and puromycin selection (2 μ g ml⁻¹) was performed in IL-3- or GM-CSF-containing medium for 7 d. Next, we washed out IL-3 or GM-CSF (day 0), and surviving cells were collected at 7 d after cytokine depletion (day 7). Cell pellets were lysed, and genomic DNA was extracted (Qiagen) and quantified by Qubit (Thermo Scientific). A quantity of gDNA covering a 1,000 \times representation of gRNA species was amplified by PCR using Q5 High-Fidelity Polymerase (New England Biolabs, M0491) to add Illumina adaptors and multiplexing barcodes. Amplicons were quantified by Qubit and Bioanalyzer (Agilent) and sequenced on the Illumina HiSeq 2500. Sequencing reads were aligned to the screened library, and counts were obtained for each gRNA. We used standard methods from the R-Bioconductor package, and the specific package was edgeR. For the probe-level analysis, we used the standard workflow with the glmLRT option for model fitting and statistical tests. For gene-level analysis, we used the camera analysis function, also from edgeR, as previously described^{46,47}.

CRISPR-directed mutations. Cas9-expressing K562 and Ba/F3 cells were transduced with the iLenti-guide-GFP vector targeting LZTR1 and *Lztr1* exonic and intronic sequences in which sgRNA expression was linked to GFP expression. For the in vivo competition assay, the percentage of GFP-expressing cells was then measured over time after infection using the BD LSRFortessa. The GFP-positive rates in living cells at each point normalized to those at day 2 after lentiviral transduction were calculated. For RT-PCR and western blot analysis, sgRNA-transduced cell lines were single-cell sorted into 96-well plates using a

BD FACSAria III cell sorter to generate single-clone-derived cells. The target sequences of sgRNA species used to induce *LZTR1* protein-coding mutations as well as intron 18 mutations in *LZTR1* and *Lztr1* are located in Supplementary Table 10. For the colony-formation assay to evaluate the in vitro effect of *Lztr1* intron 18 mutation, lineage-negative hematopoietic precursors from *Scl-Cre^{ERT2};Rosa26-Lox-STOP-Lox-Cas9-EGFP* mice were transduced with the pLKO RFP657 vector (Addgene, 57824) targeting intron 18 of *Lztr1* using RetroNectin (T100A, Takara Bio) as described above. GFP⁺RFP657⁺ double-positive cells were purified for plating using the BD FACSAria III cell sorter, followed by serial replating in vitro. Transduced cells were seeded at a density of 1,000 cells per replicate into cytokine-supplemented methylcellulose medium (MethoCult M3434, Stemcell Technologies). Colonies propagated in culture were scored at day 7. The remaining cells were resuspended and counted, and a portion was taken for replating (1,000 cells per replicate).

RT-PCR and quantitative RT-PCR. Total RNA was isolated using RNeasy Mini or Micro kits (Qiagen). For cDNA synthesis, total RNA was reverse transcribed to cDNA with the Verso cDNA kit (Thermo Scientific). Resulting cDNA was diluted 10–20-fold before use. *LZTR1* splice variants were detected via semiquantitative RT-PCR by a standard OneTaq DNA Polymerase (New England Biolabs) PCR protocol: 95 °C for 2 min and then 35 cycles of 95 °C for 30 s, 60 °C for 30 s and 72 °C for 60 s, followed by 72 °C for 7 min. Bands were visualized by ethidium bromide staining.

Quantitative RT-PCR (RT-qPCR) was performed in 10- μ l reactions with SYBR Green PCR Master Mix (Roche Life Science). All RT-qPCR analyses were performed on an Applied Biosystems QuantStudio 6 Flex Cycler (Thermo Fisher Scientific). Relative gene expression levels were calculated using the comparative C_T method, and the values were corrected with expression levels of the internal controls *GAPDH* and *Gapdh*. Primers used for RT-PCR are listed in Supplementary Table 10.

mRNA isolation and analysis. Please see full details in the Supplementary Note.

mRNA stability assays. For mRNA half-life measurement using RT-qPCR, K562 cells with *LZTR1* intron 18 mutations were infected with anti-*UPF1* shRNA species and control shRNA in the pLKO.1 puro vector (Addgene, 8453). After puromycin selection (1.0 μ g ml⁻¹ for 7 d), the transduced K562 cells were treated with 2.5 μ g ml⁻¹ actinomycin D (Life Technologies) and collected at 0, 2, 4, 6 and 8 h (using methods as described previously⁴). *LZTR1* minor intron inclusion and 18S rRNA levels were measured by RT-qPCR. Additionally, K562 cells with CRISPR-mediated *ZRSR2* KO were exposed to 2.5 μ g ml⁻¹ actinomycin D (Life Technologies) and 50 μ M of the NMD inhibitor NMDI14 (ref. ⁴⁸) (Millipore, 530838.0001), and then the relative levels of the minor intron-retained forms of *LZTR1* and *CHD4* were measured relative to 18S rRNA by RT-qPCR at 0, 1, 2, 4 and 6 h following drug exposure. Primers used in RT-PCR reactions are listed in Supplementary Table 10.

***LZTR1* minigene assay.** The *LZTR1* minigene construct was generated by inserting the DNA fragment containing the human *LZTR1* genomic sequence from exon 18 to exon 19 in between the KpnI and XhoI restriction sites of the pcDNA3.1(+) vector. The sequences of inserted fragments were verified by Sanger sequencing. Mutagenesis of minigene constructs was performed with the Agilent QuikChange II Site-Directed Mutagenesis kit according to the manufacturer's directions. Primers used for mutagenesis are listed in Supplementary Table 10. For minigene transient transfection experiments, K562 cells were seeded into a 24-well plate with culture medium 48 h before transfection with minigene constructs in the presence of X-tremeGENE HP DNA transfection reagent (Roche) according to the manufacturer's directions. Forty-eight hours after transfection, cells were collected, and RNA was extracted using the Qiagen RNeasy Mini kit. Extracted RNA was treated with DNase I (Ambion) to ensure complete removal of DNA. Minigene-derived and endogenous *LZTR1* transcripts were analyzed by RT-PCR using specific primers (Supplementary Table 10).

eCLIP library preparation. eCLIP studies were performed in duplicate by Eclipse Bioinnovations (<https://eclipsebio.com/>) according to the published single-end seCLIP protocol²⁸ with the following modifications: 10 million K562 cells expressing either a FLAG-tagged empty vector or N-terminal FLAG-tagged *ZRSR2* were UV cross-linked at 400 mJ cm⁻² with 254-nm radiation and snap frozen. Cells were then lysed and treated with RNase I to fragment RNA as previously described. Anti-FLAG antibody (F-1084, Sigma-Aldrich) was then precoupled to Protein G Dynabeads (Thermo Fisher), added to the lysate and incubated overnight at 4 °C. Before immunoprecipitation, 2% of the sample was taken as the paired input sample, with the remainder magnetically separated and washed with lysis buffer only. eCLIP was performed by excising the area from ~65 kDa to ~140 kDa. RNA adaptor ligation, immunoprecipitation–western assays, reverse transcription, DNA adaptor ligation and PCR amplification were performed as previously described²⁸.

Genome annotations. Genome annotations for mapping RNA-seq data to the human (NCBI GRCh37/UCSC hg19) and mouse (NCBI GRCm38/UCSC mm10)

genomes were generated as previously described⁴⁹. Briefly, isoform annotations from the MISO version 2.0 database⁵⁰ were merged with genome annotations from Ensembl release 71 (ref. ⁵¹) and the UCSC knownGene track⁵². Constitutively spliced introns were defined as those which exhibited no evidence of alternative splicing in the UCSC knownGene track⁵². Each intron was classified as U2- or U12-type by comparing its 5' splice site to the U2- and U12-type consensus sequences obtained from ref. ⁵³. This classification takes advantage of the highly stereotyped consensus of the U12-type 5' splice site.

RNA-seq read mapping. RNA-seq reads were first mapped to the transcriptome annotations assembled as described above with RSEM version 1.2.4 (ref. ⁵⁴), modified to invoke Bowtie version 1.0.0 (ref. ⁵⁵) with the '-v 2' option. Remaining unaligned reads were then mapped to the genome as well as a database of possible splice junctions, consisting of all possible combinations of 5' splice sites and 3' splice sites annotated for each gene, using TopHat version 2.0.8b⁵⁶. The resulting read alignments generated by TopHat were then merged with the output from RSEM. Reads were filtered to require that splice junction-spanning reads had a minimum overhang of six nucleotides.

Differential splicing analysis. Isoform expression and intron retention were estimated as previously described⁴⁹. In brief, numbers of RNA-seq reads supporting either intron excision (splice junction-spanning reads) or intron retention (reads overlapping exon–intron boundaries and reads within introns) were computed for each intron and used to estimate the fraction of mRNA for which each intron was retained (the 'isoform ratio'). Significantly retained introns were defined as those which (1) exhibited either an absolute change in retention of $\geq 10\%$ or an absolute fold change in retention of ≥ 2 and (2) had an associated Bayes factor ≥ 5 (computed using Wagenmakers' Bayesian framework⁵⁷; relevant to single-sample comparisons for mouse data) or $P \leq 0.05$ (computed using a two-sided Mann–Whitney *U* test; relevant for group comparisons for patient cohort data), and all comparisons were restricted to samples that had least 20 informative reads (reads that distinguish between isoforms) for the intron under consideration.

Determining levels of intron retention in patient cohorts. To quantify per-sample levels of intron retention in samples from patients with MDS (Extended Data Fig. 5), we calculated the ratios of the numbers of significantly retained introns (U12- or U2-type) to the numbers of introns that were removed more efficiently (U12- or U2-type) relative to the median over all samples lacking spliceosomal mutations. A pseudocount proportional to the relative abundance of each intron class was added to the numerator and denominator to regularize the computation (avoiding division by zero).

RNA-seq read coverage plots. Read coverage plots were created using the ggplot2 package⁵⁸ in R. Illustrated transcript annotations are from RefSeq⁵⁹, downloaded from the UCSC Genome Browser⁶⁰.

Sequence logo analysis. Sequence logo plots summarizing the *LZTR1* mutagenesis experiments and depicting the nucleotide consensus sequences for U12-type introns that did or did not exhibit significantly increased retention in samples from patients with WT versus mutant *ZRSR2* were created by generating and illustrating position weight matrices using the seqLogo and ggseqlogo⁶¹ packages in Bioconductor^{62,63}.

Branchpoint inference. Lariats arising from U12-type intron splicing were computationally inferred as previously described²⁹. In brief, a split read alignment to a database of 5' splice site and upstream 3' splice sequences was performed to identify reads that spanned the lariat 2'–5' linkage. High-confidence reads, defined as containing a single-nucleotide mismatch at the split read junction, were used to identify the branchpoint location. This genome-wide branchpoint annotation was expanded in this study through the inclusion of all available TCGA RNA-seq datasets from both normal and tumor samples, with a focus on identifying branchpoints within U12-type introns. This larger branchpoint annotation was used to identify sequence features correlated with response to *ZRSR2*-inactivating mutations.

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

Data availability

Genome annotations for humans and mice were from NCBI GRCh37/UCSC hg19 and NCBI GRCm38/UCSC mm10, respectively. Isoform annotations were from the MISO version 2.0 database⁵⁰ and were merged with genome annotations from Ensembl release 71 (ref. ⁵¹) and the UCSC knownGene track⁵². RNA-seq reads for the human samples reported in Madan et al.⁷ were downloaded from the Gene Expression Omnibus (accession number GSE63816). RNA-seq data generated by TCGA (dbGaP accession phs000178.v11.p8) and the Beat AML data (dbGaP accession phs001657.v1.p1) were downloaded from the National Cancer Institute Genomic Data Commons. Branchpoint-related data were obtained from a published study²⁹. RNA-seq data generated as part of this study were deposited

in the Gene Expression Omnibus (mouse data, accession [GSE149455](#)) and the human RNA-seq data were deposited in dbGaP ([phs002212.v1.p1](#)). Source data are provided with this paper.

References

42. 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).
43. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
44. Gothert, J. R. et al. In vivo fate-tracing studies using the *Scf* stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. *Blood* **105**, 2724–2732 (2005).
45. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
46. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
47. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
48. Martin, L. et al. Identification and characterization of small molecules that inhibit nonsense-mediated RNA decay and suppress nonsense *p53* mutations. *Cancer Res.* **74**, 3104–3113 (2014).
49. Dvinge, H. et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc. Natl Acad. Sci. USA* **111**, 16802–16807 (2014).
50. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
51. Flicek, P. et al. Ensembl 2013. *Nucleic Acids Res.* **41**, D48–D55 (2013).
52. Meyer, L. R. et al. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–D69 (2013).
53. Sheth, N. et al. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34**, 3955–3967 (2006).
54. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
55. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
56. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
57. Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H. & Grasman, R. Bayesian hypothesis testing for psychologists: a tutorial on the Savage–Dickey method. *Cogn. Psychol.* **60**, 158–189 (2010).
58. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
59. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
60. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
61. Wagih, O. *gseqlogo*: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).
62. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
63. Bembom, O. & Ivanek, R. *seqLogo*: sequence logos for DNA sequence alignments. R package version 1.30.0.

Acknowledgements

This work was supported by the Conquer Cancer Foundation of the American Society of Clinical Oncology (J.T.), the American Association for Cancer Research (J.T.), the American Society of Hematology (D.I., J.T., K.K.), the Robert Wood Johnson Foundation (J.T.), the ARCS Foundation (J.M.B.P.), NIH grants 1K08CA230319-01 (J.T.), T32 CA009657 (J.T.P.), R01 DK103854 (R.K.B.), R01 HL151651 (R.K.B.), R01 HL128239 (R.K.B. and O.A.-W.), R01 CA251138 (R.K.B. and O.A.-W.), the Department of Defense Bone Marrow Failure Research Program (W81XWH-12-1-0041, R.K.B. and O.A.-W.), the Leukemia & Lymphoma Society (D.I. and O.A.-W.), the Henry and Marilyn Taub Foundation (O.A.-W.), the Edward P. Evans MDS Foundation (J.T.P., J.T., O.A.-W. and R.K.B.) and the Blood Cancer Discoveries Grant program through the Leukemia & Lymphoma Society, the Mark Foundation for Cancer Research and the Paul G. Allen Frontiers Group (8023-20, R.K.B.). D.I. is supported by AMED (JP19cm0106165 and JP20cm0106165), the Takeda Science Foundation, the MDS Foundation, the Yasuda Medical Foundation, the Kanae Foundation for the Promotion of Medical Science, the MSD Life Science Foundation, the Bristol Myers Squibb Foundation, JSPS KAKENHI (JP20H00537 and JP 20H03717) and the Leukemia Research Foundation. R.K.B. is a scholar of the Leukemia & Lymphoma Society (1344-18). The results shown here are in part based upon data generated by TCGA Research Network (<https://cancergenome.nih.gov/>).

Author contributions

D.I., J.T.P., J.T., R.K.B. and O.A.-W. conceived the project and wrote the paper with editorial contributions from all authors; D.I., J.T., P.C., S.C., S.K., C.F., S.J.H., Y.H., J.M.B.P., E.E.M., C.E., K.K., M.F., H.Y., A.T., C.F., S.X.L., B.H.D., B.L., E.W., S.M., D.Z. and R.G. performed experiments; D.I., J.T.P., J.T., A.P., G.-L.C. and R.K.B. analyzed data; and F.M., R.K.B. and O.A.-W. supervised the project.

Competing interests

O.A.-W. has served as a consultant for H3B Biomedicine, Foundation Medicine Inc., Merck, Prelude Therapeutics and Janssen and is on the scientific advisory board of Envisagenics Inc., AIChem and Pfizer Boulder; O.A.-W. received prior research funding from H3B Biomedicine and Loxo Oncology unrelated to the current manuscript. The remaining authors declare no competing interests.

Additional information

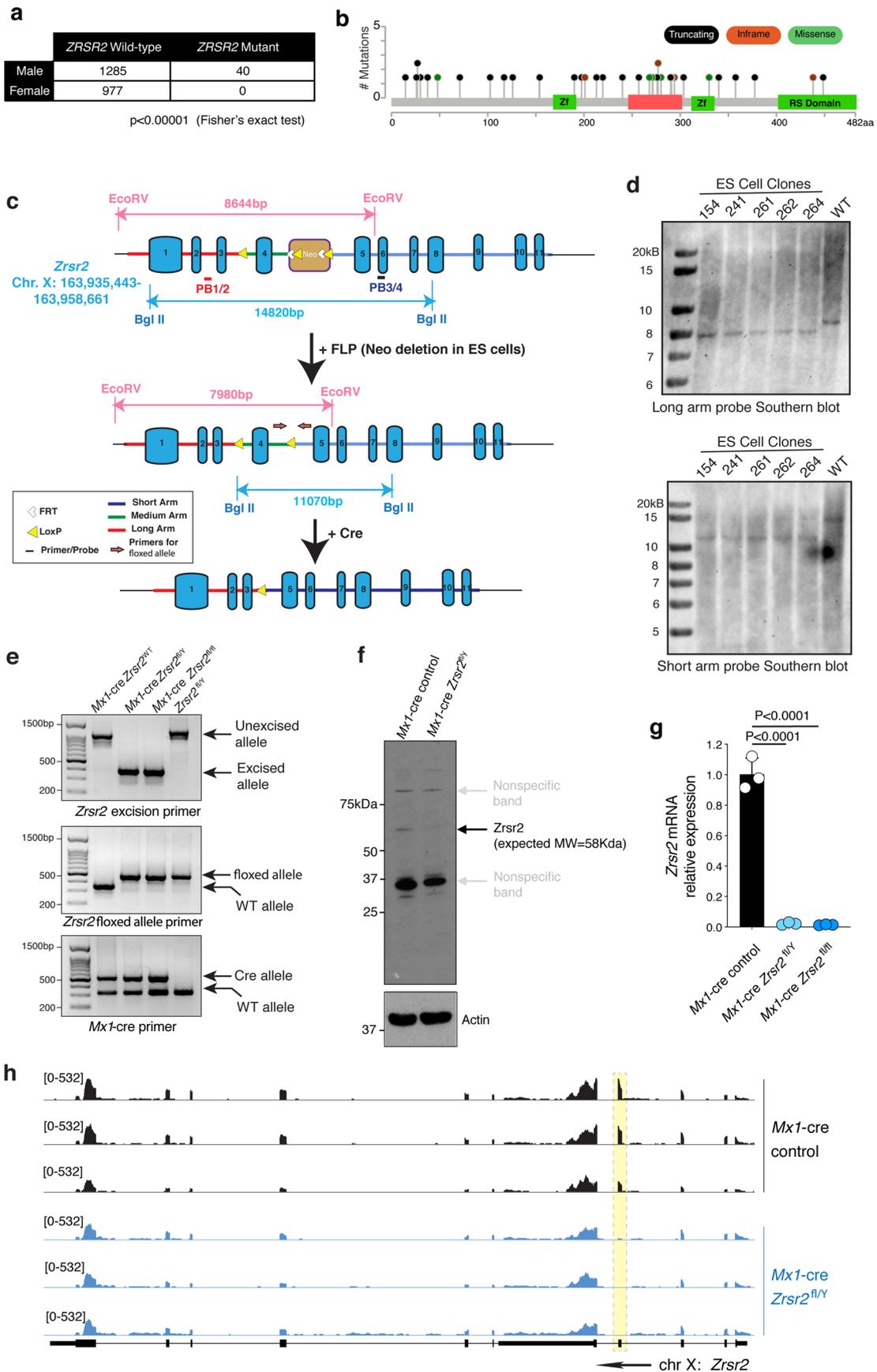
Extended data is available for this paper at <https://doi.org/10.1038/s41588-021-00828-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-021-00828-9>.

Correspondence and requests for materials should be addressed to R.K.B. or O.A.-W.

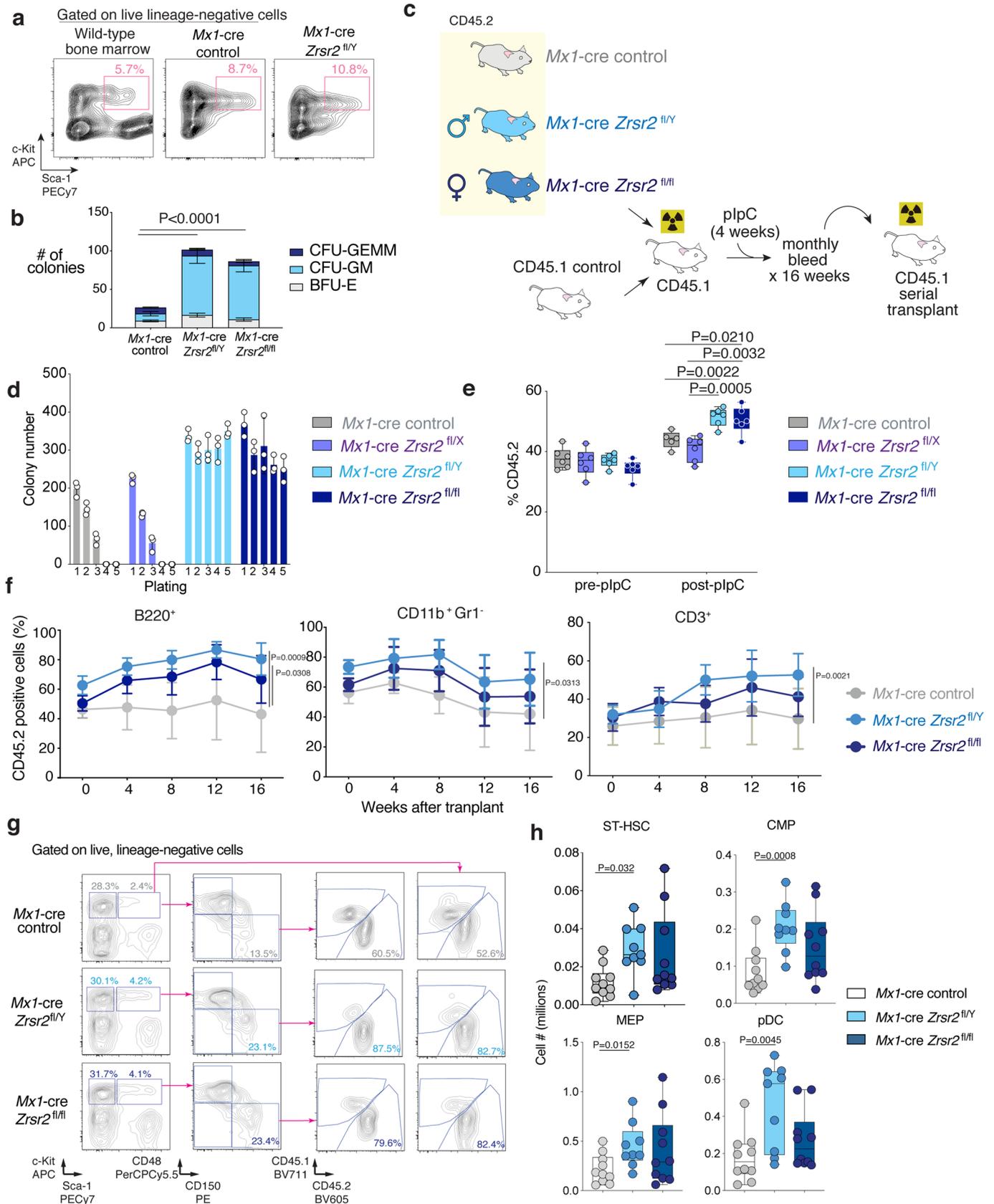
Peer review information *Nature Genetics* thanks George Vassiliou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



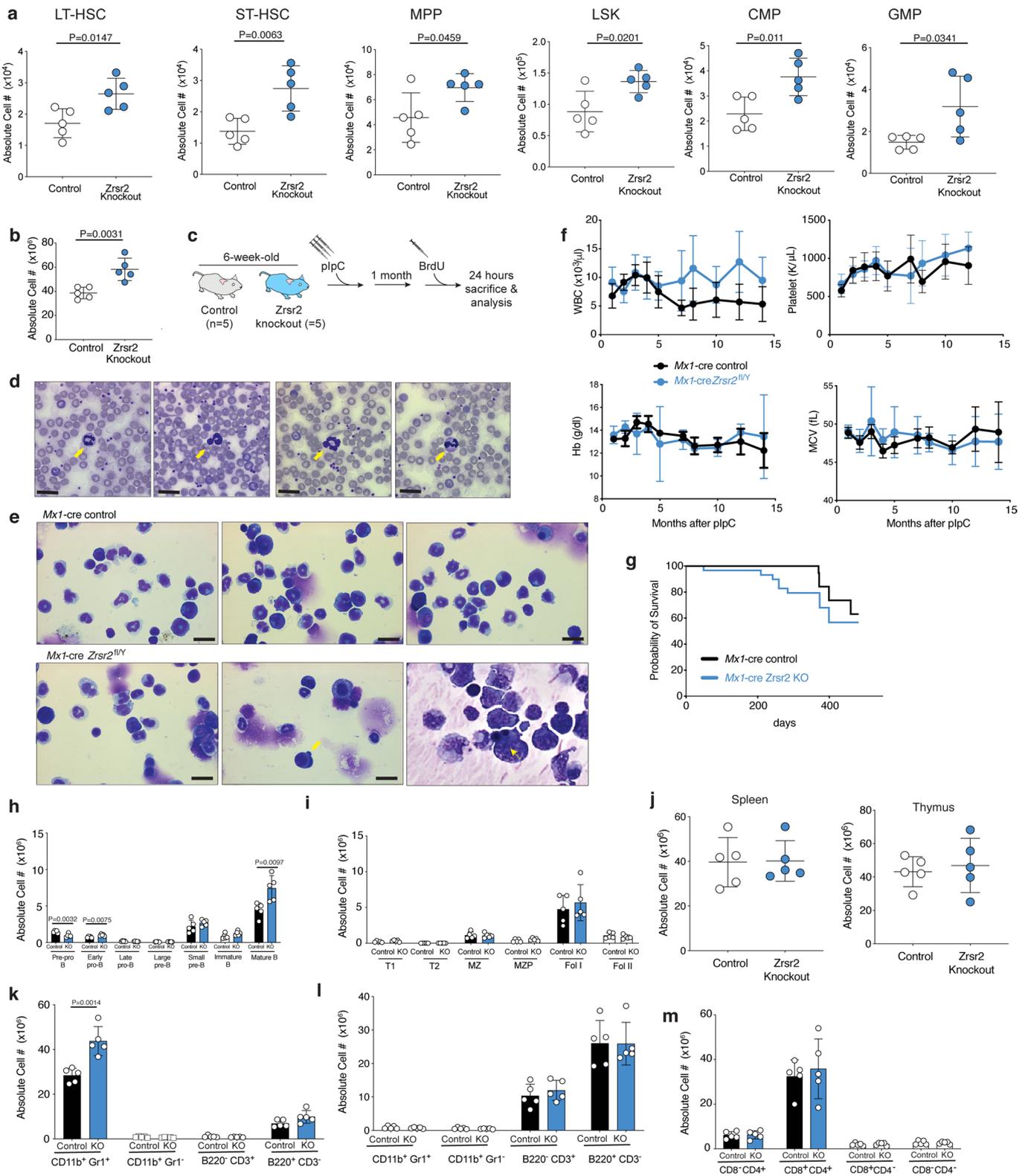
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Generation and validation of *Zrsr2* conditional knockout (cKO) mice. **a**, Punnett square enumerating male and female myeloid neoplasm patients wild-type versus mutant for *ZRSR2* across 2,302 patients. **b**, Lollipop diagram of *ZRSR2* mutations from **(a)**. **c**, Schematic depiction of the targeting strategy to generate *Zrsr2* cKO mice. The *Zrsr2* allele was deleted by targeting exon 4 in a manner that results in a frameshift following excision. Two *LoxP* sites flanking exon 4 and an *Frt*-flanked neomycin selection cassette were inserted in the downstream intron. **d**, Verification of correct homologous recombination using Southern blots from targeted embryonic stem cells. The experiment was repeated twice with similar results. **e**, Verification of the presence of *Mx1-cre* and *Zrsr2* floxed alleles as well as excision of *Zrsr2* using genomic PCR. The experiment was repeated three times with similar results. **f**, Full-length Western blot of *Zrsr2* protein in protein lysates from bone marrow mononuclear cells from *Mx1-cre* control or *Mx1-cre Zrsr2^{fl/y}* mice. Black arrow indicates full-length *Zrsr2* protein while grey arrows indicate non-specific bands. The experiment was repeated three times with similar results. **g**, *Zrsr2* expression (relative to 18S rRNA) in long-term hematopoietic stem cells (DAPI⁻ lineage-negative c-Kit⁺ Sca-1⁺ CD150⁺ CD48⁻) in *Mx1-cre* control, *Mx1-cre Zrsr2^{fl/y}*, and *Mx1-cre Zrsr2^{fl/fl}* mice 6-weeks following polyinosinic-polycytidylic acid (pIpC) administration. Mean values \pm SD. P-values calculated relative to the control group by a two-sided *t*-test. *n* = 3 biologically independent experiments. **h**, RNA-seq coverage plots from lineage-negative c-Kit⁺ cells from mice in **(g)** illustrating excision of exon 4 of *Zrsr2* following pIpC.



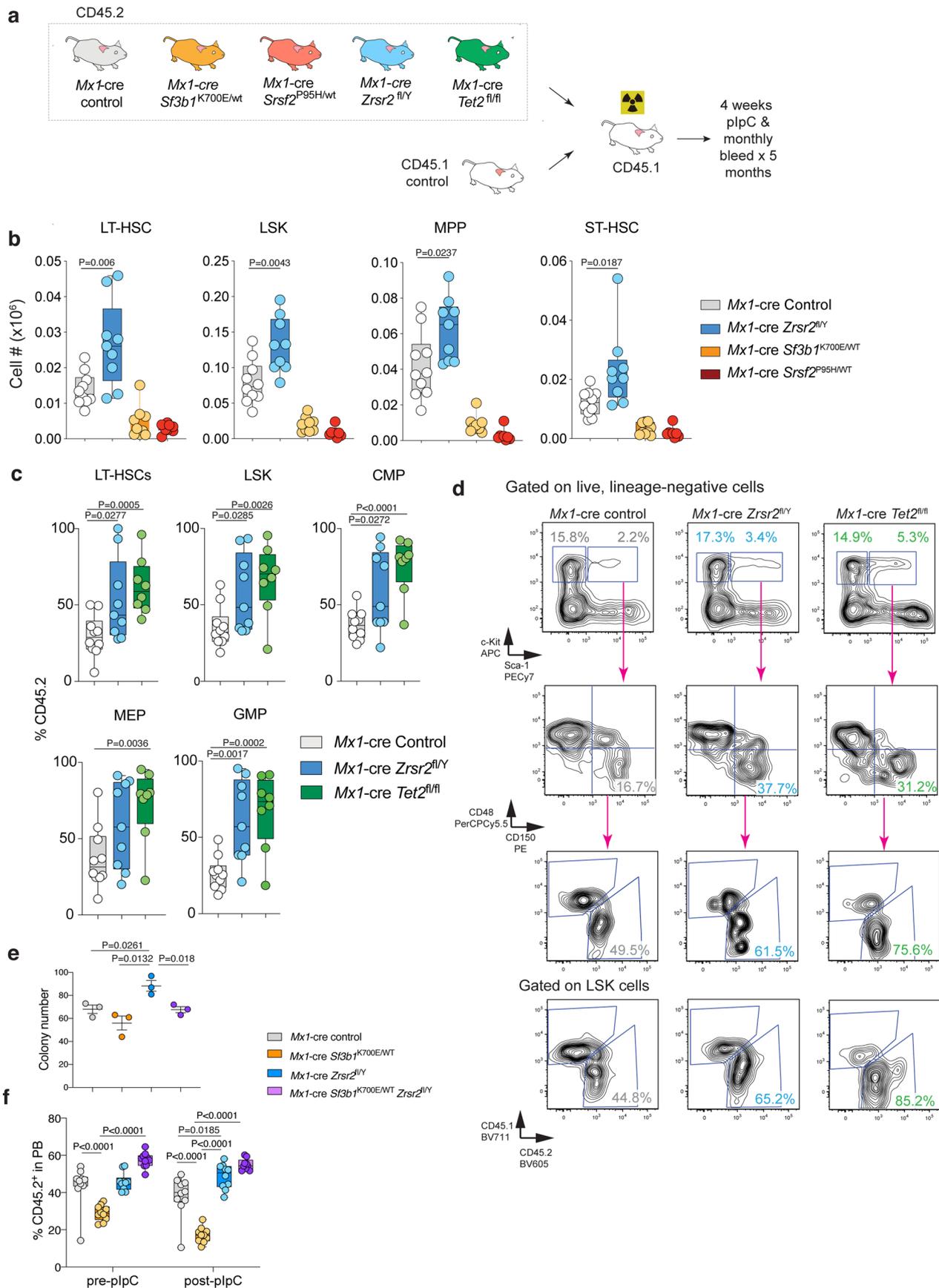
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Zrsr2 loss enhances self-renewal of hematopoietic stem cells (HSCs). **a**, FACS of cells from 5th methylcellulose plating (see Fig. 1b) for live, LSK cells. Wild-type bone marrow (BM) from 6-week old mouse (left) as a staining control. **b**, Number of CFU-GM, CFU-GEMM, and BFU-E colonies from initial plating of LT-HSCs (lin⁻ LSK CD150⁺ CD48⁻) from *Mx1-cre* control, *Mx1-cre Zrsr2^{fl/y}*, and *Mx1-cre Zrsr2^{fl/fl}* mice into methylcellulose. Mean \pm SD, two-sided *t*-test. *n* = 4 biologically independent experiments. **c**, Schematic of competitive BM transplantation. **d**, Number of methylcellulose colonies from LT-HSCs from *Mx1-cre* control, *Mx1-cre Zrsr2^{fl/x}*, *Mx1-cre Zrsr2^{fl/y}*, and *Mx1-cre Zrsr2^{fl/fl}*. Mean value \pm SD. *n* = 3 biologically independent experiments. **e**, Box-and-whisker plots of percentage of peripheral blood CD45.2⁺ cells in competitive transplantation pre- and post-plpC using CD45.2⁺ *Mx1-cre* control, *Mx1-cre Zrsr2^{fl/x}*, *Mx1-cre Zrsr2^{fl/y}*, and *Mx1-cre Zrsr2^{fl/fl}* mice. For box and whiskers plots throughout, bar indicates median, box edges first and third quartile values, and whisker edges minimum and maximum values, two-sided *t*-test. Experiment repeated three times with similar results in **(d)** and **(e)**. **f**, Percentage of CD45.2⁺ B220⁺ (left), CD45.2⁺ CD11b⁺ Gr1⁻ (middle), and CD45.2⁺ CD3⁺ cells (right) in primary competitive transplantation. Mean \pm SD. *P* values by two-sided *t*-test using the values at 16 weeks after transplant. *P* value relative to the control group at 16 weeks by a two-sided *t*-test. **g**, FACS analysis and gating strategy of BM cells from representative primary recipient mice in competitive transplantation. **h**, Box-and-whisker plots of numbers of ST-HSCs, CMPs, MEPs, and pDCs in BM of primary recipient mice in competitive transplantation. For box and whiskers plots throughout, bar indicates median, box edges first and third quartile values, and whisker edges minimum and maximum values. *P* value relative to control by a two-sided *t*-test.



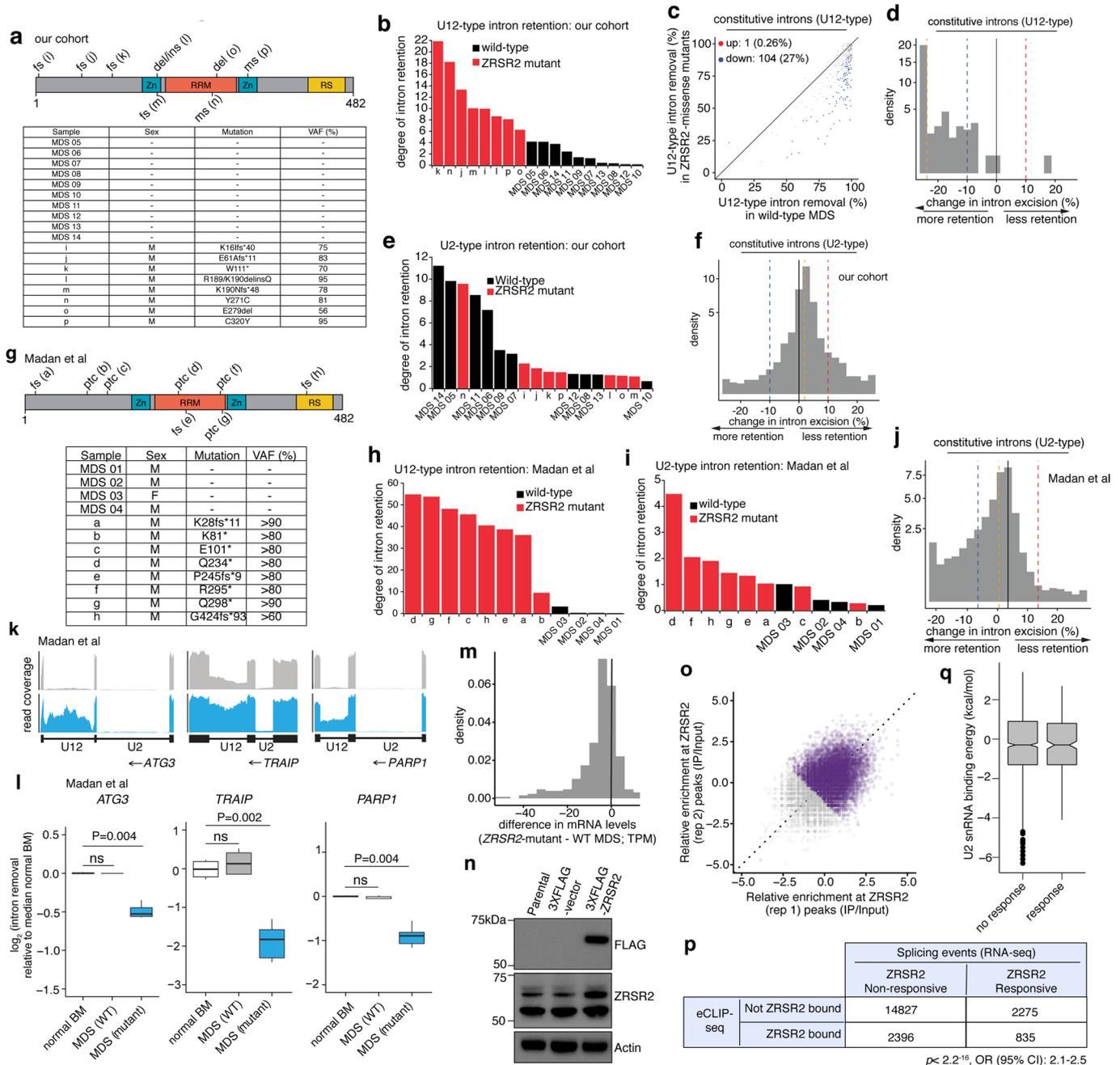
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Characterization of *Zrsr2* conditional knockout (cKO) mice. **a**, Absolute number of LT-HSCs, ST-HSCs, MPPs, LSKs, CMPs, and GMPs in primary, non-transplanted 20-week-old *Mx1-cre* control ('control') and *Mx1-cre Zrsr2^{fl/y}* mice. Mean \pm SD. $n = 5$ animals. **b**, Absolute number of live, bone marrow (BM) mononuclear cells in 10-week old *Mx1-cre* control ('control') and *Mx1-cre Zrsr2^{fl/y}* mice ('*Zrsr2* knockout') 4-weeks following *Zrsr2* excision. Mean \pm SD. $n = 5$ animals. P-values relative to control by two-sided *t*-test and indicated in figures. **c**, Schematic of BrdU analysis of hematopoietic stem cells from mice in (**b**) and Fig. 1f. **d**, Hyposegmented, hypogranular neutrophils in peripheral blood of 10-week old *Mx1-cre Zrsr2^{fl/y}* mice (yellow arrows). Bar: 10 μ m. **e**, BM cytopins indicating hyposegmented, hypogranular neutrophils (left panel) and dysplastic erythroid progenitors (middle and right panels, yellow arrows). Bar: 10 μ m. Experiment repeated three times with similar results in (**d**) and (**e**). **f**, Peripheral blood white blood cell counts (WBC), platelet count, hemoglobin (Hb), and mean corpuscular volume (MCV) in primary *Mx1-cre* control ($n = 9$) and *Mx1-cre Zrsr2^{fl/y}* ($n = 10$) mice (following *Zrsr2* excision at 6 weeks age). Mean \pm SD. **g**, Kaplan-Meier survival of primary control and *Zrsr2* KO mice (following *Zrsr2* excision at 6 weeks age). Absolute numbers of (**h**) bone marrow and (**i**) spleen B-cell subsets. **j**, Numbers of live, spleen (left) and thymic (right) mononuclear cells in mice from (**h**)-(**i**). Mean \pm SD. Absolute numbers of live mature hematopoietic cells (**k**) in marrow and (**l**) spleen of 8-week-old *Mx1-cre* control ('control') and *Mx1-cre Zrsr2^{fl/y}* ('knockout' or 'KO') mice (*Zrsr2* excision at 4 weeks). **m**, Absolute numbers of T-cell subsets in thymus of mice from (**k**). Mean \pm SD shown throughout. *Mx1-cre* control ($n = 5$) and *Mx1-cre Zrsr2^{fl/y}* ($n = 5$) mice were used in (**h**) to (**m**). P-values relative to control by two-sided *t*-test and indicated in figures.

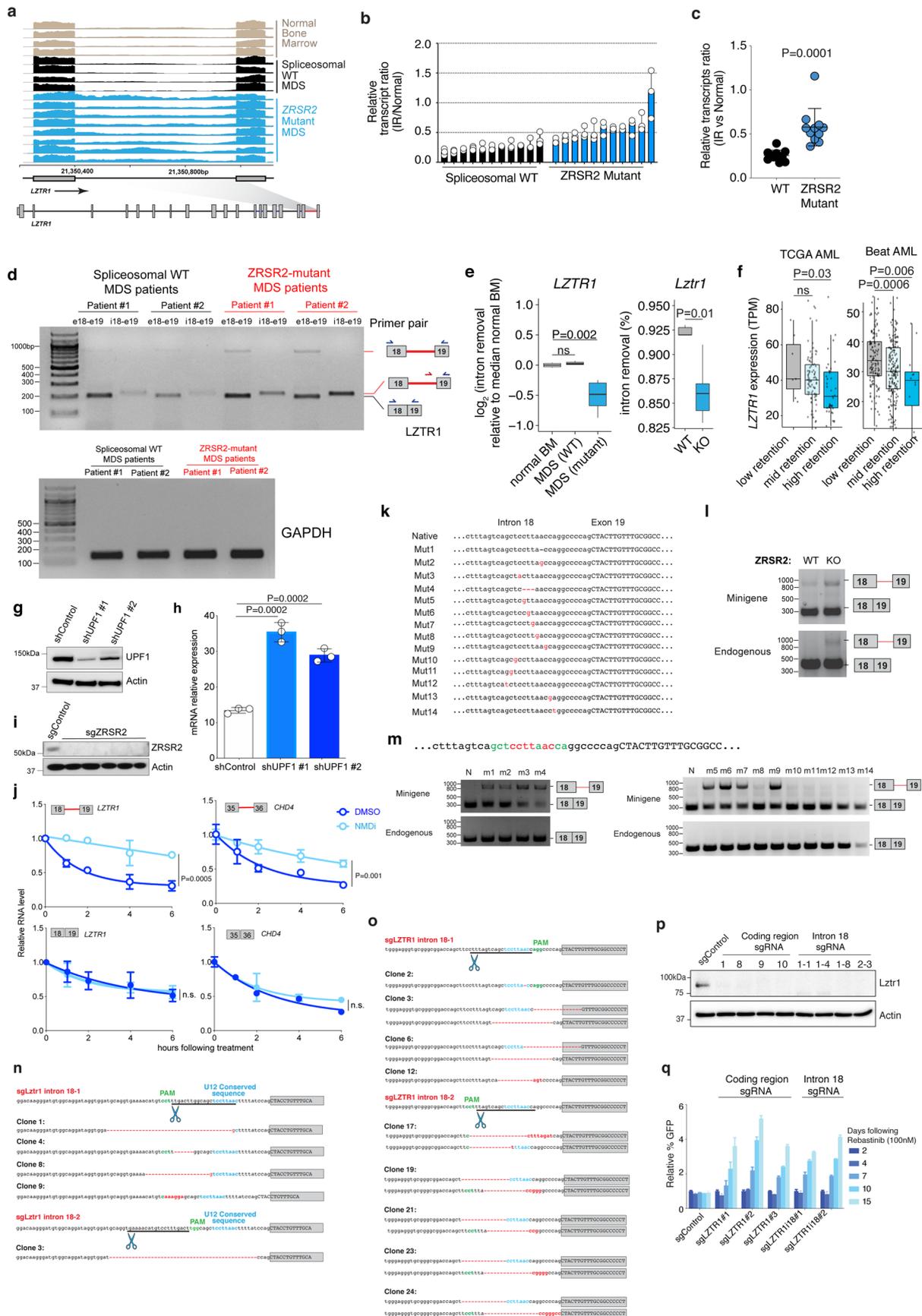


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Comparison of the effects of *Zrsr2* loss versus *Tet2* knockout or *Sf3b1*^{K700E} or *Srsrf2P95H* mutations on hematopoietic stem and progenitor cells. **a**, Schema of competitive bone marrow (BM) transplantation assays. **b**, Absolute number of CD45.2+ long-term HSCs (LT-HSCs), LSK, and MPPs in the bone marrow of CD45.1 recipient mice 16 weeks following plpC (n = 8–10 each). For box and whiskers plots throughout, bar indicates median, box edges first and third quartile values, and whisker edges minimum and maximum values. **c**, Percentage of CD45.2+ LT-HSCs, LSK, CMP, MEP, and GMP cells in the BM of CD45.1 recipient mice 16-weeks following plpC (n = 8–10 per each). P value was calculated relative to the control group by a two-sided *t*-test. **d**, Representative FACS plots of data in (c). **e**, Number of methylcellulose colonies generated from 100 sorted LT-HSCs from mice with the indicated genotype. n = 3 biologically independent experiments. Error bars, mean values ± SEM. P-values by one-way ANOVA with Tukey's multiple comparisons test. **f**, Percentage of CD45.2+ cells in the blood of recipient mice from *Zrsr2* knockout/*Sf3b1*^{K700E/WT} double mutant cells and relevant controls pre- and post-plpC administration to recipient mice (n = 10 each). P-values by two-way ANOVA with Tukey's multiple comparisons test. Data in (b), (c), and (f) are shown as box-and-whisker plots where bar indicates median, box edges first and third quartile values, and whisker edges minimum and maximum values.

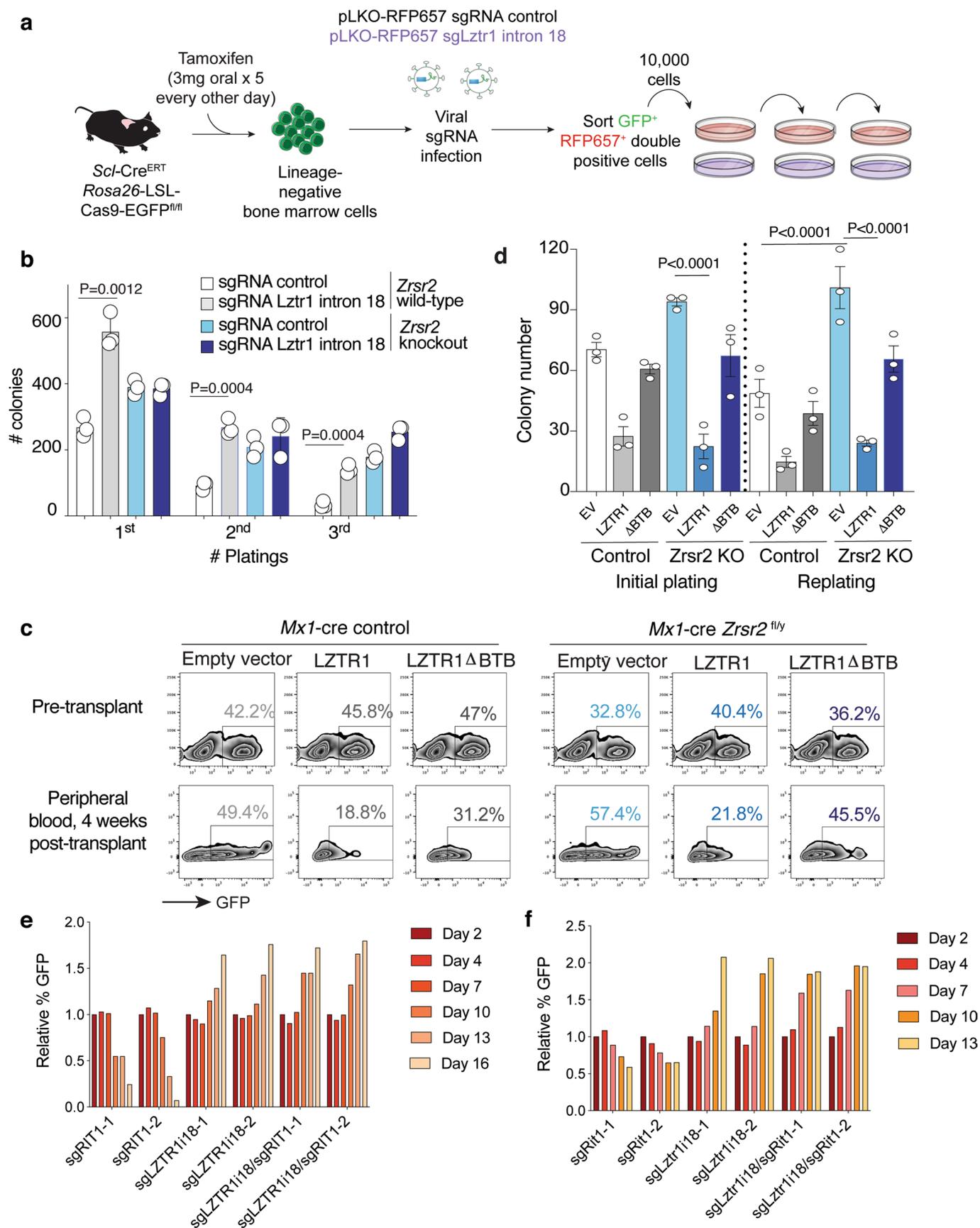


Extended Data Fig. 5 | Effect of ZRSR2 loss on minor intron splicing. **a**, ZRSR2 mutations in our cohort. 'MDS 05-14' are wild-type; 'i-p' are ZRSR2-mutant. VAF: variant allele frequency, fs: frameshift, ptc: premature termination codon, del: deletion, ins: insertion, ms: missense mutation. **b**, Comparison of U12-type intron retention in MDS samples vs. normal marrow. **c**, Differential splicing of U12-type introns. Each point corresponds to a single intron, illustrating percentage of mRNAs in which intron is spliced out. Blue/red dots: introns with significantly increased/decreased retention in ZRSR2-mutant vs. WT, with absolute change $\geq 10\%$ or absolute log fold-change of ≥ 2 with $p \leq 0.05$ (two-sided Mann-Whitney U, without adjustments for multiple comparisons). **d**, Distribution of intron retention in samples with ZRSR2 mutations. Blue/red dashed lines: thresholds of -10% and 10% for differential retention; gold line: median change in intron retention. **e**, As (**b**), and (**f**) as (**d**), for U2-type introns. **g**, As (**a**), and (**h**) as (**b**), for Madan et al. (**i**) As (**h**), for U2-type introns. **j**, As (**f**), for Madan et al. (**k**) RNA-seq coverage plots of U12-type introns averaging samples with indicated genotypes. **l**, Splicing efficiencies of introns in (**k**) relative to normal marrow (median over $n = 4$ normal samples). P-values: two-sided Mann-Whitney U. Middle line, hinges, notches, and whiskers: median, 25th/75th percentiles, 95% confidence interval and most extreme points within 1.5x interquartile range from hinge. **m**, Expression of genes with retained U12-type introns between ZRSR2-mutant vs. WT. **n**, Immunoblot in K562 cells used for eCLIP-seq (repeated twice with similar results). **o**, eCLIP of ZRSR2-binding sites. Input-normalized peak signals as log₂ fold-change. Purple points: eCLIP-enriched ZRSR2 peaks in biological replicates. **p**, Overlap of genes bound by ZRSR2 vs. differentially spliced in ZRSR2-mutant versus WT ('ZRSR2 responsive'). P-value: Fisher's exact test. **q**, U2 snRNA binding energy within ZRSR2 non-responsive and responsive minor introns.



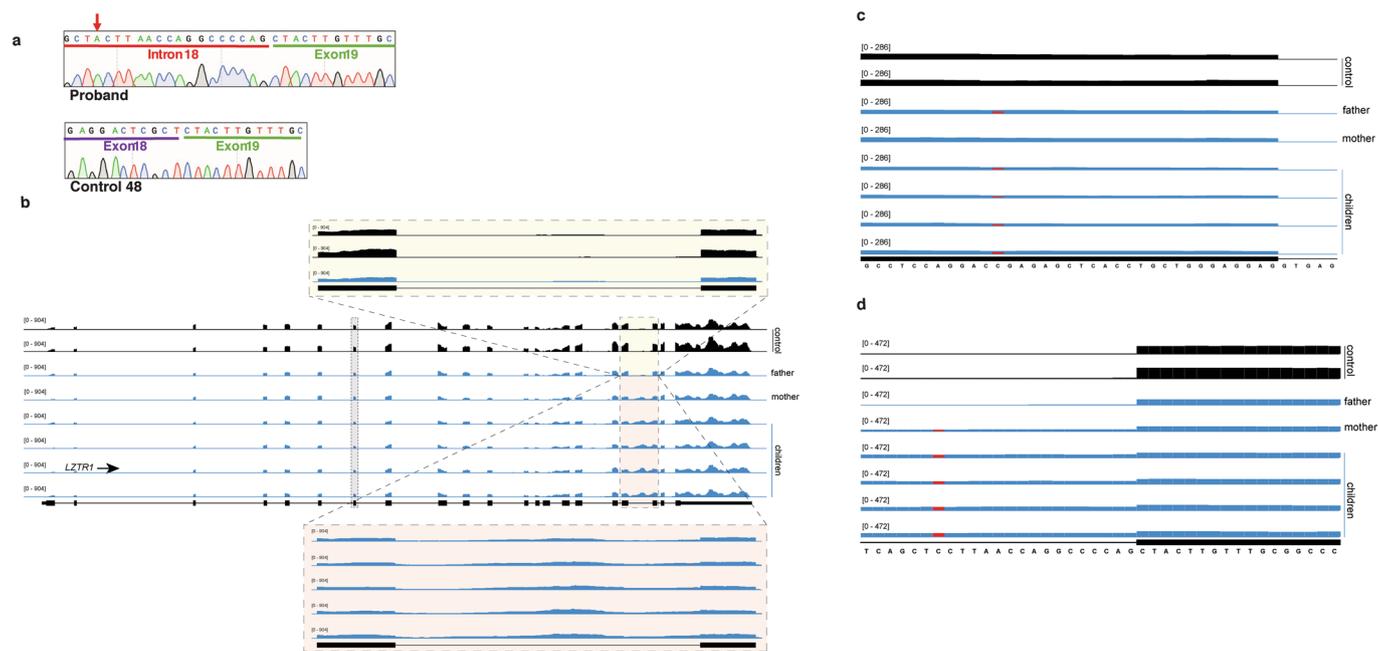
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Consequences of ZRSR2 and LZTR1 dysregulation. **a**, RNA-seq of *LZTR1*'s minor intron in normal and *ZRSR2*-mutant and WT MDS ($n=10$ each) marrow. **b**, Ratio of intron retained (IR) to normal *LZTR1* in MDS samples ($n=10$ each). Mean \pm SD. **c**, Data from **(b)** \pm SD. P-value; two-sided *t*-test. **d**, Qualitative RT-PCR using primers amplifying exons 18-19 ('e18-e19') as well as specific to IR isoform. **e**, Splicing efficiencies of *LZTR1* in patient samples (median over $n=4$ normal samples; left) or mouse lineage-negative c-Kit⁺ cells (right). P-values: two-sided Mann-Whitney U. **f**, *LZTR1* expression by level of minor intron retention ('Low': <10%; 'Mid': 10-20%; 'High-retention': >20%). P-values: one-sided Mann-Whitney U. In **(e)** and **(f)**: middle line, hinges, notches, and whiskers indicate median, 25th/75th percentiles, 95% confidence interval, and most extreme points within 1.5x interquartile range from hinge. **g**, UPF1 immunoblot in K562 cells with mutation disrupting *LZTR1*'s U12 sequence +/- anti-UPF1 shRNA. **h**, Expression of U12-retained *LZTR1* following actinomycin D +/- anti-UPF1 shRNA. $n=3$ biological replicates. Mean +/- SD. **i**, Immunoblot of K562 cells +/- *ZRSR2*-targeting sgRNA. **j**, Expression of *LZTR1* (left) or *CHD4* (right) isoforms in *ZRSR2*-null K562 cells with DMSO or NMD inhibitor (PMID 24662918). Mean +/- SD; P values: two-sided *t*-test. $n=3$ biological replicates. **k**, *LZTR1* minigene with mutations generated. **l**, RT-PCR of *LZTR1* minigene and endogenous mRNA from WT or *ZRSR2*-KO K562 cells. **m**, RT-PCR of *LZTR1* minigene and endogenous using native ('N') or mutant minigenes. **n**, *Lztr1* minor intron with location of sgRNA, PAM site, 3' U12 consequence (blue text), and sequence in individual Ba/F3 cell clones (red dash: deleted nucleotides). **o**, As (n) in K562 cells. **p**, Immunoblot of *Lztr1* in Ba/F3 single-cell clones +/- *Lztr1* protein-coding or minor intron sgRNAs. **q**, Median relative percentage of GFP-labeled K562 cells following Rebastinib. Experiments in **(d)**, **(g)**, **(i)**, and **(l)** were repeated twice with similar results.

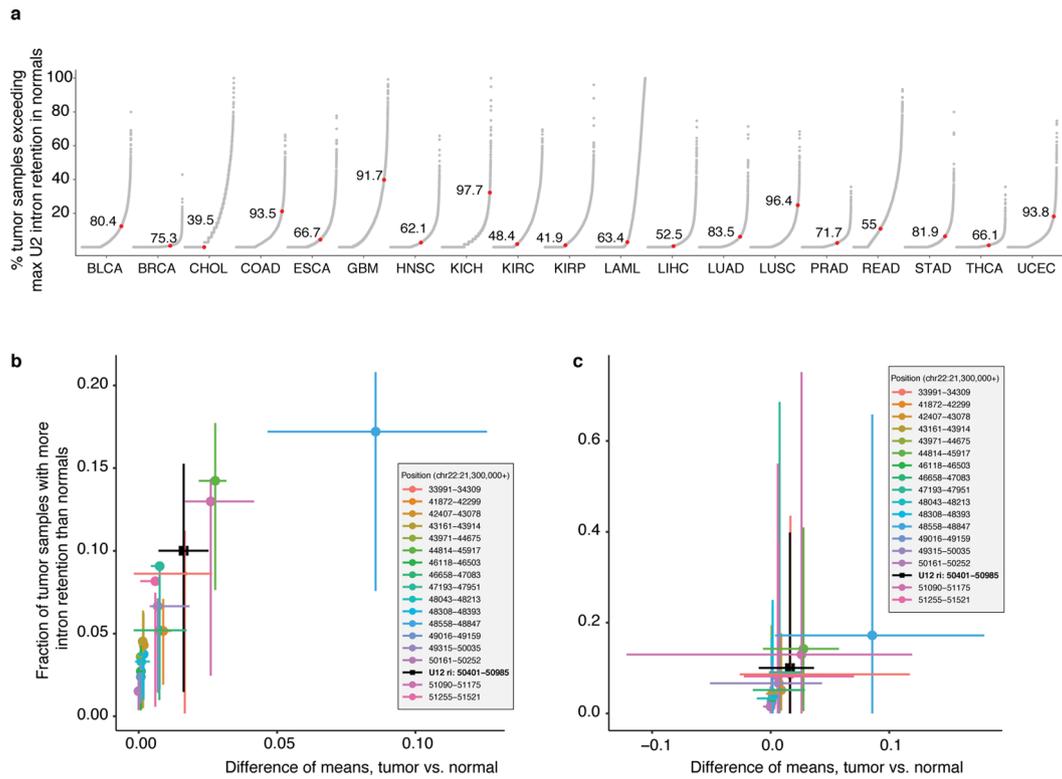


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Impaired Lztr1 minor intron splicing augments clonogenic capacity of hematopoietic precursors. **a**, Schema of experiment whereby sgRNAs targeting the conserved U12 sequence in *Lztr1*'s minor intron are delivered to lineage-negative hematopoietic precursors from *Scl-Cre^{ERT}* *Rosa26-Lox-STOP-Lox* Cas9-EGFP *Zrsr2^{fl/y}* or *Zrsr2* wild-type mice followed by serial replating *in vitro*. In this experiment, sgRNAs are encoded from an RFP657 expressing plasmid and GFP⁺/RFP657⁺ double-positive cells were purified for plating. **b**, Mean number of colonies following *Lztr1* minor intron mutagenesis versus control sgRNA treated bone marrow cells in *Zrsr2* wild-type or knockout background from **(a)**. Bars represent standard deviation. P-values calculated relative to the control group by a two-sided *t*-test. *n* = 3 biologically independent experiments. Error bars, mean values \pm SD. **c**, Representative FACS plots of GFP% in cells just before transplantation and in peripheral blood of recipient transplanted mice 4 weeks after transplantation from Fig. 5g. **d**, Number of colonies in methylcellulose CFU assays from LT-HSCs from mice in **(c)**. *n* = 3 biologically independent experiments. Error bars, mean values \pm SD. P-values by two-way ANOVA with Tukey's multiple comparisons test. **e**, Relative percentage of GFP-labeled K562 cells with knockout of RIT1 and/or mutagenesis of the minor intron in LZTR1 mixed with equal proportions of unlabeled cells to the BCR-ABL inhibitor imatinib. **f**, Relative percentage of Ba/F3 cells treated with sgRNAs targeting *Rit1* and/or the minor intron of *Lztr1* following IL-3 withdrawal (median % relative to day 2 is plotted).



Extended Data Fig. 8 | LZTR1 minor intron retention in cancer predisposition syndromes. **a**, Sanger sequence electropherogram of the *LZTR1* intron 18 retained isoform (from a representative affected family member in Fig. 6d; corresponds to the top band in the *LZTR1* RT-PCR gel in Fig. 5d) and *LZTR1* normal spliced isoform from a control fibroblast sample (corresponds to the bottom band in the *LZTR1* RT-PCR gel in Fig. 6d). Red arrow indicates mutant nucleotide in the affected family members. **b**, RNA-seq coverage plots of *LZTR1* in fibroblasts from Noonan syndrome family and controls. Zoom in magnifies the minor intron of *LZTR1*. **c**, As (**b**), but zoomed in on the region of mutation in the father. **d**, As (**b**), but zoomed in on the region of mutation within *LZTR1*'s minor intron.



Extended Data Fig. 9 | LZTR1 minor intron retention is pervasive in cancers. a, Degree of major (U2-type) intron retention across normal (N) and tumor (T) samples in cancers from TCGA. Each point corresponds to a single U2-type intron and indicates the percentage of all tumor samples in which retention of that intron exceeds the maximum corresponding retention of that intron observed in normal samples. Red dot indicates the U12-type intron of *LZTR1* for comparison. **b**, Each point illustrates the frequency of retention of a single intron of *LZTR1* (see inset for key) across all TCGA cohorts with matched normal samples. Values along the x axes represents the mean difference in intron retention in tumor versus normal samples within a cancer type, while the y axes represent the fraction of tumor samples with intron retention that exceeds that of the normal sample with the most intron retention within a cancer type. Points represent the mean value computed across all cancer types, while whiskers represent the interquartile range across cancer types. **c**, As (**b**), but whiskers represent the entire range.

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 our [Editorial Policies](#) 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 software was used for data collection.

Data analysis Publicly available software was used in this study. Specific programs are RSEM (v1.2.4), Bowtie (v1.0.0), TopHat (v2.0.8b), MISO (v2.0), Bioconductor (v3.7) within the R (v3.5.1) programming environment, seqLogo and ggseqlogo (PMID 29036507) packages in Bioconductor, the ggplot2 package (doi:10.1007/978-3-319-24277-4) in R (v3.5.1). Illustrated transcript annotations are from RefSeq, downloaded from the UCSC Genome Browser. Genome annotations were from Ensembl release 71. Flow cytometry data were analyzed using FlowJo v9.

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 and 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

Genome annotations for human and mouse were from NCBI GRCh37/UCSC hg19 and NCBI GRCm38/UCSC mm10 respectively. Isoform annotations were from the MISO v2.0 database⁵⁰ and were merged with genome annotations from Ensembl release 7151 and the UCSC knownGene track⁵². RNA-seq reads for the human samples reported in Madan et al were downloaded from the Gene Expression Omnibus (accession number GSE63816). RNA-seq data generated by TCGA (dbGaP accession phs000178.v11.p8) and the Beat AML data (dbGaP accession phs001657.v1.p1) were downloaded from the National Cancer Institute Genomic Data Commons. Branchpoint-related data were obtained from a published study (PMID 29666160). RNA-seq data generated as part of this study have been deposited in

the Gene Expression Omnibus (mouse data: accession GSE149455) and the human RNA-seq data are being deposited in dbGaP phs002212.v1.p1 and available from the investigator's upon request.

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](https://www.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	Sample size calculation was not performed. Sample sizes for mouse transplantation experiments were chosen based on published studies using similar assays and based on power calculations wherein with 10 mice in each group, we would have 80% power to detect a 20% difference in defined cell populations assuming an effect size ($\Delta\mu/\sigma$) of 2.0 between groups at the 0.05 significance level. Sample sizes for RNA-seq and functional studies were performed based on standards in the field (Kim et al. Cancer Cell 2015; Inoue et al. Nature 2019; Yoshimi et al. Nature 2019) and is indicated in figure legends.
Data exclusions	No data were excluded.
Replication	Attempts at replication were successful. All key experiments were repeated two to six times independently, as indicated in figure legends.
Randomization	Animals were randomly assigned to experimental groups. Otherwise, all experiments were performed based on disease versus no disease or based on genotype and therefore randomization was not relevant.
Blinding	The data presented did not require the use of blinding as knowledge of disease state, genetic perturbation, or genotype was required in order to perform the analysis.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

The following antibodies were used for Western Blot analysis: Zrsr2 (custom, Yenzym, 1:1000), LZTR1 (sc-390166 or sc-390166 X, Santa Cruz Biotechnology, 1:1000), RIT1 (ab53720, Abcam, 1:1000), pan-RAS antibody (Thermo Fisher Scientific, MA1-012X, 1:1000), FLAG (F-1084, Sigma-Aldrich, 1:1000), UPF1 (ab109363, Abcam, 1:1000), and Actin (A-5441, Sigma-Aldrich, 1:4000). The following antibodies were used for Flow Cytometry analysis: B220-APCCy7 (clone: RA3-6B2; BioLegend; catalog #: 103224; dilution: 1:200); B220-PerCpCy5.5 (RA3-6B2; eBioscience; 45-0452-82; 1:200); CD3-PEcy7 (17A2; BioLegend; 100220; 1:200); CD3-APCCy7 (17A2; BioLegend; 100222; 1:200); Gr1-APC (RB6-8C5; eBioscience; 25-5931-82; 1:500); CD11b-FITC (M1/70; BioLegend; 101206; 1:200); CD11b-APCCy7 (M1/70; BioLegend; 101226; 1:200); NK1.1-APCCy7 (PK136; BioLegend; 108724; 1:200); Ter119-APCCy7 (Ter119; BioLegend; 116223; 1:200); cKit-APC (2B8; BioLegend; 105812; 1:100); cKit-PerCpCy5.5 (2B8; BioLegend; 105824; 1:100); cKit-Bv605 (ACK2; BioLegend; 135120; 1:100); Sca1-PECy7 (D7; BioLegend; 108102; 1:100); CD45.1-FITC (A20; BioLegend; 110706; 1:200); CD45.1-PerCpCy5.5 (A20; BioLegend; 110728; 1:200); CD45.1-Bv711 (A20; BioLegend; 110739; 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-PerCpCy5.5 (HM48-1; BioLegend; 103422; 1:100); CD150-PE (9D1; eBioscience; 12-1501-82; 1:100); CD127 (IL-7R α)-APCCy7 (A7R34; BioLegend; 135040; 1:200); CD8-APCCy7 (53-6.7; BioLegend; 100714; 1:200); CD19-APCCy7 (6D5; BioLegend; 115529; 1:200); IgM-PE (II/41; eBioscience; 12-5790-82; 1:200); IgD-FITC (11-26c.2a; BioLegend; 405704; 1:200); CD135-APC (A2F10; BioLegend; 135310; 1:200); CD4-APCCy7 (GK1.5; BioLegend; 100413; 1:200); CD25-BV711 (PC61.5; BioLegend; 102049; 1:200); CD44-APC (IM7; BioLegend; 103012; 1:200); D43 (eBioR2/60; eBioscience; 11-0431-85; 1:200); CD24-BV605 (M1/69; BD Biosciences; 563060; 1:200); CD21/35-PE (4E3; eBioscience; 12-0212-82; 1:200); CD93-APC (AA4.1; eBioscience; 17-5892-82;

1:200); CD23-e450 (B3B4; BD Biosciences; 48-0232-82; 1:200) CD16/CD32 (FcyRII/III)-Alexa700 (93; eBioscience; 56-0161-82; 1:100); CD34-FITC (RAM34; BD Biosciences; 553731; 1:50); CD34-PerCP (8G12; BD Biosciences; 345803; 1:50); CD117-PECy7 (104D2; eBioscience; 25-1178-42; 1:100); CD45-APC-H7 (2D1; BD Biosciences; 560178; 1:200); Ly-51-PE (BP-1; BD Bioscience; 553735; 1:200).

Validation

With the exception of the novel rabbit anti-mouse Zrsr2 antibody, all antibodies were validated by the supplier for human and/or mouse samples, and were checked in the lab by Western blotting on cell lysate and by comparing to the manufacturer's or in-house results. The use of each antibody for Western blot and/or flow cytometry for mouse and/or human cells are validated by the supplier based on information provided on their websites. Specificity of the newly reported rabbit anti-mouse Zrsr2 antibody was verified using lysates from mice with genetic knockout of Zrsr2.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

K562, Ba/F3, 32D, TF1, and HEK293T cells were obtained from the American Type Culture Collection (ATCC). Human primary fibroblasts were provided by the Frank McCormick laboratory.

Authentication

An aliquot of each cell lines were authenticated using ATCC fingerprinting. Otherwise, the cells were submitted for short tandem repeat (STR) profiling to confirm their authenticity.

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 male and female Mx1-cre Zrsr2 floxed, Mx1-cre Sf3b1K700E/WT, Mx1-cre Srsf2P95H/WT, and Mx1-cre Tet2 floxed mice, all on a C57Bl/6 background were used for all experiments. The following mice were purchased from Jackson laboratory: Mx1-cre transgenic mice (stock #003556), Sf3b1K700E conditional knockin mice (stock #029766), Srsf2P95H conditional knockin mice (stock #028376), Tet2 floxed (stock #017573), and Rosa26-LSL-Cas9EGFP knockin mice (stock 026175).

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 animals were housed at Memorial Sloan Kettering Cancer Center (MSKCC). 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. All mouse experiments were performed in accordance with a protocol approved by the MSKCC Institutional Animal Care and Use Committee (11-12-029).

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

Patients with myeloid neoplasms seen at Memorial Sloan Kettering Cancer Center who provided anonymized bone marrow mononuclear cells.

Recruitment

Patients with myeloid neoplasms seen at Memorial Sloan Kettering Cancer Center (MSK) who consented to MSK IRB protocol 06-107 were eligible for inclusion regardless of race, gender, ethnicity or other characteristics. Participants are recruited from the diverse patient population served by MSK and there was no known selection bias.

Ethics oversight

Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center and Fred Hutchinson Cancer Research Center and conducted in accordance to the Declaration of Helsinki protocol. Informed consents were obtained from all human subjects. Patients provided samples after their informed consent and primary human de-identified MDS/AML/MPN samples derived from bone marrow (BM) mononuclear cells (MNCs) were utilized. Genomic alterations in patient BM MNCs were analyzed using MSK IMPACT assay. Patient samples were anonymized by the Hematologic Oncology Tissue Bank of MSK.

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

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 a lineage cocktail comprised of antibodies targeting CD3, CD4, CD8, B220, CD19, NK1.1, Gr-1, CD11b, Ter119, and IL-7R α . Cells were also stained with antibodies against c-Kit, Sca1, CD150, and CD48. Cell populations were analyzed using an LSR Fortessa (Becton Dickinson) and sorted with a FACSAria II instrument (Becton Dickinson). We used the following antibodies: c-Kit (2B8), Sca1 (D7), Mac-1/CD11b (M1/70), Gr-1 (RB6-8C5), NK1.1 (PK136), Ter119, IL-7R α (A7R34), CD4 (RM4-5), CD8 (53-6.7), CD45.1 (A20), CD45.2 (104), CD150 (9D1), CD48 (HM48-1), CD45ALL (30-F11), and CD45RB (C363-16A). The composition of immature and mature hematopoietic cell lineages in the bone marrow, spleen, and peripheral blood was assessed using a combination of antibodies against B220 (RA3-6B2), CD19 (1D3), CD3 (17A2), CD4 (GK1.5), CD8a (53-6.7), CD11b (M1/70), CD25 (PC61.5), CD44 (IM7), Gr-1, IgM (II/41), IgD (11-26c.2a), CD43 (S11), CD24 (M1/69), Ly-51 (BP-1), CD21/35 (4E3), CD93 (AA4.1), CD23 (B3B4), c-Kit (2B8), Sca-1 (D7), CD127 (A7R34) and CD135 (A2F10). For cell cycle analysis, the BrdU-APC kit was used (BD) according to the manufacturer's protocol. For evaluation of apoptosis, the Annexin V FITC apoptosis detection kit (BD) was used according to manufacturer's recommendations. DAPI was used in both BrdU and Annexin V experiments.

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 positivity in post-sort samples, the sorted samples were analyzed for GFP 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.