Cancer Cell

SRSF2 Mutations Contribute to Myelodysplasia Through Mutant-Specific Effects on Exon Recognition

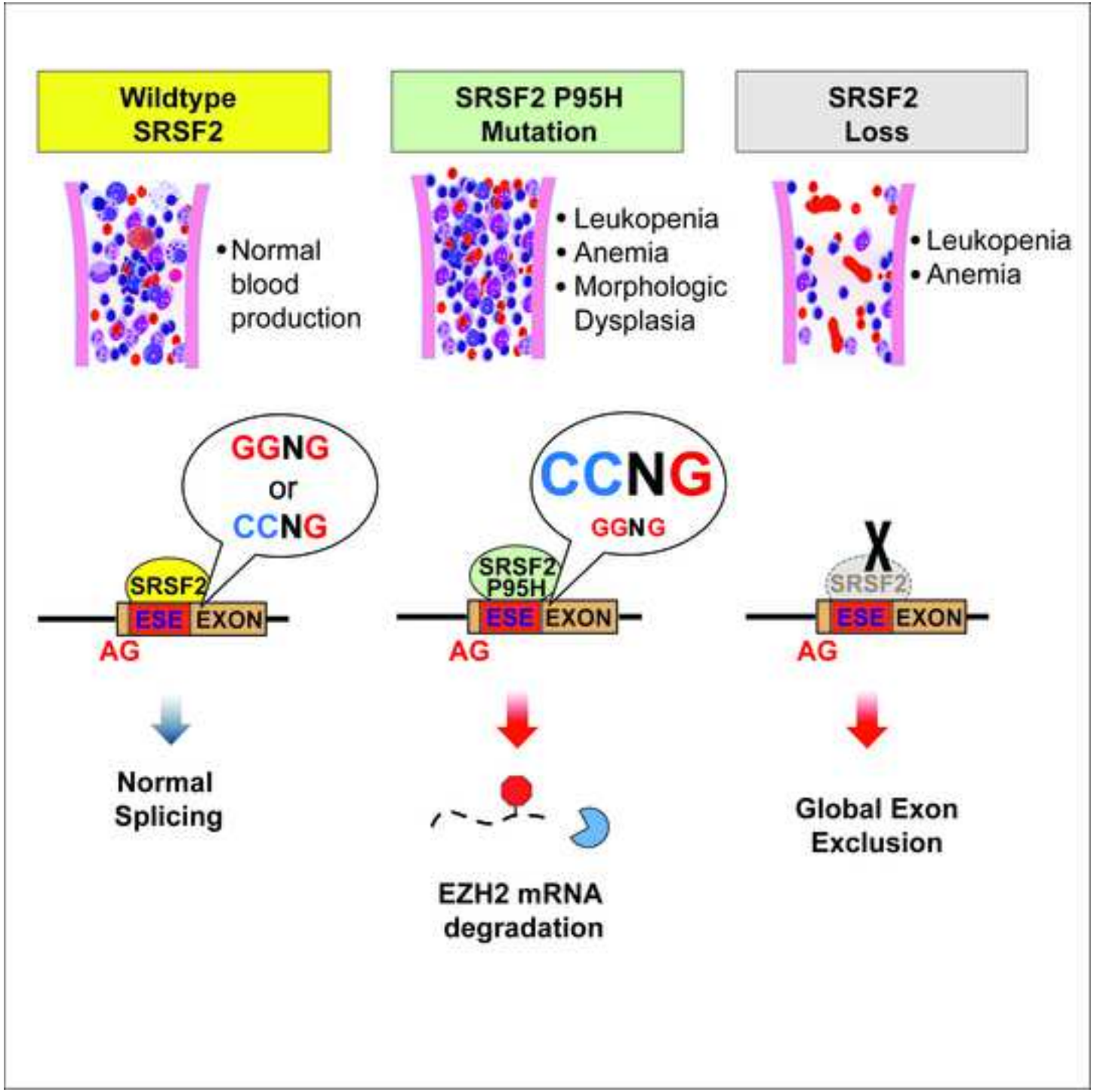
--Manuscript Draft--

Manuscript Number:	CANCER-CELL-D-14-00970R3
Full Title:	SRSF2 Mutations Contribute to Myelodysplasia Through Mutant-Specific Effects on Exon Recognition
Article Type:	Research Article
Keywords:	leukemia; myelodysplastic syndromes; Splicing; EZH2; SRSF2; SF3B1; U2AF1.
Corresponding Author:	Omar Abdel-Wahab New York, UNITED STATES
First Author:	Eunhee Kim
Order of Authors:	Eunhee Kim
	Janine O. Ilagan
	Yang Liang
	Gerrit M. Daubner
	Stanley C.-W. Lee
	Aravind Ramakrishnan
	Yue Li
	Young Rock Chung
	Jean-Baptiste Micol
	Michele Murphy
	Hana Cho
	Min-Kyung Kim
	Ahmad S. Zebari
	Shlomzion Aumann
	Christopher Y. Park
	Silvia Buonamici
	Peter G. Smith
	H. Joachim Deeg
	Camille Lobry
	Iannis Aifantis
	Yorgo Mordis
	Frederic H.-T. Allain
	Stephanie Halene
	Robert K. Bradley
	Omar Abdel-Wahab
Abstract:	Mutations affecting spliceosomal proteins are the most common class of mutations in patients with myelodysplastic syndromes (MDS), yet their role in MDS pathogenesis has not been delineated. Here we report that mutations affecting the splicing factor SRSF2 directly impair hematopoietic differentiation in vivo, which is not due to SRSF2

	loss of function. By contrast, SRSF2 mutations alter SRSF2's normal sequence-specific RNA binding activity, thereby altering recognition of specific exonic splicing enhancer motifs to drive recurrent mis-splicing of key hematopoietic regulators. This includes SRSF2 mutation-dependent splicing of EZH2 that triggers nonsense-mediated decay, which, in turn, results in impaired hematopoietic differentiation. These data provide a mechanistic link between a mutant spliceosomal protein, alterations in splicing of key regulators, and impaired hematopoiesis.
Suggested Reviewers:	Benjamin Ebert Benjamin_Ebert@dfci.harvard.edu Expert in MDS, functional genomics in leukemia, and has interest in this area of cancer biology specifically.
	Daniel Starczynowski Daniel.Starczynowski@cchmc.org Expert in MDS and murine modeling of MDS.
	Ulrich Steidl ulrich.steidl@einstein.yu.edu Expert in leukemia, murine models and transcriptional regulation in leukemia.
	Ari Melnick amm2014@med.cornell.edu Expert in leukemia, genomics of leukemia, and transcriptional regulation in leukemia.
	Thomas Look thomas_look@dfci.harvard.edu Expert in functional genomics and mouse models of leukemia.
	Benjamin Blencowe b.blencowe@utoronto.ca Expert in genomic analysis of RNA splicing in disease.
Opposed Reviewers:	Xiangdong Fu xdfu@ucsd.edu There is an element of competition in this story with Dr. Fu.
	Dong-Er Zhang d7zhang@ucsd.edu There is an element of competition in this story with Dr. Zhang.
	Seishi Ogawa dr_ogawa@mac.com There is an element of competition in this work with Dr. Ogawa.

Graphical Abstract

Click here to download Graphical Abstract: 2015_02_19 graphical abstract submitted.tif



SRSF2 Mutations Contribute to Myelodysplasia Through Mutant-Specific Effects on Exon Recognition

Authors: Eunhee Kim[1,15], Janine O. Ilagan[2,3,15], Yang Liang[4,15], Gerrit M. Daubner[5,15], Stanley C.-W. Lee[1], Aravind Ramakrishnan[6,7], Yue Li[8], Young Rock Chung[1], Jean-Baptiste Micol[1], Michele Murphy[2,3], Hana Cho[1], Min-Kyung Kim[1], Ahmad S. Zebari[2,3], Shlomzion Aumann[1], Christopher Y. Park[1,9], Silvia Buonamici[10], Peter G. Smith[10], H. Joachim Deeg[6,7], Camille Lobry[11], Iannis Aifantis[12], Yorgo Modis[8,13], Frederic H.-T. Allain[5], Stephanie Halene[4,15,16], Robert K. Bradley[2,3,15,16,17], Omar Abdel-Wahab[1,14,15,16,17]

Affiliations:
[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[2]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[3]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[4]Hematology, Yale Comprehensive Cancer Center and Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA
[5]Institute for Molecular Biology and Biophysics, ETH, Zürich, Switzerland
[6]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[7]Division of Medical Oncology, School of Medicine, University of Washington, Seattle, WA, USA
[8]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA
[9]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY USA
[10]H3 Biomedicine, Cambridge, MA 03129, USA
[11] Institut National de la Santé et de la Recherche Medicale (INSERM) U1009; Institut Gustave Roussy, Villejuif, France; Université Paris-Sud, Orsay, France.
[12]Howard Hughes Medical Institute and Department of Pathology, New York University School of Medicine, New York, NY 10016
[13]Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK
[14]Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[15]**Equal contributors.**

[16]**Co-senior authors.**

[17]**Address correspondence to:**

Robert K. Bradley
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N.
Mailstop: M1-B514
Seattle, WA 98109-1024
Phone: 206-667-5662
Fax: 206-667-1319
Email: rbradley@fhcrc.org

or

Omar Abdel-Wahab
Zuckerman 802,
Memorial Sloan Kettering Cancer Center
408 E. 69th Street
New York, NY USA 10065
Phone: 646-888-3487
Fax: 646-422-0890
Email: abdelwao@mskcc.org

Running Title: *SRSF2* Mutations Alter ESE Preference to Induce MDS

SUMMARY

Mutations affecting spliceosomal proteins are the most common class of mutations in patients with myelodysplastic syndromes (MDS), yet their role in MDS pathogenesis has not been delineated. Here we report that mutations affecting the splicing factor SRSF2 directly impair hematopoietic differentiation *in vivo,* which is not due to SRSF2 loss of function. By contrast, *SRSF2* mutations alter SRSF2's normal sequence-specific RNA binding activity, thereby altering recognition of specific exonic splicing enhancer motifs to drive recurrent mis-splicing of key hematopoietic regulators. This includes *SRSF2* mutation-dependent splicing of *EZH2* that triggers nonsense-mediated decay, which, in turn, results in impaired hematopoietic differentiation. These data provide a mechanistic link between a mutant spliceosomal protein, alterations in splicing of key regulators, and impaired hematopoiesis.

SIGNIFICANCE

Frequent somatic mutations affecting components of the spliceosome have been identified in hematologic malignancies; however, the functional role of these mutations is not known. Here, we identify that commonly occurring mutations in the spliceosomal gene *SRSF2* impair hematopoietic differentiation and promote myelodysplasia by altering SRSF2's preference for specific exonic splicing enhancer motifs. This results in consistent mis-splicing in a manner that promotes the expression of abnormal isoforms of a number of key hematopoietic regulators, some of which have previously been linked to leukemogenesis (including *BCOR* and *EZH2*). These data provide a mechanistic basis for the enrichment of spliceosomal mutations in myelodysplasia, and identify altered RNA recognition as an important driver of leukemogenesis.

INTRODUCTION

Somatic mutations in genes encoding components of the spliceosome have been identified in a spectrum of human malignancies, including ~60% of patients with myelodysplastic syndromes (MDS) (Bejar et al., 2012; Papaemmanuil et al., 2013; Yoshida et al., 2011). These mutations occur most commonly in *SF3B1 (Splicing Factor 3b Subunit 1), SRSF2 (Serine/arginine-Rich Splicing Factor 2),* and *U2AF1 (U2 Small Nuclear RNA Auxiliary Factor 1)* and almost always as heterozygous missense mutations that are mutually exclusive (Papaemmanuil et al., 2011; Wang et al., 2011; Yoshida et al., 2011). While the genetic data in MDS suggests that these alterations are critical to disease pathogenesis, it remains unknown how these mutations contribute to MDS and if they are sufficient to induce MDS.

Recent studies have suggested that mutations in the spliceosomal gene *U2AF1* alter RNA splicing (Brooks et al., 2014; Graubert et al., 2012; Ilagan et al., 2014; Przychodzen et al., 2013; Quesada et al., 2012) and studies of gene expression in primary patient samples with and without *U2AF1* mutations have been performed in an effort to identify downstream mis-spliced genes that might contribute to abnormal hematopoiesis (Brooks et al., 2014; Graubert et al., 2012; Ilagan et al., 2014). However it remains unknown how these mutations contribute to hematopoietic transformation. To date, no studies have investigated the *in vivo* effects of spliceosomal mutations expressed from the endogenous locus in the correct cellular context, which might allow delineation of how these alleles contribute to MDS pathogenesis.

To test whether spliceosomal gene mutations are sufficient to drive MDS, and determine how altered RNA splicing contributes to transformation *in vivo*, we studied the biological and transcriptional consequences of mutations in *SRSF2*. *SRSF2* mutations occur in 20-30% of MDS and ~50% of chronic myelomonocytic leukemia (CMML) patients (Papaemmanuil et al., 2013; Yoshida et al., 2011). SRSF2 is a member of the serine/arginine-rich (SR) protein family that contributes to both constitutive and alternative splicing by binding to exonic splicing enhancer (ESE) sequences within pre-mRNA through its RNA recognition motif domain (RRM)

(Graveley and Maniatis, 1998; Liu et al., 2000; Schaal and Maniatis, 1999; Zahler et al., 2004). *SRSF2* mutations are consistently associated with adverse outcome amongst MDS and AML patients (Papaemmanuil et al., 2013; Vannucchi et al., 2013; Zhang et al., 2012). Despite the clinical importance of *SRSF2* mutations, to date there have been no studies of the functional impact of *SRSF2* mutations on hematopoiesis or splicing. Here, we studied the biological and transcriptional effects of somatic expression of the common *SRSF2*P95H mutation in the hematopoietic compartment.

RESULTS

*Srsf2*P95H mutant mice develop MDS, a phenotype distinct from mice with heterozygous or homozygous loss of *Srsf2*

Given the genetic heterogeneity of primary patient samples as well as the fact that stable overexpression of spliceosomal proteins, even in wild-type (WT) form, is poorly tolerated (Lareau et al., 2007), we first generated a murine model for conditional expression of the commonly occurring *SRSF2*P95H mutation from the endogenous murine locus of *Srsf2* (**Figure 1A** and **Figure S1A-B**). Mice heterozygous for the *Srsf2*P95H allele (*Srsf2*P95H/WT) were crossed to *Mx1-cre* transgenic mice (Kuhn et al., 1995) on a C57BL/6 background to allow for inducible expression of Cre recombinase following intraperitoneal injection of polyinosine-polycytosine (pIpC) (12 μg/g every other day for three days by injection as previously described (Moran-Crusio et al., 2011)) (**Figure S1C-D** and **Supplemental Experimental Procedures**). mRNA sequencing (RNA-seq) analysis of hematopoietic stem/progenitor cells (HSPCs) two weeks after the last pIpC injection of 6-week-old *Mx1*-cre *Srsf2*P95H/WT and *Mx1*-cre *Srsf2* WT control mice confirmed heterozygous expression of the mutant allele in equal proportion to the remaining WT *Srsf2* allele in *Mx1*-cre *Srsf2*P95H/WT mice (**Figure 1B**).

It is currently unknown if the heterozygous *SRSF2*P95H mutation confers a gain-of-function, haploinsufficient loss-of-function, or dominant-negative loss-of-function. We therefore compared expression of the *Srsf2*P95H mutation with conditional loss of *Srsf2 in vivo* (Wang et al., 2001). Bone marrow (BM) mononuclear cells (MNCs) from 6-week-old CD45.2 *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/WT (heterozygous floxed mice for inducible deletion of 1 copy of *Srsf2*), *Mx1-cre Srsf2*fl/fl (homozygous floxed mice for inducible deletion of both copies of *Srsf2*), and *Mx1-cre Srsf2*P95H/WT were transplanted into lethally irradiated congenic CD45.1 recipient mice followed by pIpC injection 4 weeks later (note that all mice were treated with pIpC to control for any potential phenotypic effects of pIpC administration on biological or splicing

phenotypes). This was done to assess for the phenotypic effects of *Srsf2* deletion or mutation in a hematopoietic cell-autonomous manner. Western blot (WB) analysis revealed deletion of *Srsf2* in BM MNCs from *Mx1-cre Srsf2*fl/fl mice and normal total Srsf2 levels in *Mx1-cre Srsf2*P95H/WT BM MNCs (**Figure S1E**). Significant leukopenia and anemia were seen in mice with homozygous *Srsf2* deletion or heterozygous expression of the P95H mutation 18 weeks post-transplant (**Figure 1C-D**) that was also seen at earlier time points (**Figure S1F-G**). The presence of similar cytopenias in mice bearing homozygous *Srsf2* deletion and heterozygous *Srsf2*P95H point mutation suggested a possible dominant-negative function imposed by the P95H mutation. However, the anemia in *Srsf2*P95H mice was characterized by increased mean corpuscular volume (MCV) of red blood cells relative to WT mice or mice with loss of 1-2 copies of *Srsf2* (**Figure 1E**). Moreover, histological assessment of mice 14 weeks post-pIpC revealed prominent BM aplasia in *Srsf2* homozygous knockout (KO) mice, whereas mice expressing the heterozygous P95H mutation had normal BM cellularity (**Figure 1F**). Platelet counts were normal in *Srsf2*P95H mutant mice at all time points examined (**Figure S1H**).

Given that macrocytic anemia, a hallmark of anemia in MDS, was present in *Srsf2*P95H mutant mice we next performed cytological examination of peripheral blood and bone marrow smears from *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/fl, and *Mx1-cre Srsf2*P95H/WT mice to assess for morphologic dysplasia. This revealed prominent myeloid and erythroid dysplasia in *Srsf2*P95H mice but not in *Mx1-cre Srsf2* WT or *Mx1-cre Srsf2*fl/fl mice (**Figure 1G** and **Figure S1I**). Myeloid dysplasia was apparent based on detection of hypolobated and hypogranulated neutrophils while erythroid dysplasia was evident based on nuclear irregularities and cytoplasmic vacuolization and blebbing in erythroid precursors. Overall, these results indicate that mutations in *Srsf2*P95H result in morphologic dysplasia and cytopenias with preserved marrow cellularity, features which are characteristic of human MDS, whereas complete loss of Srsf2 is incompatible with hematopoiesis.

Given that mutations in *SRSF2* occur as early genetic events in MDS pathogenesis (Papaemmanuil et al., 2013) and that MDS is characterized by expansion of HSPCs, we next examined HSPC numbers and function in *Srsf2*P95H mice. Analysis of CD45.2+ HSPC subsets from *Mx1-cre Srsf2*P95H/WT mice and littermate controls 14 weeks after pIpC injection revealed expansion of lineage-negative Sca1+ c-Kit+ (LSK) and restricted hematopoietic progenitor cells (LSK CD48+ CD150+; hematopoietic progenitor cell fraction 2 (HPC-2) (Oguro et al., 2013)) in mutant mice relative to controls (**Figure 2A-B**). Similar LSK expansion was seen in spleens of *Srsf2*P95H mutant mice (although splenomegaly was not observed up to 20 weeks post-pIpC) (**Figure S2A-B**). Because the detection of increased HSPCs in *Srsf2*P95H mutant mice appeared paradoxical given the decreased peripheral blood counts in these same mice, we next examined cell cycle kinetics and apoptosis of *Srsf2* mutant HSPCs. Indeed, *Srsf2*P95H LSK cells were characterized by an increase in the proportion of cells in S-phase as well as in early apoptosis (**Figure 2C-E**). Despite HSPC expansion in *Srsf2*P95H mutant mice, purified LSK cells from mice with homozygous *Srsf2* deletion or heterozygous *Srsf2*P95H mutation had similarly impaired colony formation and serial re-plating capacity *in vitro* (**Figure S2C**).

To assess the functional effects of *Srsf2* alterations on HSC self-renewal *in vivo*, we next compared *Srsf2* heterozygous KO, homozygous KO, and heterozygous P95H mutant mice in competitive transplantation assays (**Figure 3A**). Equal numbers of BM MNCs from CD45.1 WT mice and CD45.2 *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl, or *Mx1-cre Srsf2*P95H/WT mice were transplanted into lethally irradiated CD45.1 mice followed by pIpC injection 4 weeks later. Assessment of peripheral blood chimerism monthly thereafter revealed a complete loss of CD45.2 chimerism in mice transplanted with *Mx1-cre Srsf2*fl/fl cells and significant decrease in chimerism in mice transplanted with *Mx1-cre Srsf2*P95H/WT cells (**Figure 3B**, **Figure S3A-B**). However, analysis of BM LSK chimerism 18 weeks post-transplant revealed an increase in CD45.2+ HSPCs derived from *Srsf2*P95H mice relative to other groups and near complete absence of CD45.2+ HSPCs from *Srsf2*fl/fl mice (**Figure 3C-D**, **Figure S3C**).

Serial competitive transplantation of whole bone marrow from *Srsf2*P95H, *Srsf2* heterozyous KO, and *Srsf2* WT primary recipient transplanted mice continued to reveal impaired reconstitution capacity of *Srsf2*P95H mutant mice relative to *Srsf2* heterozygous KO or control mice (**Figure S3D**). Of note, colony assays and competitive transplantation experiments were performed using multiple genotypes of control mice (*Cre*-negative *Srsf2*P95H mice as well as *Mx1*-cre *Srsf2* WT mice; **Figure S2C** and **S3E**) to control for any possible confounding effect of Cre expression or the presence of the unexcised P95H knockin allele.

The fact that *Mx1-cre Srsf2*P95H/WT mice had an increase in HSPCs despite impaired formation of mature peripheral blood cells suggested that mutant *Srsf2* was associated with impaired HSPC differentiation. Flow cytometric analysis of mature and intermediate precursor cell subsets in *Srsf2*P95H mice was therefore performed to identify the stage of impaired hematopoiesis. This revealed that peripheral leukopenia was predominantly due to decreased peripheral blood B-cells, evident at all stages of B lymphopoiesis following the transition of pre-proB to pro-B cells, in *Srsf2*P95H mice relative to controls (**Figure S3F**-**G**). Moreover, immunophenotypic analysis of intermediate hematopoietic progenitors (Pronk et al., 2007) revealed deficits in early erythroid progenitors in *Srsf2*P95H mice relative to controls, initiating at the pre-MegE and pre-CFUE stages (**Figure S3H-I**). Given prior data that homozygous deletion of *Srsf2* resulted in defective T cell maturation and CD45 splicing (Wang et al., 2001), we also examined thymic T cell differentiation and CD45 isoform expression in *Srsf2*P95H mice relative to controls (**Figure S3J-K**). This revealed no effect of *Srsf2*P95H mutation on thymic T cell maturation or protein expression of the specific CD45 isoforms previously identified to be down-regulated with homozygous deletion of *Srsf2* (Wang et al., 2001).

Collectively, biological analysis of *Srsf2*P95H mutant mice identified phenotypes distinct from mice with partial or complete loss of *Srsf2*, suggesting that *SRSF2* mutations alter SRSF2's normal function rather than resulting in haploinsufficiency or a dominant-negative function. Of note, despite the impaired hematopoietic differentiation, increase in HSPC subsets,

and morphologic dysplasia in *Srsf2*P95H/WT mice, no *Srsf2*P95H mutant mice developed acute myeloid leukemia in up to 70 weeks of observation.

SRSF2 Mutations Are Associated with Global Alterations of Gene Expression and Splicing

We next sought to identify the transcriptional and post-transcriptional alterations caused by *SRSF2* mutations through RNA-seq of purified LSK and myeloid progenitor (MP; lineage-negative Sca1- c-Kit+) populations. This was performed 4 weeks after pIpC administration. In unsupervised cluster analysis based on coding gene expression, samples clustered first by cell type, then genotype (**Figure S4A**). Expression of several hematopoietic regulators was altered in *Srsf2*P95H mutant cells, including up-regulation of *Gfi1*, *Cebpe*, and *Hoxb2* in LSK cells, down-regulation of *Gata1* and *Gata2* in MP cells, and down-regulation of *Cdkn1a* in both populations. In addition, we observed preferential down- versus up-regulation of the expression of coding genes in *Srsf2* mutant cells relative to WT (**Figure S4B-C**). Gene Ontology (GO) analysis revealed an enrichment for down-regulation of genes in both LSK and MP cells involved in the regulation of cell cycle, proliferation, differentiation, and apoptosis (up-regulated genes were not enriched for these processes; **Figure S4D**).

To identify changes in splicing driven by *SRSF2* mutations that might contribute to disease, we augmented our mouse data with RNA-seq data from primary CMML (n=13; 3 with *SRSF2* mutation) and acute myeloid leukemia (AML; n=9; 5 with *SRSF2* mutation) patient samples (**Table S1**), as well as K562 cells ectopically expressing an empty vector or a single allele of *SRSF2* (WT, P95H, P95L, P95R). In all sequenced patients with *SRSF2* mutations, the WT and mutant alleles were expressed at similar levels (**Table S1**), as was the case for the *Srsf2*P95H mouse cells (**Figure 1**). Similarly, isogenic K562 cells with lentiviral expression of WT or mutant SRSF2 cells expressed WT and mutant SRSF2 at roughly equal levels (**Figure S4E-G**). We quantified global changes in splicing of ~125,000 alternative splicing events and

~160,000 constitutive splice junctions associated with *SRSF2* mutations in these five datasets (LSK, MP, CMML, AML, K562). We required a minimum change in isoform ratio of 10% to call an event differentially spliced (where a change in isoform ratio is defined as an absolute, rather than relative, quantity as the increase or decrease in the percentage of all mRNAs transcribed from the parent gene that follow a given splicing pattern). In all datasets, *SRSF2* mutations were associated with differential splicing of all classes of splicing events, as well as novel alternative splicing and intron retention of splice junctions annotated as constitutively spliced. However, only a relatively small fraction of alternatively spliced events of any class were affected by *SRSF2* mutations (**Figure S4H**). *SRSF2* mutations were associated with a mild bias towards exon skipping, but did not lead to globally increased levels of predicted substrates for degradation by nonsense-mediated decay.

SRSF2 Mutations Alter Exonic Splicing Enhancer Preference While SRSF2 Loss Does Not

Since SRSF2 normally recognizes ESE elements within the pre-mRNA to promote exon recognition (Graveley and Maniatis, 1998; Liu et al., 2000; Schaal and Maniatis, 1999; Zahler et al., 2004), we hypothesized that *SRSF2* mutations might alter its normal sequence-specific activity. To test this, we performed an *ab initio* motif identification screen. We quantified the occurrence of each possible k-mer (k = 4, 5, 6) within cassette exons that were differentially spliced in *Srsf2*P95H MP cells, and identified k-mers that were enriched or depleted in cassette exons promoted versus repressed in *Srsf2*P95H cells. We identified enriched and depleted motifs using a non-parametric (Kolmogorov-Smirnov) statistical test with a *p*-value threshold of 0.05. Significantly enriched k-mers were C-rich, while depleted k-mers were G-rich (**Figure S4I-J**). We then performed an identical analysis using our K562 data, which likewise identified CCAG and GGTG as the most enriched and depleted consensus motifs, respectively (**Figure 4A-B**). A recent solution structure of SRSF2 in complex with RNA revealed that SRSF2 has a consensus motif of SSNG (where “S” represents C or G) and efficiently recognizes both CCNG

and GGNG (Daubner et al., 2012). Therefore, our *ab initio* analysis suggested that mutations affecting the P95 residue may alter SRSF2's ability to recognize variants of its normal SSNG motif.

To further explore this hypothesis, we compared the relative enrichment of all four SSNG variants in cassette exons that were differentially spliced upon depletion of *SRSF2*, overexpression of WT *SRSF2*, or expression of mutant *SRSF2*. *SRSF2* depletion—achieved by knock down of endogenous *SRSF2* in the absence of mutant protein expression (**Figure S4K**)—caused preferential skipping of cassette exons, consistent with SRSF2's canonical role in promoting exon recognition (**Figure 4C**). *Ab initio* motif analyses identified both C- and G-rich variants of the SSNG motif as the most enriched motifs in cassette exons that were repressed following SRSF2 depletion (**Figure 4D**). Quantitation of the enrichment of each SSNG variant revealed that all were associated with exon repression following knock down; in contrast, overexpression of WT SRSF2 was associated with enrichment of each SSNG variant (**Figure 4E**). This data suggests that different SSNG variants function as equally efficacious SRSF2-dependent ESEs, consistent with SRSF2's *in vitro* binding specificity (Daubner et al., 2012). In contrast, K562 cells as well as LSK and MP cells expressing mutant *Srsf2* exhibited enrichment for CCNG and depletion for GGNG in exons that were promoted versus repressed (**Figure 4E**).

To test whether this motif enrichment and depletion was due to ESE activity, we computed the spatial distribution of CCNG and GGNG motifs across genomic loci containing cassette exons that were promoted or repressed in association with *SRSF2* mutations. CCNG and GGNG were respectively enriched and depleted specifically over cassette exons, and not over the flanking introns or exons. We observed similar motif preferences and distributions in patient transcriptomes (**Figure 4F**). As CCNG/GGNG motifs were not consistently enriched/depleted in introns flanking differentially spliced cassette exons, and we were unable to identify enriched motifs with *ab initio* searches in introns, we conclude that differential cassette exon splicing is likely due primarily to altered recognition of exonic motifs. Together,

these data reveal spatially restricted enrichment of specific ESEs in association with *SRSF2* mutations, and suggest that *SRSF2* mutations cause alteration rather than loss of normal ESE recognition activity.

SRSF2 Proline 95 Mutations Alter RNA Binding Specificity By Changing the Conformation of Both RRM Termini

We next tested whether this association between *SRSF2* mutations and enrichment/depletion of specific ESEs was due to altered SRSF2:RNA interactions. We purified SRSF2's RNA recognition motif domain (RRM) as previously described and performed isothermal titration calorimetry (ITC) with the RNA ligand 5'-uCCAGu-3', an optimal SRSF2 target according to the SSNG consensus sequence (Daubner et al., 2012). All three P95 mutations resulted in an increase in binding affinity of 3.9-4.5-fold relative to WT SRSF2 (**Figure 5A-B** and **Figure S5A**), consistent with the enrichment for CCNG motifs that we observed in exons promoted by *SRSF2* mutations (**Figure 4B**). We next tested whether P95 mutations resulted in altered RNA binding specificity. In contrast to 5'-uCCAGu-3' RNA, ITC measurements revealed that all three P95 mutants exhibited a 1.2-2.1-fold decrease in binding affinity to the 5'-uGGAGu-3' RNA relative to WT SRSF2 (**Figure 5A-B** and **Figure S5B**). ITC measurements using the RNA sequences 5'-uGCAGu-3' and 5'-uCGAGu-3' revealed that G>C substitutions at the second motif position resulted in larger increases in binding affinity than at the first motif position (2.6-3.4-fold versus 1.1-1.8-fold; **Figure 5B and Figure S5C-D**). The RNA binding preferences measured by ITC were remarkably consistent with ESE enrichment identified by RNA-seq. For each mutant, the level of motif enrichment (**Figure 4E**) was roughly proportional to the affinity increase (**Figure 5C**), and the enrichment and affinity measurement supported the same relative preference for each specific motif (CC > GC > CG > GG). This strongly supports the notion that splicing changes caused by P95 mutations are the result of an altered sequence-specificity of the SRSF2 RRM.

P95 is located at the C-terminal end of the SRSF2 RRM and the published solution structure of SRSF2 in complex with 5'-uCCAGu-3' revealed extensive contacts of P95 with the second cytosine (**Figure S5E**), emphasized by several intermolecular Nuclear Overhauser Effects (NOEs) (Daubner et al., 2012). To test whether SRSF2's RNA binding surface was altered by P95 mutations, we conducted nuclear magnetic resonance (NMR) titration with the SRSF2 P95H RRM and the 5'-uCCAGu-3' RNA, and assigned the backbone of this complex using standard heteronuclear NMR experiments. Mapping of the chemical shift perturbations revealed that the RNA-binding surface of the RRM is not disturbed by the P95H mutation. However, both termini experienced large changes in their environment (**Figure 5D**), an observation that held true for all three P95 mutations (**Figure S5F**). Consistent with our ESE and ITC analyses, this relocation of termini primarily affected the second cytosine, which exhibited the largest chemical shift perturbations of its proton resonances (**Figure S5G-H**). Smaller changes of chemical shifts were observed when P95 mutants were bound to 5'-uGGAGu-3' (**Figure S5H**). Together, our experiments indicate that *SRSF2* mutations change SRSF2's normal RNA-binding affinity and specificity *in vitro*, likely explaining the widespread alterations in ESE preference that we observed *in vivo*.

Mutant *SRSF2* Promotes Mis-splicing and Degradation of *EZH2*

We next used our transcriptome data to identify common changes in splicing driven by *SRSF2* mutations that might contribute to disease. Intersection of differentially spliced genes in LSK, MP, CMML, and AML samples identified 75 genes differentially spliced in association with *SRSF2* mutations in both LSK and MP cells and at least one primary patient cohort, as well as an additional 97 (LSK) and 87 (MP) genes differentially spliced in one mouse cell population, but not the other, as well as a patient cohort (**Figure 6A**, **Table S2-5**). Many of these genes have known importance in myeloid malignancies. For example, *SRSF2* mutations promoted inclusion of a highly conserved "poison" cassette exon of *EZH2 (Enhancer of zeste homolog 2)*,

and repressed a frame-preserving cassette exon of *BCOR* (*BCL6 corepressor*) (**Figure 6A** and **Figure S6A-B**). Of note, we did not identify altered splicing of CD45 in *SRSF2* mutant cells (**Supplemental Tables 2-5**) which was previously noted as altered in murine *Srsf2* KO hematopoietic cells (Wang et al., 2001).

In order to identify potential functional consequences of recurrent mis-splicing, we focused on the splicing event in *EZH2*. *SRSF2* mutant cells exhibited preferential inclusion of a "poison" cassette exon, which introduces a premature termination codon predicted to result in nonsense-mediated decay (NMD) of *EZH2* (**Figure 6B-C**). Both the poison exon itself and its flanking intronic sequences exhibited high sequence conservation across vertebrates- exceeding the sequence conservation exhibited by the upstream and downstream constitutive coding exons themselves- which is a common feature of physiologically important splicing events (Lareau et al., 2007; Ni et al., 2007) (**Figure 6B**).

We validated this *EZH2* splicing change using both qualitative and quantitative isoform-specific RT-PCR in leukemia cell lines that were WT or mutant for *SRSF2* (**Figure S6C-D**), as well as in an independent panel of primary AML patient samples with or without *SRSF2* mutations (n = 8; 4 with *SRSF2* mutations; **Figure 6D; Figure S6E**).

Next, to confirm whether the cassette exon promoted by *SRSF2* mutations triggers degradation by NMD, we measured the half-life of the inclusion isoform of *EZH2* in *SRSF2*P95H cells transfected with a control or anti-*UPF1* (a required NMD factor) shRNA following transcriptional shutoff with actinomycin D (t Hoen et al., 2011) (**Figure 6E** and **Figure S6F-G**). The fact that the mRNA half-life of the inclusion isoform of *EZH2* was lengthened by *UPF1* knockdown in these experiments suggests that this particular isoform of *EZH2*, which is promoted by mutant SRSF2, undergoes NMD. The half-life of a well-characterized NMD substrate of *SRSF3* (Lareau et al., 2007; Ni et al., 2007) similarly increased following *UPF1* knockdown, confirming that *UPF1* knockdown effectively inhibited NMD (**Figure S6H**).

Next, to identify whether the protein product of *EZH2* is altered in *SRSF2* mutant cells, we performed WB analysis of a panel of human AML cell lines WT (TF-1, K562) or mutant for *SRSF2* (K052) (all WT for *EZH2*). This revealed lower EZH2 protein levels as well as lower global levels of histone H3 lysine 27 trimethylation (H3K27me3; a methylation mark placed by EZH2) in *SRSF2* mutant K052 cells (**Figure 6F**). To further validate this finding in an isogenic context, we performed WB analysis in K562 cells ectopically expressing WT *SRSF2*, or *SRSF2*P95H/L/R mutant cDNA. This analysis revealed consistent down-regulation of EZH2 protein expression as well as global H3K27me3 in all three *SRSF2*-mutant samples compared with *SRSF2* WT K562 cells (**Figure 6G**).

Consistent with *SRSF2* mutations promoting a disabling splicing change in *EZH2*, *EZH2* loss-of-function mutations are common in MDS. In an analysis of >1,800 MDS patients where *EZH2* and *SRSF2* were both sequenced, *EZH2* loss-of-function mutations were mutually exclusive with *SRSF2* mutations ($p<0.0001$; (Bejar et al., 2012; Ernst et al., 2010; Haferlach et al., 2014; Muto et al., 2013; Papaemmanuil et al., 2013) (**Figure 6H**).

The above data strongly link *SRSF2* mutations to disabling splicing of *EZH2*. We next sought to examine whether the change in RNA ESE preference induced by *SRSF2* mutations caused *EZH2* mis-splicing. We therefore cloned the genomic locus containing the *EZH2* poison exon and flanking introns and constitutive exons to create a minigene that recapitulates this splicing event. We identified three potential SRSF2-dependent SSNG motifs in the poison exon (CCTG, CCTG, GCAG), one or more of which we expected to be better recognized by mutant SRSF2 than WT SRSF2. We then mutated each motif to the corresponding GG equivalent, both separately and in combination (**Figure 6I**). Measuring cassette exon recognition in K562 cells expressing WT or mutant *SRSF2*, we found that the first motif was required for robust splicing change in *SRSF2*-mutant cells, such that the mutation CCTG>GGTG prevented an increase in poison exon recognition (**Figure 6J**). We conclude that *SRSF2* mutations induce a disabling

splicing change in *EZH2* in an ESE-dependent manner consistent with altered RNA recognition activity.

We next sought to test whether restoring normally spliced *EZH2* mRNA could rescue hematopoiesis in *SRSF2*-mutant cells. *EZH2* full-length cDNA or an empty vector (both in a retroviral ZsGreen1 vector) were overexpressed in c-Kit+ *Srsf2*P95H or WT cells followed by assessment of methylcellulose colony formation of c-Kit+/ZsGreen1+ cells. *EZH2* cDNA was equally overexpressed in *Srsf2* mutant and WT cells (**Figure S6I**) and *Srsf2*P95H mutant cells overexpressing full-length *EZH2* experienced a ~50% increase in colony formation relative to *Srsf2*P95H mutant cells expressing an empty vector (**Figure 6K** and **Figure S6J**). In contrast, *EZH2* overexpression had no substantial effect on initial colony formation in *Srsf2* WT cells (**Figure 6K** and **Figure S6I**). These data identify that restoration of normally spliced *EZH2* mRNA in *SRSF2* mutant cells at least partially rescues the hematopoietic defects induced by mutant *SRSF2.*

DISCUSSION

The consistent occurrence of heterozygous point mutations affecting highly restricted residues of spliceosomal proteins strongly suggests a gain-of-function or dominant-negative activity for these mutations in malignant transformation. Here we identify an effect of the *SRSF2*P95H mutation distinct from loss of *SRSF2*, and reveal that mutations in *SRSF2* confer an alteration in function that results in key aspects of MDS. This includes an increase in HSPCs in *Srsf2*P95H mutant mice with impaired differentiation, altered cell cycle kinetics, and increased apoptosis resulting in peripheral cytopenias and morphologic dysplasia. By contrast, WT *Srsf2* appears to be constitutively required for hematopoiesis.

Transcriptional analysis of *SRSF2* mutant cells revealed that *SRSF2* mutations result in genome-wide alterations in ESE preference in both human and murine cells. Biochemical analysis of the interaction of SRSF2 with RNA in cell-free *in vitro* assays identified an analogous change in specificity of interactions between SRSF2 and pre-mRNA induced by *SRSF2* mutations. This altered interaction of mutant SRSF2 with RNA appears to be due to an effect of *SRSF2*P95H/L/R mutations on the conformations of the termini of SRSF2's RRM domain as revealed by NMR spectroscopy. Our genomic and biochemical assays indicate that *SRSF2* mutations cause alteration, rather than loss-of-function, driving preferential recognition of cassette exons containing C- versus G-rich ESEs.

The altered pre-mRNA recognition activity of mutant SRSF2 likely underlies the mis-splicing of key transcriptional regulators—several of which have previously been implicated in MDS pathogenesis. This includes promotion of a "poison" exon of *EZH2* that undergoes NMD and results in reduced EZH2 protein expression in *SRSF2* mutant cells. Loss-of-function mutations in *EZH2* occur in the same exact spectrum of myeloid malignancies as *SRSF2* mutations (Ernst et al., 2010; Nikoloski et al., 2010) and loss of *Ezh2* has been functionally linked to MDS development *in vivo* (Muto et al., 2013). Moreover, *SRSF2* and *EZH2* mutations are mutually exclusive in MDS patients (Haferlach et al., 2014; Papaemmanuil et al., 2013), but

the basis for this observation was previously unknown. The data here provide a mechanistic basis for this mutual exclusivity as *SRSF2* mutations functionally reduce EZH2 protein expression.

In addition to the effects of mutant *SRSF2* on *EZH2* splicing and protein expression, a number of other genes of known importance in hematopoiesis and malignancy were also consistently differentially spliced in isogenic human cells, primary patient samples, and murine cells bearing mutant *SRSF2*. These include additional genes mutated in MDS (such as *BCOR*), genes with importance in hematopoietic stem cell self-renewal (such as *IKAROS*), and genes critical for cell survival (such as *CASPASE 8*). Future efforts to understand the functional effects of each of these specific splicing events will be important in further delineating the effects of mutant *SRSF2* on MDS pathogenesis as well as possibly providing novel means for therapeutic targeting of *SRSF2*-mutant cells.

Our studies, which reveal both mechanistic splicing alterations and specific mis-spliced isoforms in *SRSF2*-mutant cells, may inform therapeutic opportunities for targeting *SRSF2*-mutant cells. For example, the observations that mutant *SRSF2* promotes inclusion of a poison exon in an ESE-dependent manner, and that restoration of normally spliced *EZH2* mRNA partially rescues defective hematopoiesis in *SRSF2*-mutant cells, suggest that normal cellular function may be at least partially restored by manipulating specific pathologic splicing events.

EXPERIMENTAL PROCEDURES

Generation of the *Srsf2*P95H conditional knock-in mice is described in **Supplemental Experimental Procedures.** All animal procedures were conducted 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 Memorial Sloan Kettering Cancer Center.

Patient samples. 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.

mRNA sequencing. For sorted mouse cell populations, K562 cells, and primary AML and CMML samples, RNA was extracted using Qiagen RNeasy columns. poly(A)-selected, unstranded Illumina libraries were prepared with a modified TruSeq protocol. 0.5X AMPure XP beads were added to the sample library to select for fragments <400 bp, followed by 1X beads to select for fragments >100 bp. These fragments were then amplified with PCR (15 cycles) and separated by gel electrophoresis (2% agarose). 300 bp DNA fragments were isolated and sequenced on the Illumina HiSeq 2000 (~100M 2x49 bp reads per sample).

RNA-seq read mapping. Reads were mapped to the UCSC hg19 (NCBI GRCh37) human genome or UCSC mm10 (NCBI GRCm38) genome assemblies. First, a modified version of RSEM that called Bowtie v1.0.0 with the -v 2 argument was created. This modified RSEM was then called with the arguments --bowtie-m 100 --bowtie-chunkmbs 500 --calc-ci --output-genome-bam on the gene annotation file. Read alignments with mapq scores of 0 and or a splice junction overhang of less than 6 bp were then filtered out. Remaining unaligned reads were then aligned TopHat v2.0.8b with the arguments --bowtie1 --read-mismatches 2 --read-

edit-dist 2 --no-mixed --no-discordant --min-anchor-length 6 --splice-mismatches 0 --min- intron-length 10 --max-intron-length 1000000 --min-isoform-fraction 0.0 --no-novel-juncs --no-novel-indels --raw-juncs on the splice junction file (--mate-inner-dist and --mate-std-dev were calculated by mapping to constitutive coding exons with MISO's exon_utils.py utility). The resulting TopHat alignments were then filtered as for the RSEM-generated alignments. Finally, the RSEM and TopHat aligned were merged to create final BAM files.

Isoform expression measurements. Two different methods were used to quantify isoform ratios. For alternative splicing events from MISO's v2.0 annotation, MISO was used to estimate isoform ratios. For alternative splicing or intron retention of annotated constitutive junctions, junction reads alone were used as previously described (Hubert et al., 2013). To identify differentially expressed events, we required a minimum of 20 identifying reads (supporting either, but not both, isoforms) per event, as well as a change in isoform ratio >= 10%. For the LSK, MP, and K562 data, we used two-sample statistical comparisons (Wagenmakers's framework; Bayes factor >= 5); for the AML and CMML data, we used group statistical comparisons (Mann-Whitney U test; p-value <= 0.05). Real-time PCR used to measure *EZH2* cassette exon inclusion described in Supplementary Methods.

RNA-seq data deposition. RNA-seq data from this study are deposited in the Gene Expression Omnibus (GEO) under accession number: GSE65349.

AUTHOR CONTRIBUTIONS

E.K., J.I., Y.L., G.M.D., F.H.-T.A., S.H., R.K.B, and O.A.-W. designed the study. E.K., S.L., Y.R.C., J.B.M., H.C., M.-K.K., and O.A.-W. performed animal experiments and generated mice. J.I., A.R., M.M. and A.S.Z. generated SRSF2 constructs, K562 cell lines, and CMML RNA-seq data. S.B. and P.S. provided additional SRSF2 constructs. R.K.B. performed RNA-seq analysis. A.R., J.D., and O.A.-W. provided primary patient leukemia samples. C.L. and I.A. provided advice on animal experiments and helped generate RNA-seq data. S.A. and C.Y.P. performed cytopathologic and histopathologic analyses. Y.L., G.M.D., Y.L., Y.M., F.H.-T.A., and S.H. prepared protein and RNA samples for NMR and ITC studies and performed analysis; J.I. created minigenes and conducted splicing assays. E.K., G.M.D., F.H.-T.A., S.H., R.K.B., and O.A.-W. prepared the manuscript with help from all co-authors.

ACKNOWLEDGMENTS

E.K. is supported by the Worldwide Cancer Research Fund. AR was supported by the NIH/NHLBI (U01 HL099993), NIH/NIDDK (K08 DK082783), the J.P. McCarthy Foundation, and the Storb Foundation. S.H. and O.A.-W. are supported by grants from the Edward P. Evans Foundation. SH was supported by Yale Comprehensive Cancer Center institutional funds. RKB was supported by the Hartwell Innovation Fund, Damon Runyon Cancer Research Foundation (DFS 04-12), Ellison Medical Foundation (AG-NS-1030-13), NIH/NIDDK (R56 DK103854), NIH/NCI recruitment support (P30 CA015704), and Fred Hutchinson Cancer Research Center institutional funds. J.O.I. was supported by a NIH/NCI training grant (T32 CA009657) and NIH/NIDDK pilot study (P30 DK056465). C.L. is supported by a Career Development Award Grant from the Leukemia and Lymphoma Society and an ATIP-Avenir grant from the French Government. O.A.-W. is supported by an NIH K08 Clinical Investigator Award (1K08CA160647-

01), a US Department of Defense Postdoctoral Fellow Award in Bone Marrow Failure Research (W81XWH-12-1-0041), the Josie Robertson Investigator Program, and a Damon Runyon Clinical Investigator Award with support from the Evans Foundation. F.H.-T.A. acknowledges support from the NCCR RNA & Disease, funded by the Swiss National Science Foundation and the SNF Sinergia CRSII3_127454. Y.L. and Y.M. were supported by NIH/NIGMS grant R01 GM102869, and by a Senior Research Fellowship grant number 101908/Z/13/Z to Y.M. from the Wellcome Trust. J.D. acknowledges assistance from Dr. Nezih Cereb, HistoGenetics (Ossining, NY).

REFERENCES

Bejar, R., Stevenson, K., Caughey, B., Abdel-Wahab, O., Steensma, D., Galili, N., Raza, A., Kantarjian, H., Levine, R., Neuberg, D.*, et al.* (2012). Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *30*, 3376-3382.

Brooks, A.N., Choi, P.S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., Pedamallu, C.S., Sivachenko, A., Rosenberg, M., Chmielecki, J.*, et al.* (2014). A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. PloS one *9*, e87361.

Daubner, G.M., Clery, A., Jayne, S., Stevenin, J., and Allain, F.H. (2012). A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. EMBO J *31*, 162-174.

Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A.*, et al.* (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet *42*, 722-726.

Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E.*, et al.* (2012). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet *44*, 53-57.

Graveley, B.R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Molecular cell *1*, 765-771.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T.*, et al.* (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia *28*, 241-247.

Hubert, C.G., Bradley, R.K., Ding, Y., Toledo, C.M., Herman, J., Skutt-Kakaria, K., Girard, E.J., Davison, J., Berndt, J., Corrin, P.*, et al.* (2013). Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. Genes Dev *27*, 1032-1045.

Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2014). U2AF1 mutations alter splice site recognition in hematological malignancies. Genome research.

Kuhn, R., Schwenk, F., Aguet, M., and Rajewsky, K. (1995). Inducible gene targeting in mice. Science *269*, 1427-1429.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature *446*, 926-929.

Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2000). Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. Molecular and cellular biology *20*, 1063-1071.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer cell *20*, 11-24.

Muto, T., Sashida, G., Oshima, M., Wendt, G.R., Mochizuki-Kashio, M., Nagata, Y., Sanada, M., Miyagi, S., Saraya, A., Kamio, A.*, et al.* (2013). Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. The Journal of experimental medicine *210*, 2627-2639.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev *21*, 708-718.

Nikoloski, G., Langemeijer, S.M., Kuiper, R.P., Knops, R., Massop, M., Tonnissen, E.R., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T.*, et al.* (2010). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nat Genet *42*, 665-667.

Oguro, H., Ding, L., and Morrison, S.J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell stem cell *13*, 102-116.

Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C.*, et al.* (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med *365*, 1384-1395.

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A.*, et al.* (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616-3627; quiz 3699.

Peterson, R.D., Theimer, C.A., Wu, H., and Feigon, J. (2004). New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. Journal of biomolecular NMR *28*, 59-67.

Pronk, C.J., Rossi, D.J., Mansson, R., Attema, J.L., Norddahl, G.L., Chan, C.K., Sigvardsson, M., Weissman, I.L., and Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell stem cell *1*, 428-442.

Przychodzen, B., Jerez, A., Guinta, K., Sekeres, M.A., Padgett, R., Maciejewski, J.P., and Makishima, H. (2013). Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. Blood *122*, 999-1006.

Quesada, V., Conde, L., Villamor, N., Ordóñez, G., Jares, P., Bassaganyas, L., Ramsay, A., Beà, S., Pinyol, M., Martínez-Trillos, A.*, et al.* (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature genetics *44*, 47-52.

Schaal, T.D., and Maniatis, T. (1999). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Molecular and cellular biology *19*, 261-273.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S.*, et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome research *15*, 1034-1050.

t Hoen, P.A., Hirsch, M., de Meijer, E.J., de Menezes, R.X., van Ommen, G.J., and den Dunnen, J.T. (2011). mRNA degradation controls differentiation state-dependent differences in transcript and splice variant abundance. Nucleic acids research *39*, 556-566.

Vannucchi, A.M., Lasho, T.L., Guglielmelli, P., Biamonte, F., Pardanani, A., Pereira, A., Finke, C., Score, J., Gangat, N., Mannarelli, C.*, et al.* (2013). Mutations and prognosis in primary myelofibrosis. Leukemia *27*, 1861-1869.

Wagenmakers, E.J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. Cognitive psychology *60*, 158-189.

Wang, H.Y., Xu, X., Ding, J.H., Bermingham, J.R., Jr., and Fu, X.D. (2001). SC35 plays a role in T cell development and alternative splicing of CD45. Molecular cell *7*, 331-342.

Wang, L., Lawrence, M., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D., Zhang, L.*, et al.* (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. The New England journal of medicine *365*, 2497-2506.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M.*, et al.* (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64-69.

Zahler, A.M., Damgaard, C.K., Kjems, J., and Caputi, M. (2004). SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. The Journal of biological chemistry *279*, 10077-10084.

Zhang, S.J., Rampal, R., Manshouri, T., Patel, J., Mensah, N., Kayserian, A., Hricik, T., Heguy, A., Hedvat, C., Gonen, M.*, et al.* (2012). Genetic analysis of patients with leukemic

transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood *119*, 4480-4485.

FIGURE LEGENDS

Figure 1. **Conditional expression of *Srsf2*P95H results in myeloid dysplasia, a phenotype distinct from heterozygous or homozygous loss of *Srsf2*.** (A) Depiction of the *Srsf2*P95H allele. (B) RNA-seq of LSK cells in *Mx1-cre Srsf2*WT and *Mx1-cre Srsf2*P95H/WT mice. (C) White blood cell (WBC) count, (D) hemoglobin (Hb), and (E) mean corpuscular volume (MCV) of red blood cells of CD45.1 recipient mice 18 weeks following noncompetitive transplantation of bone marrow from CD45.2+ *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl, and *Mx1-cre Srsf2*P95H/WT mice (n=10 mice/genotype for all genotypes except *Mx1-cre Srsf2*fl/WT where n=5; pIpC was administered to recipient mice 4 weeks following transplantation). (F) H&E staining of femurs (Bars: 50 μm) and (G) peripheral blood smears from *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/fl or *Mx1-cre Srsf2*P95H/WT mice (Bars: 10 μm). A representative neutrophil (left) and erythroid precursor (right) is shown for *Srsf2* WT and KO mice. *Mx1*-cre *Srsf2*P95H cells were marked by hypolobated and hypogranulated neutrophils (left 2 photos) and nuclear irregularities as well as cytoplasmic vacuolization and blebbing of erythroid precursors (2 rightmost photos). Error bars represent mean ± SD; ***p < 0.001; ****p < 0.0001. See also Figure S1.

Figure 2. **Conditional expression of *Srsf2*P95H results in expansion of hematopoietic stem and progenitor cells with increased cell proliferation and apoptosis.** (A) Enumeration and (B) FACS analysis of BM LSK cells, long-term hematopoietic stem cells (LT-HSC), restricted hematopoietic progenitor cell fractions 1 (HPC-1) and 2 (HPC-2) and multipotent progenitor (MPP) cells (Oguro et al., 2013) in 12-week old *Mx1-cre Srsf2* WT and *Mx1-cre Srsf2*P95H/WT mice (n=5 mice/genotype). (C) Cell cycle analysis of LSK cells from *Mx1-cre Srsf2*WT or *Mx1-cre Srsf2*P95H/WT mice with *in vivo* BrdU administration. Representative

FACS plot analysis showing gating on LSK cells followed by BrdU versus DAPI stain is shown on the left. (D) Relative quantification of the percentage of LSK cells in S, G2M, and G1 phase is shown on the right (n=8 mice per group). (E) Relative quantification of the percentage of Annexin V+/DAPI- LSK cells (n=8 mice/genotype). Error bars represent mean ± SD; *p < 0.05; **p < 0.01; ****p < 0.0001. See also Figure S2.

Figure 3. *Srsf2*P95H mutation impairs hematopoietic stem cell self-renewal in a manner distinct from *Srsf2* loss. (A) Depiction of competitive bone marrow (BM) transplantation assay. (B) Percentage of CD45.2+ chimerism in the peripheral blood of recipient mice (n=10 mice/genotype). (C) Chimerism and (D) flow cytometric enumeration of CD45.2+ LSK (left) and myeloid progenitor (MP; lineage-negative Sca1-c-Kit+) (right) cells in BM of *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl and *Mx1-cre Srsf2*P95H/WT mice 14 weeks after pIpC injection. Error bars represent mean ± SD; **p< 0.001, ***p < 0.0002; ****p< 0.0001. See also Figure S3.

Figure 4. *SRSF2* mutations alter exonic splicing enhancer preference. (A) Scatter plot of cassette exon inclusion in K562 cells expressing empty vector or *SRSF2*P95R. Percentages, percent of alternatively spliced cassette exons with increased or decreased inclusion. Red and blue dots represent individual cassette exons that are promoted or repressed in *SRSF2*P95R versus empty vector cells, respectively. Promoted and repressed cassette exons are defined as those whose inclusion levels are increased or decreased by ≥10% with a Bayes factor ≥5, as estimated by Wagenmakers's framework (Wagenmakers et al., 2010). (B) Enriched (right) and depleted (left) k-mers in cassette exons promoted versus repressed in *SRSF2*P95R versus WT cells. (C) Scatter plot of cassette exon inclusion in TF-1 cells following transfection with a siRNA

against *SRSF2* or a control non-targeting siRNA ("KD", knockdown). Percentages, percent of alternatively spliced cassette exons with increased or decreased inclusion. (D) Enriched (right) and depleted (left) k-mers in cassette exons promoted versus repressed in *SRSF2* KD versus control cells. (E) Mean enrichment of all variants of the SSNG motif in cassette exons promoted versus repressed in TF-1 cells following *SRSF2* knockdown and K562, LSK, and MP cells expressing WT or mutant *SRSF2*. Error bars, 95% confidence intervals estimated by bootstrapping. (F) Relative frequency of CCNG and GGNG motifs in cassette exons promoted versus repressed by *SRSF2* mutations in LSK and MP cells (top), K562 cells (left), and primary AML and CMML samples with or without *SRSF2* mutations (right; sample numbers correspond to patient identifiers in Table S1). Shading, 95% confidence interval by bootstrapping. The cartoon illustrates a portion of a meta-gene containing the differentially spliced cassette exon; from left to right, the features are the upstream exon (gray box) and intron (black line), the cassette exon (black box, vertical dashed lines), and the downstream intron (black line) and exon (gray box). Horizontal axis, genomic coordinates defined with respect to the 5' and 3' splice sites, where 0 is the splice site itself. Vertical axis, relative frequency of the indicated motifs over genomic loci containing cassette exons promoted versus repressed by *SRSF2* mutations (log scale). See also Figure S4.

Figure 5. **Proline 95 mutations change RNA-binding specificity of the SRSF2 RNA recognition motif domain (RRM) *in vitro* and lead to relocation of the N- and C-termini.** (A) ITC raw data and binding curve for SRSF2 RRM P95H mutant with 5'-uCCAGu-3' and 5'-uGGAGu-3' RNA. (B) Change in RNA-binding affinity (%) for SRSF2 RRM P95H (blue), P95L (green) and P95R (black) mutants compared to WT (red) (Daubner et al., 2012), using RNA targets 5'-uCCAGu-3', 5'-uGCAGu-3', 5'-uCGAGu-3' and 5'-uGGAGu-3'. (C) Change in RNA-binding specificity of SRSF2 RRM WT, P95H, P95L and P95R with 5'-UCCAGU-3' (blue), 5'-

UGCAGU-3' (dark grey), 5'-UCGAGU-3' (light grey) and 5'-UGGAGU-3' RNA (orange). (D) (left) Overlay of 2D [^{15}N-^{1}H] HSQCs of wild type (red) and P95H mutant (blue) bound to 5'-UCCAGU-3' RNA, with negative peaks in green (WT) and light-green (mutant). (right) Difference of the chemical shift perturbations of P95H mutant and wild type. Positive values (blue) with a higher perturbation with the P95H mutant and negative values (red) with a higher perturbation with the WT are shown. Missing assignments are marked with grey bars and proline with a grey P. Residues with the highest difference are depicted in both graph and spectra. See also Figure S5.

Figure 6. *SRSF2*-mutant primary murine and patient samples exhibit convergent splicing alterations. (A) Intersection of genes exhibiting differential splicing in *SRSF2* mutant versus WT mouse LSK and MP cells and primary AML and CMML samples (restricted to orthologous genes). (B) IGV/sashimi plot illustrating the *EZH2* cassette exon promoted by *SRSF2* mutations in multiple datasets analyzed here (top; patient numbers listed in the Sashimi plot correspond to numbers in Table S1 detailing patient characteristics). The DNA sequence conservation of the locus, as estimated by phastCons (Siepel et al., 2005), across 30 vertebrate species is shown in the track below the Sashimi plot. (C) Bar plot describing the percentage of *EZH2* transcripts harboring a specific cassette exon in *SRSF2* mutant relative to WT primary AML samples from RNA-seq data. Error bars, 95% confidence intervals. (D) RT-PCR of *EZH2* exon inclusion event in an independent set of *SRSF2* WT and mutant AML samples. (E) qRT-PCR of *EZH2* inclusion isoform in *SRSF2*P95H mutant cell line K052 cells with or without UPF1 knockdown and actinomycin D treatment. (F) Western blot (WB) analysis for EZH2 and histone H3 lysine 27 trimethylation (H3K27me3) in *SRSF2/EZH2* WT (TF-1, K562) and *SRSF2*P95H mutant*/EZH2* WT (K052) AML cell lines. (G) WB analysis for EZH2, H3K27me3, and FLAG epitope in K562 cells with lentiviral overexpression of N-terminal FLAG tagged SRSF2 WT, SRSF2P95H,

SRSF2P95L, or SRSF2P95R (left). Relative quantification of EZH2 protein expression by WB to total histone H3 expression in K562 cells expressing SRSF2 mutants relative to WT is shown on right. (H) *EZH2* and *SRSF2* mutations are mutually exclusive in sequencing of >1,000 MDS patients (Bejar et al., 2012; Ernst et al., 2010; Haferlach et al., 2014; Muto et al., 2013; Papaemmanuil et al., 2013). (I) Cartoon of *EZH2* cassette exon, with SSNG motifs highlighted and mutations to GG equivalents shown. (J) *EZH2* cassette exon inclusion for minigenes containing the endogenous cassette exon or a cassette exon with mutation of motifs 1, 2, and/or 3 to the GG equivalent. (K) Photographs (left) and enumeration (right) of c-Kit+/ZsGreen1+ cells from *Srsf2* WT or *Srsf2*P95H mice 14 days after overexpression of empty vector or *EZH2* cDNA and plating in methylcellulose media. See also Figure S6 and Table S1-S5.

Figure
Figure 1

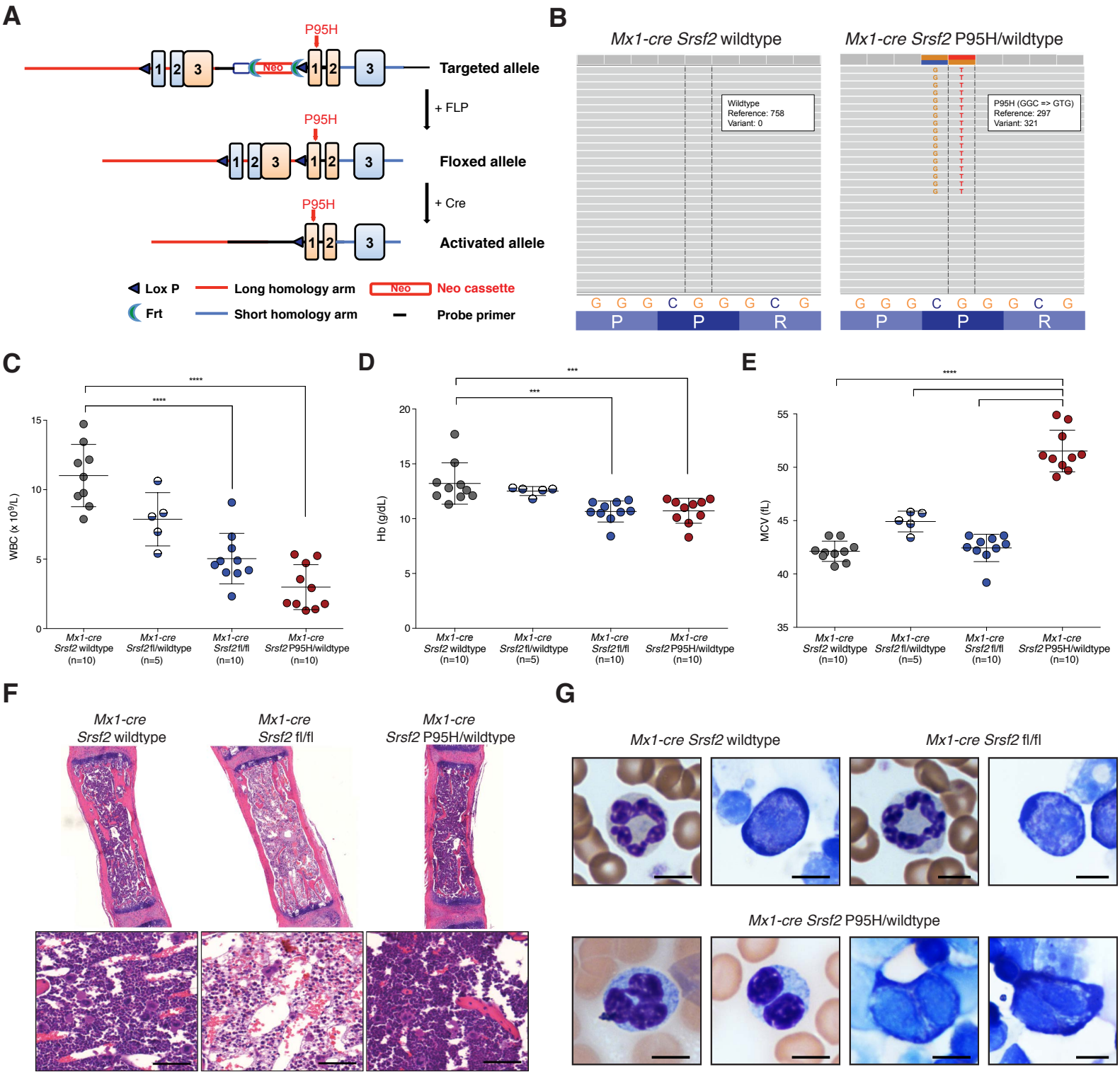


Figure
Figure 2

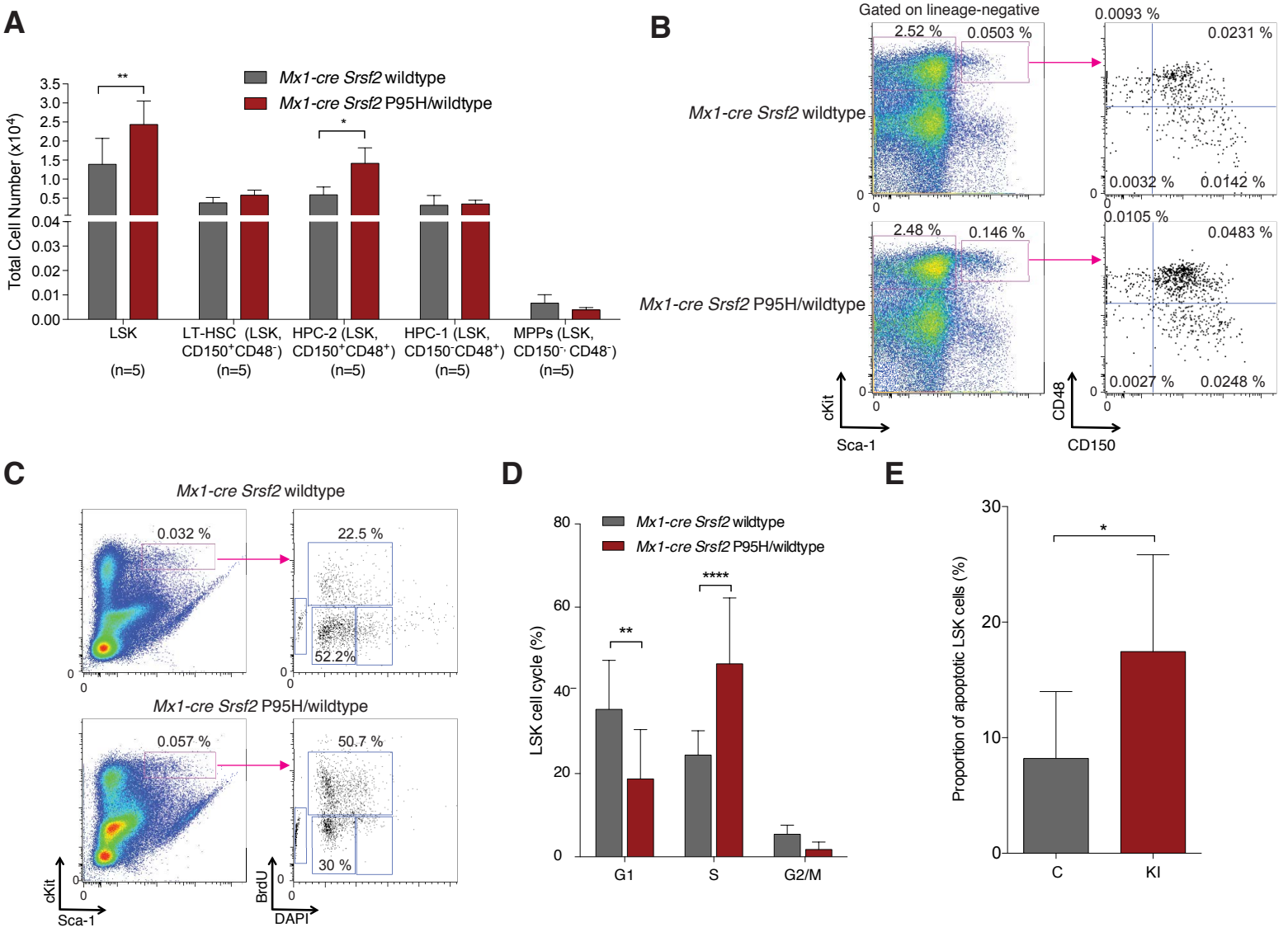
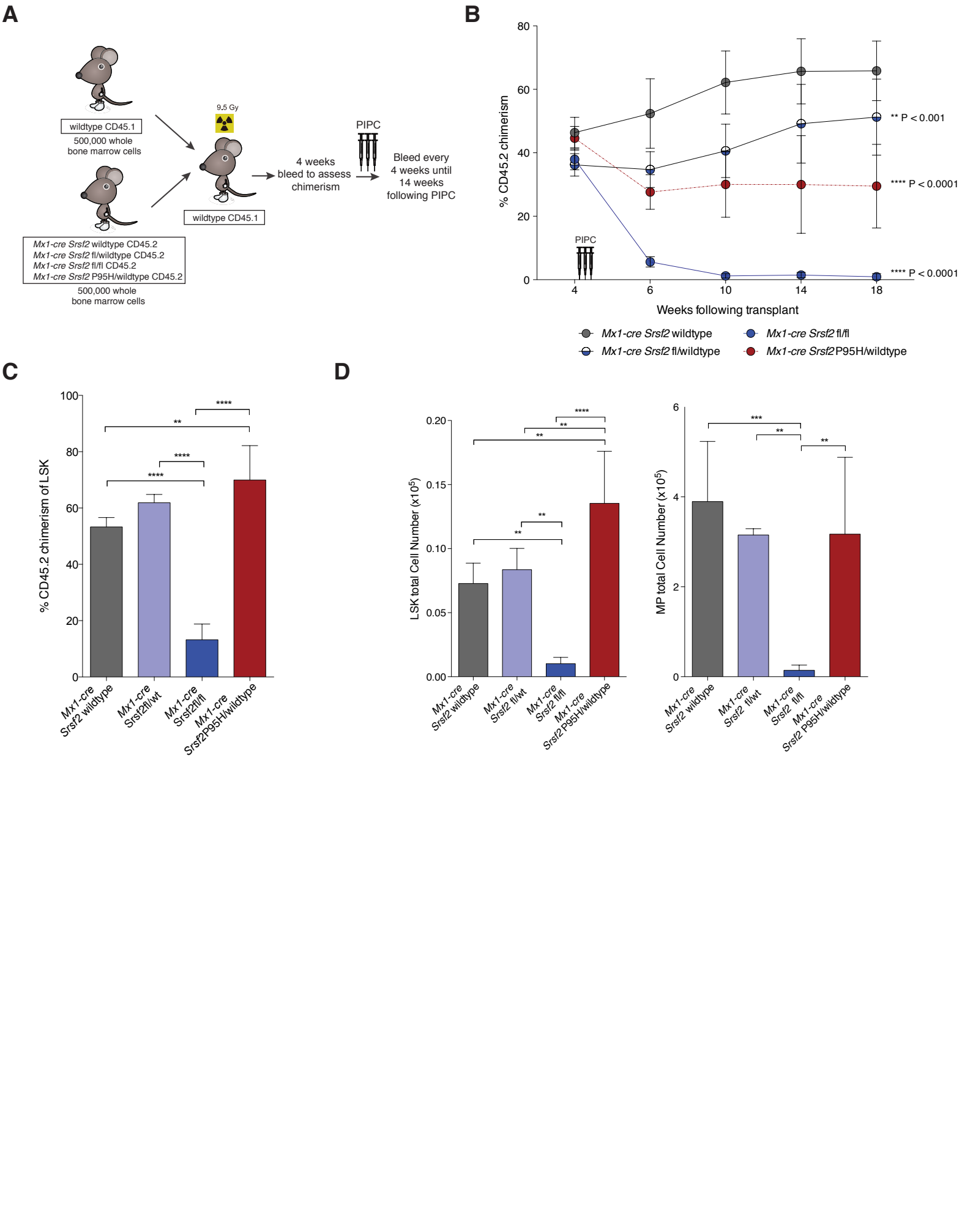
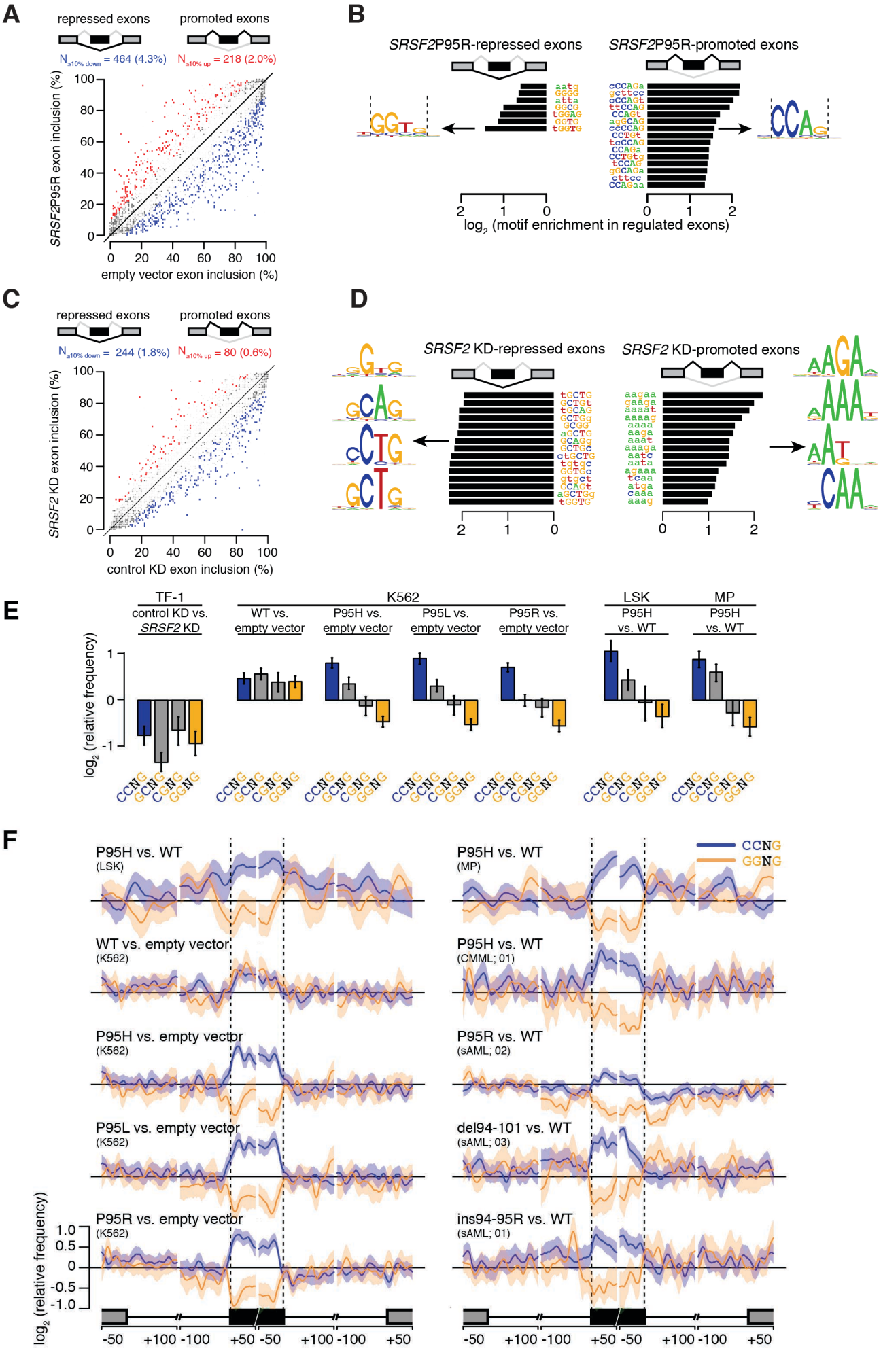


Figure
Figure 3



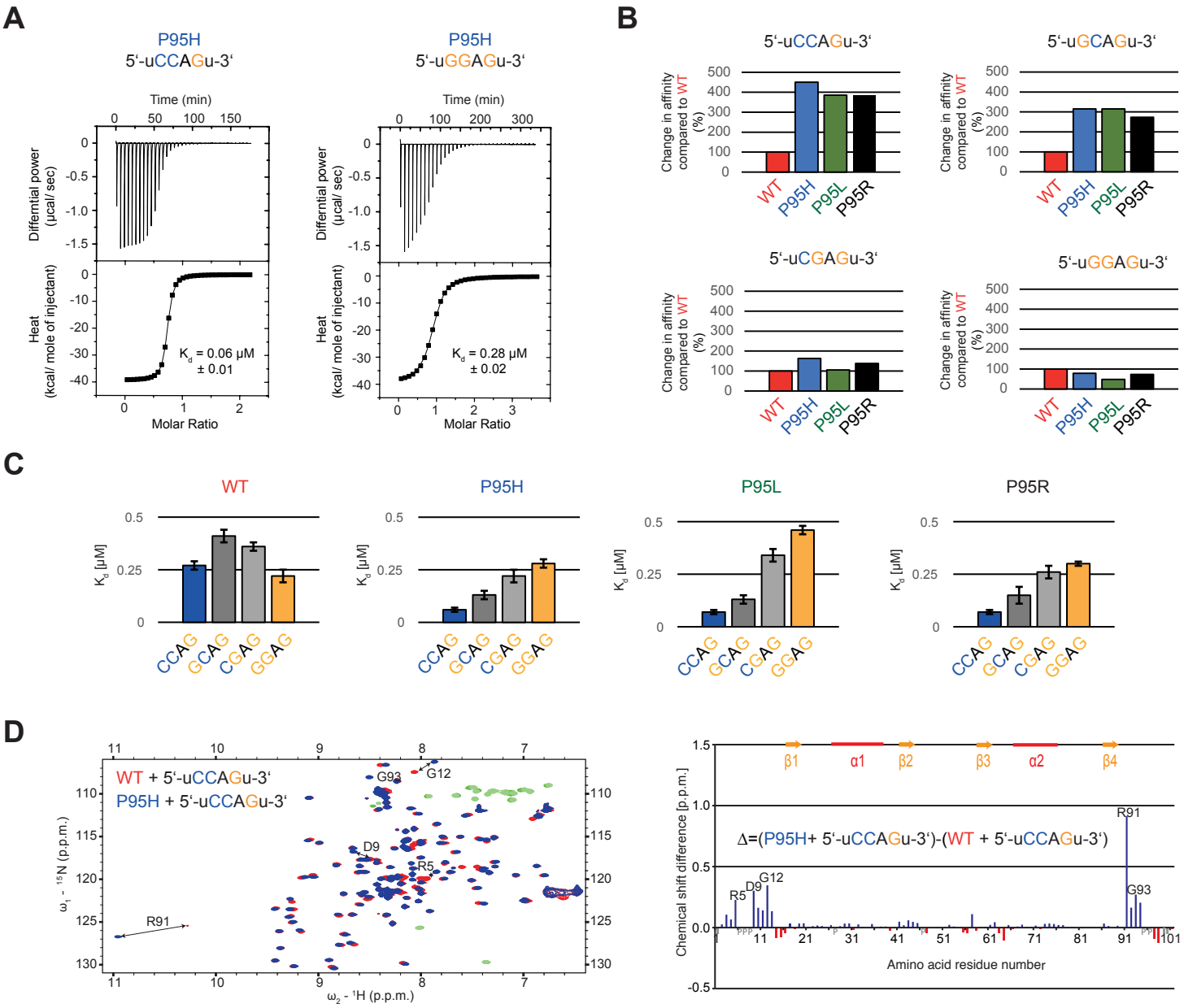
Figure

Figure 4



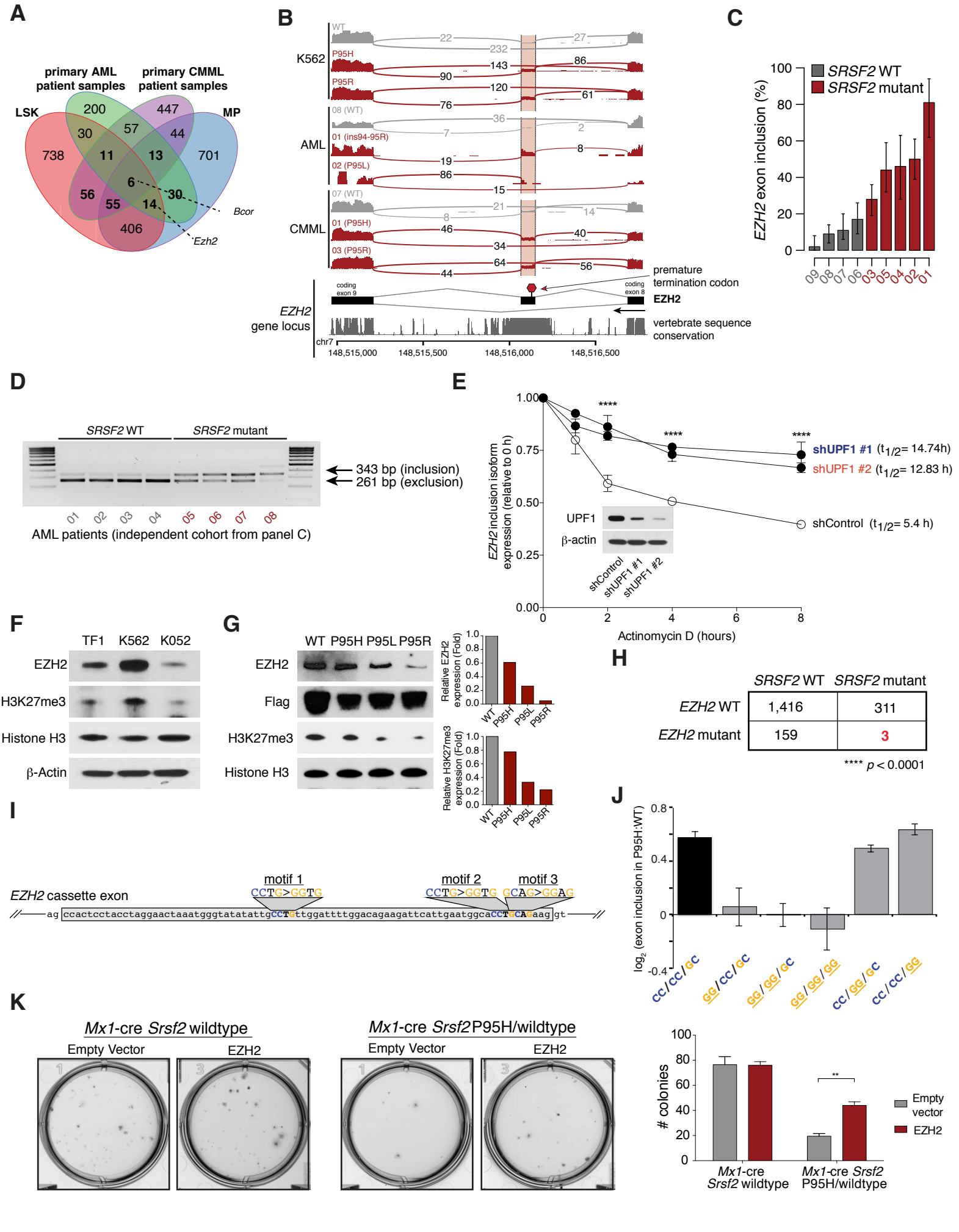
Figure

Figure 5



Figure

Figure 6



Click here to download Supplemental Text & Figures: 2015_04_10 Supplemental Information RESUBMITTED final.pdf

Supplemental Data

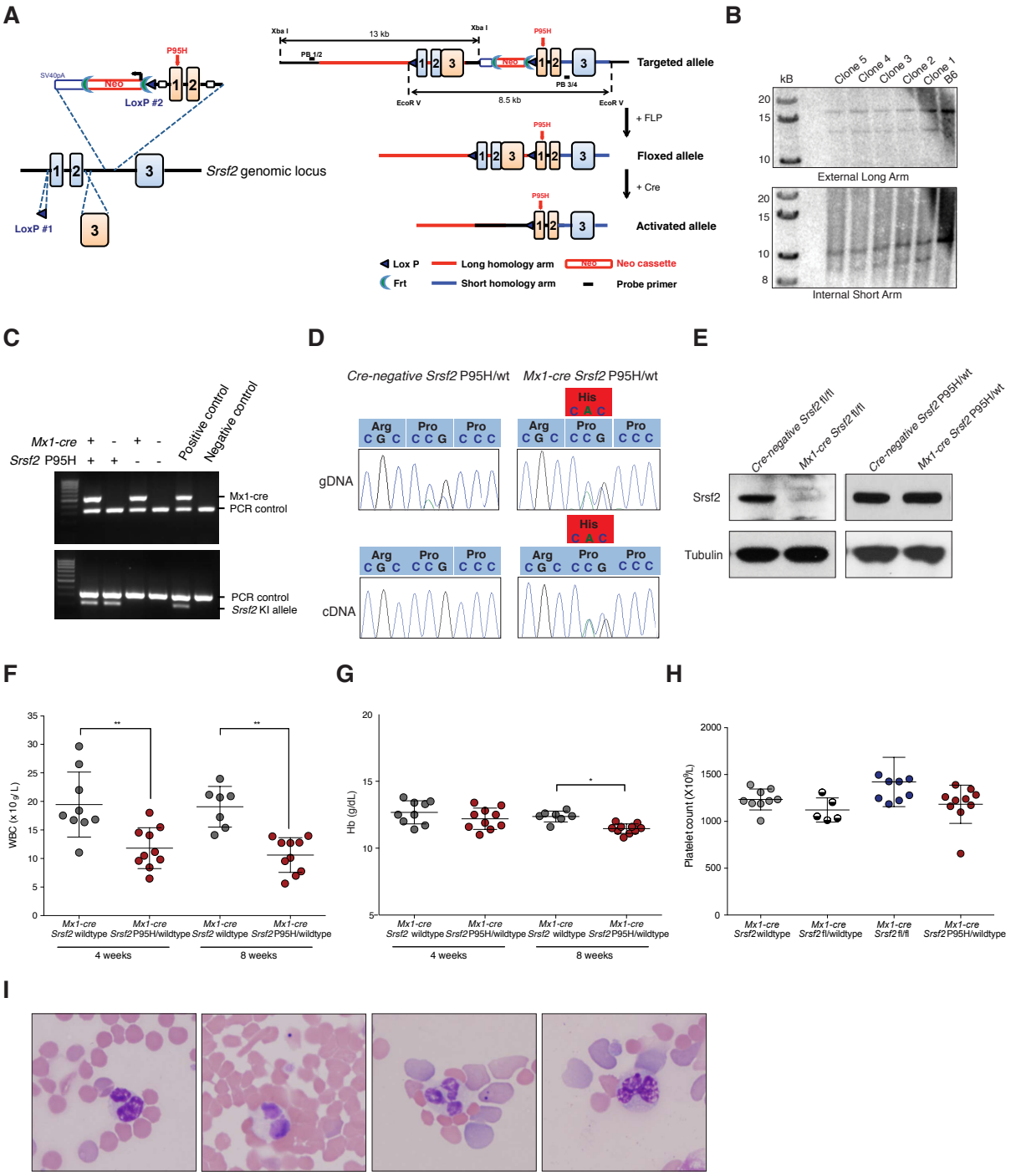


Figure S1, Related to Figure 1. *Srsf2*P95H mutation results in cytopenias and dysplasia *in vivo*. (A) Schematic depiction of the targeting strategy to generate the *Srsf2*P95H conditional knock-in allele. The *Srsf2*P95H mutation (c.284_5 CG > AC) was created within exon 1 of *Srsf2* in an *Srsf2* mini-gene following an SV40pA-FRT-Neo-FRTLoxP cassette. The entire construct was inserted in intron 2-3 of *Srsf2*. A second copy of exon 3 was fused to the end of exon 2 to enable appropriate expression of WT *Srsf2*. Additionally a distal LoxP cassette was inserted 87 bp upstream of exon 1. In the presence of cre recombinase, WT *Srsf2* is excised allowing in-frame expression of *Srsf2* P95H. (B) Verification of correct homologous recombination using Southern blots from targeted embryonic stem cells. (C) Verification of the presence of *Mx1-cre* and *Srsf2*P95H alleles using genomic PCR. (D) Sequencing electropherograms of genomic DNA and cDNA from nucleated bone marrow (BM) cells of *cre*-negative *Srsf2*P95H/WT and *Mx1-cre Srsf2*P95H/WT mice 2 weeks following pIpC injection. (E) Immunoblot analysis of Srsf2 in nucleated BM cells from *cre*-negative *Srsf2*fl/fl, *Mx1-cre Srsf2*fl/fl, *cre*-negative *Srsf2*P95H/WT, and *Mx1-cre Srsf2*P95H/WT mice from same time point following pIpC injection as shown in (D). Enumeration of white blood cells (WBCs) (F) and hemoglobin (Hb) (G) of mice transplanted in a noncompetitive manner with bone marrow (BM) cells from *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl or *Mx1-cre Srsf2*P95H/WT mice at 4 weeks and 8 weeks following pIpC injection. N=7-10 mice/genotype for all data shown here. (H) Platelet counts of *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/fl or *Mx1-cre Srsf2*P95H/WT mice 14 weeks post-pIpC injection. (I) Peripheral blood smears of *Mx1-cre* Srsf2P95H/WT mice 14 weeks after pIpC injection revealing hypolobated and hypogranulated neutrophils

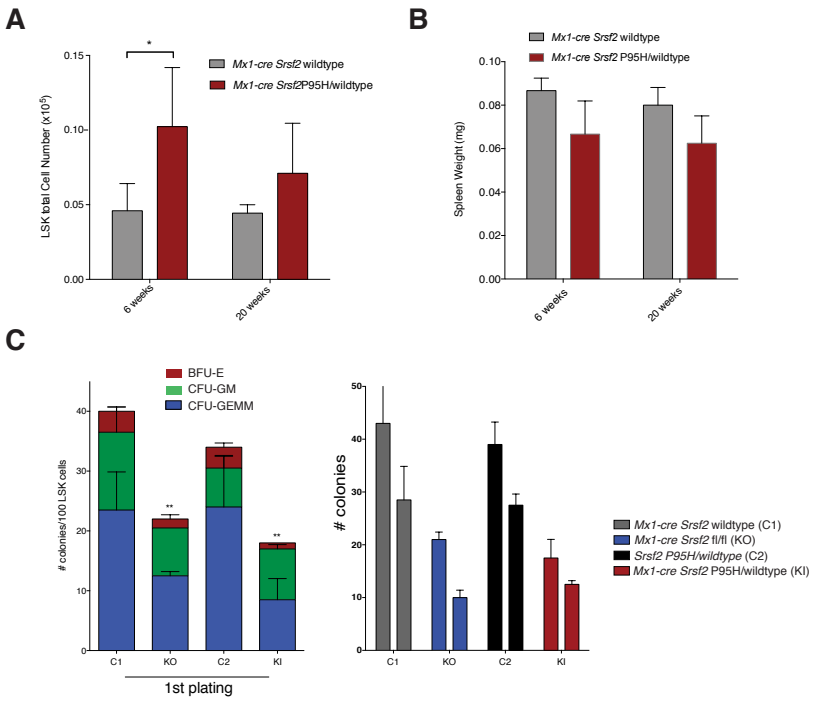


Figure S2, Related to Figure 2. Conditional expression of *Srsf2*P95H results in expansion of hematopoietic stem and progenitor cells. (A) Quantification of total numbers of lineage-negative Sca1+ c-Kit+ (LSK) cells in the spleen of 6 and 20-week-old *Mx1*-cre *Srsf2*P95H and *Mx1*-cre control mice (n=5 mice per genotype and age). (B) Spleen weight of mice shown in (A). (C) Assessment of initial colony formation (left) and serial replating capacity (right) of 100 purified lineage-negative Sca1+ cKit+ (LSK) cells from 6-week-old primary *Mx1-cre Srsf2* WT ("C1"), *Mx1-cre Srsf2*fl/fl ("KO"), *Cre*-negative *Srsf2*P95H ("C2"), and *Mx1-cre Srsf2*P95H ("KI") mice 2 weeks following pIpC injection. Error bars represent mean ± SD; *p < 0.05; **p < 0.001.

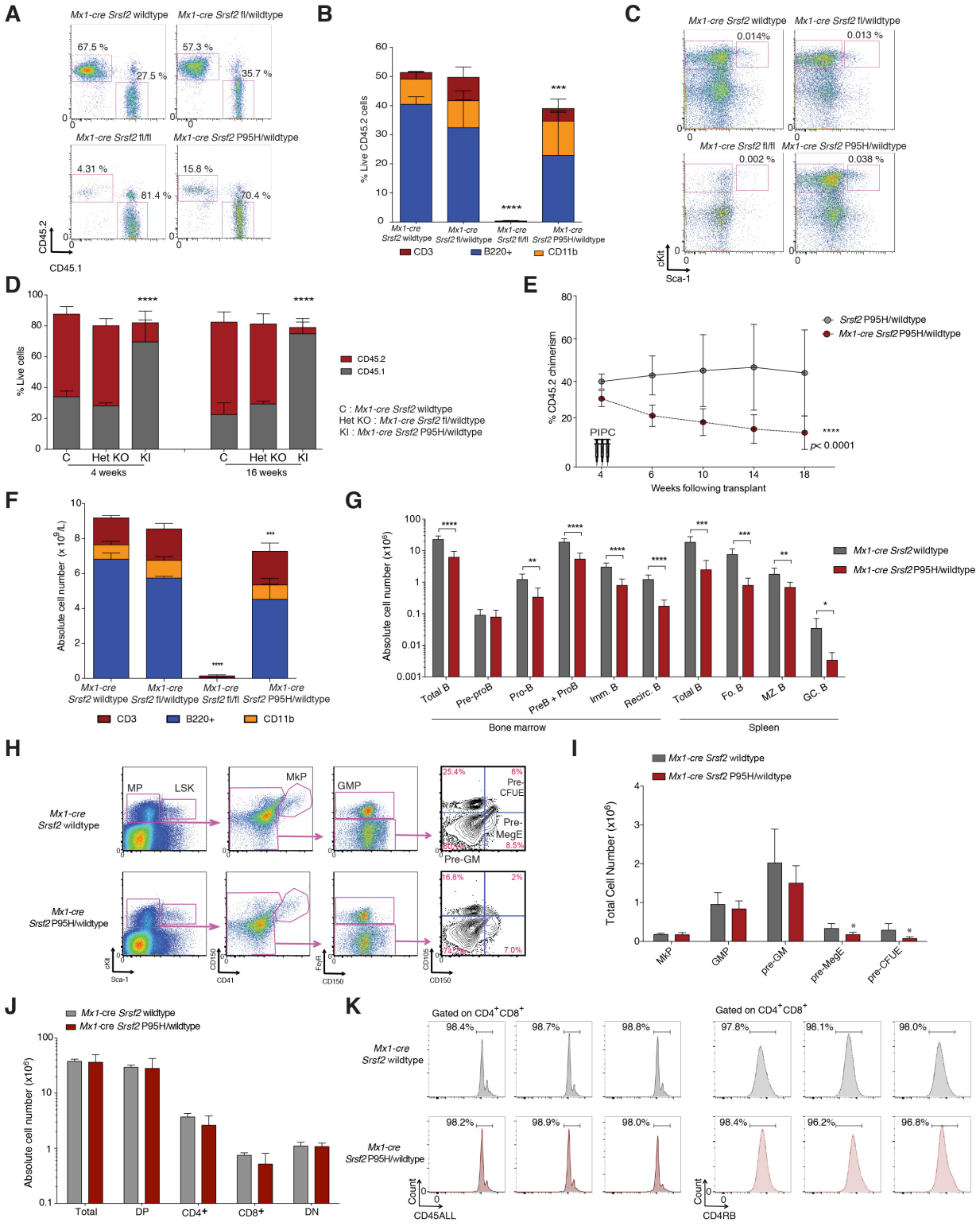

A
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 fl/wildtype
67.5 %
27.5 %
57.3 %
35.7 %
Mx1-cre Srsf2 fl/fl
Mx1-cre Srsf2 P95H/wildtype
4.31 %
81.4 %
15.8 %
70.4 %
CD45.2
CD45.1
B
% Live CD45.2 cells
***
****
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 fl/wildtype
Mx1-cre Srsf2 fl/fl
Mx1-cre Srsf2 P95H/wildtype
CD3
B220+
CD11b
C
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 fl/wildtype
0.014%
0.013 %
Mx1-cre Srsf2 fl/fl
Mx1-cre Srsf2 P95H/wildtype
0.002 %
0.038 %
cKit
Sca-1
D
% Live cells
****
****
CD45.2
CD45.1
C
Het KO
KI
4 weeks
16 weeks
C : Mx1-cre Srsf2 wildtype
Het KO : Mx1-cre Srsf2 fl/wildtype
KI : Mx1-cre Srsf2 P95H/wildtype
E
% CD45.2 chimerism
Srsf2 P95H/wildtype
Mx1-cre Srsf2 P95H/wildtype
PIPC
****
p< 0.0001
Weeks following transplant
F
Absolute cell number (x 10^9/L)
***
****
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 fl/wildtype
Mx1-cre Srsf2 fl/fl
Mx1-cre Srsf2 P95H/wildtype
CD3
B220+
CD11b
G
Absolute cell number (x10^6)
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 P95H/wildtype
Total B
Pre-proB
Pro-B
PreB + ProB
Imm. B
Recirc. B
Total B
Fo. B
MZ. B
GC. B
Bone marrow
Spleen
H
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 P95H/wildtype
MP
LSK
MkP
GMP
25.4%
6%
Pre-CFUE
Pre-MegE
8.5%
Pre-GM
16.8%
2%
7.0%
cKit
Sca-1
CD150
CD41
FcγR
CD150
CD105
CD150
I
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 P95H/wildtype
Total Cell Number (x10^6)
MkP
GMP
pre-GM
pre-MegE
pre-CFUE
J
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 P95H/wildtype
Absolute cell number (x10^6)
Total
DP
CD4+
CD8+
DN
K
Gated on CD4+CD8+
Gated on CD4+CD8+
Mx1-cre Srsf2 wildtype
Mx1-cre Srsf2 P95H/wildtype
98.4%
98.7%
98.8%
97.8%
98.1%
98.0%
98.2%
98.9%
98.0%
98.4%
96.2%
96.8%
Count
CD45ALL
Count
CD4RB

Figure S3, Related to Figure 3. *Srsf2*P95H mutation impairs hematopoietic stem cell self-renewal and differentiation. (A) Representative FACS plots showing CD45.1/CD45.2 staining and chimerism in peripheral blood mononuclear cells 20 weeks after pIpC injection. (B) Proportion of CD45.2+ peripheral blood cells of each lineage 24 weeks after competitive transplantation (n=10 mice/genotype) as determined by FACS analysis. (C) Representative FACS analysis of bone marrow LSK population in mice transplanted with each genotype. Antibody stains are as indicated and parent gate is live, CD45.2+, lineage-negative cells. (D) Percentage of CD45.2+ chimerism in the peripheral blood of recipient CD45.1 mice (n=10 mice/genotype) following serial secondary competitive transplantation with equal amounts of CD45.1 bone marrow mononuclear cells or CD45.2+ *Mx1*-cre *Srsf2* WT, *Mx1*-cre *Srsf2*fl/WT or *Mx1*-cre *Srsf2*P95H/WT mice following 18 weeks of transplantation in primary recipients. (E) Percentage of CD45.2+ chimerism in the peripheral blood of recipient CD45.1 mice (n=10 mice/genotype) following competitive transplantation with equal amounts of CD45.1 bone marrow mononuclear cells or CD45.2+ *Cre*-negative *Srsf2*P95H/WT or *Mx1*-cre *Srsf2*P95H/WT mice. pIpC was administered at 4 weeks following transplantation. (F) Flow cytometric enumeration of donor-derived (CD45.2+) B220+, CD11b+ and CD3+ cells in the peripheral blood of *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl or *Mx1-cre Srsf2*P95H/WT mice 14 weeks after noncompetitive transplantation. (G) Number of viable cells for B–lymphoid populations in the BM and spleen from *Mx1-cre Srsf2* WT and *Mx1-cre Srsf2*P95H mice. Cell surface markers used to define various subsets are as follows: total B, B220+CD19+; pre-pro B, CD11c−NK1.1−CD19−B220+CD43+; pro-B, B220+CD19+cKit+IgM-; preB+proB, B220+CD19+cKit−IgM−; immature B (Imm. B), B220lowCD19+IgM+; recirculating B (Recirc. B), B220highCD19+IgM+; follicular B, B220+CD23+CD21high; marginal zone B, B220+CD23-CD21high; germinal center B, B220+CD19+Fas+GL7+. (H) Representative FACS plots of immunophenotyping of intermediate myeloid progenitors (Pronk et al., 2007) in 6-week-old *Mx1-cre* control and *Srsf2*P95H mice. Highlighted are the pre-MegE (lineage-negative (LN) Sca1- c-

Kit+ CD41- FcγR- CD150+ CD105-) and pre-CFUE (LN Sca1- c-Kit+ CD41- FcγR- CD150+ CD105+) populations. (I) Quantification of total numbers of intermediate myeloid progenitors based on FACS analysis shown in (H) (n=7 mice per genotype). (J) Number of viable cells from thymic T cell populations in *Mx1-cre Srsf2* WT or *Mx1-cre Srsf2*P95H mice 24 weeks after non-competitive transplantation (DP: CD4/CD8 double-positive cells; DN: CD4/CD8 double-negative cells). (K) Histogram of CD45ALL and CD45RB expression on CD4/CD8 double positive thymic T-cells by FACS from the mice shown in (J). Each histogram reveals data for an individual control or *Srsf2*P95H mutant mouse. Error bars represent mean ± SD; *p < 0.05; **p < 0.001; ***p < 0.0002; ****p < 0.0001.

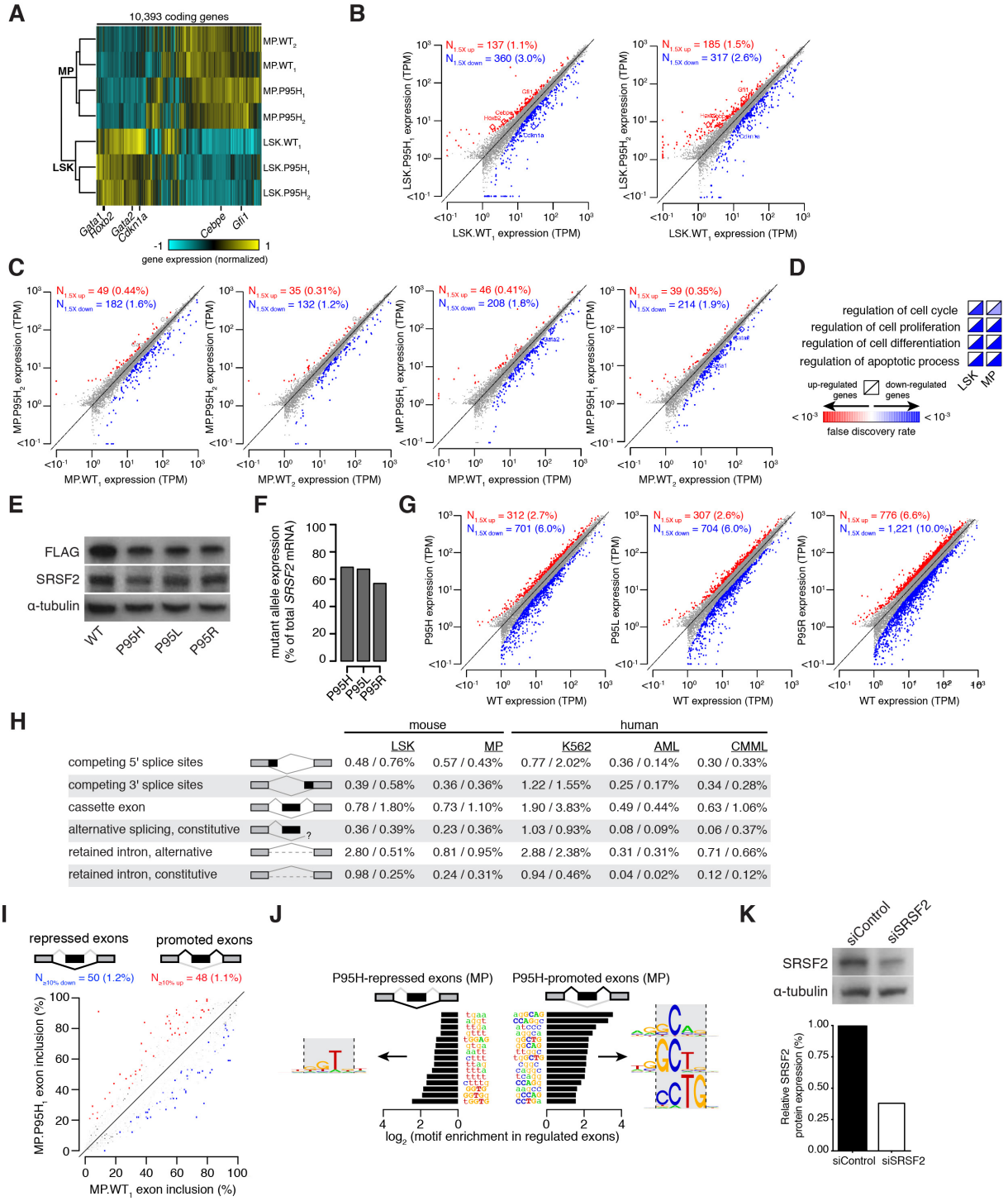


Figure S4, Related to Figure 4. *SRSF2* mutations cause a bias towards gene repression in hematopoietic cells and are associated with sequence-specific alterations in splicing of cassette exons. (A) Heat map based on coding gene expression of sorted lineage-negative

Sca1+ c-Kit+ (LSK) and lineage-negative Sca1- c-Kit+ myeloid progenitor (MP) cells from *Mx1-cre Srsf2*P95H/WT or *Mx1-cre* control mice. Gene expression measured by RNA-seq and normalized with the TMM method (Robinson and Oshlack, 2010). Ward's method, z-score normalized. (B) Scatter plots of coding gene expression in lineage-negative Sca1+ c-Kit+ (LSK) cells from *Mx1-cre Srsf2*P95H (P95H) mice relative to *Mx1*-cre controls (WT). Red and blue dots represent up- and down-regulated coding genes in P95H vs. WT cells, respectively, where up/down-regulation are defined as increases/decreases in gene expression of magnitude ≥1.5-fold with a Bayes factor ≥100 as estimated by Wagenmakers's framework. Percentages indicate the fraction of coding genes that are up or down-regulated. Plot restricted to coding genes with expression ≥1 TPM in at least one sample. TPM, transcripts per million as estimated by RSEM. Each graph displays data from a single biological replicate of the cell population sequenced. (C) As (B), but for myeloid progenitor cells (MP; lineage-negative Sca1- c-Kit+). (D) Gene Ontology (GO) enrichment for differentially expressed genes in *Mx1-cre Srsf2*P95H/WT versus *Mx1-cre* control primary murine LSK and MP cells. (E) Western blot showing levels of FLAG-tagged SRSF2 protein in K562 cells expressing the indicated allele. (F) Bar plot indicating the ratio of mutant to WT *SRSF2* mRNA levels in K562 cells as measured by RNA-seq. (G) As (B), but for K562 cells expressing the *SRSF2*P95H, P95L, and P95R alleles relative to expression of *SRSF2* WT (WT). (H) Frequency of differential splicing for different classes of splicing events associated with *SRSF2* mutations, normalized to the number of alternatively spliced events of each class. From top to bottom, numbers indicate the percentage of increases/decreases in usage of intron-proximal 5' splice sites, usage of intron-proximal 3' splice sites, inclusion of cassette exons, canonical splicing of annotated constitutive splice junctions, removal of introns annotated as alternative, and removal of introns annotated as constitutive. AML, primary acute myeloid leukemia samples. CMML, primary chronic myelomonocytic leukemia samples. (I)-(J), As Fig. 4A-B, but for myeloid progenitor (MP; lineage-negative Sca1- c-Kit+) cells from *Mx1*-cre

*Srsf2*P95H and *Mx1*-cre control cell population. (K) Western blot analysis and quantification of *SRSF2* knockdown in TF-1 cells.

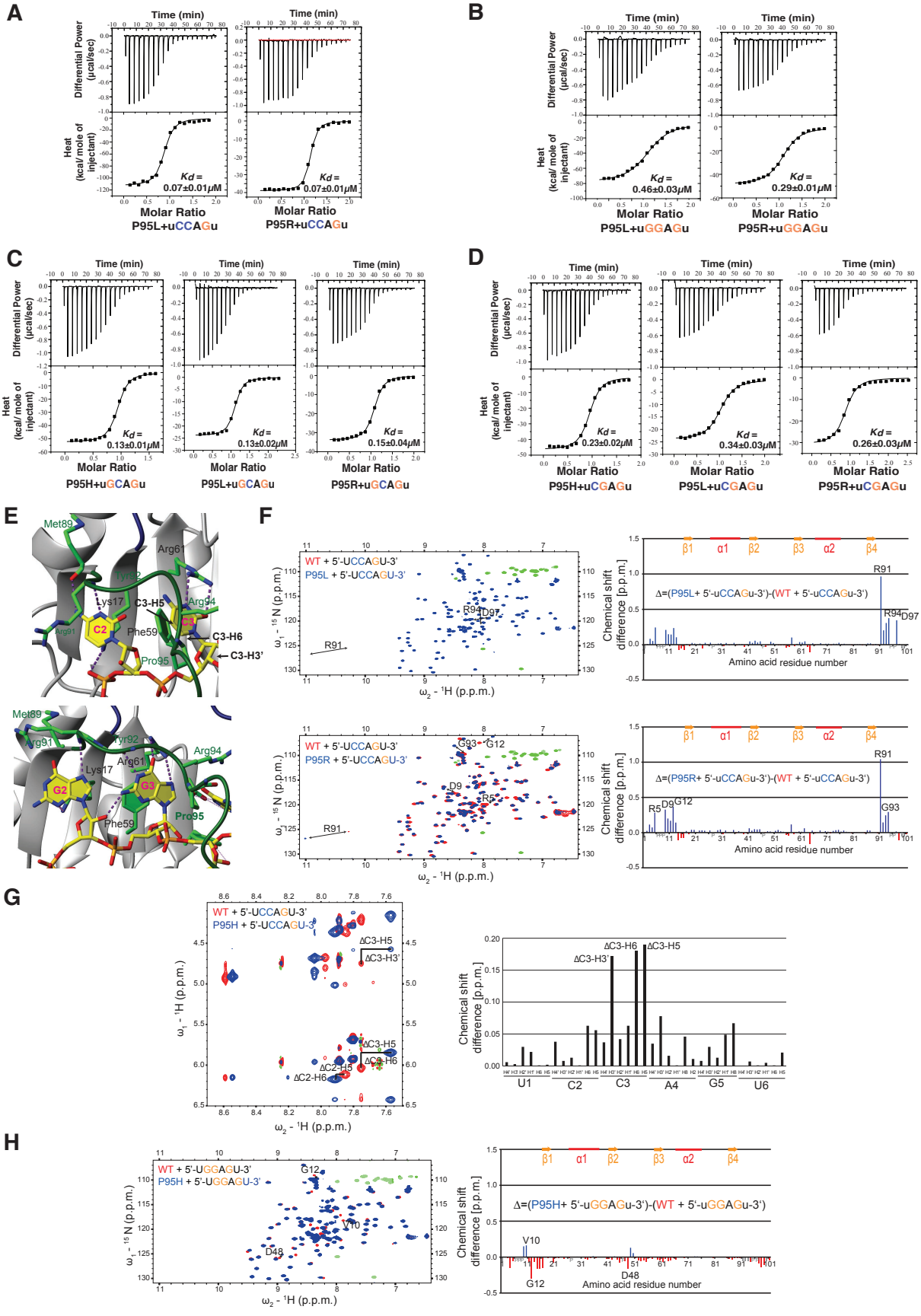

A
Time (min)
Differential Power (μcal/sec)
Heat (kcal/ mole of injectant)
Kd = 0.07±0.01μM
Molar Ratio
P95L+uCCAGu
Kd = 0.07±0.01μM
Molar Ratio
P95R+uCCAGu
B
Kd = 0.46±0.03μM
Molar Ratio
P95L+uGGAGu
Kd = 0.29±0.01μM
Molar Ratio
P95R+uGGAGu
C
Kd = 0.13±0.01μM
Molar Ratio
P95H+uGCAGu
Kd = 0.13±0.02μM
Molar Ratio
P95L+uGCAGu
Kd = 0.15±0.04μM
Molar Ratio
P95R+uGCAGu
D
Kd = 0.23±0.02μM
Molar Ratio
P95H+uCGAGu
Kd = 0.34±0.03μM
Molar Ratio
P95L+uCGAGu
Kd = 0.26±0.03μM
Molar Ratio
P95R+uCGAGu
E
Met89
Arg61
Tyr92
Lys17
C3-H5
Arg94
Arg91
C2
Phe59
C3-H6
C3-H3'
Pro95
Met89
Arg91
Tyr92
Arg94
Arg61
Lys17
G2
G3
Phe59
Pro95
F
WT + 5'-UCCAGU-3'
P95L + 5'-UCCAGU-3'
R91
R94
D97
ω1 - 15N (p.p.m.)
ω2 - 1H (p.p.m.)
β1 α1 β2 β3 α2 β4
Δ=(P95L+ 5'-uCCAGu-3')-(WT + 5'-uCCAGu-3')
Chemical shift difference [p.p.m.]
Amino acid residue number
WT + 5'-UCCAGU-3'
P95R + 5'-UCCAGU-3'
G93
G12
D9
R5
Δ=(P95R+ 5'-uCCAGu-3')-(WT + 5'-uCCAGu-3')
G
WT + 5'-UCCAGU-3'
P95H + 5'-UCCAGU-3'
ΔC3-H5
ΔC3-H3'
ΔC2-H5
ΔC2-H6
ΔC3-H6
ω1 - 1H (p.p.m.)
ω2 - 1H (p.p.m.)
Chemical shift difference [p.p.m.]
U1 C2 C3 A4 G5 U6
H
WT + 5'-UGGAGU-3'
P95H + 5'-UGGAGU-3'
G12
V10
D48
ω1 - 15N (p.p.m.)
ω2 - 1H (p.p.m.)
Δ=(P95H+ 5'-uGGAGu-3')-(WT + 5'-uGGAGu-3')
Chemical shift difference [p.p.m.]
Amino acid residue number

Figure S5, Related to Figure 5. *SRSF2*P95 mutations result in altered RNA-binding specificity and lead to relocation of termini with the central cytosine most affected. ITC raw data and corresponding binding curves depicted for *SRSF2* wild type and mutant (P95H, P95R, P95L) SRSF2 RNA recognition motif domains with (A) 5'-uCCAGu-3', (B) 5'-uGGAGu-3', (C) 5'-uGCAGu-3', and (D) 5'-uCGAGu-3'. (E) Close up on P95 as observed in the solution structure of SRSF2 RRM bound to 5'-uCCAGu-3' (2leb, top) and 5'-uGGAGu-3' (2lec, bottom) (Daubner et al., 2012). The backbone is in cartoon representation and the amino acids with important roles in RNA binding are shown in stick representation, with carbon atoms in green (protein) and yellow (RNA), nitrogen in blue, oxygen in red and phosphate in orange (Koradi et al., 1996). (F) (left) Overlay of 2D [^{15}N-^{1}H] HSQCs for wild type (red) and P95H/L/R mutants (blue) bound to 5'-uCCAGu-3' RNA, with negative peaks in green (WT) and light-green (mutant). (right) Difference of the chemical shift perturbations of P95H/L/R and WT SRSF2. (G) (left) Overlay of 2D F1-filtered F2-filtered [^{1}H-^{1}H] nuclear Overhauser effect spectroscopy (NOESYs) of WT (red) and P95H SRSF2 (blue) bound to 5'-uCCAGu-3' RNA. The three RNA resonances with the highest chemical shift perturbations are depicted. See also (E) for a depiction of RNA resonances in the WT solution structure. Mapping of the chemical shift perturbations calculated as $\Delta\delta=[(\delta H)^2+(\delta H)^2]^{1/2}$. (H) (left) Overlay of 2D [^{15}N-^{1}H] HSQCs for wild type (red) and P95H mutant (blue) bound to 5'-uGGAGu-3' RNA, with negative peaks in green (WT) and light-green (mutant). (right) Difference of the chemical shift perturbations for P95H mutant and wild type. Note that some of the C-terminal residues could not be assigned due to chemical exchange broadening.

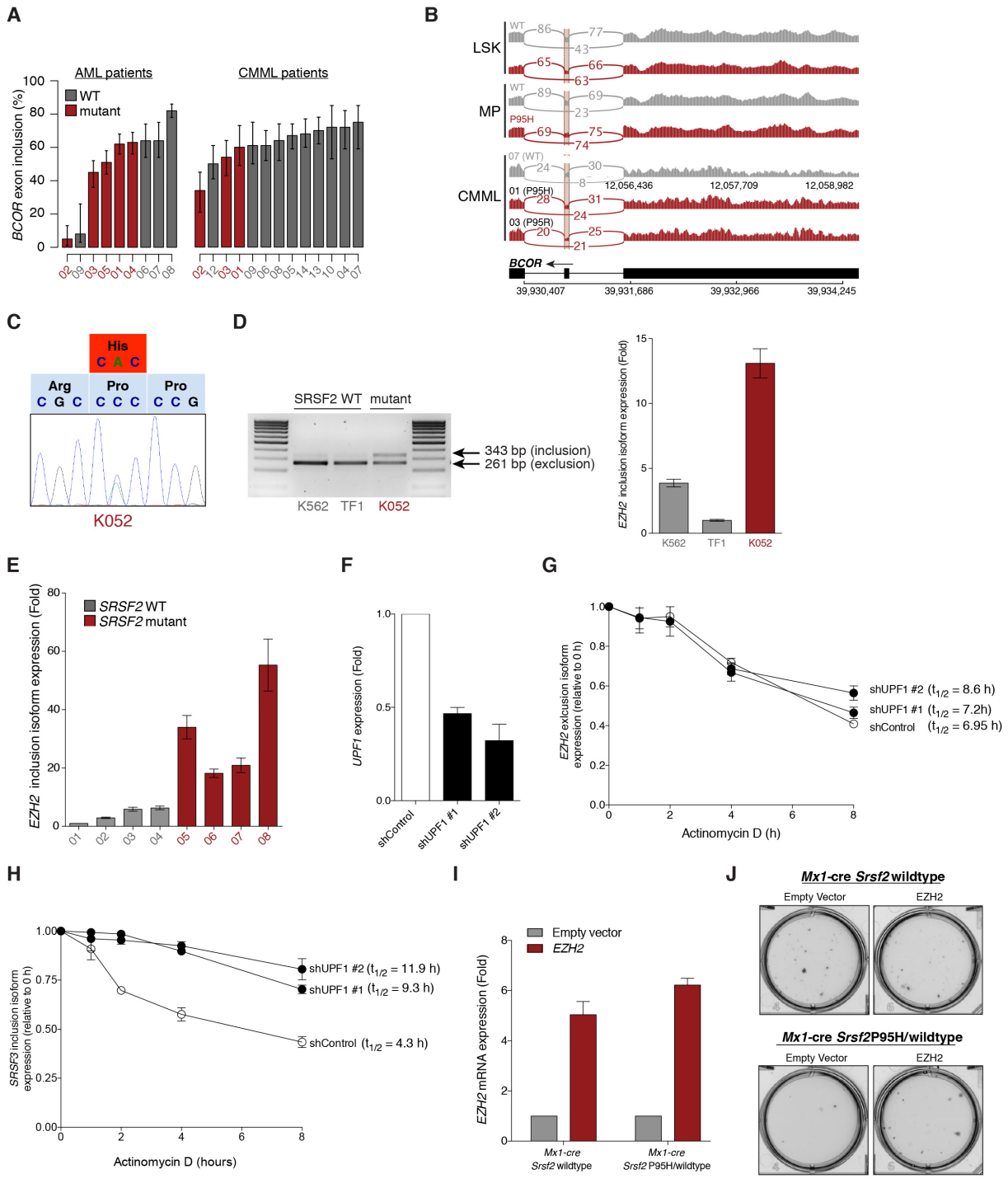


Figure S6, Related to Figure 6. Key splicing alterations affected by *SRSF2* mutations. (A) Bar plot describing the percentage of *BCOR* transcripts harboring a specific cassette exon in the indicated primary patient acute myeloid leukemia (AML) and chronic myelomonocytic leukemia (CMML) samples. (B) IGV/sashimi plots illustrating the cassette exon event repressed by

SRSF2 mutations in the multiple datasets analyzed here (as quantified in patient samples in (A)). Shown are data from primary murine LSK (lineage-negative Sca1+ c-Kit+) and myeloid progenitor (MP; lineage-negative Sca1- c-Kit+) cells, as well as *SRSF2* WT and mutant primary patient CMML samples. The exon repressed by *SRSF2* mutations is shaded in red. The numbers labeling each arc correspond to the number of reads aligned to the junction spanning the exons connected by the arc. For primary patient samples, the sample numbers correspond to patient identifiers listed in Table S1, and the exact mutation in *SRSF2* in each sample is shown in parentheses. (C) Sanger sequencing identifying the *SRSF2*P95H mutation in the human leukemia cell line K052 cells. (D) Gel picture of RT-PCR (left) and qRT-PCR data using primers specific for the *EZH2* isoform inclusion event (right) in *SRSF2* WT (K562, TF1) and mutant (K052) human leukemia cell lines. (E) qRT-PCR using primers specific for the *EZH2* isoform inclusion event in a panel of primary human AML samples WT or mutant for *SRSF2* (n=4 *SRSR2* mutant samples; corresponds to Figure 6D). (F) qRT-PCR of *UPF1* shRNA-mediated knockdown in K052 cells. (G) qRT-PCR of *EZH2* exclusion isoform event (from the same experiments as shown in Figure 6D) in K052 cells following control or *UPF1* knockdown and actinomycin D treatment. (H) qRT-PCR of an *SRSF3* exon inclusion isoform known to induce nonsense-mediate decay (Lareau et al., 2007; Ni et al., 2007) in the same experiment as shown in (G). (I) qRT-PCR of c-Kit+/ZsGreen1+ cells following overexpression of empty vector (pRetroX-IRES-ZsGreen1) or *EZH2* full-length cDNA (expressed from a pRetroX-IRES-ZsGreen1 vector backbone). (J) Photographs of methylcellulose colony assays 14 days after plating of 20,000 c-Kit+/ZsGreen1+ cells from *Srsf2* WT and mutant mice overexpressing empty vector or *EZH2* (biological duplicate of the experiment shown in Figure 6K).

Tables S1-S5 (Please see separately uploaded Excel files for Tables S2-S5)

Table S1, Related to Figure 4. Characteristics of primary patient leukemia samples with *SRSF2* genotype, mutant allele expression level, and clinical characteristics used for RNA-seq.

Table S2, Related to Figures 4 and 6. Differentially spliced events in LSK *Srsf2*P95H vs. WT cells. Splicing events that are differentially spliced in LSK *Srsf2*P95H vs. WT cells. Each row of the table corresponds to isoform 1 of a splicing event, where isoform 1 is defined as: inclusion isoform for cassette exons ("se"), intron-proximal isoform for competing 3' or 5' splice site events ("a3ss", "a5ss"), inclusion of upstream exon for mutually exclusive exons ("mxe"), spliced isoform for intron retention events annotated as alternative ("ri") or constitutive ("ci"), and canonical splicing of constitutive junctions ("cj"). The table columns are defined as: "coords", genomic coordinates of splicing events; "gene", Ensembl gene ID; "geneName", name of gene; "geneDescription", description of gene obtained from Ensembl; "type", class of splicing event; "nmdTarget", whether isoform 1 of the event is a predicted NMD substrate (NA indicates that neither or all isoforms of the event are predicted NMD substrates); "deltaPsi_XX", difference in isoform ratio between the samples indicated by "XX". Table restricted to events that are differentially spliced in at least one comparison.

Table S3, Related to Figures 4 and 6. Differentially spliced events in MP *Srsf2*P95H vs. WT cells. As Table S2, but for MP cells.

Table S4, Related to Figure 6. Differentially spliced events in CMML *SRSF2* mutant vs. WT samples. As Table S2, but for CMML *SRSF2* mutant vs. WT samples. "deltaPsi", difference in median isoform ratios for mutant and WT samples; "pval", p-value based on the Mann-Whitney

U test.

Table S5, Related to Figures 4 and 6. Differentially spliced events in AML *SRSF2* mutant vs. WT samples. As Table S4, but for AML samples.

Supplemental Experimental Procedures

Animals. All animals were housed at Memorial Sloan Kettering Cancer Center. All animal procedures were conducted 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 Memorial Sloan Kettering Cancer Center.

Generation of the *Srsf2*P95H conditional knock-in mice. A 10.65 kb genomic DNA used to construct the targeting vector was first subcloned from a positively identified B6 BAC clone (RP23: 363E6) using a homologous recombination-based technique. The region was designed such that the long homology arm (LA) extends ~6.74 kb 5' to the distal LoxP site, and the short homology arm (SA) extends 2.57 kb 3' to the mini-gene that contains the duplicated genomic sequence starting at 87 bp upstream of exon 1 and ending at 185 bp downstream of exon 2. The distal LoxP cassette was inserted 87 bp upstream of exon 1. The point mutation, which alters a two-nucleotide sequence from CG to AC, was engineered in exon 1 of the mini-gene. The mutation modifies the codon sequence to change the amino acid from Proline to Histidine. The cassette LoxP-FRT-Neo-FRT was inserted immediately upstream of the mini-gene insertion site in intron 2-3. A second copy of exon 3 was fused to the end of exon 2 to enable an appropriate expression of the WT (wt) *Srsf2* gene. The mutation CCG > CAC (resulting in an change of amino acid Proline 95 to Histidine) within exon 1 of the mini-gene was generated by 3-step PCR mutagenesis. 4 mutagenesis primers (Ex1KpnFw1, PT2, PT3 and Ex2EcoRIrev2) were designed to amplify a fragment with the size of ~1.34 kb, which incorporates the mutation at the desired position. The PCR fragment carrying the point mutation was then inserted right after the cassette SV40pA-FRT-Neo-FRTLoxP in a shuttle vector. Then, the whole sequence including Neo cassette and the mini-gene was inserted in the location of the targeting vector as described above using the engineered enzyme site, BsiWI. Fusion of exons 2 and 3 was also

accomplished by over-lapping PCR. An endogenous enzyme site SacII and the engineered enzyme BsiWI site (the same site used for the insertion of Neo and mini-gene) were used for replacing the correspondent sequence by the PCR synthesized fragment containing exon2-3 fusion.

The targeting vector was confirmed by restriction analysis and sequencing after each modification. The boundaries of the 2 homology arms were confirmed by sequencing with P6 and T73 primers that read through both sides of the backbone vector into the genomic sequence. The LoxPFRT-Neo-FRT-SV40pA cassette was confirmed by sequencing with N1 and N2 primers that read from the 5'- and 3'- ends of the Neo cassette, respectively, into the mini-gene (N1) and intron 2. The mini-gene was sequencing confirmed with primers FRSSQ2 and N1. The distal LoxP site was sequencing confirmed with primer Lox-PT1. The fusion of exons 2 and 3 was sequencing confirmed with primers fusion-PT1 and N2. SRSF-PT2 was used to sequencing confirm the remaining PCR modified region.

The BAC was sub cloned into a ~2.4kb pSP72 (Promega) backbone vector containing an ampicillin selection cassette for retransformation of the construct prior to electroporation. A pGK-gb2 FRT-flanked Neomycin cassette with a LoxP site attached on the 5' end was inserted into the gene as described in the project schematic.

10 mg of the targeting vector was linearized by NotI and then transfected by electroporation of IC1 (C57BL/6) embryonic stem (ES) cells. After selection with G418 antibiotic, surviving clones were expanded for PCR analysis to identify recombinant ES clones. Secondary confirmation of positive clones identified by PCR was performed by Southern Blotting analysis. DNA was digested with XbaI (External long arm) or EcoRV (Internal short arm), and electrophoretically separated on a 0.8% agarose gel. After transfer to a nylon membrane, the digested DNA was hybridized with a probe targeted against the 5' external region. DNA from a normal C57BL/6 (B6) mouse was used as WT control. Positive ES clones were expanded and microinjected into Balb/c blastocysts. Resulting chimeras with a high percentage black coat

color were mated to C57BL/6 FLP mice (The Jackson Laboratory) to remove the Neo cassette, and subsequently to IFNα-inducible *Mx1-cre* transgenic mice (Kuhn et al., 1995). Mice were backcrossed for six generations to C57BL/6 mice (also previously described; (Mead et al., 2013; Moran-Crusio et al., 2011)). Polyinosine-polycytosine (pIpC) (GE Healthcare (catalog # 27-4732-01)) was administered to 6-week-old mice at 12μg/g intraperitoneally every other day for 3 doses (as previously described (Mead et al., 2013; Moran-Crusio et al., 2011)).

Genotyping of *Srsf2*P95H conditional knock-in mice. *Srsf2* WT and *Srsf2* P95H/WT littermate mice were genotyped by PCR with primers loxP-F (5'-CAACACGGCCGATATCATAA-3') and Exon1-R (5'- TACAGTCCTCGTGGGTAGGG-3') using the following parameters: 95°C for 2 min, flowed by 35 cycles of 95°C for 45s, 55°C for 45s, and 72°C for 1 min, and then 72°C for 5 min. The P95H allele was detected as a band of 225 bp and the WT allele was not detected.

***Srsf2* conditional KO mice.** *Srsf2* floxed mice were obtained from The Jackson Laboratory and were previously generated and described by Dr. Xiang-Dong Fu (Wang et al., 2001).

Bone marrow (BM) transplantation. Freshly dissected femora and tibiae were isolated from *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl or *Mx1-cre Srsf2*P95H/WT CD45.2+ mice. Bone marrow (BM) was flushed with a 3-cc insulin syringe into PBS supplemented with 3 % fetal bovine serum. The BM was spun at 0.5 g by centrifugation and RBCs were lysed in ammonium chloride-potassiume bicarbonate lysis buffer for 5 min. After centrifugation, cells were resuspended in PBS plus 3 % FBS, passed through a cell strainer, and counted. Finally, 0.5 x 10^6 total BM cells of *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl or *Mx1-cre Srsf2*P95H/WT CD45.2+ mice were mixed with 0.5 x 10^6 WT CD45.1+ support BM and transplanted via tail vein injection into lethally irradiated (two times 450 cGy) CD45.1+ recipient

mice. Chimerism was measured by FACS from the peripheral blood at 4 weeks after transplant (week 0, pre-polyI:polyC). Chimerism was followed via FACS in the peripheral blood every 4 weeks (week 0, 4, 6, 8,12, and 16 after polyI:polyC injection). Additionally, for each bleeding, whole blood cell counts were measured on a blood analyzer, and peripheral blood smears were scored. For noncompetitive transplantation experiments, 10^6 total BM cells of *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl or *Mx1-cre Srsf2* P95H/WT CD45.2+ mice were injected into lethally irradiated (two times 450 cGy) CD45.1+ recipient mice.

***In vitro* colony-forming assays.** LSK cells were FACS-sorted from the BM of *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/fl, *cre*-negative *Srsf2*P95H, and *Mx1-cre Srsf2*P95H mice and seeded at a density of 500 cells/replicate for LSK cells 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 re-plating (20,000 cells/replicate).

EZH2 cDNA overexpression experiments. c-Kit+ cells were selected by anti-mouse CD117 MicroBeads (Miltenyi Biotech) from 6-week-old primary *Mx1*-cre *Srsf2* WT or *Mx1*-cre *Srsf2*P95H/WT mice 2 weeks following pIpC injection. Two million c-Kit+ cells were then infected with retrovirus expressing pRetroX-IRES-ZsGreen1 empty vector or full-length *EZH2* cDNA (NM_004456) in the same pRetroX-IRES-ZsGreen1 backbone. c-Kit+/ZsGreen+ cells were then FACS-sorted and 20,000 cells were plated in cytokine-supplemented methylcellulose medium (Methocult M3434; STEMCELL Technologies) in duplicate. Colonies were then enumerated 14 days later. Expression of *EZH2* cDNA was performed by qRT-PCR for *EZH2* using the following qRT-PCR primers: forward 5 - ACGCTTTTCTGTAGGCGATGT - 3 ; reverse 5 - TGTTGCATGAAAAGAATAATTGCAC - 3 .

Antibodies, FACS, and Western blot analysis. All FACS antibodies were purchased from BD Pharmingen, eBioscience, or BioLegend. BM mononuclear cells were stained with a lineage cocktail comprised of antibodies targeting CD3, CD4, CD8, B220, CD19, NK1.1, Gr-1, CD11b, Ter119, and IL-7Rα. 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-7Rα (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 following antibodies were used for Western Blot analysis: SRSF2 (Abcam), α-Tubulin (Sigma-Aldrich), histone H3K27me3 (Millipore), and total histone H3 (Cell Signaling).

Cell cycle and apoptosis analyses. For cell cycle analysis, the BrdU-APC kit was used (BD) according to the manufacturer's protocol. Mice were treated with 1 mg BrdU via intraperitoneal injection, followed by harvest of BM cells 24 h later. 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.

Histological analysis. Mice were sacrificed and autopsied, and the dissected tissue samples were fixed for 24 h in 4% paraformaldehyde, dehydrated, and embedded in paraffin. Paraffin blocks were sectioned at 4 mm and stained with H&E. Images were acquired using an Axio Observer A1 microscope (Carl Zeiss).

Peripheral blood analysis. Blood was collected by retro-orbital bleeding using heparinized microhematocrit capillary tubes (Thermo Fisher Scientific). Automated peripheral blood counts were obtained using a HemaVet 950 (Drew Scientific) according to standard manufacturer's

instruction. Differential blood counts were realized on blood smears stained using Wright-Giemsa staining and visualized using an Axio Observer A1 microscope.

Vector construction, viral infection, and cell culture (K562 cells). The *SRSF2* cDNA sequence (NCBI identifier NM_003016) was used to generate the construct SRSF2 + Gly + Gly + FLAG + T2A + mCherry, where T2A is the cleavage sequence EGRGSLLTCGDVEENPGP. Site-directed mutagenesis was used to generate the P95H, P95L, and P95R alleles. The resulting constructs were then cloned into the lentiviral vector pRRLSIN.cPPT.PGK-GFP.WPRE (Addgene Plasmid 12252). These vectors were transfected into 293T cells with the packaging vector PsPAX2 (Addgene plasmid 12260) and envelope vector pMD2.G (Addgene plasmid 12259) and viral supernatant was collected. K562 cells were grown in RPMI-1640 supplemented with 10% FCS. To generate stable cell lines, K562 cells were infected with concentrated lentiviral supernatant at a MOI of ~5. K562 cells were then expanded and mCherry+ cells were isolated by FACS.

Genome annotations. The UCSC knownGene and Ensembl 71 genome annotations were combined to generate a single genome annotation. This was then augmented by MISO v2.0 annotations of alternative splicing events. Constitutive junctions were defined as those splice junctions exhibiting no evidence of alternative splicing in the UCSC knownGene track. The procedure was identical for the human and mouse genomes. For each genome, we created two additional annotation files for the purpose of read mapping: a gene annotation file, created from the merge of UCSC and Ensembl, and a splice junction annotation file, created from a merge of UCSC, Ensembl, MISO, and all possible junctions between annotated splice sites of each gene (for identifying novel alternative splicing of constitutively spliced junctions).

Gene ontology enrichment analysis. We used the GOseq method to perform gene ontology enrichment analysis, and performed a bias correction corresponding to the total number of RNA-seq reads mapping to each gene. This corrects for biases arising from both gene length and expression level as described in GOseq.

Cluster analysis. Cluster analysis was performed with a modified version of heatmap.2 in the R gplots package. All coding genes with expression >= 1 TPM (transcripts per million) in at least one sample were used to compute the Euclidean distance between samples. Following a z-score normalization, Ward's method was used to cluster the samples.

Motif analysis. Statistically enriched motifs were identified with an *ab initio* approach by querying all k-mers of length 4, 5, or 6. An enrichment statistic was defined for each k-mer as the occurrence of each k-mer in all cassette exons that were promoted in *SRSF2*-mutant versus WT cells, divided by the occurrence of each k-mer in corresponding repressed cassette exons. This method is equivalent to identifying k-mers that maximally distinguish between promoted and repressed cassette exons. Statistically significant enrichment was identified with the non-parametric Kolmogorov-Smirnov test, performed using the distributions of occurrence for each k-mer in promoted versus repressed exons. Differentially enriched motifs with a p-value<0.05 were considered significant. Consensus motifs (**Figure 4B**) were then identified by performing k-means clustering on significantly enriched or depleted motifs with n = 5 centers. An identical *ab initio* analysis was performed using the intronic regions flanking differentially spliced cassette exons (taking 100 nt upstream or downstream of the exons), but no significantly enriched motifs were identified.

Isothermal titration calorimetry. The different protein mutants were purified with the same protocol as previously described (Daubner et al., 2012). Measurements were conducted either

on a MicroCal VP-ITC or MicroCal iTC200 instrument, which were calibrated according to the manufacturer's protocol. Concentrations of RNA and protein were calculated based on their optical density absorbance at 260 nm or 280 nm, respectively. The sample cell was loaded with either 1.4ml (VP-ITC) of 10 μM RNA and the syringe with 200 μM of protein, or 0.25ml (iTC200) of 10 μM protein and the syringe with 200μM RNA. Measurements were done at 37°C in the final buffer using either 35 injections of 6 μl protein (VP-ITC) or 21 injections of 2 μl RNA (iTC200). Data was integrated and normalized using the Origin 7.0 software according to a 1:1 RNA:protein ratio binding model.

NMR measurements and resonance assignments. All measurements were conducted in the same buffer (50 mM arginine, 50 mM glutamine, 20 mM Na_2HPO_4/NaH_2PO_4, pH = 5.5) at a temperature of 37°C. We recorded our data on Bruker spectrometers ranging from 500 Mhz to 900 Mhz field strength, all equipped with cryo-probes.

The backbone assignment of the SRSF2 P95H mutant was facilitated by the previously published assignment of the wild type (Daubner et al., 2012). We used the following additional experiments to complete the assignment of the P95H mutant: 2D [^{15}N-^{1}H] HSQC, 3D HNCA, 3D HNCO and 3D CBCA(CO)NH, all recorded in 90% H_2O, 10% D_2O.

The chemical shift differences were calculated for free and bound form of SRSF2 RRM with the formula $\Delta\delta=[(\delta HN)^2+(\delta N/6.41)^2]^{1/2}$. To show the difference in chemical shift perturbations, values were derived by $\Delta=\Delta\delta$(mutant)-$\Delta\delta$(wild type).

RNA was purchased from Dharmacon, deprotected according to the manufacturer's protocol, lyophilized and resuspended in the NMR buffer. To assign the resonances of the unlabeled RNA bound to SRSF2 RRM P95H mutant, we recorded: 2D [^{1}H-^{1}H] TOCSY, 2D [^{1}H-^{1}H] NOESY, 2D F1-filtered F2-filtered [^{1}H-^{1}H] NOESY and 2D F2 filtered [^{1}H-^{1}H] NOESY, all in 100% D_2O (Peterson et al., 2004). We used a mixing time of 120 ms for NOESY and 10, 30 and 60 ms for TOCSY spectra.

Human leukemia cell line sequencing. In order to identify a human leukemia cell line with an endogenous *SRSF2* mutation and to delineate *SRSF2* mutant and WT cell lines, genomic DNA from the following human leukemia cell lines were sequenced for the regions of recurrent mutation in *SRSF2*: HEL, HL-60, Kasumi-1, KBM5, K562, K052, KG-1, KU812, MOLM-13, MV4;11, Monomac-6, MUTZ3, NB4, NOMO1, OCI/AML-3, OCI/AML-4, SET2, SKNO1, TF1, THP-1, U937.

Quantitative RT-PCR. Total RNA was isolated using RNeasy Mini kit (Qiagen). For cDNA synthesis, total RNA was reverse transcribed with Verso cDNA Kit (Thermo scientific). Quantitative PCR of the target cDNAs was performed on Applied Biosystems QuantStudio 6 Flex cycler using FastStart SYBR Green Master mix (Roche Life Science) with the following primers: *EZH2* inclusion: forward: 5 - CAGCATTTGCCACTCCTACC - 3 ; reverse: 5 - AGAGCAGCAGCAAACTCCTTT - 3 . *EZH2* exclusion: forward: 5 -CAGCATTTGGAGGGAGCA - 3 ; reverse: 5 - GCTGGGCCTGCTACTGTTATT - 3 . *UPF1:* forward: 5 - AATTTGGTTAAGAGACATGCGG - 3 ; reverse: 5 - TCAGGGACCTTGATGACGTG - 3 . *18s rRNA:* forward: 5 - GTAACCCGTTGAACCCCATT - 3 ; reverse: 5 - CCATCCAATCGGTAGTAGCG - 3 ; *SRSF3* inclusion: forward: 5' – TGGAACTGTCGAATGGTGAA – 3'; reverse: 5' – GGGTGGTGAGAAGAGACATGA – 3'.

All samples including the template controls were assayed in triplicate. The relative number of target transcripts was normalized to the number of human 18s rRNA transcripts found in the same sample. The relative quantification of target gene expression was performed with the standard curve or comparative cycle threshold (CT) method.

Semiquantitative RT-PCR. Total RNA was isolated using RNease Mini kit (Qiagen). For cDNA synthesis, total RNA was reverse transcribed with Verso cDNA kit (Thermo scientific). Human *EZH2* exon 9-10 region was amplified using the forward 5 - TTTCATGCAACACCCAACACT -3 and reverse 5 - CCCTGCTTCCCTATCACTGT – 3' primer. The PCR cycling conditions (28 cycles) chosen were as follows: (1) 45s at 95 °C (2) 45s at 52 °C (3) 45s at 72 °C with a subsequent 5 min extension at 72 °C. Reaction products were analyzed on 2 % agarose gels. The bands were visualized by ethidium bromide staining.

mRNA stability assay. For mRNA half-life measurement using qRT-PCR, *UPF1* shRNA and shControl lentivirus infected K052 cells were treated with 2.5 μg/ml Actinomycin D and harvested at the indicated time points. *EZH2* inclusion, *EZH2* exclusion, *SRSF3* inclusion and *18s rRNA* mRNA levels were measured by RT-PCR as described above.

***EZH2* minigene.** An insert containing the *EZH2* genomic locus (chr7:148515109-148516788) was cloned into the EcoRV site of pUB6/V5-HisA vector (Invitrogen) by Gibson assembly cloning (NEB). Site-directed mutagenesis was used to introduce C>G mutations at each SSNG motif within the cassette exon. Plasmids containing minigenes were transfected using the Nucleofector II device from Lonza with the Cell Line Nucleofector Kit V (program T16). RNA was collected after 48h. We isolated total RNA from K562 cells using TRIzol and extracted RNA using the Qiagen RNeasy kit. Complementary DNA (cDNA) was generated using 1 ug of total RNA with a primer specific to the minigene transcript immediately upstream of the poly(A) tail (ACAACAGATGGCTGGCAACTAGAAG). Assays were performed in biological triplicate. Triplicates of equal amounts of 6 ng cDNA were used in a 5 uL reaction with 2.5 uL 2x SYBR Green PCR Master Mix (Life Technologies), and 50 nM forward and reverse primers. *EZH2* primers: inclusion (forward CAGCATTTGCCACTCCTACC, reverse

AGAGCAGCAGCAAACTCCTTT) and exclusion (forward CAGCATTTGGAGGGAGCA, reverse GCTGGGCCTGCTACTGTTATT).

References

Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. Journal of molecular graphics *14*, 51-55, 29-32.

Mead, A.J., Kharazi, S., Atkinson, D., Macaulay, I., Pecquet, C., Loughran, S., Lutteropp, M., Woll, P., Chowdhury, O., Luc, S.*, et al.* (2013). FLT3-ITDs instruct a myeloid differentiation and transformation bias in lymphomyeloid multipotent progenitors. Cell reports *3*, 1766-1776.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology *11*, R25.

Click here to download Supplemental Movies & Spreadsheets: Tables S1 to S5.xlsx