Mechanistic consequences of splicing factor mutations and their use for targeted cancer therapy

Khrystyna North

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Robert K. Bradley, Chair

Alice Berger

Stan Fields

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Mechanistic consequences of splicing factor mutations and their use for targeted cancer therapy

Khrystyna North

Chair of the Supervisory Committee:
Full Professor Robert K. Bradley
Fred Hutchinson Cancer Research Center

In 2011 Yoshida et al. identified mutations in RNA splicing factors as some of the most common mutations in patients with myelodysplastic syndrome (MDS). Mutations in splicing machinery result in the dysregulation of hundreds of isoforms, which lead to disease.

The dysregulation of these isoforms is consistent and splicing factor specific. The most commonly mutated splicing factors are SF3B1 and SRSF2. While we know that mutations in each have distinct consequences on splicing, the molecular drivers of disease and sequence characteristics which influence mutant promoted alternative splicing remain unknown. To address this, I first show through RNA-seq analysis that co-occurring mutations in SRSF2 and SF3B1 are synthetically lethal, and both converge on the NF-kB signaling pathway dysregulation. I then show that a less common mutation in SF3B1 allows cells to escape synthetic lethality through allele-specific splicing. Finally, I use the splicing patterns specific to SF3B1 we learned in the previous studies to identify introns that respond to the most common SF3B1 mutations across more than 12 different cancer types. I inserted these introns into a synthetic killer gene regulated by splicing and then systematically redesigned the introns for a high throughput screen to identify sequence variants that allow the introns to only be spliced in the presence of an SF3B1 mutation and be used in a therapeutic context.

TABLE OF CONTENTS

Chapter 1. Introduction 1

Chapter 2. Synthetic lethal and convergent biological effects of cancer assosiated spliceosomal mutations 4

2.1 Summary and Significance 5

2.2 Introduction 5

2.3 Results 7

2.4 Discussion 19

2.5 Figures and Legends 22

2.6 Supplemental Figures 39

2.7 Methods 55

2.8 Acknowledgements 66

2.9 Author contributions 67

2.10 Competing financial Interests 67

2.11 References 67

Chapter 3. Single-cell genomics reveals the genetic and molecular bases for escape from mutational epistasis in myeloid neoplasms 74

3.1 Abstract 75

3.2 Introduction 75

3.3 Results 77

3.4 Discussion 82

3.5 Figures and Figure Legends 86
3.6 Methods 99
3.7 Acknowledgements 103
3.8 Author contributions 104
3.9 Disclosure of conflicts of interest 104
3.10 References 104
Chapter 4. Synthetic introns enable mutation-dependent targeting of cancer cells 106
4.1 Abstract 107
4.2 Introduction 107
4.3 Results 108
4.4 Discussion 113
4.5 Figures and Legends 114
4.6 Methods 124
4.7 Acknowledgements 129
4.8 Competing Interests 129
4.9 References 129
Chapter 5. Discussion 131

LIST OF FIGURES

Figure 2:1 Simultaneous expression of cancer-associated mutations in Srsf2 and Sf3b1 or expression of a single RNA splicing factor mutation in the homozygous state is incompatible with hematopoiesis 23

Figure 2:2 Combined expression of mutations in *Srsf2* and *Sf3b1* results in hematopoietic stem and progenitor cell apoptosis and loss of quiescence 25

Figure 2:3: *Srsf2* and *Sf3b1* mutations have distinct and independent effects on gene dysregulation 27

Figure 2:4 Srsf2 and Sf3b1 mutations have distinct and independent effects on RNA splicing 29

Figure 2:5 Co-expression of Srsf2 and Sf3b1 mutations results in aberrant splicing and impaired expression of key regulators of hematopoietic stem cell survival and quiescence ... 31

Figure 2:6 SF3B1 mutations promote mis-splicing of MAP3K7, resulting in hyperactivation of NF-κB signaling. 33

Figure 2:7 MAP3K7 loss is responsible for hyperactivated NF-κB signaling in SF3B1-mutant cells. 35

Figure 2:8 SRSF2 mutations promote aberrant splicing of caspase-8 (CASP8), resulting in expression of a truncated protein that hyperactivates NF-κB signaling. 37

Figure 2:9 (Supplementary) Severe self-renewal disadvantage of hematopoietic stem cells from Srsf2P95H/+ Sf3b1K700E/+ double-mutant mice 40

Figure 2:10 (Supplementary) Co-expression of Srsf2 and Sf3b1 mutations results in death before birth in Vav-Cre Srsf2P95H/+ Sf3b1K700E/+ mice. 42

Figure 2:11 (Supplementary) *Srsf2* and *Sf3b1* mutations have distinct and independent effects on gene dysregulation 43

Figure 2:12 (Supplementary) *Srsf2* and *Sf3b1* mutations have distinct and independent effects on RNA splicing. 45

Figure 2:13 (Supplementary) Srsf2 and Sf3b1 mutations have distinct and independent effects on RNA splicing 48

Figure 2:14 (Supplementary) SF3B1 mutations promote mis-splicing of MAP3K7 in SF3B1-mutant cells 50
Figure 2:15 (Supplementary) MAP3K7 loss results in hyperactive NF-κB signaling in *SF3B1*-mutant cells. 51
Figure 2:16 (Supplemenray) *SRSF2* mutations promote aberrant splicing of caspase-8 (CASP8) resulting in expression of a stable truncated protein that hyperactivates NF-κB signaling 53
Figure 3:1 Genetic features of patients harboring two concomitant mutations in RNA splicing factors at the bulk and single cell level. 86
Figure 3:2 Allele- specific mutational co-occurrences in RNA splicing factor mutations88
Figure 3:3 Allele-specific effects on RNA binding and splicing in splicing factor mutations seen in patients harboring two concomitant mutations in splicing factors 91
Figure 3:4 (Supplementary) : Characteristics of study cohort and genetic alterations in splicing factors analyzed 92
Figure 3:5 (Supplementary) Single cell genomic analysis of myeloid neoplasm patients bearing two mutations in RNA splicing factors. 94
Figure 3:6 (Supplementary) Oncoprints of mutations in individual cells of dual splicing factor mutant myeloid neoplasm patients 95
Figure 3:7 (Supplementary) Clonal structure of RNA splicing factor mutations in double mutant patients before and after adjustments for allele dropout (ADO) 97
Figure 3:8 (Supplementary) Allele-specific differences in splicing based on SF3B1 mutational hotspot and *U2AF1* isoform usage in $U2AF1^{WT}$ versus $U2AF1^{S34F/Q157R}$ double mutant cells. 98
Figure 4:1 Figure 1. Synthetic introns can mimic *SF3B1* mutation-dependent mis-splicing in cancers. 114
Figure 4:2 Synthetic introns enable mutation-dependent cancer cell killing. 116
Figure 4:3 Massively parallel screening reveals critical elements governing synthetic intron function. 118
Figure 4:4 Synthetic introns enable mutation-dependent cancer cell targeting *in vivo*. . 120

Figure 4:5 (Supplementary) Validation of *SF3B1* mutation-dependent differential splicing for endogenous and synthetic introns. 121
Figure 4:6 (Supplementary) PCR validation of the pilot screen and alternate construct design 122
Figure 4:7 (Supplementary) Massively parallel screening reveals critical elements governing the function of very short synthetic introns 123

ACKNOWLEDGEMENTS

During my time at Fred Hutch and UW Genome Sciences, I was fortunate to be mentored and supported by many people, without whom this work would not have been possible.

First, I would like to thank Rob for his mentorship and for fostering a supportive and exciting lab environment. One of the most amazing things about my time in Rob's lab was that I was able to merge a training in basic biology with technology development and engineering. This wasn't easy an easy journey, but Rob has a way of seeing things in a very positive light. My favorite example of this was when we concluded that the flow cytometry plots, I had showing how to not make a genotype specific reporter were actually a positive result motivating the entire project. The synthetic intron project was a much more exciting project than I had ever dreamed of before starting my graduate work and I am very grateful for Rob's advice, excitement and support.

I also would like to thank all the members of the Bradley Lab during my time here. Heidi Dvinge helped me with learning RNA-seq analysis and introduced me to RShiny. Janine Ilagan explained many questions about splicing. Qing Feng, Heather Johns, and Sujatha Jagannathan taught me how to design vectors and clone. Joey Pangallo introduced me to tissue culture. Guo-Liang "Chewie" Chew, Dylan Udy, and Jose Pineda were the best "bay mates" and spent countless hours discussing scientific challenges on the whiteboards with me for when our bay rubber duck was not enough. Emma Hoppe lent an ear and some advice when the cloning got tough. Jake Polaski and James Thomas helped me learn high throughput library prep and took over as amazing post-doc mentors. Emma De Neef acknowledged that sometimes the best troubleshooting happens with some good endorphins while running. Austin Gabel, Andrea Belleville, Siegen McKellar, and Taylor Nicholas joined recently are continuing the friendly environment and exciting science.

Next, I want to thank my committee: Alice Berger, Stan Fields, and Alan Weiner. It was such a pleasure and honor to have Alice, Stan, and Alan meet with me and help guide my scientific work. I really appreciate the time they took to listen to my progress reports, ask me challenging questions, and offer great advice. Additionally, I want to thank Brian Giebel, the UW Genome Sciences Program, the Fred Hutch Computational Biology program, and my classmates.

Finally, I would like to thank my family, my husband Nate, and, of course, our pup Camper for the support and encouragement during this journey.

Chapter 1. INTRODUCTION

In 2011, Yoshida et al. identified mutations in RNA splicing factors as some of the most common mutations in patients with hematopoietic malignancies.(1) Specifically, four splicing factors were identified as the most commonly mutated: SRSF2, SF3B1, U2AF1, and ZRSR2. By 2016, about the time when we began the work in this dissertation, it was known that the mutation in these splicing factors were always heterozygous point mutations, never co-occurred, and resulted in a modest disruption in splicing. Furthermore, a growing number of solid tumors were identified with these mutations, increasing the importance of understanding how these mutations contribute to tumor development and the significance of developing a potential treatment. (2)
In hematopoietic malignancies such as MDS, the most frequently mutated splicing factors are SF3B1 and SRSF2. Both genes tend to have recurrent, heterozygous point mutations resulting in gain-of-function or change-of-function phenotypes. Mutations in SF3B1 tend to have altered 3's s recognition. (3) Mutations in SRSF2 disrupt exon inclusion by decreasing its affinity to the "GGNG" motif. (4) Despite leading to very different splicing consequences, the mutants result in the development of similar diseases.

Unlike SRSF2, SF3B1 mis-splicing is much more complicated and much less understood. SF3B1 is mutated across the most diverse set of tumors, including uveal melanoma (UVM), breast cancer (BRCA), urothelial bladder carcinoma (BLCA), skin cutaneous melanoma (SKCM), and pancreatic adenocarcinoma (PAAD), in addition to hematopoietic malignancies. (5) The different diseases tend to have different SF3B1 hotspot mutations implying that there may be different mutant-specific consequences.

In this dissertation I address three questions:

1. **What is the mechanism behind SF3B1 and SRSF2 mutual exclusivity?**

Chapter 2 focuses on our study showing that SRSF2P95H and SF3B1K700E, the most common hotspot mutations in each protein, are synthetically lethal when expressed in the same cell using isogenic mouse models. The study highlights that co-expression of SF3B1K700E and SRSF2P95H

mutations is not compatible with hematopoiesis but surprisingly does not result in a widespread disruption in splicing when expressed simultaneously. Based on RNA-seq analysis, both mutants have entirely distinct consequences on gene expression and splicing, while co-expression results in more additive dysregulation.

The study also shows that mutations in SF3B1 and SRSF2 converge on hyperactive NF-kB signaling through distinct alternative splicing events, elucidating how they contribute to similar diseases. SF3B1K700E mutants alter the splicing of MAP3K7 with alternative 3' splice site usage, which leads to a premature termination codon and nonsense-mediated decay. SRSF2P95H mutants result in the increased expression of an alternative CASP8 isoform by changing the inclusion of two mutually exclusive exons.

2. **How can some SF3B1 mutants co-exist with SRSF2 mutations at higher frequencies?**

With increased patient tumor sequencing and increased interest in splicing factor mutations, it became clearer that some splicing factor mutations can escape epistasis and co-occur more often than others. Chapter 3 confirms, using bulk single-cell sequencing analysis of patients with a myeloid malignancy and more than one splicing factor mutation, SF3B1K700E and SRSF2P95H mutations do not co-exist in the same cell while a less common SF3B1K666 mutation can. Further RNA-seq analysis of patients with either an SF3B1 K666 mutation or an SF3B1 K700E mutation shows that the minor mutant has a reduced effect on RNA splicing.

3. **What specific sequence characteristics make genes vulnerable to mis-slicing in the presence of SF3B1 mutations, and can we use this vulnerability to design a treatment?**

SF3B1 mutations are present among the most diverse set of cancer types. This makes these mutations an attractive therapeutic target. Chapter 4 describes a method to design synthetic introns that enable mutation-dependent targeting of cancer cells. Specifically, the study focuses on developing synthetic introns that are only spliced out by cells with an SF3B1 mutation and retained in cells with a wild-type SF3B1. The study shows that the most common SF3B1 mutations across numerous cancer types result in the mis-splicing of the same select introns. We use these introns as a base for designing a synthetic killer gene with HSV-TK. We then show that we can optimize the synthetic intron's splicing to be more mutant specific by modifying varying intronic components and performing a high throughput massively-parallel screen of the variants. Finally,

we show that when this construct is expressed in vivo in leukemia xenografts, the tumor sizes decrease significantly after treatment.

REFERENCES

1. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011 Sep 11;478(7367):64-9.
2. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer. 2016;16(7):413-430.
3. Alsafadi S, et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat Commun. 2016 Feb 4;7:10615.
4. Kim E, Ilagan J, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell. 2015 May 11;27(5):617-30.
5. Seiler M, Peng S, Agrawal AA, et al. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. Cell Rep. 2018;23(1):282-296.e4.

Chapter 2. SYNTHETIC LETHAL AND CONVERGENT BIOLOGICAL EFFECTS OF CANCER ASSOSIATED SPLICEOSOMAL MUTATIONS

This research was originally published in *Cancer Cell*. Lee SC*, North K*, Kim E*, Jang E, Obeng E, Lu SX, Liu B, Inoue D, Yoshimi A, Ki M, Yeo M, Zhang XJ, Kim MK, Cho H, Chung YR, Taylor J, Durham BH, Kim YJ, Pastore A, Monette S, Palacino J, Seiler M, Buonamici S, Smith PG, Ebert BL, Bradley RK, Abdel-Wahab O. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. Cancer Cell. 2018 Aug 13;34(2):225-241.e8
*co-first authors

Contributions by Khrystyna North:

In this work I mapped, quantified and analyzed RNA-seq data from the mouse and human data shown.

Specific analyses contributed are displayed in the following figures:
Figure 2:3 B, C, D, E, F: I analyzed differential gene expression and pathway dysregulation in single mutant vs. double mutant mouse data.
Figure 2:4 A, B, C, D, E, F: I analyzed differential splicing and motif usage in single mutant vs. double mutant mouse data.
Figure 2:5 A and C: I analyzed the overlap of differentially spliced genes and differentially expressed genes in mouse data identifying the genes tested in this figure. I also performed the gene ontology enrichment analysis.
Figure 2:6 A and B: I analyzed the overlap of human and mouse differentially expressed genes and the conservation in the MAP3K7 intron shown.
Figure 2:8 A: I analyzed CASP8 alternative splicing in patient data.
Figure 2:11 A, B, C: I compared the expected vs. observed gene dysregulation in mouse data.
Figure 2:12 A, B, C, D, E, F, G, H, I: I analyzed differential splicing and motif usage in single mutant vs. double mutant mouse data in Mx1-Cre and Vav-Cre mice.
Figure 2:13 A: I identified the three miss spliced events.
Figure 2:16 A and B: I analyzed the alternative splicing of MAP3K7 and CASP8 in SRSF2 and SF3B1 mutant patient data and cell lines, respectively.

2.1 SUMMARY AND SIGNIFICANCE

Mutations affecting RNA splicing factors are the most common genetic alterations in myelodysplastic syndrome (MDS) patients and occur in a mutually exclusive manner. The basis for the mutual exclusivity of these mutations and how they contribute to MDS is not well understood. Here we report that although different spliceosome gene mutations impart distinct effects on splicing, they are negatively selected for when co-expressed due to aberrant splicing and downregulation of regulators of hematopoietic stem cell survival and quiescence. In addition to this synthetic lethal interaction, mutations in the splicing factors SF3B1 and SRSF2 share convergent effects on aberrant splicing of mRNAs which promote NF-κB signaling. These data identify shared consequences of splicing factor mutations and the basis for their mutual exclusivity.

Mutual exclusivity of different mutations that affect a single pathway in cancer is commonly thought to indicate convergent effects of these mutations. RNA splicing factor mutations constitute the most common class of alterations in patients with myelodysplastic syndromes and are also frequent in several additional cancer types and occur as heterozygous mutations at restricted residues in a mutually exclusive manner. The mutual exclusivity of spliceosomal mutations suggests synthetic lethal and/or convergent biological effects of these mutations; however, there is currently no functional evidence supporting either of these possibilities. Here we report that individual spliceosomal mutations have non-overlapping effects on splicing and are mutually exclusive due to both synthetic lethal interactions and convergent effects on hyperactivation of innate immune signaling.

2.2 INTRODUCTION

Recently, recurrent somatic mutations in genes encoding components of the spliceosome have been identified through large-scale DNA sequencing across multiple cancer types. Spliceosomal gene mutations most commonly occur in myeloid neoplasms including myelodysplastic syndromes (MDS), chronic myelomonocytic leukemia (CMML), and acute myeloid leukemia (AML) (Graubert et al., 2012; Papaemmanuil et al., 2011; Yoshida et al., 2011), chronic lymphocytic leukemia (CLL) (Quesada et al., 2012; Wang et al., 2011), and uveal melanoma

(Harbour et al., 2013; Martin et al., 2013). *SF3B1, U2AF1,* and *SRSF2*, encoding RNA splicing factors that promote recognition of 3' splice sites and exons, are the most commonly mutated genes. Mutations in each of these genes are consistently heterozygous and occur as point mutations at highly restricted residues, suggesting that these are oncogenic gain-of-function alterations. In addition, mutations in splicing factor genes are mutually exclusive with one another, presumably due to redundant oncogenic effects and/or a limit on cellular tolerance of disrupted spliceosome function.

Despite recent insights into the individual effects of each of these mutations on pre-mRNA splicing, gene expression, and hematopoiesis, the mechanistic basis for the mutual exclusivity of these mutations and any functionally convergent effects they may have are currently unknown. For example, recent work has identified that mutations affecting the core RNA splicing factor SF3B1 are associated with cryptic 3’ splice site (ss) selection and altered branchpoint recognition (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015; Obeng et al., 2016; Wang et al., 2016). In contrast, mutations affecting SRSF2, an auxiliary splicing factor that binds to exonic splicing enhancers to promote splicing, alter its RNA binding preferences in a sequence-specific manner and thereby alter the efficiency of exon inclusion (Kim et al., 2015; Zhang et al., 2015). Finally, mutations affecting U2AF1, the small subunit of the U2AF heterodimer that binds the AG dinucleotide at the 3' ss, promote or repress 3' ss based on sequences flanking the AG dinucleotide (Fei et al., 2016; Ilagan et al., 2014; Okeyo-Owuor et al., 2015; Shirai et al., 2015). Given that the reported effects of *SF3B1*, *U2AF1*, and *SRSF2* mutations on splicing mechanisms are distinct, it is unclear why mutations affecting these three factors are mutually exclusive with one another. Moreover, a direct comparison of the effects of these individual mutations on RNA splicing and gene expression in an isogenic manner has not been performed.

Here we directly evaluated the functional basis for the mutually exclusive and heterozygous occurrence of mutations affecting RNA splicing factors using murine models expressing single and/or multiple splicing factor mutations simultaneously. Evaluation of mice with inducible expression of mutations in *Srsf2* or *Sf3b1*, alone or together, and in a heterozygous, homozygous, or hemizygous state, definitively demonstrated that each mutation affects RNA splicing in a distinct manner and are only tolerated when expressed alone in a heterozygous state.

Mutations in *SF3B1* and *SRSF2* were not tolerated when co-expressed in the same cell due to aberrant splicing and impaired expression of key regulators of hematopoietic stem cell survival and quiescence. Remarkably, although mutant SRSF2 and SF3B1 alter RNA splicing through different mechanisms, hematopoietic cells expressing either mutant splicing factor resulted in hyperactivated NF-κB signaling via aberrant splicing of distinct signaling intermediates. These data provide a functional explanation for the genetic configuration of mutations affecting RNA splicing factors and identify a convergent biological effect of these mutations in a pathway of established relevance to MDS pathogenesis.

2.3 RESULTS

2.3.1 *Simultaneous expression of cancer associated mutations in Srsf2 and Sf3b1 is incompatible with hematopoiesis*

Evaluation of sequencing data from >4,000 patients with MDS and related myeloid neoplasms (Bejar et al., 2012; Damm et al., 2012; Haferlach et al., 2014; Lasho et al., 2012; Makishima et al., 2012; Meggendorfer et al., 2012; Papaemmanuil et al., 2013; Patnaik et al., 2013; Thol et al., 2012; Yoshida et al., 2011; Zhang et al., 2012) revealed that while ~48% of patients (1,935/4,032) have a mutation in an RNA splicing factor, only ~2% of MDS patients (86/4,032) have mutations in >1 splicing factor (**Figure 2:1A**). To understand the functional basis of this apparent mutual exclusivity, we generated mice to permit inducible heterozygous expression of two of the most common mutations in RNA splicing factors (the *SF3B1*K700E and *SRSF2*P95H mutations) simultaneously (*Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$) (**Figure 2:1B**). We performed noncompetitive bone marrow transplantation (BMT), wherein each mutation was induced, either alone or together, following stable engraftment in recipient mice (**Figure 2:1B**). Bone marrow (BM) cells from mice co-expressing *Srsf2*^{P95H} and *Sf3b1*^{K700E} mutations had severe defects in multi-lineage reconstitution and severe leukopenia compared to other groups (**Figure 2:1C-D** and **Figure 2:9A-D)**; chimerism analysis and evaluation for recombination prior to administration of polyinosinic-polycytidylic acid (pIpC) confirmed minimal spontaneous excision occurred prior to transplantation and that pre-transplant chimerism was indistinguishable between groups (**Figure 2:9E-F**)). Similarly, in competitive BMT assays (**Figure 2:9E**), simultaneous expression of *Sf3b1*^{K700E} and *Srsf2*^{P95H} mutations resulted in severe defects in the repopulating capacity of hematopoietic stem and

progenitor cells (HSPCs), which were more readily outcompeted by CD45.1 competitor bone marrow cells relative to wild-type (WT) and single-mutant controls. Analyses of hematopoietic organs six months post-BMT revealed a near-complete absence of *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ double-mutant cells, which was distinct from expression of the *Srsf2*^{P95H} or *Sf3b1*^{K700E} mutation alone (**Figure 2:1E-F** and **Figure 2:9C**). These data provide functional evidence that co-expression of these change-of-function mutations affecting RNA splicing factors are not tolerated in the same cell.

2.3.2 *Expression of cancer-associated RNA splicing factor mutations in a homozygous or hemizygous state is incompatible with hematopoiesis*

Prior data from epithelial and hematopoietic cells identified that cells bearing mutant splicing factors require the WT allele for survival (Fei et al., 2016; Lee et al., 2016; Zhou et al., 2015). While these observations potentially explain the heterozygous nature of RNA splicing factor mutations in patients, the copy number of the RNA splicing factor was not maintained in these experiments. To test whether these mutations can exist in a homozygous state, we generated mice with conditional homozygous expression of the *SRSF2*^{P95H} mutation (*Mx1*-Cre *Srsf2*$^{P95H/P95H}$). In competitive BMT assays, HSPCs from these mice showed severe defects in self-renewal as well as multi-lineage reconstitution relative to heterozygous mutant (*Mx1*-Cre *Srsf2*$^{P95H/+}$) or control (*Mx1*-Cre *Srsf2*$^{+/+}$) HSPCs, analogous to the defects seen with hemizygous *Srsf2*^{P95H} expression *(Mx1*-Cre *Srsf2*$^{P95H/fl}$) (**Figure 2:1G-H** and **Figure 2:10A-C**). These data firmly establish that neither hemizygous nor homozygous expression of splicing factor mutations is tolerated by hematopoietic cells, further highlighting the unique dependency of spliceosome-mutant cells on residual function of WT spliceosome components.

2.3.3 *Simultaneous expression of mutations in Srsf2 and Sf3b1 resulted in increased apoptosis and reduced quiescence of HSPCs*

In order to identify the cellular basis for why simultaneous expression of *Srsf2*^{P95H} and *Sf3b1*^{K700E} resulted in severe disadvantage of HSPC function, we next examined the effects of expression of both mutations together versus single-mutant and WT controls on cell cycle and cell death. Two-weeks following induction of recombination *in vivo*, there was a significant increase in the proportion of lineage-negative Sca-1^{+} c-Kit^{+} (LSK) cells undergoing apoptosis (as indicated by

Annexin V positivity) and in the S-phase of cell cycle in double-mutant mice (*Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$) relative to all controls **(**Figure 2:2**A-B)**. These data suggest that co-expression of *Srsf2*^{P95H} and *Sf3b1*^{K700E} mutations severely impairs hematopoiesis due to a simultaneous loss of HSPC quiescence as well as increased apoptosis.

Selection against co-expression of two splicing factor mutations was also strongly supported by the fact that mice with hematopoietic-specific expression of both *Srsf2*^{P95H} and *Sf3b1*^{K700E} (*Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$) mutations were associated with 100% lethality at weaning, in contrast to control mice with expression of either mutation alone, which were present at expected Mendelian ratios (**Figure 2:2C**). Timed sacrifice of mid-gestation female mice and genotyping of 139 embryos revealed that *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ fetuses were detectable at the expected frequency at embryonic day 14.5 (E14.5) (**Figure 2:11A-B**). Analysis of HSPCs from E14.5 embryos revealed that *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ fetal livers were marked by increased frequency and number of LSK cells along with phenotypically-defined long-term hematopoietic stem cell (LT-HSC, LSK CD150^{+} CD48^{-}), multi-potent progenitor (MPP, LSK CD150^{-} CD48^{-}), HPC-1 (LSK CD150^{-} CD48^{+}) and HPC-2 (LSK CD150^{+} CD48^{+}) populations compared to all other genotypes (**Figure 2:2 D-E** and **Figure 2:11C-G**). Despite a significant increase in the number of phenotypically defined HSPCs, E14.5 fetal liver cells from *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ embryos demonstrated significantly impaired ability to form progeny *in vitro* (**Figure 2:2F**) and were marked by increased proportion of cells in the active phase of cell cycle as well as undergoing apoptosis (**Figures 2:2G-I**). These derangements in HSPC survival and clonogenic capacity were manifest in embryos by E18.5 where *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ fetuses were characterized by near absence of myeloid hematopoietic cells in the developing bone marrow space as well as increased proportions of apoptotic cells expressing cleaved Caspase-3 in the fetal liver (**Figure 2:11H-I**). Taken together, multiple complementary genetic systems here conclusively demonstrate that co-expression of MDS-associated RNA splicing factor mutations severely disrupts hematopoiesis, driven in part by aberrant cell cycle progression and increased apoptosis.

2.3.4 *Srsf2 and Sf3b1 mutations have largely distinct effects on gene expression*

To understand the mechanistic basis for mutual exclusivity of *SF3B1* and *SRSF2* mutations, we performed RNA-seq of lineage$^-$ c-Kit$^+$ (LK) cells from *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice, as well as the respective single-mutant and WT control cells, two weeks after conditionally expressing each mutation alone or together (**Figure 2:3A**). The mean allelic ratio of *Sf3b1*^{K700E} and *Srsf2*^{P95H} expressed in double-mutant mice was 20.7% and 33.5%, markedly lower than the near 50% expression seen in single-mutant controls, illustrating the intolerability of combined *SF3B1* and *SRSF2* mutations (**Figure 2:3B**). Expression of either mutation resulted in dysregulation of hundreds of coding genes (**Figure 2:3C**). Double-mutant (*Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$) cells exhibited significantly more marked gene dysregulation than did single-mutant cells, consistent with their more dramatic biological phenotype (**Figure 2:3D**). Double-mutant cells shared a greater proportion of differentially expressed genes with each single-mutant genotype than single-mutants shared with one another (**Figure 2:3E**). Gene Ontology (GO) enrichment analysis of coding genes that were differentially expressed within each genotype relative to WT control revealed strong signatures of processes associated with impaired hematopoiesis in double-mutant cells, consistent with the multi-lineage defects in hematopoietic reconstitution evident in these cells noted above (**Figure 2:3F**).

We next tested whether mutations in *SF3B1* and *SRSF2* have independent effects on gene expression, in the sense that double-mutant cells recapitulated the gene dysregulation observed for each single-mutant genotype relative to WT cells. Approximately 80% and 40% of genes that were dysregulated in *Sf3b1*$^{K700E/+}$ and *Srsf2*$^{P95H/+}$ single-mutant cells were also dysregulated in *Sf3b1*$^{K700E/+}$ *Srsf2*$^{P95H/+}$ double-mutant cells (**Figure 2:3E**). That degree of recapitulation of gene dysregulation was highly similar to and statistically indistinguishable ($p = 0.90$ and 0.44 for *Sf3b1* and *Srsf2* mutations, respectively) from the expected recapitulation under the assumption of independence (**Figure 2:12A**). We therefore conclude that *SF3B1* and *SRSF2* mutations have independent effects on gene expression, even when present in the same cell. Interestingly, in addition to recapitulating the dysregulation of specific genes expected based on single-mutant cells, double-mutant cells exhibited additional gene dysregulation (**Figures 2:12B-C**), consistent with the dramatic hematopoietic phenotype of these cells. These results suggest that *SF3B1* and

SRSF2 mutations have independent but compound effects on gene expression when present in the same cell.

2.3.5 *Srsf2 and Sf3b1 mutations have distinct and independent effects on RNA splicing*

We next assessed the consequences of single versus double mutations in *SF3B1* and *SRSF2* on RNA splicing. Mutations in *SF3B1* have been proposed to preferentially alter 3' ss recognition (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015). In contrast, mutations in *SRSF2* alter exon recognition via preferential recognition of C-rich exonic splicing enhancer (ESE) motifs relative to G-rich ESEs, while WT SRSF2 recognizes both classes of ESEs (Kim et al., 2015; Zhang et al., 2015). Our current understanding of *SF3B1* and *SRSF2* mutations therefore indicates that they induce distinct changes in RNA splicing. However, the consequences of these mutations for splicing have not yet been compared in an isogenic context.

We used our double-mutant system to directly compare the consequences of *SF3B1* and *SRSF2* mutations on RNA splicing mechanisms. Consistent with prior studies (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015; Kim et al., 2015; Zhang et al., 2015), *SF3B1* and *SRSF2* mutations preferentially affected 3' ss and cassette exon recognition, respectively (**Figure 2:4A-B**). These preferential effects of *SF3B1* and *SRSF2* mutations on 3' ss recognition and cassette exon recognition were statistically significant (**Figure 2:12D-E**). We observed a modest enrichment of adenosines upstream of intron-proximal 3' ss promoted by the *Sf3b1*K700E mutation, independent of the presence or absence of *Srsf2*P95H. This enrichment was absent from cells expressing *Srsf2*P95H alone (**Figure 2:12F**). Cassette exons that were promoted versus repressed in cells expressing *Srsf2*P95H were respectively enriched for the CCNG and GGNG ESEs, independent of the presence or absence of *Sf3b1*K700E, but were absent from cells expressing *Sf3b1*K700E alone (**Figure 2:4C** and **Figure 2:12G**). These sequence analyses are consistent with previous reports that *SF3B1* and *SRSF2* mutations affect branchpoint recognition and ESE preference, respectively, and furthermore demonstrate that these changes in splicing mechanisms are specific to *SF3B1* and *SRSF2* mutations.

As *SF3B1* and *SRSF2* mutations cause distinct changes in splicing mechanisms, even when simultaneously present in a cell, we hypothesized that the effects of *SF3B1* and *SRSF2* mutations on splicing were likely independent. Consistent with this hypothesis, double-mutant cells recapitulated the preferential effects of *SF3B1* and *SRSF2* mutations on 3' ss recognition and cassette exon recognition, respectively, that we observed in single-mutant cells (**Figure 2:12D**). We did not observe signs of widespread decreases in splicing efficiency, such as increased retention of constitutive introns, in double-mutant cells (**Figure 2:12H**). Instead, *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells exhibited modest increases in the number of mis-spliced genes relative to cells bearing single mutations in *Sf3b1* or *Srsf2* (**Figures 2:4A-B**). *Sf3b1*$^{K700E/+}$ *Srsf2*$^{P95H/+}$ cells recapitulated approximately 40% and 32% of splicing dysregulation driven by expression of the single mutations *Sf3b1*K700E and *Srsf2*P95H (**Figure 2:13A**). We identified only seven genes that were mis-spliced in both single-mutant as well as double-mutant cells, and only six genes that were mis-spliced in both single-mutant but not double-mutant cells. Such genes might represent cases where mutant SF3B1 and SRSF2 have opposing or convergent consequences for splicing. However, these genes were ten-fold less abundant than genes that were shared between either single-mutant and the double-mutant cells. We therefore conclude that *SF3B1* and *SRSF2* mutations typically have distinct and independent consequences for RNA splicing.

In order to confirm the effects of co-expressing *Srsf2* P95H and *Sf3b1* K700E mutations on splicing and gene expression using an alternate system, we next performed RNA-seq analysis on purified LK cells from E14.5 *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ double-mutant fetal liver and all controls. At this timepoint, the allelic frequencies of both mutations in double-mutant cells were near 50% across all replicates and comparable to that seen in *Vav*-Cre single-mutant controls (**Figure 2:4D**). Similar to the effects seen in the *Mx1*-Cre system, the greatest alterations in gene expression relative to *Vav*-Cre WT controls was seen in *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ HSPCs compared to *Vav*-Cre single-mutant controls, and double-mutant cells recapitulated much of the gene expression dysregulation observed in single-mutant cells (**Figures 2:4E-F**). GO enrichment analysis again revealed strong signatures of impaired hematopoiesis in double-mutant cells, consistent with those seen in *Mx1*-Cre mice and in the functional analyses of HSPCs from *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ embryos (**Figure 2:13B**). Moreover, similar effects on splicing were seen in HSPCs from *Vav*-Cre mice as were seen in the *Mx1*-Cre system, including preferential

effects of *SF3B1* and *SRSF2* mutations on 3' ss recognition and cassette exon recognition (**Figure 2:13C-D**).

2.3.6 *Combined expression of Srsf2 and Sf3b1 mutations impairs expression of regulators of HSPC survival and increases sensitivity to inflammatory activation*

We next sought to understand the mechanistic basis for the impaired HSPC survival, increased apoptosis, and increased cell cycling seen when *Srsf2*$^{P95H/+}$ and *Sf3b1*$^{K700E/+}$ mutations were co-expressed. Given the largely non-overlapping effects of *Srsf2*$^{P95H/+}$ and *Sf3b1*$^{K700E/+}$ on gene expression and splicing, we systematically identified and evaluated transcripts that exhibited concomitant dysregulation of both gene expression and splicing in the *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ double-mutant state (**Figure 2:5A**). This analysis revealed that multiple well-known regulators of HSPC survival and quiescence were subject to severe alterations in splicing as well as significantly reduced expression in *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ double-mutant HSPCs. This included aberrant splicing and decreased expression of the thrombopoietin receptor *Mpl,* integrin αIIb (*Itga2b*, CD41, GPIIb), and the transcription factor *Pbx1* (**Figure 2:5A** and **2:13E**). After confirming reduced expression of each of these mRNAs in LK cells in an independent cohort of *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice (**Figure 2:5B**), we evaluated the functional consequences of shRNA-mediated depletion of these mRNAs in WT HSPCs after confirmation of efficient knockdown of each of these shRNAs (**Figure 2:13F-G**). Consistent with prior published data on the effects of loss of these factors on HSPC function (Alexander et al., 1996; Ficara et al., 2008; Gekas and Graf, 2013; Qian et al., 2007; Yoshihara et al., 2007), silencing any of these individual genes was associated with severe impairment of HSPC clonogenicity (**Figure 2:13H**). Overall, these data reveal that the inability of the hematopoietic system to tolerate combined expression of mutations in *SRSF2* and *SF3B1* is not due to widespread failure of splicing catalysis, but rather due to impaired splicing and expression of key regulators of HSPC survival and quiescence.

In addition to gene expression changes consistent with impaired hematopoietic processes in *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ double-mutant cells (**Figure 2:3F**), GO analysis also revealed a strong signature associated with immune signaling that was augmented in double- relative to single-mutant cells (**Figure 2:5C** and **Figure 2:13B**). To functionally evaluate these gene expression

changes we studied the response of *Srsf2* and *Sf3b1* mutant cells to inflammatory stimulus from lipopolysaccharide (LPS). HSPCs from *Vav*-Cre spliceosomal mutant mice exhibited increased NF-κ activation (marked by increased nuclear accumulation of phosphorylated p65 (p-p65)) in *Srsf2* or *Sf3b1* single-mutant relative to WT cells, which was further enhanced in double- relative to single-mutant cells at baseline and following *ex vivo* LPS administration (**Figure 2:5D-E** and **Figure 2:13I-J**).

Given prior data identifying that chronic exposure to inflammation impairs quiescence and repopulation potential of HSCs (Esplin et al., 2011; Zhao et al., 2013), we next evaluated the effects of chronic inflammatory insult on the repopulating potential of *Srsf2* or *Sf3b1* mutant BM HSPCs (**Figure 2:5F**). *Vav*-Cre WT, *Vav*-Cre *Srsf2*$^{P95H/+}$ and *Vav*-Cre *Sf3b1*$^{K700E/+}$ mice were treated with LPS (1 mg/kg) via intra-peritoneal injection every second day for 30 days (Esplin et al., 2011) followed by serial BMT into lethally irradiated syngeneic recipient mice. Chronic LPS treatment had a mild effect on the repopulating potential of WT HSPCs evidenced by similar peripheral blood and BM HSPC chimerism in primary transplant recipients **(Figure 2:5G-H)**. In contrast, LPS-treated *Srsf2*$^{P95H/+}$ and *Sf3b1*$^{K700E/+}$ BM HSPCs showed significant reduction in their repopulating ability relative to the respective vehicle-treated counterparts, and this defect was further exacerbated following serial transplantation into secondary recipient mice **(Figure 2:5G-H)**.

The above observations suggest that *Srsf2*- and *Sf3b1*-mutant HSPCs exhibit intrinsic hyperactivation to inflammatory signaling which contributes to defective hematopoietic function similar to features found in human MDS patients. In addition, chronic LPS administration resulted in splenomegaly in both *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ mice while BM cellularity was significantly reduced in *Srsf2*$^{+/+}$ mice (**Figure 2:14A**). *Vav*-Cre *Srsf2*$^{P95H/+}$ mice treated with LPS showed enhanced myeloid skewing with reduced B-lymphopoiesis relative to either vehicle treated mice or *Vav*-Cre *Srsf2*$^{+/+}$ mice similarly treated with LPS (**Figure 2:14A**). These data reveal that chronic inflammatory stimulation promotes myeloid disease phenotypes in *Vav*-Cre *Srsf2*$^{P95H/+}$ mice *in vivo*. Of note, acute LPS stimulation *ex vivo* of purified BM LSK cells from *Vav*-Cre *Srsf2*$^{P95H/+}$ mice also resulted in significant increase in nuclear translocation of p-65 relative to *Vav*-Cre *Srsf2*$^{+/+}$ LSK cells (**Figure 2:14B-C**).

The enhanced response to inflammatory stimuli across *Srsf2* and *Sf3b1* mutant mice was evident *in vivo* upon exposure to LPS-induced sepsis. Intraperitoneal injection of primary *Mx1*-Cre WT, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ mice with LPS (15 mg/kg) 8 weeks following induction of mutant allele expression resulted in accelerated death in both *Srsf2*- and *Sf3b1*-mutant mice relative to WT controls (**Figure 2:5I**). Overall, these data identify that spliceosomal mutant HSPCs exhibit activated immune signaling linked with impaired quiescence and self-renewal. Combined expression of these mutations is not tolerated due to further exacerbation of activated immune signaling in addition to impaired splicing of critical mediators of quiescence and self-renewal in the double-mutant state.

2.3.7 *SF3B1 mutations promote mis-splicing of MAP3K7, resulting in hyperactivation of NF-κB signaling*

The above results identify that spliceosomal gene mutations are intolerable when co-expressed but that they also share convergent effects on activation of NF-κB signaling. As innate immune signaling has been implicated in MDS pathogenesis (Basiorka et al., 2016; Fang et al., 2014; Fang et al., 2017; Varney et al., 2015; Wei et al., 2013), we next attempted to understand precisely how *SF3B1* and *SRSF2* mutations activate convergent immune signaling pathways. We first focused on events that were mis-spliced within *SF3B1*-mutant hematopoietic cells in both human and mouse. Although recent studies reported very few such shared mis-spliced transcripts (Mupo et al., 2016; Obeng et al., 2016), we identified a larger set of 205 mRNAs that were mis-spliced in both *Sf3b1*-mutant murine hematopoietic cells and *SF3B1*-mutant MDS patient samples (**Figure 2:6A**). One of the most dramatically mis-spliced mRNAs in MDS patients with *SF3B1* mutations was *MAP3K7* (also known as TAK1; TGFβ-activated kinase 1), for which an intron-proximal 3' ss was promoted by mutant SF3B1 in both human and mouse hematopoietic cells. This same aberrant isoform of *MAP3K7* has been previously noted in *SF3B1*-mutant CLL (DeBoever et al., 2015) and uveal melanoma (Darman et al., 2015) patient samples. Our re-analysis of RNA-seq data from these cohorts and isogenic human leukemia cell lines with endogenous mutations in *SF3B1*, as well as directed RT-PCR and Sanger sequencing of cDNA from *SF3B1*-mutant cells, validated this aberrant isoform (**Figures 2:6B-C** and **Figure 2:14D**). This aberrant 3' ss occurred in exon 5 of *MAP3K7*, which encodes part of the kinase domain, and is predicted to result in an

out-of-frame transcript that undergoes nonsense-mediated decay (NMD; **Figure 2:14E**). Consistent with induction of NMD (which was experimentally confirmed through mRNA half life measurements in isogenic cells with or without endogenous *SF3B1* K700E mutation; **Figure 2:14F**), we observed reduced MAP3K7 protein levels in isogenic cell lines expressing *SF3B1*K700E (**Figure 2:6D-E**) as well as primary patient BM MNCs from MDS patients (**Figure 2:6F**), and peripheral blood MNCs from CLL patients (**Figure 2:6G**) with *SF3B1* mutations versus patients lacking any splicing factor mutations.

MAP3K7 encodes a kinase that mediates TNFα, IL-1β, LPS, TRAIL, and toll-like receptor (TLR) signaling through the NF-κB, JNK, and p38 MAPK pathways. The biological effects of MAP3K7 loss have been extensively studied across a variety of immune cell types and can result in either loss or promotion of inflammatory signaling depending on the cellular context (Ajibade et al., 2012; Lamothe et al., 2012; Sato et al., 2005; Tang et al., 2008; Vink et al., 2013; Xin et al., 2017). Here, we observed that *SF3B1*K700E human myeloid (K562) or lymphoid (NALM-6) leukemia cells stimulated with LPS had enhanced NF-κ activation (marked by p-p65 and simultaneous IκB-α downregulation) compared with *SF3B1* WT controls (**Figure 2:6D-E** and **Figure 2:14G-H**). Increased p-p65 was also evident in baseline in these isogenic cells as well as in primary patient BM MNCs from *SF3B1*-mutant MDS and PB MNCs from *SF3B1*-mutant CLL relative to spliceosomal WT counterpart (**Figures 2:6F-G**). Enhanced NF-κB transcriptional activation and p-p65 levels in *SF3B1*K700E cells compared with *SF3B1* WT and silent mutant (*SF3B1* K700K) controls were also evident in the same cells by increased p-p65 nuclear translocation (**Figure 2:7A**) and NF-κB transcriptional activation using an NF-κB transcriptional reporter assay (DeDiego et al., 2014; Hernandez et al., 2015; Simoes et al., 2015) (**Figures 2:7B;** background leakiness of this reporter was excluded using a version of the reporter where NF-κB binding elements in the reporter were mutagenized (**Figure 2:14I**)). We next assessed whether hyperactive NF-κB signaling in *SF3B1*-mutant cells resulted in the induction of known NF-κB target genes that are associated with inflammatory responses. Quantitative RT-PCR (qRT-PCR) analysis revealed marked induction in NF-κB target genes IL-1β and TNF in response to LPS stimulation in *SF3B1*K700E NALM-6 cells relative to parental and *SF3B1*K700K control cells (**Figure 2:7C**).

Given that a number of aberrant splicing and gene expression events occur in *SF3B1*-mutant cells beyond aberrant splicing of *MAP3K7*, we next sought to understand the contribution of MAP3K7 loss to hyperactivated NF-κB signaling. Firstly, hyperactive NF-κB signaling was confirmed using shRNA-mediated downregulation of *MAP3K7* in NALM-6 and K562 parental cells at both the level of p-p65 signaling and NF-κB transcriptional activity (**Figure 2:7B** and **Figure 2:14J**). In addition, restored expression of *MAP3K7* cDNA in K562 *SF3B1*K700E cells resulted in a significant decrease in p65 phosphorylation following LPS exposure based on immunoblot analysis, NF-κB luciferase reporter, and immunofluorescence of nuclear p-p65 (**Figures 2:7D-F** and **Figure 2:14K**). At a biological level, however, restoration of *Map3k7* expression in *Sf3b1*$^{K700E/+}$ HSPCs resulted in only mild rescue of HSPC clonogenic capacity (**Figure 2:14L**). Overall, these data suggest that the effects of *SF3B1*K700E mutation on induction of NF-κB signaling are indeed, in part, mediated through aberrant splicing of *MAP3K7*.

2.3.8 *SRSF2 mutations promote aberrant splicing of caspase-8 (CASP8), resulting in expression of a truncated protein that hyperactivates NF-κB signaling*

Consistent with the distinct effects of *SF3B1* and *SRSF2* mutations on splicing, *MAP3K7* aberrant splicing was restricted to *SF3B1*-mutant cells and not seen in *SRSF2*-mutant cells (**Figure 2:14M**). We therefore searched for aberrant splicing events in *SRSF2*-mutant cells that might impact NF-κ signaling. One such event was aberrant splicing of *caspase-8* (*CASP8*), which was recurrently mis-spliced in AML and CMML patients with *SRSF2* mutations, but not in cells bearing *SF3B1* mutations (**Figure 2:8A** and **Figure 2:14N**). Caspase-8 is a cysteine protease that initiates death receptor-mediated apoptosis (Shu et al., 1997; Thome et al., 1997) in addition to playing a role in necroptosis (Kearney and Martin, 2017; Weinlich et al., 2017) and serving as a key activator of NF-κ during the early response to antigenic stimuli (Chaudhary et al., 2000; Hu et al., 2000; Shikama et al., 2003). *SRSF2* mutations repressed a cassette exon of *CASP8*, as was evident by RNA-seq, RT-PCR, and direct sequencing of cDNA from cell lines and primary patient samples (**Figures 2:8A-B** and **Figures 2:15A-C**). Exclusion of this cassette exon results in an mRNA encoding a truncated caspase-8 protein lacking the C-terminal catalytic domains, which was readily detectable in *SRSF2*-mutant cells by immunoblotting using an N-terminal anti-caspase-8 antibody, including a newly generated genetically-modified myeloid leukemia cell line (K562

cells) with knockin of the P95H mutation in the *SRSF2* endogenous locus, and primary patient samples bearing endogenous mutations in *SRSF2* (**Figure 2:8C-D**). *CASP8* normally encodes a 54/55 kDa protein of 480 amino acids containing two death-effector domains (DED) at the N-terminus and a C-terminal catalytic protease domain. Although several caspase-8 isoforms have been described which lack the C-terminal catalytic domain (Xu et al., 2009; Yuan et al., 2012), the specific isoform of *CASP8* detected in *SRSF2*-mutant cells is distinct from those previously described.

Previously described DED-only forms of caspase-8 have been suggested to serve either as competitive inhibitors or promoters of apoptosis (Xu et al., 2009; Yuan et al., 2012). Given this, we first tested the effect of expressing this *SRSF2*-mutant-specific truncated isoform (hereafter referred to as *CASP8 TR*) in an *SRSF2* WT cell line (K562 cells) followed by stimulation with TRAIL, a potent agonist of death-receptor mediated cell death. Overexpression of both *CASP8* full-length (*CASP8 FL*) and *CASP8 TR* isoforms resulted in the robust expression of caspase-8 proteins at the expected size (**Figure 2:8E** and **Figure 2:15D**) with no apparent effect on growth of K562 cells (**Figure 2:15E**). Moreover, both isoforms promoted TRAIL-mediated cell death in K562 cells, suggesting that the truncated isoform did not impart a clear differential effect of modulating cell death relative to the *CASP8 FL* isoform (**Figure 2:15F**). Given this, we next evaluated the effect of overexpressing *CASP8 TR* relative to *CASP8 FL* or an empty vector control on NF-κ transcriptional activity, signaling, and nuclear localization. In cells with endogenous *CASP8* expression, overexpression of the *CASP8 TR*, but not the *CASP8 FL* isoform, resulted in the robust induction of NF-κ activity upon increasing concentration of TRAIL exposure. This was demonstrated in both K562 and HAP1 (a near-haploid cell line derived from chronic myeloid leukemia cells KBM7 (Carette et al., 2011)) cells with overexpression of *CASP8 TR* or *CASP8 FL* (**Figure 2:8E-F** and **Figures 2:15G-H**). These data were also confirmed in K562 cells with or without endogenous *SRSF2*P95H mutation (**Figure 2:8G** and **Figure 2:16A-B**). To rule out the effect of the endogenous caspase-8 in K562 and HAP1 cells, we next assessed the effect of *CASP8 TR* on the cell death and NF-κB activity in *CASP8* knockout (*CASP8* KO) HAP1 cells. Consistent with the lack of the catalytic domain, HAP1 cells expressing *CASP8 TR* isoform were unable to undergo TRAIL-mediated cell death (**Figure 2:16C-D**); however, the *CASP8 TR* isoform was able to induce robust NF-κB signaling in the absence of WT caspase-8, evidenced by elevated NF-κB

transcriptional activity, increased p-p65 and a concomitant decrease in total IκB-α level (**Figure 2:8F** and **Figures 2:16F**).

2.4 DISCUSSION

Mutations affecting RNA splicing factors are the most common genetic alterations in patients with MDS (Graubert et al., 2012; Papaemmanuil et al., 2011; Yoshida et al., 2011) but the biological basis for their significant enrichment in this disease remains largely unexplained. Here we identify that two of the most commonly mutated RNA splicing factors in MDS converge on activation of innate immune signaling through aberrant splicing of mRNAs encoding distinct enzymes. These findings were identified using genetically engineered mouse models and cell lines to evaluate the effect of expressing each RNA splicing factor mutation in an isogenic context, and by comparing the effects of each mutation on gene expression and splicing across mouse and human cells.

Although the initial description of spliceosomal gene mutations predicted that the mutually exclusive pattern of these mutations might be due to a common impact of these mutations on MDS pathogenesis (Yoshida et al., 2011), the data presented here identify that these mutations are mutually exclusive, in part due to synthetic lethal interaction. This provides one of the few examples of mutually exclusive oncogenic alterations due to synthetic lethal interactions in cancer. Interestingly, co-expression of *SF3B1*K700E and *SRSF2*P95H mutations resulted in impaired HSPC self-renewal, differentiation and survival, not because of widespread inhibition of splicing, but due to aberrant splicing and impaired expression of key regulators of HSPC function, including the thrombopoietin receptor *Mpl,* the homeodomain transcription factor *Pbx1* and integrin αIIb. Hematopoietic-specific deletion of these factors individually has previously been shown to result in failure of hematopoiesis due to reduced HSC self-renewal, increased apoptosis, and loss of quiescence (Ficara et al., 2008; Gekas and Graf, 2013; Qian et al., 2007; Yoshihara et al., 2007), all features which were characteristic of the *Srsf2*$^{P95H/+}$ *Sfb31*$^{K700E/+}$ double-mutant state. Of note, these specific aberrant splicing events in the double-mutant state were from mice and thereby may not address the precise basis of potential synthetic lethal effects of combined SRSF2/SF3B1 mutations in patients.

Multiple lines of evidence implicate innate immune signaling in MDS pathogenesis (Basiorka et al., 2016; Fang et al., 2014; Fang et al., 2017; Varney et al., 2015; Wei et al., 2013), including experiments demonstrating that increased innate immune signaling contributes to MDS development *in vivo* (Fang et al., 2017; Varney et al., 2015). In particular, increased activation of TLR and IL-1R signaling with downstream activation of the MAPK and NF-κB pathways is widely reported in MDS, but studies of the mechanistic basis for this activation have mostly been restricted to MDS with deletion of chromosome 5q (Fang et al., 2014; Fang et al., 2017; Starczynowski, 2014; Varney et al., 2015). Here we provide data identifying a mechanism for hyperactivated NF-κB signaling in a wider spectrum of MDS patients carrying mutations in *SF3B1* and *SRSF2*. Physiologic effects of hyperactivated innate immune signaling driven by mutant splicing factors was associated with both enhanced myeloid skewing of hematopoiesis upon chronic exposure to inflammatory stimuli as well as accelerated death due to septic shock from acute inflammatory stimuli. It will be interesting to understand if *EZH2* downregulation also contributes to aberrant NF-κB activation as reported in other cancer types recently (Lee et al., 2011; Yamagishi et al., 2012), given the *EZH2* mis-splicing driven by mutant SRSF2 as well as loss-of-functions mutations in *EZH2* in myeloid neoplasms (Kim et al., 2015).

Although recent studies of *SF3B1*K700E mutations have identified relatively few mRNAs that are mis-spliced in both mouse and human cells (Mupo et al., 2016; Obeng et al., 2016), we identified here a much greater overlap of aberrantly spliced transcripts shared across mouse and human hematopoietic cells with an *SF3B1*K700E mutation than has been previously reported. This includes aberrant splicing of *MAP3K7* through the use of an alternative 3' ss promoted by mutant SF3B1, an event that was conserved between mouse and human cells (Mupo et al., 2016; Obeng et al., 2016). As partial downregulation of *Map3k7* within myeloid cells *in vivo* has been previously shown to promote development of myeloid neoplasms (Ajibade et al., 2012; Eftychi et al., 2012; Lamothe et al., 2012; Vink et al., 2013), this finding is likely to be relevant to the role of mutant SF3B1 in MDS pathogenesis. For example, downregulation of *Map3k7* through shRNA knockdown *in vivo* resulted in splenomegaly, myeloproliferation, extramedullary hematopoiesis as well as increased immune activation that was exacerbated by LPS challenge (Vink et al., 2013). Similarly, myeloid-specific deletion of *Map3k7* results in heightened response to inflammatory stimuli and these mice succumb to a clonal myeloid leukemia (Ajibade et al., 2012; Eftychi et al.,

2012; Lamothe et al., 2012; Xin et al., 2017). Interestingly, the pro-inflammatory and leukemogenic effects of Map3k7 loss appear to be restricted to the myeloid lineages (Ajibade et al., 2012), which is in stark contrast to pan-hematopoietic-deletion of *Map3k7*, which results in complete failure of hematopoiesis (Tang et al., 2008). Exactly how *Map3k7* loss results in cell-type specific effects on signaling and cell survival is currently not defined.

While aberrant splicing of *MAP3K7* and consequent downregulation of MAP3K7 expression was specific to cells expressing mutant SF3B1, cells expressing mutant SRSF2 shared a similar transcriptional profile of elevated innate immune signaling and heightened sensitivity to LPS-induced inflammation as did *SF3B1*-mutant cells. This led us to identify a novel gain-of-function effect of *SRSF2* mutations through the generation of a novel *CASP8* isoform that promotes NF-κB signaling. This observation can be partly attributed to a specific splicing aberration in *CASP8*, whereby mutant SRSF2 promoted a cassette exon exclusion event resulting in the production of a previously unannotated transcript that stably expresses a truncated form of caspase-8 lacking the C-terminal effector domains. This is consistent with prior work showing that N-terminal prodomain-only containing isoforms of procaspase-8 activate NF-κB signaling through interactions with upstream regulators of NF-κB, a function not mediated by the canonical full-length caspase-8 (Chaudhary et al., 2000; Hu et al., 2000; Shikama et al., 2003). While the experiments here are focused on mutations in *SF3B1* and *SRSF2*, it is also noteworthy that mutations in *U2AF1* also drive *CASP8* mis-splicing (Ilagan et al., 2014).

Together, our data demonstrate that different splicing factor mutations alter an entirely distinct set of targets at the level of pre-mRNA splicing that nonetheless converge on the same downstream signaling node to drive hyperactive innate immune signaling, a phenomenon that has been previously implicated in MDS pathogenesis. In addition to identifying the functional basis for the mutual exclusivity and convergent consequences of splicing factor mutations, our study also highlights new avenues for potential therapeutic development. The universal requirement for the expression of the WT allele of a splicing factor in cells with an existing spliceosomal mutation motivated the use of compounds that inhibit basal splicing catalysis to specifically target mutant cells (Fei et al., 2016; Lee et al., 2016; Obeng et al., 2016; Shirai et al., 2017; Zhou et al., 2015). The identification that hematopoietic cells with expression of either of two mutated splicing factors

simultaneously or in the homozygous state, further reinforces the requirement of WT splicing catalysis in cells with heterozygous expression of a mutant RNA splicing factor.

In addition to therapeutic implications regarding targeting RNA splicing, it is important to note that while partial downregulation of *Map3k7* can drive the development of myeloid neoplasms in murine models, complete *Map3k7* loss across hematopoietic cells results in eradication of hematopoiesis (Tang et al., 2008). Thus, we speculate that the convergent effects downstream of splicing factor mutations on immune signaling may suggest new therapeutic opportunities through inhibition of residual MAP3K7 function in *SF3B1*-mutant cells or NF-κB signaling in *SF3B1*- or *SRSF2*-mutant cells. In particular, efforts to develop potent and specific small-molecule inhibitors of MAP3K7 may be especially important to study in the context of spliceosomal mutant cancers (Bosman et al., 2014; Kilty and Jones, 2015; McNew et al., 2016; Singh et al., 2012). Finally, this study also highlights the advantage of functionally interrogating aberrantly spliced transcripts to identify the potentially pathogenic events that can be targeted by novel strategies. The use of oligonucleotide-based approaches to target aberrant splicing choices mediated by mutant splicing factors or specific mRNAs encoding aberrant isoforms, such as the *CASP8 TR* isoform, may be another therapeutic avenue for targeting cells bearing RNA splicing factor mutations.

2.5 Figures and Legends

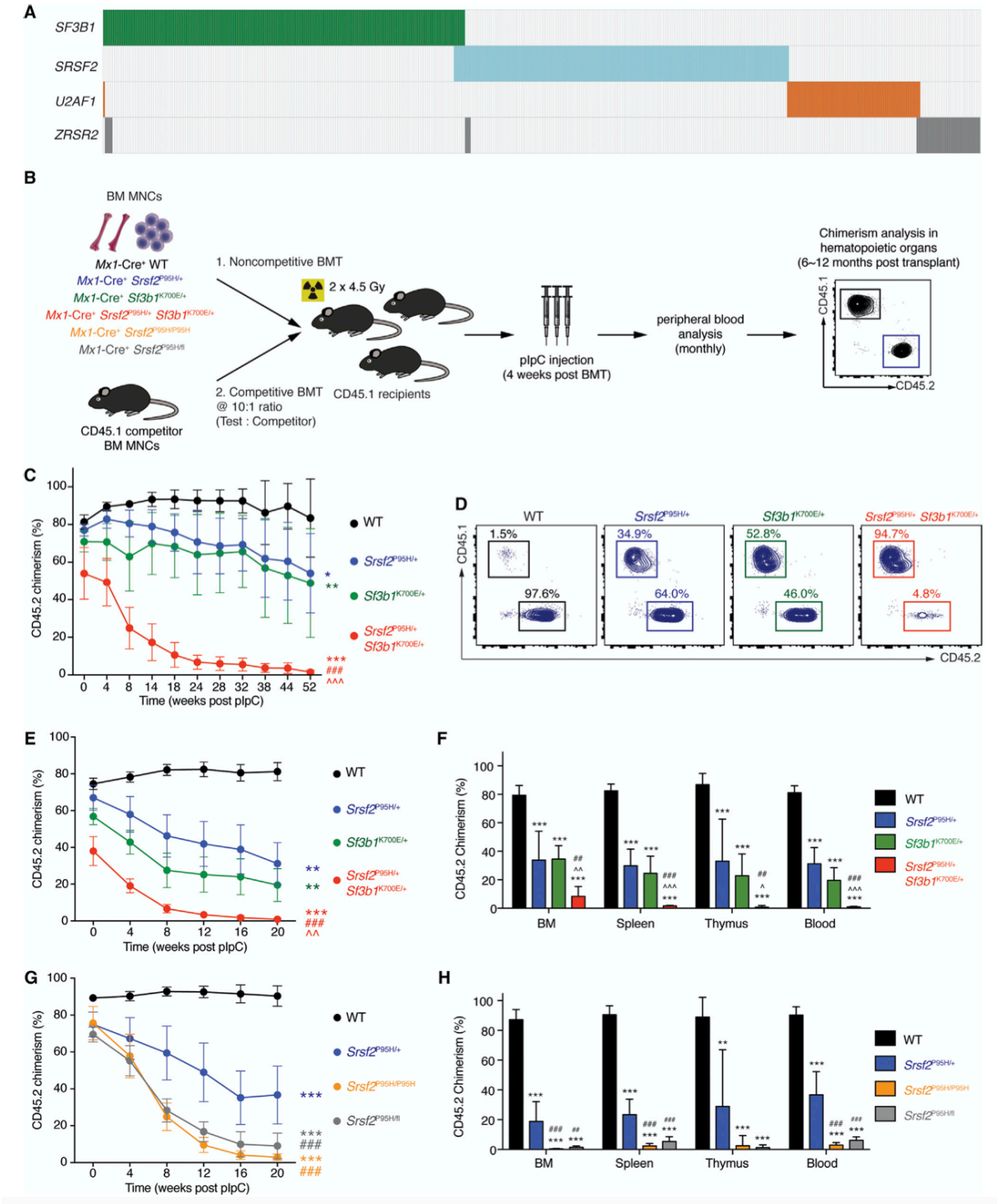


Figure 2:1 Simultaneous expression of cancer-associated mutations in Srsf2 and Sf3b1 or expression of a single RNA splicing factor mutation in the homozygous state is incompatible with hematopoiesis

Figure 1. Simultaneous expression of cancer-associated mutations in Srsf2 and Sf3b1 or expression of a single RNA splicing factor mutation in the homozygous state is incompatible with hematopoiesis. (A) Heatmap representation of the four most commonly mutated genes encoding pre-mRNA splicing factors across 11 sequencing studies in patients with myeloid malignancies (Bejar et al., 2012; Damm et al., 2012; Haferlach et al., 2014; Lasho et al., 2012; Makishima et al., 2012; Meggendorfer et al., 2012; Papaemmanuil et al., 2013; Patnaik et al., 2013; Thol et al., 2012; Yoshida et al., 2011; Zhang et al., 2012). Each column represents a single patient, and each colored bar represents the presence of the specified mutation in an RNA splicing factor. (B) Experimental schema of competitive and noncompetitive bone marrow transplantation (BMT) experiments using bone marrow mononuclear cells (BM MNCs) from 8-12 week old *Mx1*-Cre wild-type (WT), *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$, *Mx1*-Cre *Srsf2*$^{P95H/P95H}$ and *Mx1*-Cre *Srsf2*$^{P95H/fl}$ mice. Polyinosinic-polycytidylic acid (pIpC) was administered to recipient mice 4-weeks post-transplantation to induce expression of mutant alleles. (C) Percentage of CD45.2+ (donor-derived) chimerism in the peripheral blood of CD45.1 recipient mice (n=8-10 mice per genotype) in noncompetitive BM transplantation. (D) Representative FACS plots of CD45.2^{+} (donor-derived) cells in the peripheral blood of CD45.1 recipient mice 52 weeks post pIpC administration. (E) Percentage of CD45.2^{+} (donor-derived) chimerism in the peripheral blood of CD45.1 recipient mice (n=10 mice per genotype) in competitive BM transplantation from *Mx1*-Cre WT, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice. (F) Analysis of CD45.2 chimerism in the BM, spleen, thymus and blood 20 weeks post-pIpC administration. (G) Percentage of CD45.2^{+} (donor-derived) chimerism in the peripheral blood of CD45.1 recipient mice (n= 5-10 mice per genotype) in competitive BM transplantation from *Mx1*-Cre *Srsf2*$^{+/+}$, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Srsf2*$^{P95H/P95H}$ and *Mx1*-Cre *Srsf2*$^{P95H/fl}$ mice. (H) Analysis of CD45.2 chimerism in the BM, spleen, thymus, and blood 20 weeks post pIpC administration. Error bars represent means ± standard deviation. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each *p*-value shown has been adjusted to account for multiple comparisons. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ versus WT group; ^ $p<0.05$, ^^ $p<0.01$, ^^^ $p<0.001$ versus *Mx1*-Cre *Sf3b1*$^{K700E/+}$ mice; $^{\#\#}$ $p<0.01$, $^{\#\#\#}$ $p<0.001$ versus *Mx1*-Cre *Srsf2*$^{P95H/+}$ mice.

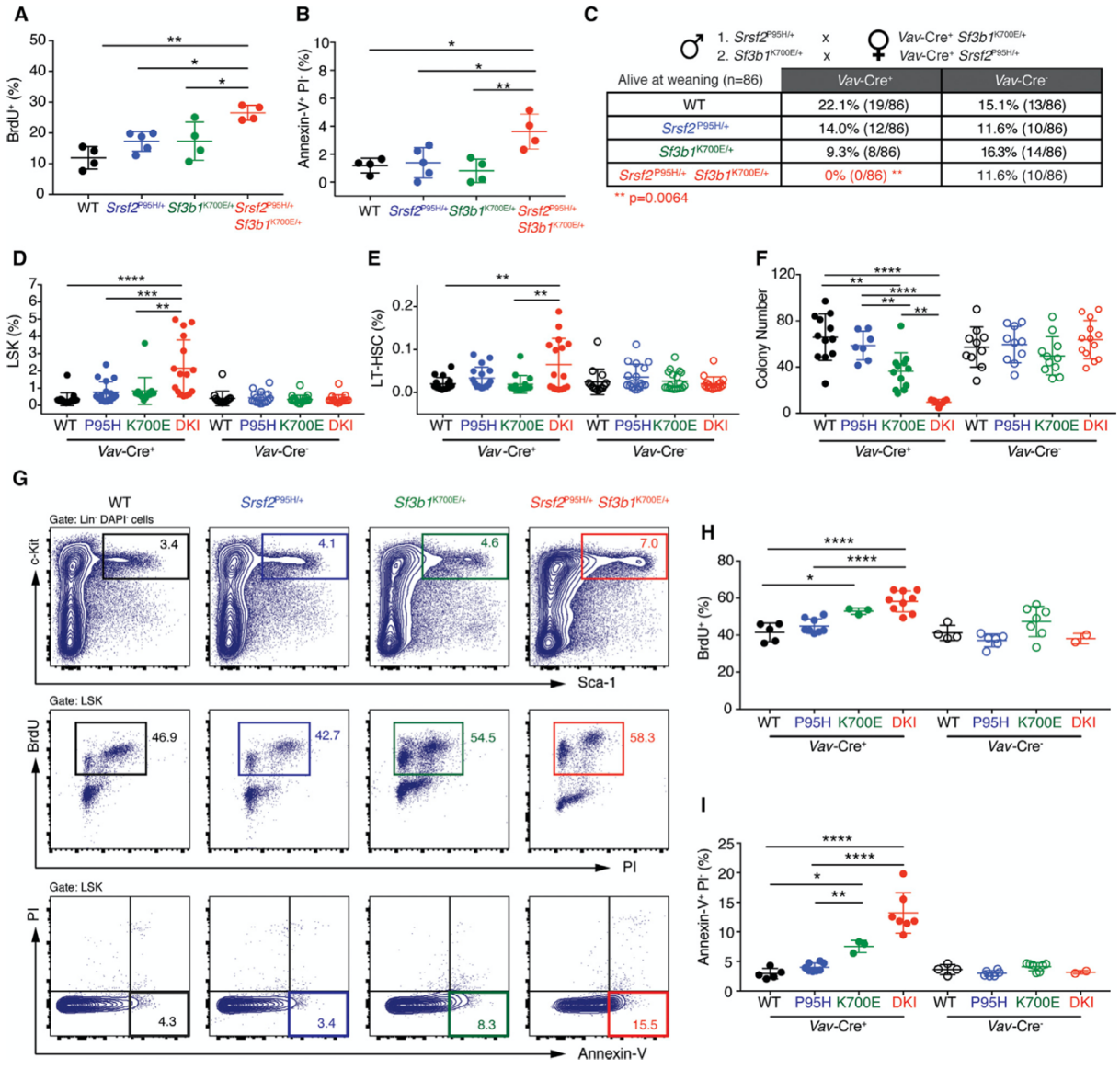


Figure 2:2 Combined expression of mutations in *Srsf2* and *Sf3b1* results in hematopoietic stem and progenitor cell apoptosis and loss of quiescence

Figure 2:2. Combined expression of mutations in *Srsf2* and *Sf3b1* results in hematopoietic stem and progenitor cell apoptosis and loss of quiescence. Percentage of (A) Bromodeoxyuridine^{+} (BrdU^{+}) or (B) Annexin-V^{+} PI^{-} lineage-negative c-Kit^{+} Sca-1^{+} (LSK) cells from *Mx1*-Cre wild-type (WT), *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$ and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice 2-weeks following pIpC administration *in vivo*. (C) Number of mice

present at weaning (Day 14) per genotype from crossing *Sf3b1*$^{K700E/+}$ x *Vav*-Cre *Srsf2*$^{P95H/+}$ or *Srsf2*$^{P95H/+}$ x *Vav*-Cre *Sf3b1*$^{K700E/+}$ mice. A two-sided Chi-square test was used to determine statistical significance (** p=0.0064). Percentage of phenotypic (D) LSK and (E) long-term hematopoietic stem cells (LT-HSC; LSK CD150^{+} CD48^{-}) in embryonic day 14.5 (E14.5) fetal livers from *Vav*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT), *Vav*-Cre *Srsf2*$^{P95H/+}$ (P95H), *Vav*-Cre *Sf3b1*$^{K700E/+}$ (K700E) and *Vav-Cre Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ double knock-in (DKI) fetuses and from *Vav*-Cre-negative controls. (F) Numbers of colonies formed from E14.5 fetal liver cells from WT, P95H, K700E, and DKI fetuses and from *Vav*-Cre-negative controls. (G) Representative flow cytometric plots and quantitation of (H) BrdU^{+} and (I) Annexin V^{+} PI^{-} LSK cells from WT, P95H, K700E, and DKI fetuses. Error bars represent means ± standard deviation. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each p-value shown has been adjusted to account for multiple comparisons. * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001.

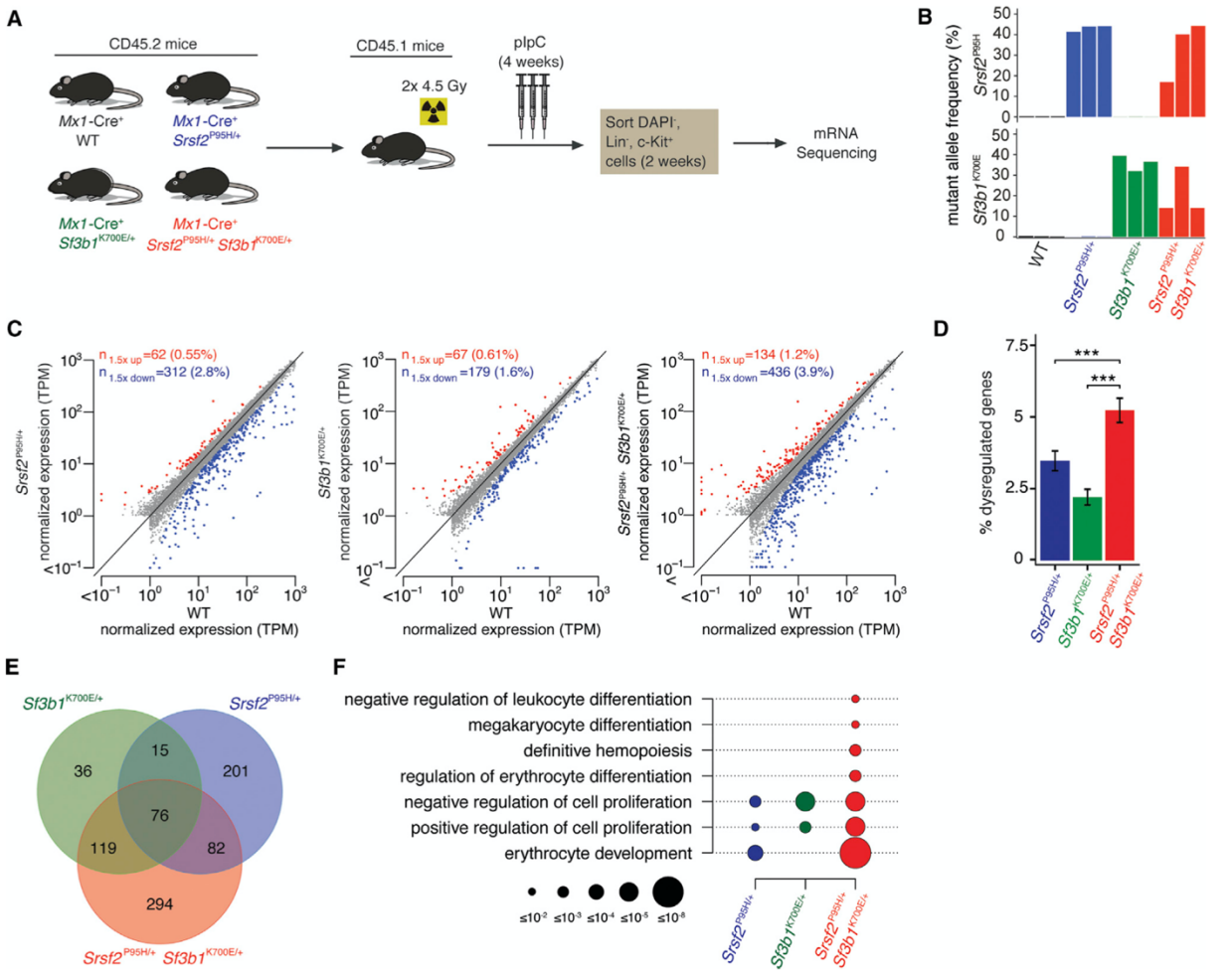


Figure 2:3: *Srsf2* and *Sf3b1* mutations have distinct and independent effects on gene dysregulation

Figure 2:3. *Srsf2* and *Sf3b1* mutations have distinct and independent effects on gene dysregulation. (A) Experimental schema of sample collection for RNA-seq performed in biological triplicate. (B) Expression of *Srsf2*^{P95H} and *Sf3b1*^{K700E} alleles as percentage of mRNAs expressed from *Srsf2* and *Sf3b1*. Color indicates genotype; the three biological replicates are A-C from left to right. (C) Scatter plots comparing coding gene expression in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells for replicate B. Red and blue indicate coding genes that were significantly up- or down-regulated in mutant relative to wild-type cells. TPM, transcripts per million (TMM-normalized). (D) Bar plots comparing percentage of significantly dysregulated coding genes (% of total coding genes expressed) in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells for replicate

B. A two sided binomial proportion test was used to compare differences between groups. *** $p<0.0001$. (E) Venn diagram showing the overlap between coding genes that were significantly dysregulated in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells for replicate B. (F) Gene Ontology (GO) enrichment analysis of *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to wild-type *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells for replicate B. Circle size indicates the magnitude of the *p*-value for each term and genotype comparison.

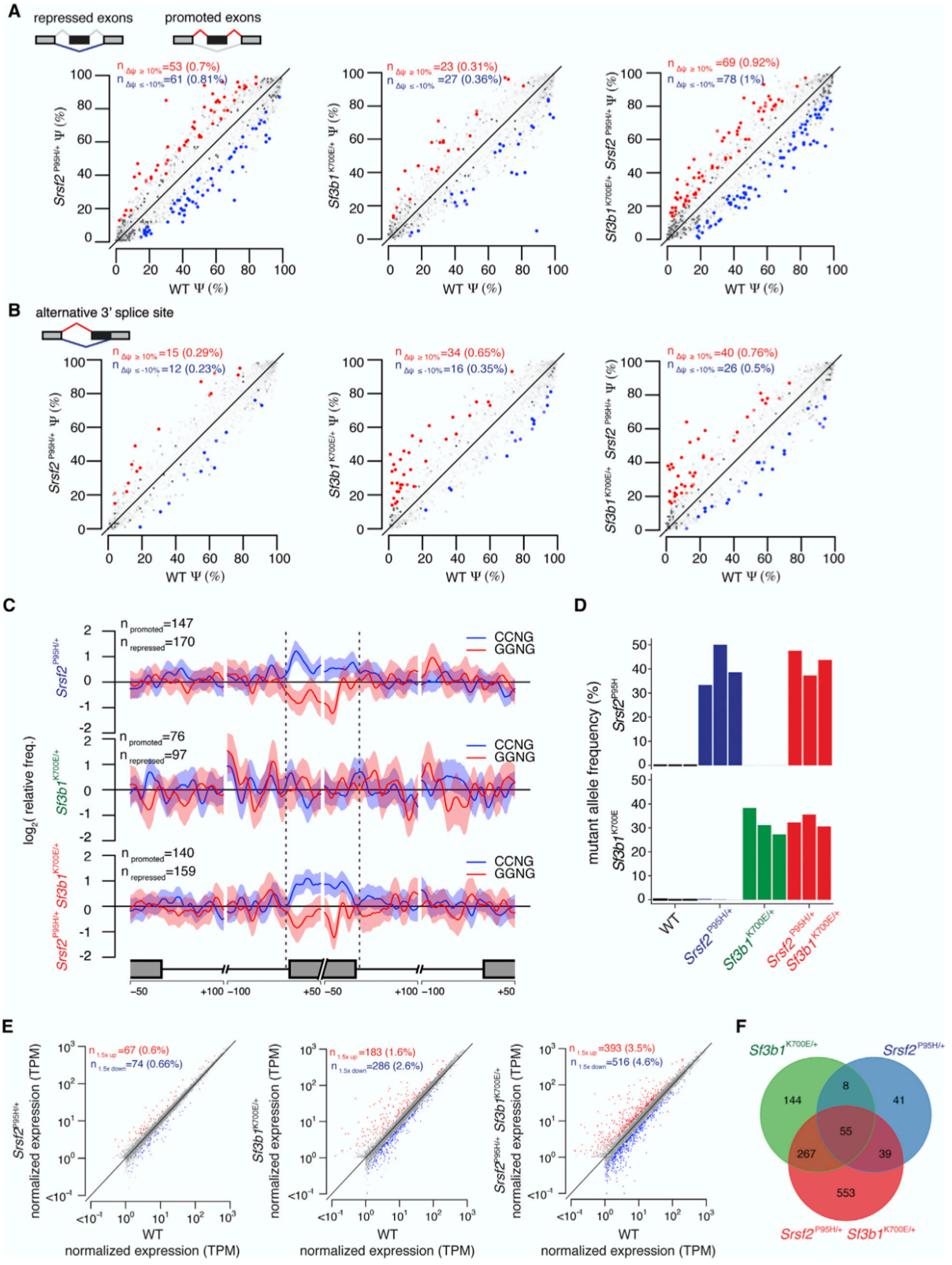


Figure 2:4 Srsf2 and Sf3b1 mutations have distinct and independent effects on RNA splicing

Figure 2:4. *Srsf2* and *Sf3b1* mutations have distinct and independent effects on RNA splicing. (A) Scatter plots of cassette exon inclusion in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells for replicate B. Axes indicate the fraction of mRNAs containing each cassette exon in the indicated sample. Red and blue indicate cassette exons whose inclusion is promoted and repressed in mutant relative to wild-type cells. (B) As (A), but for competing/alternative 3' splice site events. Axes indicate the fraction of mRNAs which use the intron-proximal 3' splice site in the indicated sample. Red and blue indicate intron-proximal 3' splice sites whose usage is promoted and repressed in mutant relative to wild-type cells. (C) Plots illustrating the spatial distribution of the CCNG and GGNG (N = any nucleotide) exonic splicing enhancers adjacent to differentially spliced cassette exons that are promoted versus repressed in mutant relative to wild-type cells. Vertical axis indicates the relative frequency of each motif, averaged over all promoted versus repressed cassette exons for the indicated genotype comparisons. (D) Expression of *Srsf2*^{P95H} and *Sf3b1*^{K700E} alleles as percentage of mRNAs expressed from *Srsf2* and *Sf3b1* in lineage-negative (Lin$^-$) c-Kit$^+$ cells from fetal livers of *Vav*-Cre mice at embryonic day 14.5 (E14.5). Color indicates genotype; the three biological replicates are A-C from left to right. (E) Scatter plots comparing mean coding gene expression for all replicates in *Vav*-Cre *Srsf2*$^{P95H/+}$, *Vav*-Cre *Sf3b1*$^{K700E/+}$, and *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Vav*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells. Red and blue indicate coding genes that were significantly up- or down-regulated in mutant relative to wild-type cells. TPM, transcripts per million (TMM-normalized). (F) Venn diagram showing the overlap between coding genes that were significantly dysregulated in *Vav*-Cre *Srsf2*$^{P95H/+}$, *Vav*-Cre *Sf3b1*$^{K700E/+}$, and *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Vav*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells in all replicates.

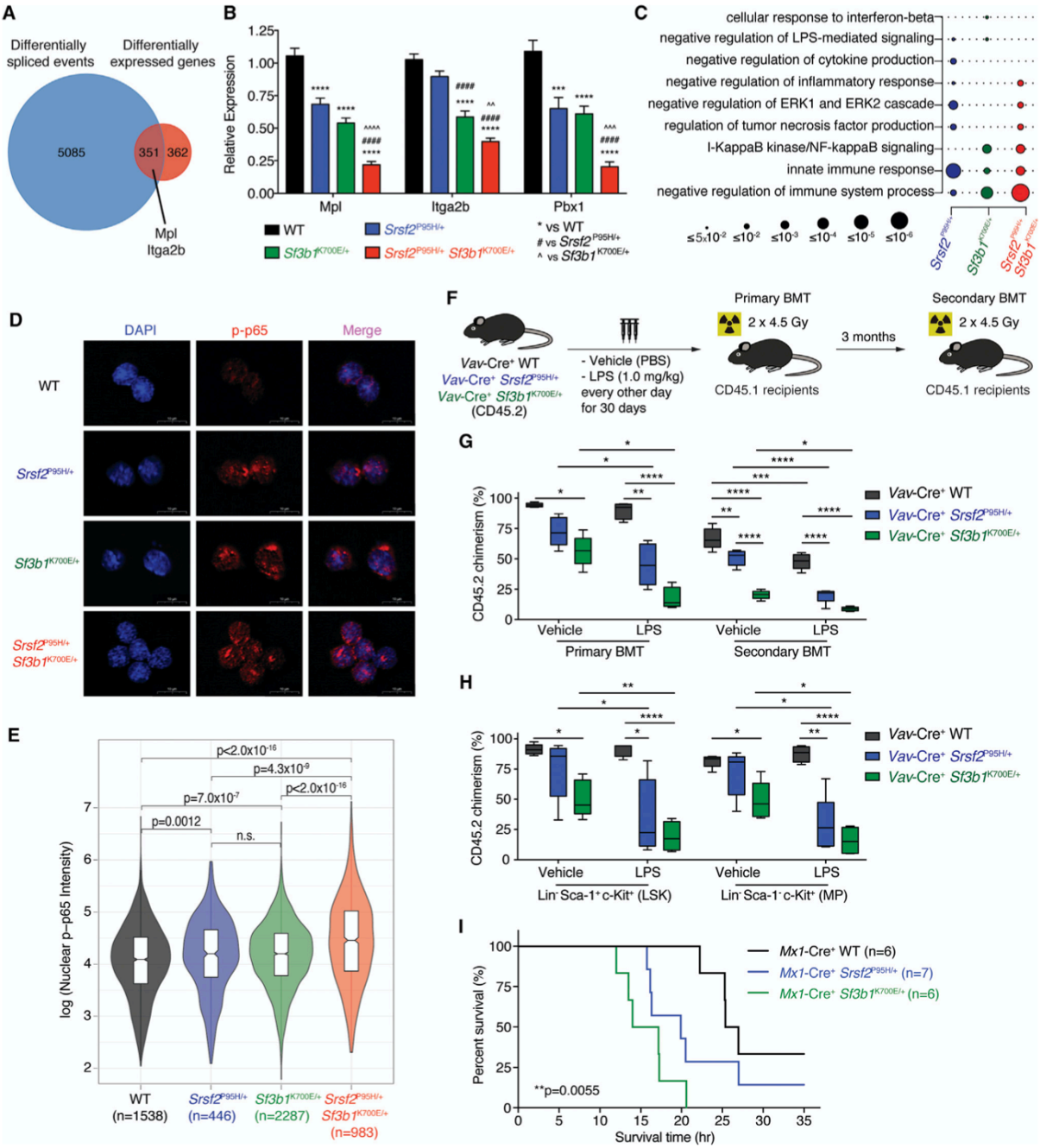


Figure 2:5 Co-expression of Srsf2 and Sf3b1 mutations results in aberrant splicing and impaired expression of key regulators of hematopoietic stem cell survival and quiescence

Figure 2:5. Co-expression of *Srsf2* and *Sf3b1* mutations results in aberrant splicing and impaired expression of key regulators of hematopoietic stem cell survival and quiescence. (A) Venn diagram of the overlap between genes that were differentially expressed and spliced in *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells

in any replicate. (B) Relative expression by quantitative RT-PCR of *Mpl,* and *Itga2b* and *Pbx1* in purified lineage-negative c-Kit^{+} (LK) cells from *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice relative to other control groups (n = 8-10 mice per genotype). (C) Gene Ontology enrichment analysis of mouse hematopoietic stem and progenitor cells from *Mx1*-Cre mice revealed that differentially expressed genes were enriched for innate immune and inflammatory pathways in *Sf3b1*$^{K700E/+}$, *Srsf2*$^{P95H/+}$ and *Sf3b1*$^{K700E/+}$ *Srsf2*$^{P95H/+}$ cells relative to wild-type control. (D) Immunofluorescence analysis of nuclear phosphorylated p65 (P-p65) in LK cells from *Vav*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type, *Vav*-Cre *Srsf2*$^{P95H/+}$, *Vav*-Cre *Sf3b1*$^{K700E/+}$, and *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice *ex vivo*. (E) Violin plots showing quantitation of P-p65 intensity in the nucleus of LK cells from (D). Analysis of variance using Kruskal-Wallis ranked test was performed. Multiple pair-wise comparisons were adjusted by false-discovery rate. (F) Experimental schema of competitive (1:10) bone marrow (BM) transplantation experiments using bone marrow mononuclear cells from *Vav*-Cre WT, *Vav*-Cre *Srsf2*$^{P95H/+}$ and *Vav*-Cre *Sf3b1*$^{K700E/+}$ mice that were subjected to chronic LPS treatment (1 mg/kg; every second day for 30 days). (G) Percentage of CD45.2^{+} (donor-derived) chimerism in the peripheral blood of CD45.1 recipient mice (n=4-5 mice per group) from primary and secondary competitive BM transplantation. Donor mice were treated with either vehicle (PBS) or LPS (1 mg/kg) outlined in (F). (H) Chimerism in BM lineage^{-} Sca-1^{+} c-Kit^{+} (LSK) or myeloid progenitor (MP; lineage^{-} Sca-1^{-} c-Kit^{+}) fractions from primary transplant recipients 14-weeks post BM transplantation. (I) Kaplan-Meier survival analysis of *Mx1*-Cre wild-type, *Mx1*-Cre *Srsf2*$^{P95H/+}$ and *Mx1*-Cre *Sf3b1*$^{K700E/+}$ mice following a single dose of LPS (15 mg/kg) *in vivo*. Log-ranked Mantel-Cox test was used to compare median survival difference. ** p=0.0055. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each p-value shown has been adjusted to account for multiple comparisons. * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001 versus wild-type ("#" is versus *Srsf2*P95H and "^" is versus *Sf3b1*K700E.

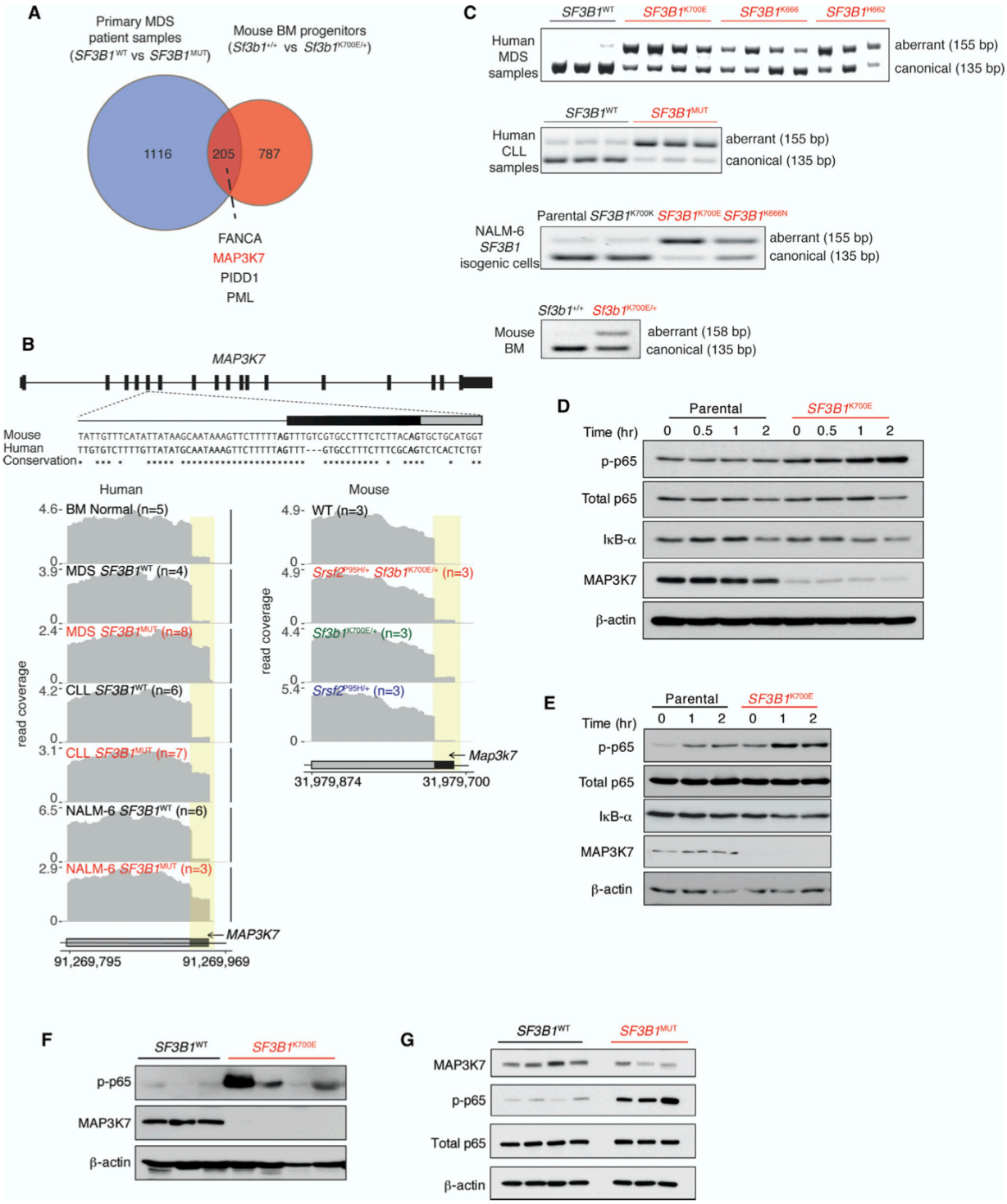


Figure 2:6 SF3B1 mutations promote mis-splicing of MAP3K7, resulting in hyperactivation of NF-κB signaling.

Figure 2:6. SF3B1 mutations promote mis-splicing of MAP3K7, resulting in hyperactivation of NF-κB signaling. (A) Venn diagram illustrating the overlap of differentially spliced genes across MDS patient samples that were mutant versus wild-type for *SF3B1* and murine bone marrow hematopoietic progenitors (lineage$^-$ c-Kit$^+$) in *Sf3b1*$^{K700E/+}$ versus *Sf3b1*$^{+/+}$ mice. (B) From top to bottom, exon-intron structure of *MAP3K7,* conservation of mouse and human sequences adjacent to the competing 3' splice site affected by *SF3B1* mutations, and RNA-seq coverage plots of this region in human and mouse samples with or without *SF3B1* mutations. (C) RT-PCR of the *MAP3K7* competing 3' splice site in primary patient MDS and CLL samples with or without *SF3B1* mutations as well as in isogenic human and mouse cells. Immunoblot analysis of phosphorylated p65 (phospho-p65), IκB-α, MAP3K7 (using a C-terminal anti-MAP3K7 antibody), and loading controls in isogenic (D) K562 and (E) NALM-6 leukemia cells expressing wild-type or mutant *SF3B1* from its endogenous locus ("Time (h)" refers to hours following LPS (5 μg/mL) exposure). Immunoblot analysis of phospho-p65, MAP3K7, and loading controls in (F) bone marrow mononuclear cells (MNCs) from MDS patients with or without *SF3B1* mutation and (G) peripheral blood MNCs from CLL patients with or without *SF3B1* mutation.

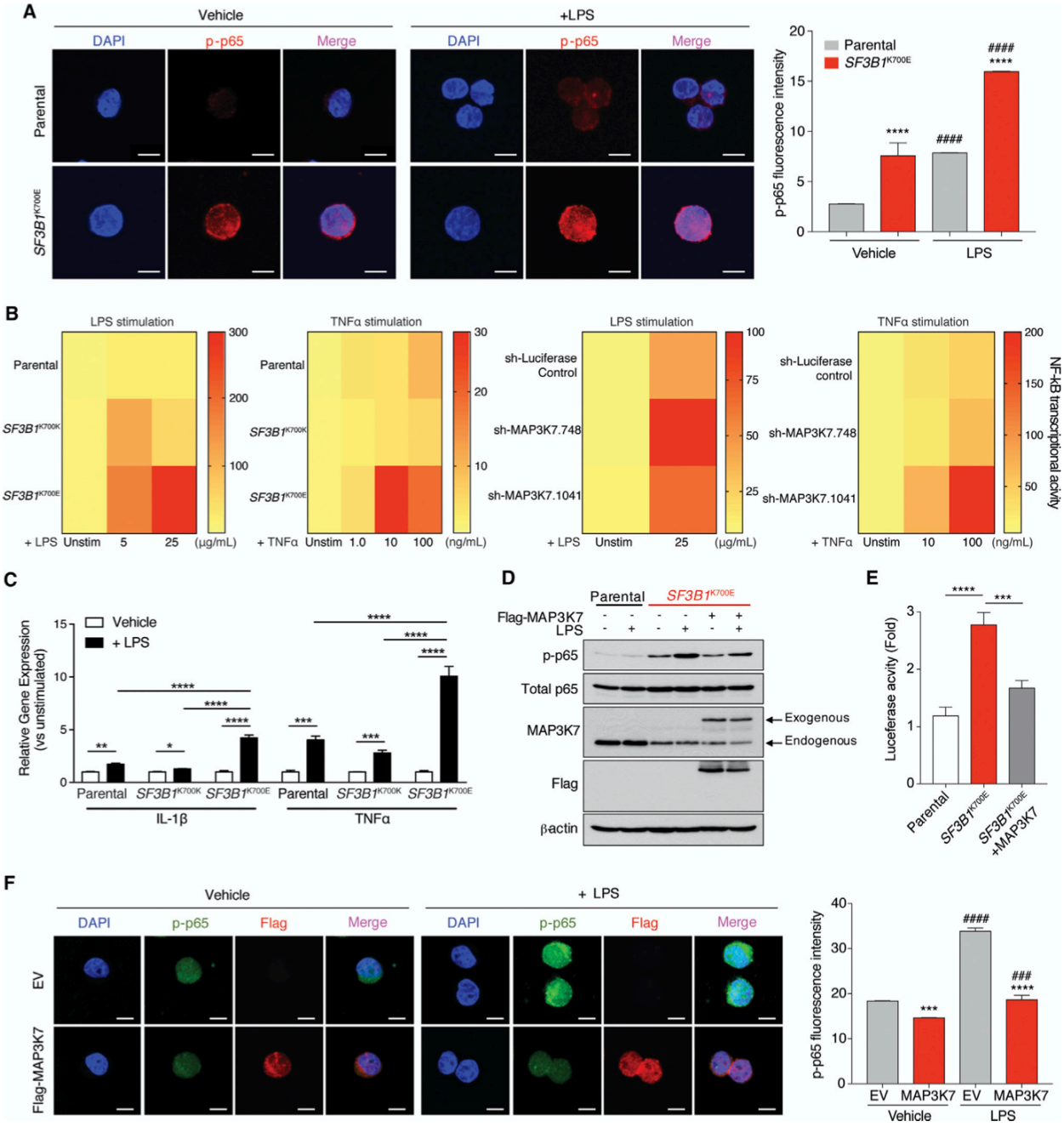


Figure 2:7 MAP3K7 loss is responsible for hyperactivated NF-κB signaling in SF3B1-mutant cells.

Figure 2:7. MAP3K7 loss is responsible for hyperactivated NF-κB signaling in *SF3B1*-mutant cells. (A) Immunofluorescence analysis of phosphorylated p65 (P-p65) in NALM-6 cells with or without *SF3B1*K700E mutation following stimulation with lipopolysaccharide (LPS). Quantitation of phospho-p65 intensity is shown on the right. (B) Heatmap representation of NF-

κB luciferase reporter signal in SF3B1 mutant knockin NALM-6 cells (left two panels) or parental NALM-6 cells with anti-*MAP3K7* shRNAs (right two panels) following stimulation with LPS or TNFα. (C) Quantitative RT-PCR (qRT-PCR) analysis of NF-κB target genes IL-1β and TNF 8 hours post LPS stimulation in NALM-6 *SF3B1*-isogenic cells (two independent experiments performed, each in technical triplicate). (D) Immunoblot analysis of phospho-p65 in isogenic K562 cells with and without *SF3B1*K700E knockin and with and without Flag-MAP3K7 cDNA expression and/or LPS (5 μg/mL) exposure. (E) Quantitation of NF-κB luciferase reporter signal in cells from (D; from a biological replicate of three independent experiments). (F) Immunofluorescence analysis of phosphorylated p65 (phospho-p65), Flag, and DAPI in cells from (D; quantitation of phospho-p65 intensity is shown on the right). Error bars represent means ± standard deviation. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each *p*-value shown has been adjusted to account for multiple comparisons. * $p<0.05$, ** $p<0.005$, *** $p<0.0002$, **** $p<0.0001$. See also Figure S6.

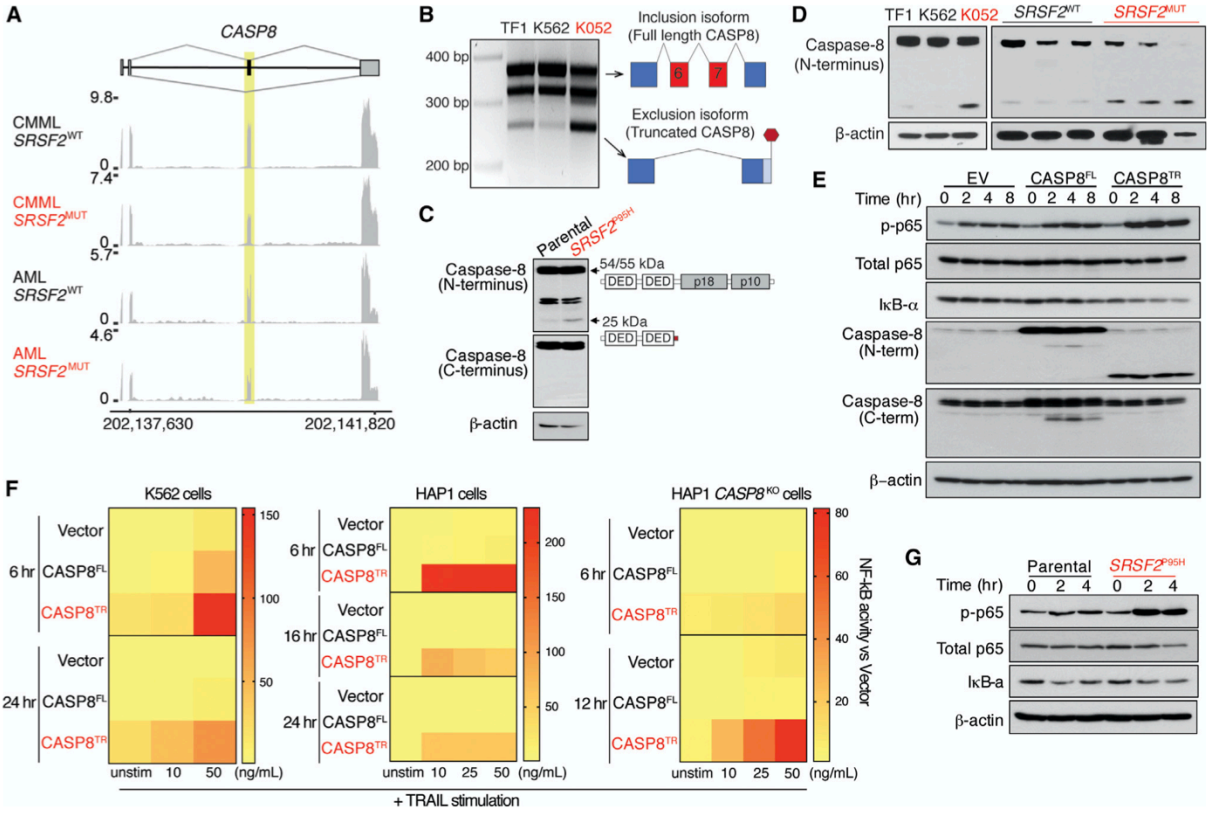
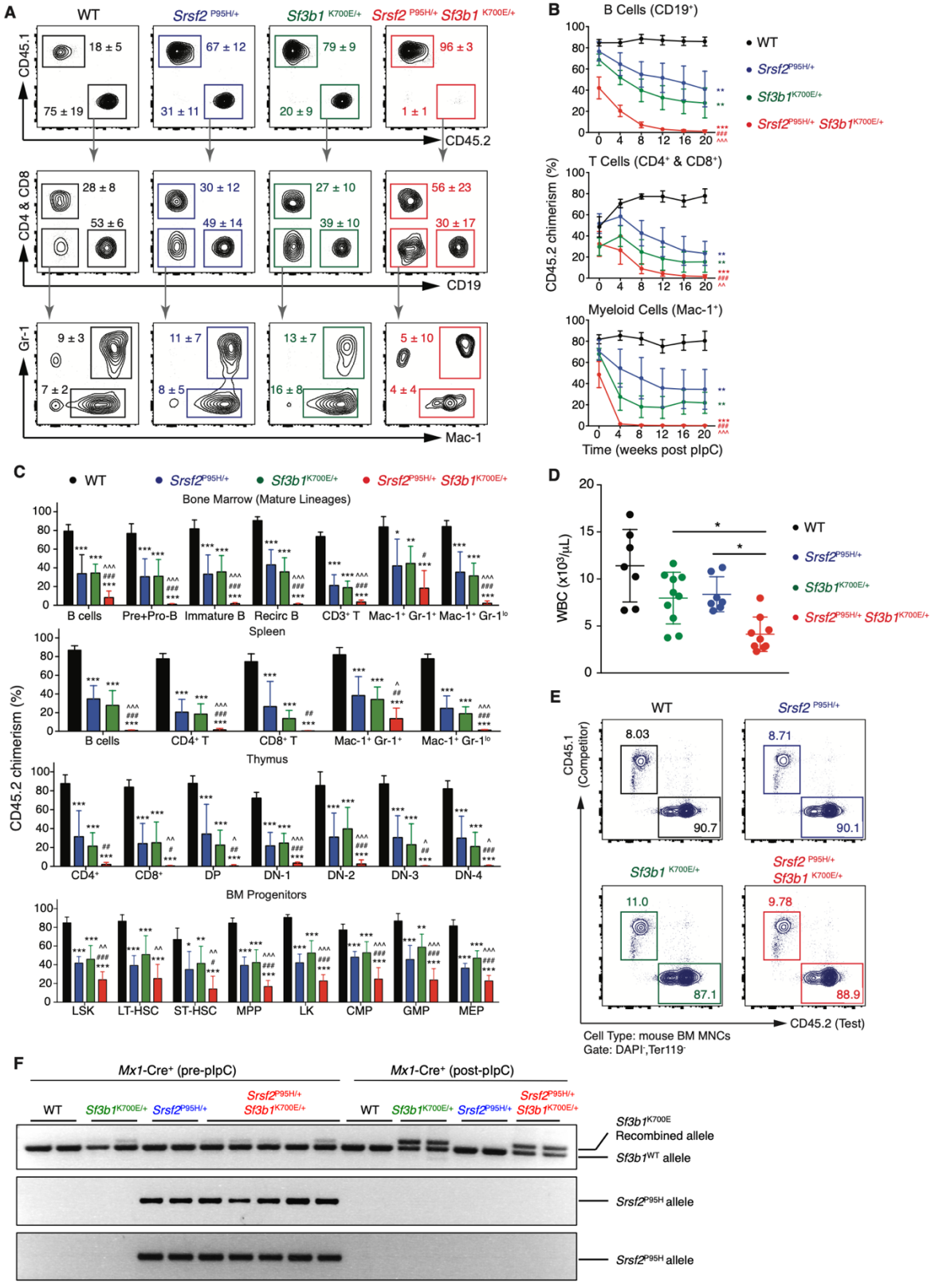


Figure 2:8 SRSF2 mutations promote aberrant splicing of caspase-8 (CASP8), resulting in expression of a truncated protein that hyperactivates NF-κB signaling.

Figure 2:8. SRSF2 mutations promote aberrant splicing of caspase-8 (CASP8), resulting in expression of a truncated protein that hyperactivates NF-κB signaling. (A) RNA-seq coverage plots of *CASP8* splicing in chronic myelomonocytic leukemia (CMML) and acute myeloid leukemia (AML) patients wild-type or mutant for *SRSF2*. *SRSF2* mutations repress the illustrated *CASP8* cassette exon. AML and CMML RNA-seq data are from Kim *et al* (Kim et al., 2015). (B) RT-PCR analysis of the *CASP8* splicing event in human leukemia cell lines wild-type (TF1 and K562) or mutant (K052) for *SRSF2* and a diagram of the corresponding *CASP8* isoforms. Immunoblot analysis of caspase-8 using an N-terminal anti-caspase-8 antibody in (C) isogenic K562 cells with or without *SRSF2*P95H mutation expressed from the endogenous *SRSF2* locus, or (D) human leukemia cells (left panel) and primary AML patient samples (right panel) wild-type or mutant for *SRSF2*. (E) Immunoblot analysis of phosphorylated p65, IκB-α, Caspase-8, and loading controls in K562 cells expressing cDNA constructs encoding empty vector (EV), full-length *CASP8* (*CASP8*^{FL}), or the aberrantly spliced *caspase-8* truncated isoform (*CASP8*^{TR})

following exposure to TRAIL (50 ng/mL). (F) Heatmap representation of NF-κB luciferase reporter signal following TRAIL stimulation in K562, *CASP8* wild-type HAP1, or *CASP8* KO HAP1 cells expressing cDNA constructs encoding empty vector, full-length *caspase-8* (*CASP8*^{FL}), or the aberrantly spliced *caspase-8* truncated isoform (*CASP8*^{TR}). (G) Immunoblot analysis of phosphorylated p65, IκB-α, and loading controls in K562 cells with or without *SRSF2*P95H mutation expressed from the endogenous *SRSF2* locus following exposure to TRAIL (50 ng/mL).

2.6 SUPPLEMENTAL FIGURES



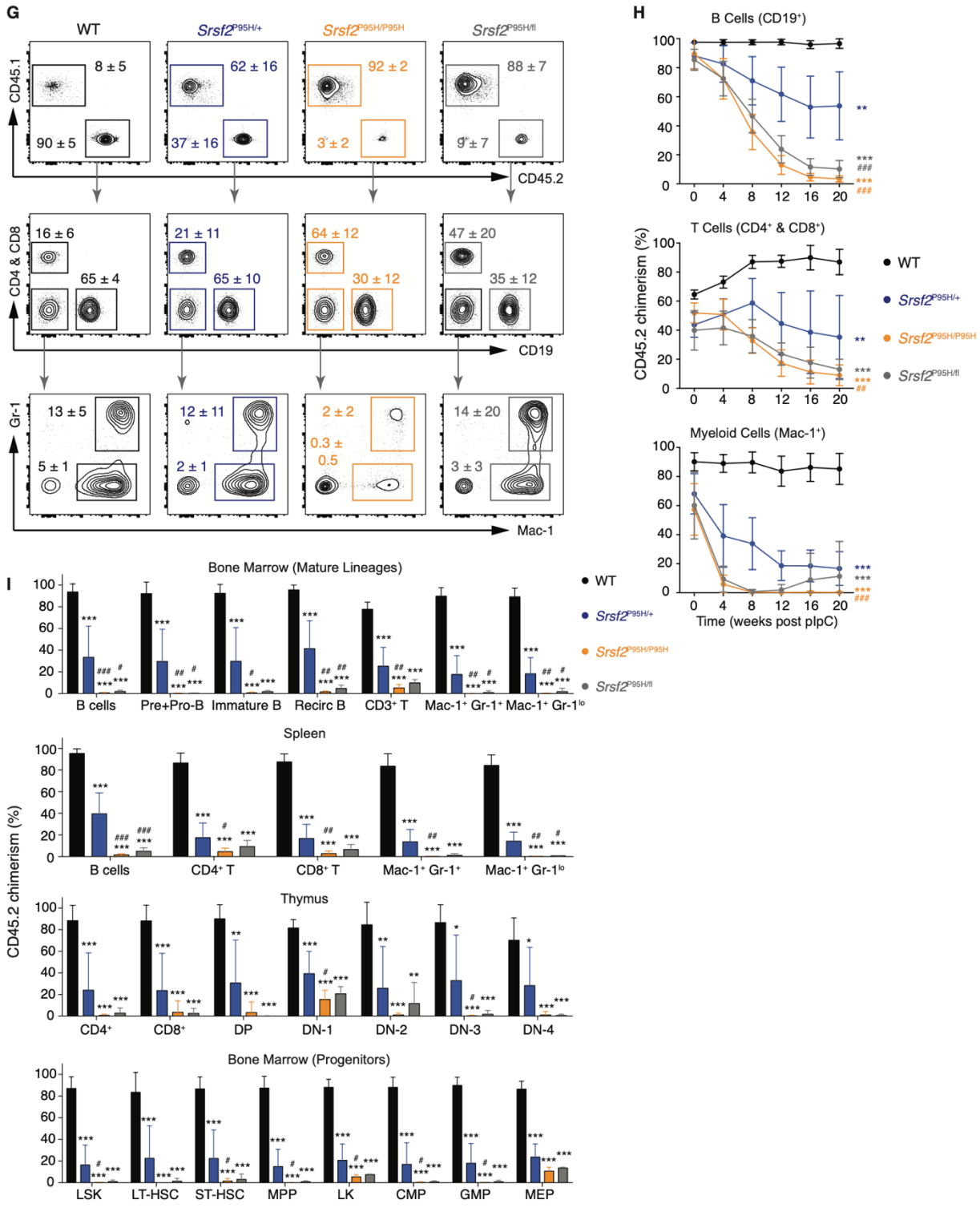


Figure 2:9 (Supplementary) Severe self-renewal disadvantage of hematopoietic stem cells from Srsf2P95H/+ Sf3b1K700E/+ double-mutant mice

Figure 2:9 Severe self-renewal disadvantage of hematopoietic stem cells from *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ double-mutant mice and *Srsf2*$^{P95H/P95H}$ (A) Representative FACS plots of CD45.2^{+} (donor-derived) cells of both lymphoid and myeloid lineages in the peripheral blood of CD45.1 recipient mice reconstituted with bone marrow (BM) mononuclear cells (MNCs) from *Mx1*-Cre, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$ and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice 20 weeks post pIpC administration. **(B)** Percentage of CD45.2 (donor-derived) chimerism in the B-, T- or myeloid lineages in peripheral blood (PB) of CD45.1 recipient mice (n=10 mice per genotype) in competitive BM transplantation. **(C)** Analysis of CD45.2 chimerism in the various cellular lineages in the bone marrow, spleen and thymus 20 weeks post pIpC administration in CD45.1 recipient mice. Error bars represent means ± standard deviation. **(D)** White blood cell (WBC) count of *Mx1*-Cre, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$ and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice 52 weeks post pIpC administration. **(E)** Representative FACS plots of BM MNC samples prior to noncompetitive transplantation. **(F)** Genotyping PCR of PB MNCs of *Mx1*-Cre, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$ and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice pre- and post-pIpC administration. **(G)** Representative FACS plots of CD45.2 (donor-derived) cells of both lymphoid and myeloid lineages in the peripheral blood of CD45.1 recipient mice reconstituted with bone marrow (BM) mononuclear cells(MNCs) from Mx1-Cre+ WT,Mx1-Cre^{+} *Srsf2*$^{P95H/+}$,Mx1-Cre *Srsf2*$^{P95H/P95H}$ and Mx1-Cre *Srsf2*$^{P95H/lf}$ mice 20 weeks post pIpC administration. **(H)** Percentage of CD45.2 (donor-derived) chimerism in the B, T or myeloid lineages in peripheral blood of CD45.1 recipient mice (n=5-10 mice per genotype) in competitive BMT**. (I)** Analysis of CD45.2 chimerism in the various cellular lineages in the bone marrow, spleen and thymus 20 weeks post pIpC administration in CD45.1 recipient mice. Error bars represent mean±standard deviation. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each *p*-value shown has been adjusted to account for multiple comparisons. *p<0.05, **p< 0.01 and ***p< 0.001 versus *Mx1*-Cre mice; ^p<0.05, ^^p<0.01 and ^^^p<0.001 versus *Mx1*-Cre *Sf3b1*$^{K700E/+}$ mice; $^{\#}p$< 0.05, $^{\#\#}p$< 0.01 and $^{\#\#\#}p$< 0.001 versus *Mx1*-Cre *Srsf2*$^{P95H/+}$ mice.

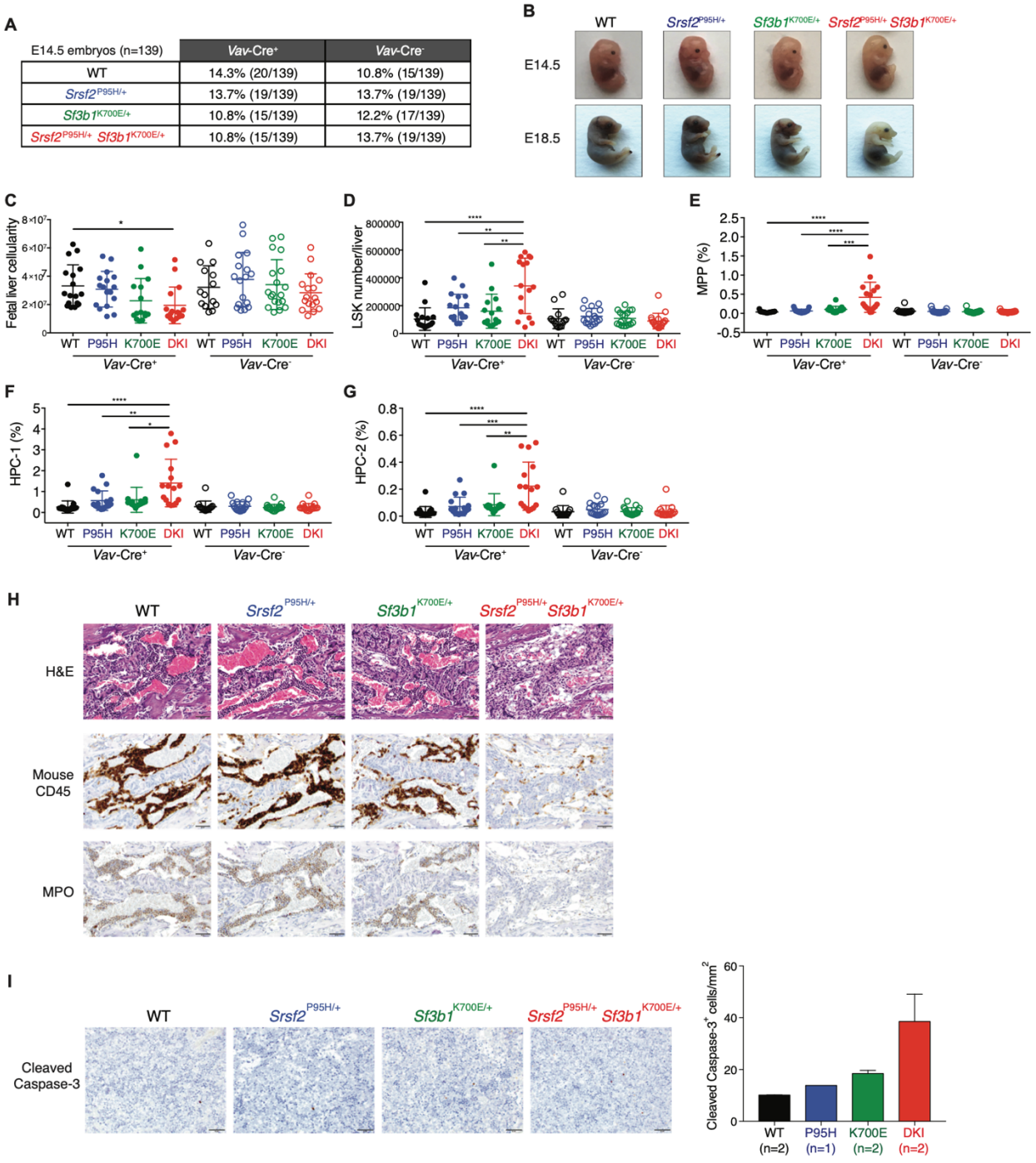


Figure 2:10 (Supplementary) Co-expression of Srsf2 and Sf3b1 mutations results in death before birth in Vav-Cre Srsf2P95H/+ Sf3b1K700E/+ mice.

Figure 2:10. Co-expression of *Srsf2* and *Sf3b1* mutations results in death before birth in *Vav*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice. (A) Number of embryos detected at embryonic day 14.5 (E14.5) per genotype from crossing *Sf3b1*$^{K700E/+}$ x *Vav*-Cre *Srsf2*$^{P95H/+}$ or *Srsf2*$^{P95H/+}$ x *Vav*-Cre *Sf3b1*$^{K700E/+}$ mice. **(B)** Photographs of representative fetuses with each genotype at E14.5 and E18.5. **(C)** Number of viable fetal liver mononuclear cells (MNCs) and **(D)** LSK (lineage-negative Sca-1$^+$ c-Kit$^+$) cells from E14.5 embryos from (A). Frequency of **(E)** MPP (LSK CD48$^-$ CD150$^-$), **(F)** HPC-1 (LSK CD48$^+$ CD150$^-$), and **(G)** HPC-2 (LSK CD48$^+$ CD150$^+$) cells amongst fetal liver MNCs at E14.5. **(H)** Photomicrographs of hematoxylin and eosin (H&E) and anti-CD45, and myeloperoxidase (MPO) immunohistochemical stains of E18.5 marrow cavity cells of the mice in (B). **(I)** Photomicrographs of cleaved caspase-3 immunohistochemical stains of E18.5 fetal liver cells of the mice in (B) with quantitation on right. Error bars represent means ± standard deviation. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each *p*-value shown has been adjusted to account for multiple comparisons. **p*< 0.05, ***p*< 0.01, ****p*< 0.001, and *****p*< 0.0001.

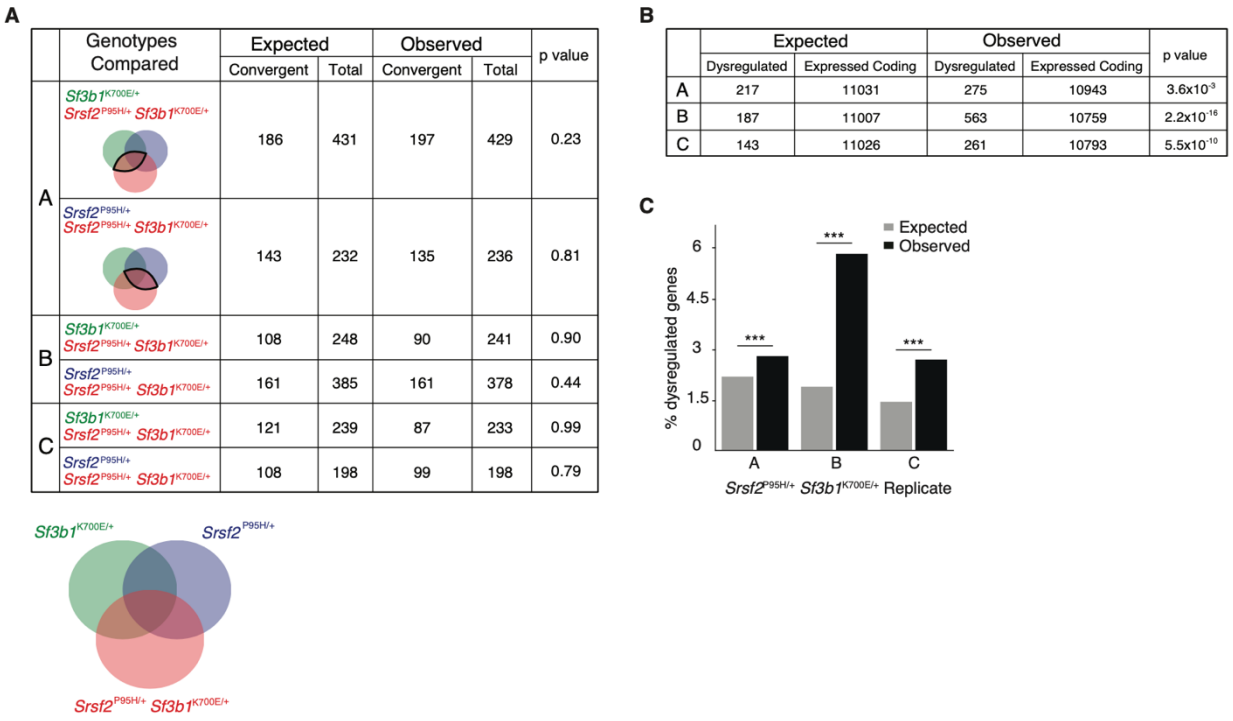


Figure 2:11 (Supplementary) *Srsf2* and *Sf3b1* mutations have distinct and independent effects on gene dysregulation

Figure 2:11 (Supplementary). *Srsf2* and *Sf3b1* mutations have distinct and independent effects on gene dysregulation. (A) Table illustrating the numbers of expected versus observed coding genes exhibiting dysregulated expression that were shared between the illustrated comparisons between genotypes ("A", "B", and "C" refer to replicates of RNA-seq). The set of expected coding genes exhibiting dysregulation was defined assuming that the *Sf3b1* and *Srsf2* mutations were independent, such that the double-mutant gene expression program was simply the mean of the two single-mutant gene expression programs. "Convergent" is defined as the number of differentially expressed genes shared by the genotypes compared. "Total" is defined as the number of differentially expressed genes in the first genotype listed per category. p values were computed using the one-sided binomial proportion test. **(B)** Table illustrating the numbers of expected versus observed coding genes that were dysregulated in double-mutant cells relative to wild-type cells. The set of expected coding genes exhibiting dysregulation was defined as in (A). **(C)** Bar plots showing percentage (%) of expected versus observed coding genes exhibiting dysregulated expression in each replicate of *Mx1*-Cre^{+} *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre^{+} WT cells. ***p<0.0001.

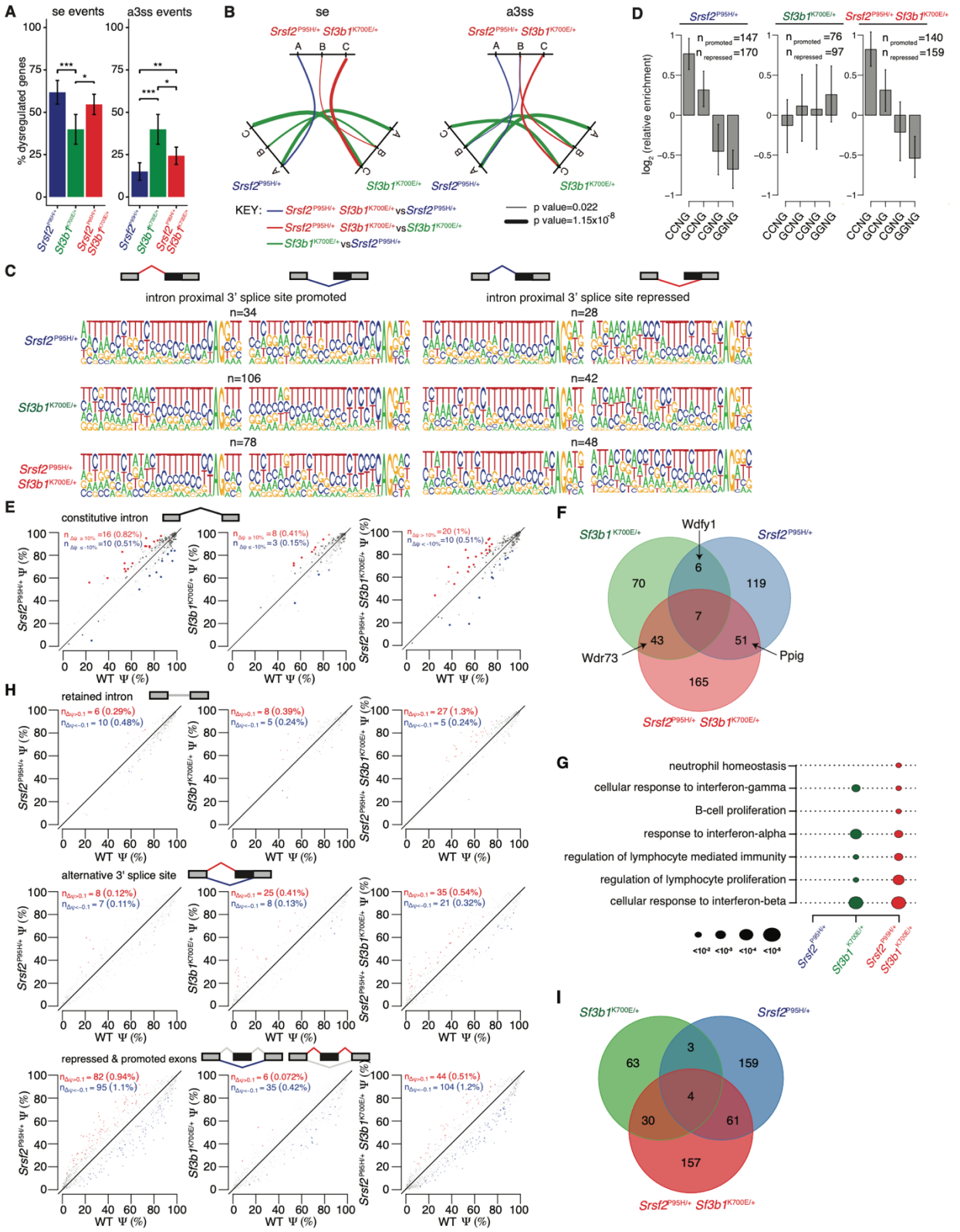


Figure 2:12 (Supplementary) *Srsf2* and *Sf3b1* mutations have distinct and independent effects on RNA splicing.

Figure 2:12. *Srsf2* and *Sf3b1* mutations have distinct and independent effects on RNA splicing. (A) Bar plot illustrating the percentage of dysregulated splicing events that corresponded to cassette exon events (se; left) or competing 3' splice site events (a3ss; right) for each illustrated genotype relative to wild-type cells. *p*-values were computed using a two sided binomial proportion test where $^{**}p < 0.01$ and $^{***}p < 0.001$. **(B)** Diagram illustrating whether there was a statistically significant difference in the fraction of dysregulated splicing events corresponding to cassette exon events (se; left) or competing 3' splice site events (a3ss; right). *p*-values were computed using the binomial proportion test and are represented by the thickness of the lines. **(C)** Sequence logos associated with competing 3' splice sites that are differentially spliced in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wildtype (WT) cells in any replicate. The left- and right-hand columns contain sequence logos created for the intron-proximal and intron-distal 3' splice sites. **(D)** Bar plots illustrating enrichment of different variants of the SSNG (S = C or G; N = any nucleotide) exonic splicing enhancer in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wildtype (WT) cells. Enrichment computed using the alternatively spliced sequence for cassette exons that are differentially spliced in the indicated genotype comparisons. **(E)** Scatter plots of constitutive intron splicing in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells for replicate B. Axes indicate the fraction of mRNAs for which each constitutive intron is removed. Red and blue indicate constitutive introns that are preferentially spliced and retained in mutant relative to wild-type cells. **(F)** Venn diagram showing the overlap between splicing events that were significantly dysregulated in in *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, and *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to *Mx1*-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) cells for replicate B. **(G)** Gene Ontology enrichment analysis of Vav-Cre *Srsf2*$^{P95H/+}$, Vav-Cre *Sf3b1*$^{K700E/+}$, and Vav-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to Vav-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT). Circle size indicates the magnitude of the p-value. **(H)** Scatter plots of retained introns (top), competing 3' splice cites (middle), and cassette exon inclusion (bottom) in Vav-Cre *Srsf2*$^{P95H/+}$, Vav-Cre *Sf3b1*$^{K700E/+}$, and Vav-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to Vav-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT) **(I)** Venn diagram showing the overlap between splicing events that were significantly dysregulated in Vav-Cre *Srsf2*$^{P95H/+}$, Vav-Cre

Sf3b1$^{K700E/+}$, and Vav-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ cells relative to Vav-Cre *Srsf2*$^{+/+}$ *Sf3b1*$^{+/+}$ wild-type (WT)

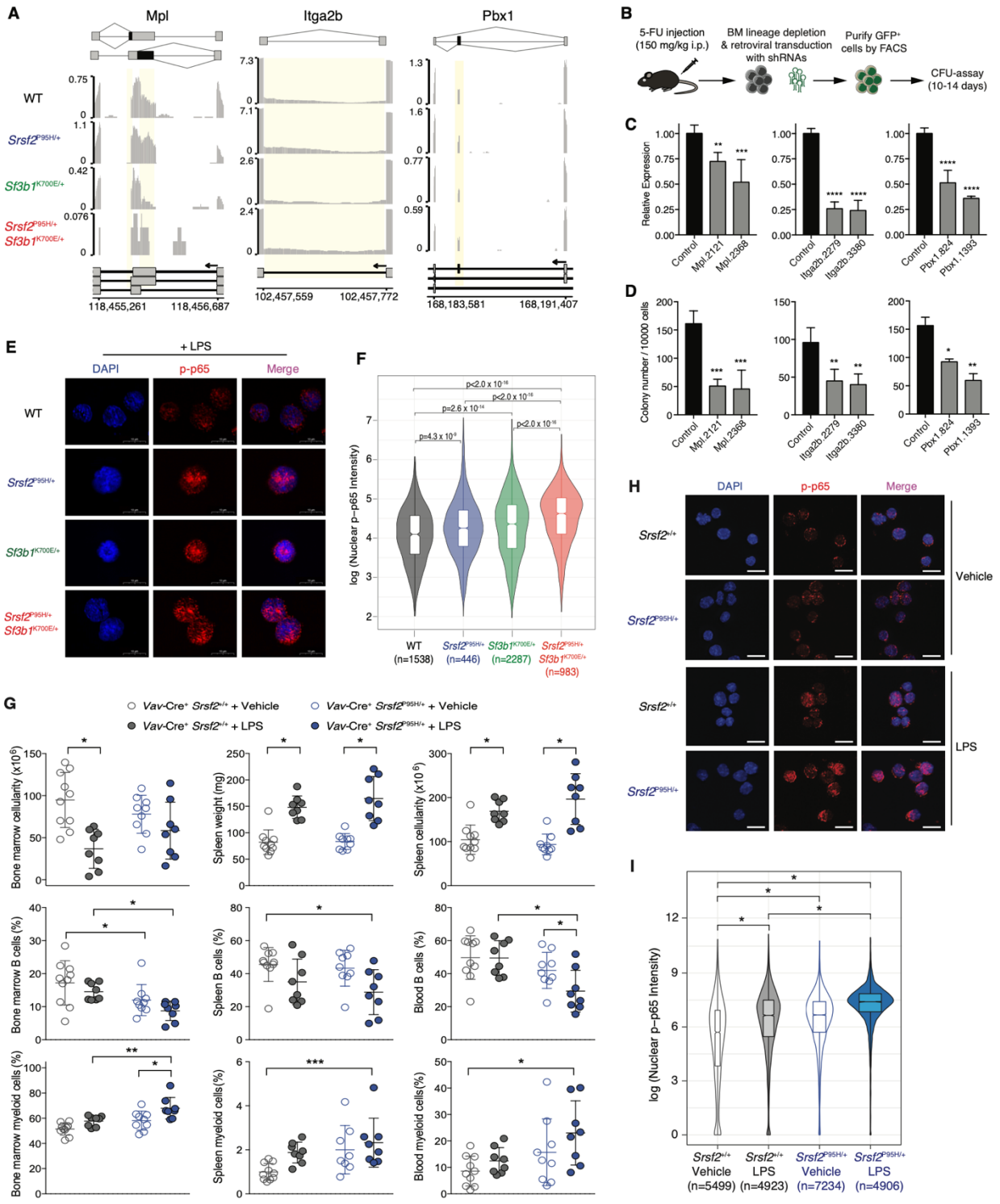


Figure 2:13 (Supplementary) Srsf2 and Sf3b1 mutations have distinct and independent effects on RNA splicing

Figure 2:13. Co-expression of *Srsf2* and *Sf3b1* mutations results in aberrant splicing and expression of regulators of hematopoietic stem cell survival and quiescence, as well as increased sensitivity to LPS-induced NF-κB signaling **(A)** RNA-seq read coverage across the genomic loci containing the illustrated differentially spliced events in *Mpl, Itga2b* and *Pbx1* for all genotypes. Yellow indicates the differentially spliced sequence for each event. **(B)** Schema of evaluation of the effects of shRNA mediated knockdown of *Mpl, Itga2b* and *Pbx1* on 5-fluorouracil-treated, lineage-depleted bone marrow (BM) cells from 8-12 week-old C57BL/6 WT mice. **(C)** Quantitative RT-PCR (qRT-PCR) for *Mpl, Itga2b* or *Pbx1* in lineage-negative (lin^-) bone marrow cells treated with shRNA or renilla-luciferase control (sh-Control). **(D)** Colony numbers from lin^- HSPCs following shRNA-mediated knockodown of *Mpl*, *Itga2b* and *Pbx1* cells. Error bars represent means ± standard deviation. Error bars represent means ± standard deviation. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each *p*-value shown has been adjusted to account for multiple comparisons. ** $p<0.01$ and **** $p<0.0001$. **(E)** Immunofluorescence analysis of nuclear P-p65 in LK cells from *Vav*-Cre $Srsf2^{+/+}$ $Sf3b1^{+/+}$ wild-type, *Vav*-Cre $Srsf2^{P95H/+}$, *Vav*-Cre $Sf3b1^{K700E/+}$, and *Vav*-Cre $Srsf2^{P95H/+}$ $Sf3b1^{K700E/+}$ mice following stimulation with lipopolysaccharide (LPS; 200 ng/mL; 15 minutes) *ex vivo*. **(F)** Violin plots showing quantitation of phosphorylated p-65 intensity in the nucleus of LK cells following LPS stimulation from (H). Analysis of variance using Kruskal-Wallis ranked test was performed. Multiple pair-wise comparisons were adjusted by false-discovery rate. **(G)** Phenotypic analysis of bone marrow, spleen, and blood from *Vav*-Cre $Srsf2^{+/+}$ and *Vav*-Cre $Srsf2^{P95H/+}$ mice following chronic LPS or vehicle exposure. Error bars represent means ± standard deviation. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups, and each *p*-value shown has been adjusted to account for multiple comparisons. *$p< 0.05$ and ***$p< 0.001$. **(H)** Immunofluorescence of phosphorylated p65 (phospho-p65) intensity from FACS-purified LSK cells following acute exposure to LPS *ex vivo*. Scale bar: 10 μm. **(I)** Violin plots showing quantitation of phospho-p65 intensity in the nucleus of LSK cells following LPS stimulation of the following groups: $Srsf2^{+/+}$ Veh (n = 5499 cells), $Srsf2^{+/+}$ LPS (n = 4923 cells), $Srsf2^{P95H/+}$ Veh (n = 7234 cells), $Srsf2^{P95H/+}$ LPS (n = 4906 cells). Analysis of variance using Kruskal-Wallis ranked test was performed. Multiple pair-wise comparisons were adjusted by false-discovery rate. *$p < 2.2e^{-16}$.

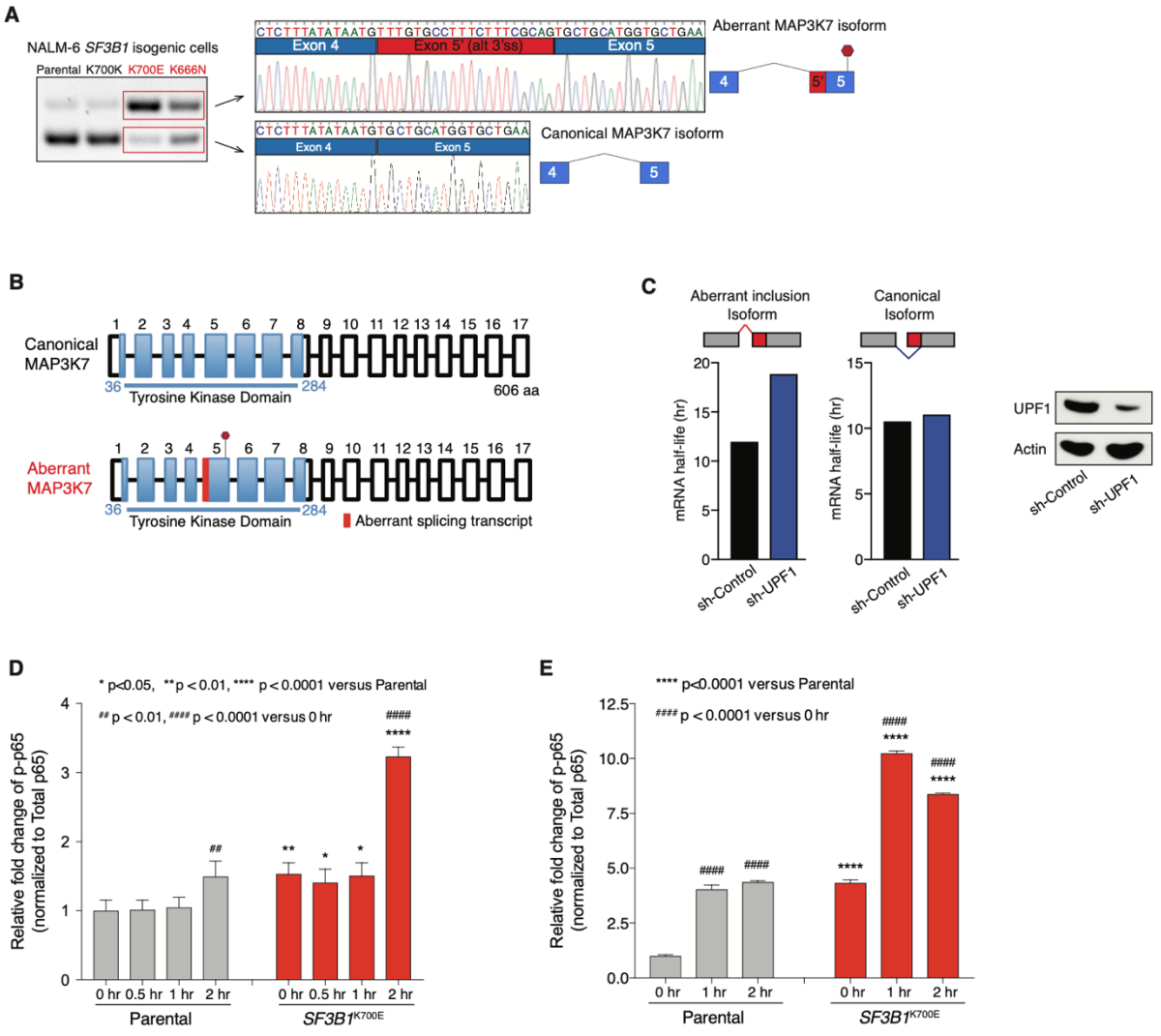


Figure 2:14 (Supplementary) SF3B1 mutations promote mis-splicing of MAP3K7 in SF3B1-mutant cells

Figure 2:14. *SF3B1* mutations promote mis-splicing of *MAP3K7* in *SF3B1*-mutant cells. (A) RT-PCR analysis and Sanger sequencing traces of the normal and aberrantly spliced *MAP3K7* isoforms in NALM6 *SF3B1* isogenic cells. **(B)** Protein diagram of MAP3K7 indicating region of aberrant 3' splice site usage in *SF3B1* mutant cells. **(C)** mRNA half-life measurements of *MAP3K7* 3' splice site inclusion (left) and exclusion (right) isoforms in *SF3B1*$^{K700E/WT}$ knockin K562 cells. UPF1 knockdown efficiency in these cells is shown. **(D)** Quantitation of immunoblot densitometry of Figure 6E.

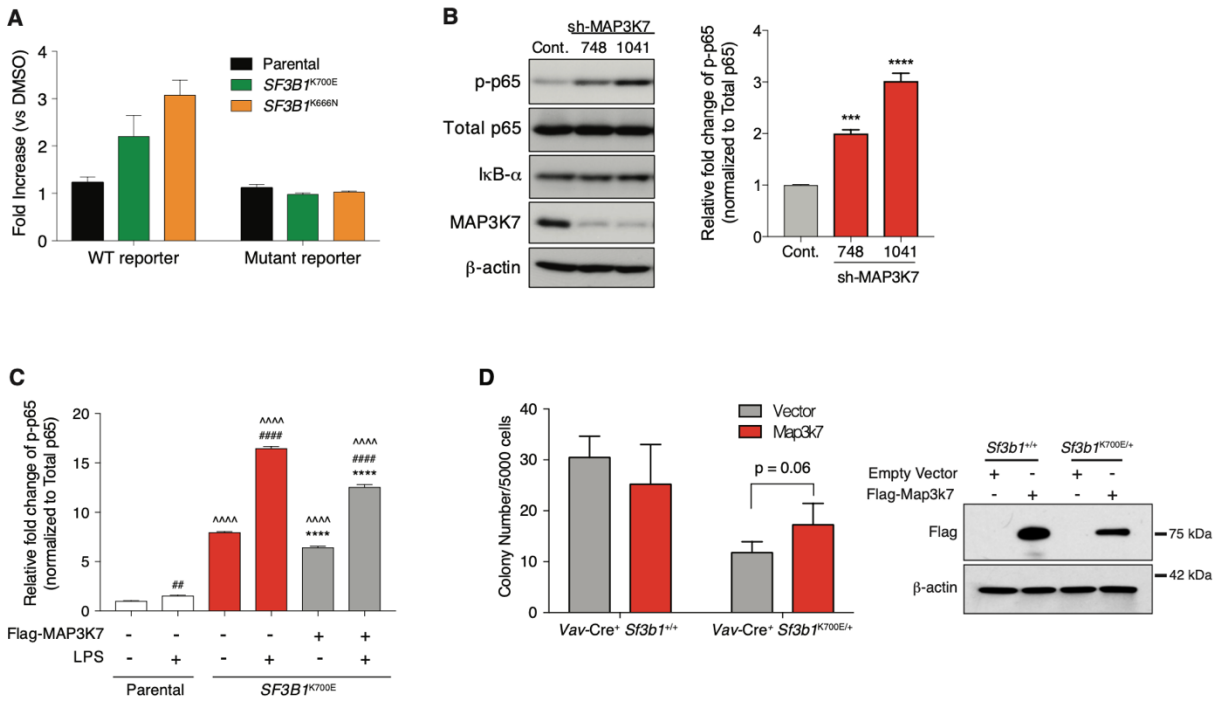
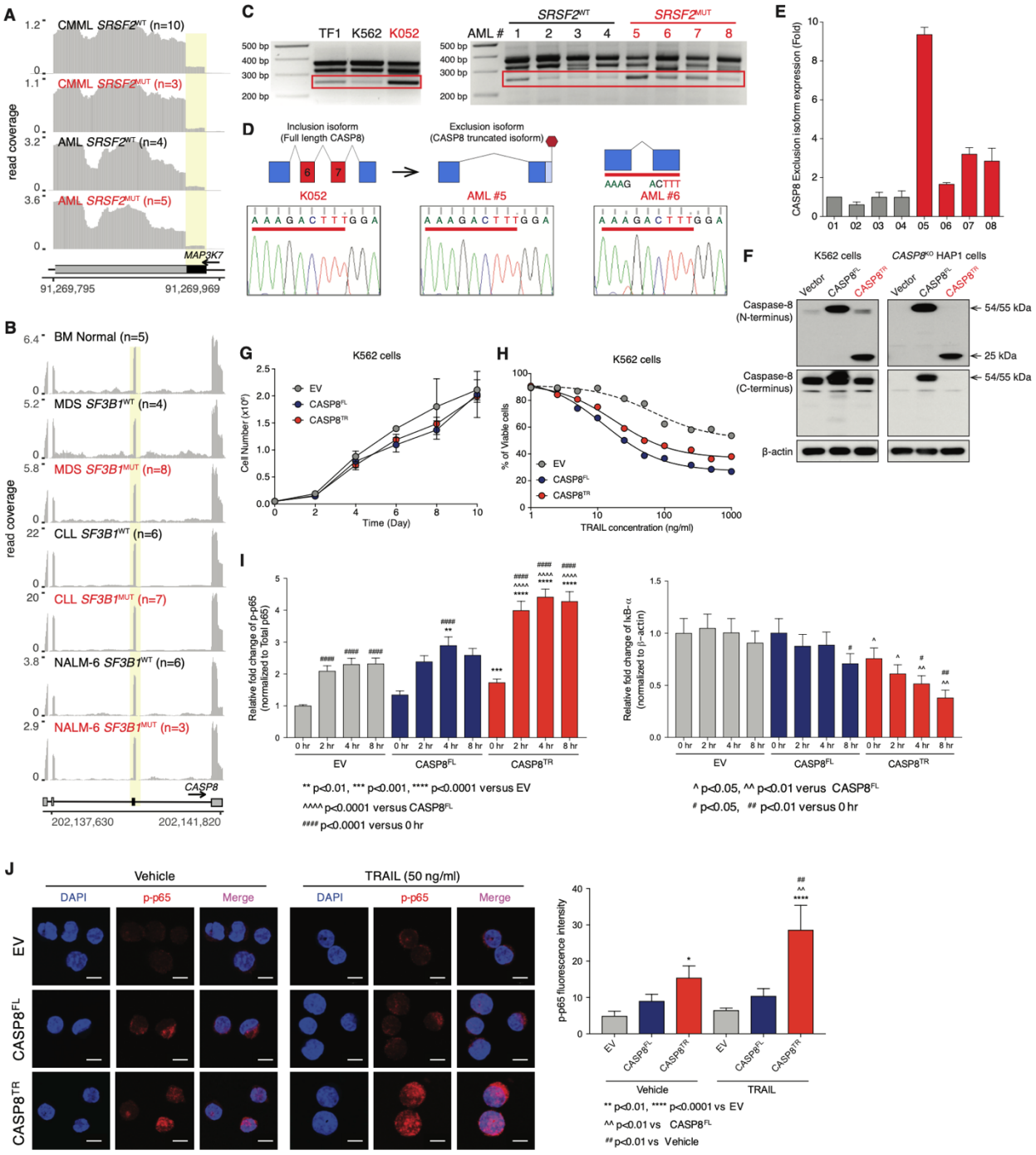


Figure 2:15 (Supplementary) MAP3K7 loss results in hyperactive NF-κB signaling in *SF3B1*-mutant cells.

Figure 2:15. MAP3K7 loss results in hyperactive NF- κB signaling. Fold-change in luciferase activity of NALM-6 isogenic cells transfected with a wildtype NF-κB luciferase reporter or a version of the reporter where all five NF-κB response elements were mutagenized. **(J)** Immunoblot of phospho-p65, IκB-□, MAP3K7, and loading controls in K562 cells treated with anti-*MAP3K7* shRNAs. Quantitation of immunoblot densitometry (bottom). **(K)** Quantitation of immunoblot densitometry of Figure 7D. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups. **(L)** Numbers of hematopoietic colonies 10-14 days after *Map3k7* cDNA expression in lineage-negative cells from *Vav*-Cre *Sf3b1*$^{K700E/+}$ and *Vav*-Cre WT E14.5 fetal liver cells (Left). Immunoblot analysis of Flag-tagged *Map3k7* cDNA from hematopoietic colonies derived from *Sf3b1*$^{+/+}$ and *Sf3b1*$^{K700E/+}$ cells (Right).


A
CMML SRSF2WT (n=10)
CMML SRSF2MUT (n=3)
AML SRSF2WT (n=4)
AML SRSF2MUT (n=5)
read coverage
MAP3K7
91,269,795
91,269,969
B
BM Normal (n=5)
MDS SF3B1WT (n=4)
MDS SF3B1MUT (n=8)
CLL SF3B1WT (n=6)
CLL SF3B1MUT (n=7)
NALM-6 SF3B1WT (n=6)
NALM-6 SF3B1MUT (n=3)
CASP8
202,137,630
202,141,820
C
TF1 K562 K052
SRSF2WT
SRSF2MUT
AML # 1 2 3 4 5 6 7 8
500 bp
400 bp
300 bp
200 bp
D
Inclusion isoform (Full length CASP8)
Exclusion isoform (CASP8 truncated isoform)
K052
AML #5
AML #6
AAAG
ACTTT
E
CASP8 Exclusion isoform expression (Fold)
F
K562 cells
CASP8KO HAP1 cells
Vector
CASP8FL
CASP8TR
Caspase-8 (N-terminus)
Caspase-8 (C-terminus)
β-actin
54/55 kDa
25 kDa
G
K562 cells
Cell Number (x10⁶)
Time (Day)
EV
CASP8FL
CASP8TR
H
K562 cells
% of Viable cells
TRAIL concentration (ng/ml)
I
Relative fold change of p-p65 (normalized to Total p65)
Relative fold change of IκB-α (normalized to β-actin)
0 hr 2 hr 4 hr 8 hr
** p<0.01, *** p<0.001, **** p<0.0001 versus EV
^^^^ p<0.0001 versus CASP8FL
p<0.0001 versus 0 hr
^ p<0.05, ^^ p<0.01 verus CASP8FL
# p<0.05, ## p<0.01 versus 0 hr
J
Vehicle
TRAIL (50 ng/ml)
DAPI
p-p65
Merge
EV
CASP8FL
CASP8TR
p-p65 fluorescence intensity
** p<0.01, **** p<0.0001 vs EV
^^ p<0.01 vs CASP8FL
p<0.01 vs Vehicle

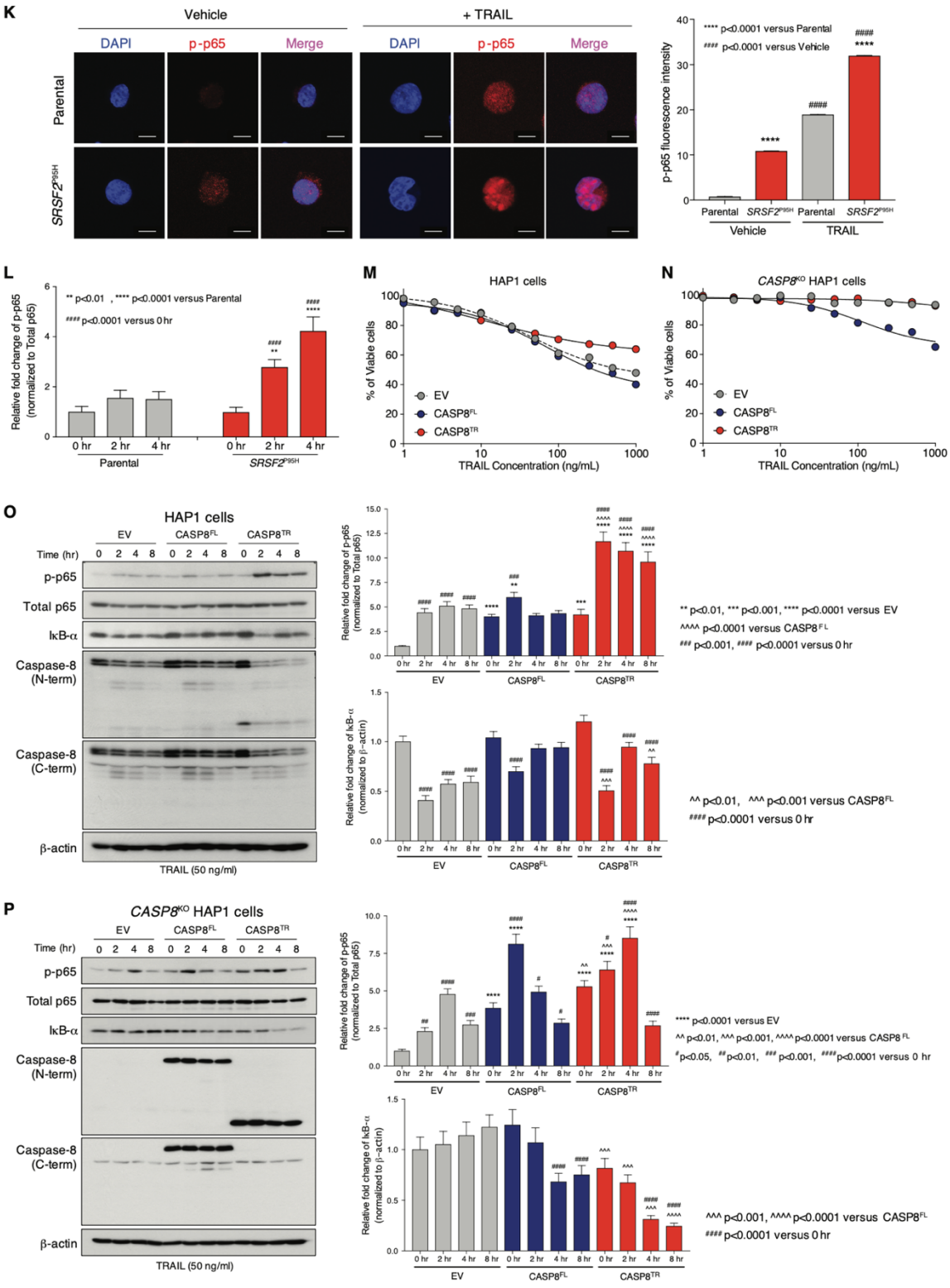


Figure 2:16 (Supplementary) *SRSF2* mutations promote aberrant splicing of caspase-8 (CASP8) resulting in expression of a stable truncated protein that hyperactivates NF-κB signaling

Figure 2:16 SRSF2 mutations promote aberrant splicing of caspase-8 (CASP8) resulting in expression of a stable truncated protein that hyperactivates NF-κB signaling. (A) RNA-seq coverage plots of *MAP3K7* in acute myeloid leukemia (AML) and chronic myelomonocytic leukemia (CMML) patient samples wild-type (WT) or mutant (MUT) for *SRSF2*. **(B)** RNA-seq coverage plots of *CASP8* in normal human bone marrow (BM) cells as well as the NALM-6 cell line and cells from myelodysplastic syndrome (MDS) and chronic lymphocytic leukemia (CLL) patients wild-type (WT) or mutant (MUT) for *SF3B1.* **(C)** RT-PCR of the *CASP8* splicing event in human leukemia cell lines and primary human AML patient samples wild-type or mutant for *SRSF2* with diagram of predicted effects of *CASP8* mRNA. The red box highlights the *CASP8* isoform promoted by mutant SRSF2. **(D)** Sanger sequencing traces of the aberrant *CASP8* isoform (*CASP8 TR*) in *SRSF2*-mutant leukemia cell line (K052) and *SRSF2*-mutant primary AML patient samples. **(E)** Quantitative RT-PCR (qRT-PCR) analysis of the aberrant *CASP8* isoform expression in *SRSF2* wild-type and *SRSF2* mutant patient AML samples. **(F)** Immunoblot analysis of caspase-8 in K562 cells (left panel) and *CASP8* knockout (KO) HAP1 cells (right panel) expressing cDNA constructs encoding empty vector (EV), full-length *CASP8* ($CASP8^{FL}$), or the aberrantly spliced *caspase-8* truncated isoform ($CASP8^{TR}$). **(G)** Growth assay of K562 cells overexpressing empty vector control (EV), full-length *caspase-8* ($CASP8^{FL}$) and the aberrant *caspase-8* truncated isoform ($CASP8^{TR}$). **(H)** Cell viability assay using CellTiter-Glo in K562 cells expressing empty vector control (EV), $CASP8^{FL}$, and $CASP8^{TR}$ isoforms 48 hours post-TRAIL stimulation. **(I)** Quantitation of Western blot densitometry of Figure 8E. **(J)** Immunofluorescence analysis of phosphorylated p65 (phospho-p65) in K562 cells from Figure 8E with quantitation of results shown on right. Scale bar: 10 μm. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups. **(K)** Immunofluorescence analysis of phosphorylated p65 in isogenic genetically modified K562 cells with or without SRSF2P95H mutation expressed from the endogenous SRSF2 locus with and

without TRAIL stimulation (50 ng/mL). Scale bar: 10 μm. **(L)** Quantitation of immunoblot densitometry of Figure 8G. Cell viability assay using CellTiter-Glo in **(M)** CASP8 wild-type (WT) and **(N)** CASP8 knockout (KO) HAP1 cells expressing EV control, CASP8FL, and CASP8TR isoforms 48 hours post TRAIL stimulation. Analysis of NF-κB signaling following TRAIL stimulation (50 ng/mL) in **(O)** CASP8 wild-type and **(P)** CASP8 KO HAP1 cells expressing EV, CASP8FL, and CASP8TR isoforms. Quantification of immunoblot signals using densitometry is shown on right. Analysis of variance followed by Tukey's post-hoc test was used to compare differences between groups.

2.7 METHODS

2.7.1 *EXPERIMENTAL MODEL AND SUBJECT DETAILS*

2.7.1.1 Animals

All animals were housed at Memorial Sloan Kettering Cancer Center (MSKCC). All animal procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSKCC. Generation and genotyping of the *Srsf2*$^{P95H/+}$, *Sf3b1*$^{K700E/+}$ and *Srsf2*$^{fl/+}$ as well as the *Mx1*-Cre and *Vav*-Cre transgenic mice have been previously described (Kim et al., 2015; Obeng et al., 2016). 8-week-old female CD45.1 C57BL/6J mice (The Jackson Laboratory) were used as recipients for bone marrow transplantation assays.

2.7.1.2 Primary human MDS and CLL samples

Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center and conducted in accordance to the Declaration of Helsinki protocol. Primary human de-identified MDS and CLL samples derived from whole peripheral blood or BM mononuclear cells were utilized.

2.7.1.3 Cell lines

The NALM-6 isogenic cell lines (NALM-6 cells engineered to express the single mutations *SF3B1*K700E or *SF3B1*K700K from the endogenous locus) were cultured in RPMI/10% FCS and K562 isogenic cell lines (engineered to express *SF3B1*K700E or *SRSF2*P95H mutations from each respective endogenous locus) were cultured in IMDM/10% FCS. HAP1 and *CASP8-KO* HAP1 cells (obtained from Horizon Discovery) were cultured in IMDM/10% FCS.

2.7.2 *METHOD DETAILS*

2.7.2.1 Peripheral blood analysis

Blood was collected by submandibular bleeding using heparinized microhematocrit capillary tubes (Thermo Fisher Scientific). Automated peripheral blood counts were obtained using a ProCyte Dx Hematology Analyzer (IDEXX).

2.7.2.2 Bone marrow transplantation assays

Primary mouse bone marrow (BM) cells were isolated from *Mx1*-Cre, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$, *Mx1*-Cre *Srsf2*$^{P95H/P95H}$ or *Mx1*-Cre *Srsf2*$^{P95H/fl}$ mice (aged 8-12 weeks) into cold PBS (without Ca^{2+} and Mg^{2+}) supplemented with 2% bovine serum albumin (BSA) to generate single cell suspensions. Red blood cells (RBCs) were removed using ammonium chloride-potassium bicarbonate (ACK) lysis buffer, resuspended in PBS/2% BSA, and filtered through a 40µm cell strainer. Total nucleated cells were quantified by the Vi-Cell XR cell counter (Beckman Coulter). For competitive transplantation experiments, a total of 1.8 x 10^6 BM cells from donor mice (*Mx1*-Cre, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$, *Mx1*-Cre *Srsf2*$^{P95H/P95H}$, *Mx1*-Cre *Srsf2*$^{P95H/fl}$, *Vav*-Cre WT, *Vav*-Cre *Srsf2*$^{P95H/+}$, and *Vav*-Cre *Sf3b1*$^{K700E/+}$ CD45.2$^+$ mice were mixed with 0.2 x 10^6 wild-type CD45.1$^+$ BM and transplanted via tail vein injection into 8-week old lethally irradiated (2x 450 cGy) CD45.1$^+$ recipient mice. For noncompetitive transplantation experiments, 2 x 10^6 total BM cells from *Mx1*-Cre, *Mx1*-Cre *Srsf2*$^{P95H/+}$, *Mx1*-Cre *Sf3b1*$^{K700E/+}$, or *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Sf3b1*$^{K700E/+}$ mice were injected into lethally irradiated (2 x 450 cGy) CD45.1$^+$ recipient mice. To induce the conditional alleles on *Mx1*-Cre background, mice were treated with 3 doses of polyinosinic:polycytidylic acid (pIpC; 12mg/kg/day; GE Healthcare) every other day via intra-

peritoneal injection. Peripheral blood chimerism of mature blood cell lineages was assessed routinely by flow cytometry.

2.7.2.3 *In vivo* LPS stimulation experiment

For *in vivo* LPS stimulation, *Escherichia coli* 055:B5 LPS (Sigma Aldrich) was used. For chronic LPS exposure, *Vav*-Cre WT, *Vav*-Cre *Srsf2*$^{P95H/+}$ and *Vav*-Cre *Sf3b1*$^{K700E/+}$ received intra-peritoneal injection of LPS (1 mg/kg) every other day for 30 days. For acute LPS exposure, *Mx1*-Cre wild-type (WT), *Mx1*-Cre *Srsf2*$^{P95H/+}$ and *Mx1*-Cre *Sf3b1*$^{K700E/+}$ mice were given a single dose of LPS (15 mg/kg) via intra-peritoneal injection.

2.7.2.4 *In vitro* colony-forming assays

Single-cell suspension was prepared from E14.5 fetal livers, and 25,000 cells from each embryo were plated in duplicates in cytokine-supplemented methylcellulose medium (MethoCult^{TM} GF M3434; StemCell Technologies), and colonies were enumerated 10-14 days later. To assess the effect of shRNA-mediated knockdown of target genes, 8-12 week-old C57BL/6 male mice were treated with 5-fluorouracil (150 mg/kg) via intra-peritoneal injection. Five days after injection, BM cells were harvested from the legs (femora and tibiae) and hip bones, and lineage-depletion was performed with biotin-conjugated antibodies against B220 (RA3-6B2), CD19 (1D3), CD3 (17A2), CD4 (GK1.5), CD8a (53-6.7), Gr-1 (RB6-8C5), Mac-1 (M1/70), NK1.1 (PK136) and Ter119, labeled with anti-bioin MicroBeads (130-090-485; Miltenyi Biotec), and lineage-negative (Lin$^-$) cells were magnetically separated using MACS columns according manufacturer's instructions. Lin$^-$ BM cells were cultured overnight in IMDM/10% FCS supplemented with mIL-3 (10 ng/mL), mIL-6 (10 ng/mL) and mSCF (50 ng/mL). The next day, cells were subjected to spinfection (2,700 rpm for 1 hour) with retroviral supernatants containing shRNAs or cDNAs of interests in the presence of polybrene (5 μg/mL). 24 hours after spinfection, cells that were successfully infected with retrovirus were marked with GFP, and were purified by flow cytometry. FACS-sorted cells were cultured in cytokine-supplemented methylcellulose medium (MethoCult^{TM} GF M3434; StemCell Technologies), and colonies were enumerated 10-14 days later.

2.7.2.5 Flow cytometry analyses

Cells were incubated with antibodies in PBS/2% BSA (without Ca^{2+} and Mg^{2+}) for 30-45 minutes on ice. For hematopoietic stem and progenitor cell analysis from adult mouse bone marrow, cells were stained with a lineage cocktail of monoclonal antibodies including B220 (RA3-6B2), CD19 (1D3), CD3 (17A2), CD4 (GK1.5), CD8a (53-6.7), Gr-1 (RB6-8C5), Mac-1 (M1/70), NK1.1 (PK136) and Ter119, allowing for mature lineage exclusion from the analysis. For fetal liver analysis, Mac-1 (M1/70) was excluded from the lineage-depletion cocktail. Cells were also stained with antibodies against c-Kit (2B8), Sca-1 (D7), FcγRII/III (93), CD34 (RAM34), CD45.1 (A20), CD45.2 (104), CD48 (HM48-1) and CD150 (9D1). DAPI was used to exclude dead cells. The composition of mature hematopoietic cell lineages in the bone marrow, spleen, thymus and peripheral blood was assessed using a combination of antibodies against Mac-1, Gr-1, B220, CD19, CD3, CD4, CD8a, CD25 (PC61.5), CD44 (IM7), IgM (Il/41), CD43 (S11). All FACS sorting was performed on FACS Aria, and analysis was performed on an LSR Fortessa (BD Biosciences). Data analysis was performed using the FlowJo software.

2.7.2.6 Cell cycle and apoptosis analyses

For apoptosis assays, freshly harvested bone marrow or fetal liver cells were first stained with antibodies against cell surface markers of interests, and then stained with FITC-conjugated Annexin V in Annexin V binding buffer (BD Pharmingen) according to manufacturer instructions. For assessment of cell cycle status in adult bone marrow HSPCs, BrdU (1 mg/kg) was administered via intra-peritoneal injection to adult mice 48 hours prior to sacrifice. For cell cycle analysis of E14.5 fetal HSPCs, BrdU (1 mg/kg) was administered to pregnant mice via intra-peritoneal injection 3 hours prior to harvesting fetal livers. Assessment of BrdU incorporation was performed following manufacturer instructions (BD Pharmingen) and data was acquired on a LSR Fortessa (BD Biosciences).

2.7.2.7 Histological analysis

Tissues, embryos and pups were fixed in 4% paraformaldehyde, processed routinely in alcohol and xylene, embedded in paraffin, sectioned at 5-micron thickness, and stained with hematoxylin-eosin (H&E). Multiple sections were obtained through the head in the coronal plane, trunk in the transverse plane, and fore and hind limbs in the longitudinal plane. Immunohistochemistry (IHC) was performed on a Leica Bond RX automated stainer (Leica Biosystems, Buffalo Grove, IL).

Following HIER at pH 6.0, the primary antibody against mouse CD45 (BD Pharmingen; 550539), myeloperoxidase (Dako; A0398), and cleaved caspase-3 (Cell Signaling; 9661) were applied at a concentration of 1:250, 1:1000, and 1:250 respectively, followed by application of a polymer detection system (DS9800, Novocastra Bond Polymer Refine Detection, Leica Biosystems) in which the chromogen was 3,3 diaminobenzidine tetrachloride (DAB) and counterstain was hematoxylin. For quantification of cleaved caspase-3 by image analysis, whole slide digital images were generated on a slide scanner (Pannoramic 250 Flash III, 3DHistech, 20x/0.8NA objective, Budapest, Hungary) at a resolution of 0.2431 μm per pixel. Staining quantification was performed with QuPath 0.1.2 software (Centre for Cancer Research & Cell Biology, Queen's University Belfast, UK). The region of interest (ROI) was defined as the liver parenchyma. The number of DAB positive cells per mm^2 was measured with the positive cell detection algorithm. ROI selection and algorithm optimization and validation, and qualitative examination of all H&E and IHC slides were performed by a board-certified veterinary pathologist (S.M.).

2.7.2.8 Immunoblot

For immunoblotting, the following antibodies were used: NF-κB/p65 (CST; 8242), phospho NF-κB/p65-Ser536 (CST; 3033), IκB-α (CST; 9242), TAK1/MAP3K7 (CST; 5206), Caspase-8, N-terminal (Abcam; clone E6), Caspase-8, C-terminal (Enzo Life Science; clone 12F5), Flag (Sigma-Aldrich; F-1804), β-actin (Sigma-Aldrich; A-5441).

2.7.2.9 Immunofluorescence

Following stimulation with LPS (Sigma-Aldrich), TNFα (PeproTech), or TRAIL (Enzo Life Science), cells were fixed with 4% paraformaldehyde/PBS for 10 minutes at room temperature (RT), permeabilized with PBS-T (PBS/1% BSA/0.2% Triton-X) for 15 minutes, blocked with PBS-T/5% goat serum (ThermoFisher PCN500) for 1 hour at RT, and incubated with primary antibodies (1:50 dilution for phospho-p65 and 1:100 for Flag) in PBS-T/5% goat serum overnight at 4°C. Cells were washed three times with PBS-T for 10 minutes at RT with gentle agitation, and were incubated with goat anti-rabbit Alexa Fluor 594 (ThermoFisher Scientific; A-11012) or goat anti-mouse Alexa Fluor 488 (ThermoFisher Scientific; A-11001) secondary antibodies (1:500 dilution in PBS-T/5/% normal goat serum) for 2 hours at RT in the dark. Cells were then washed twice with PBS-T for 10 minutes at RT, and counter-stained with DAPI (0.5 μg/mL; Sigma-Aldrich; D-9542) for 20 minutes at RT. Cells were coverslipped with ProLong Gold anti-fade

(ThermoFisher Scientific; P36930). Images were captured using either a confocal microscope (Leica TCS SP5, Upright; Leica MicroSystems), or were digitally scanned with Pannoramic Confocal Scanner (3DHistech, Budapest Hungary) using a 20x/0.8NA objective. The projected images were exported into tif format using CaseViewer software (3DHistech) and analyzed with ImageJ/FIJI. A macro was written that segments each nucleus using the DAPI channel and measures the nuclear phospho-p65 intensity within the nucleus after appropriate threshold was set.

2.7.2.10 Caspase 8 and MAP3K7 constructs

MSCV-Flag-CASP8^{FL}-IRES-GFP, MSCV-Flag-CASP8^{TR}-IRES-GFP, MSCV-Flag-hMAK3K7-IRES-GFP, MSCV-Flag-mMap3k7-IRES-GFP and MSCV-IRES-GFP empty vector constructs were used for overexpression studies. Retroviral supernatants were produced by transfecting 293 GPII cells with cDNA constructs and the packaging plasmid VSV.G using XtremeGene9 (Roche), and were used to transduce HAP1, *CASP8* KO HAP1 and K562 parental and K562 *SF3B1*K700E isogenic cells in the presence of polybrene (5 μg/mL; Millipore). Successfully transduced cells expressing GFP were purified by flow cytometry. Cells were stimulated with LPS (Sigma-Aldrich), TNFα (PeproTech) or TRAIL (Enzo Life Science).

2.7.2.11 mRNA stability assay

For mRNA half-life measurement using qRT-PCR, *UPF1* shRNA and shControl lentivirus infected K562 *SF3B1*K700E cells were treated with 2.5 μg/ml Actinomycin D (Life Technologies) and harvested at 0, 2, 4, 6, and 8 hours (following protocols established previously (t Hoen et al., 2011)). *MAP3K7* inclusion, *MAP3K7* exclusion and *18s rRNA* mRNA levels were measured by qRT-PCR.

2.7.2.12 Luciferase reporter assay

We generated K562, NALM-6 *SF3B1* isogenic cells, HAP1 and *CASP8* knockout (KO) HAP1 cells expressing the luciferase reporter for NF-κB response elements by following the manufacturer instructions (Cignal^{TM} Reporter Assay, Qiagen). Cells were stimulated with LPS, TNFα or TRAIL as described above, and NF-κB activity was assessed by luciferase intensity using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer instructions. To verify that the luciferase reporter assay was not aberrantly activated by basal leakiness, we used

another NF-κB reporter plasmids (Promega; N1111) with known NF-κB response elements (RE), and performed mutagenesis in the NF-κB-RE using the QuiKChange II Site-Directed Mutagenesis Kit (Agilent Technologies; #200522). There are five putative NF-κB-RE binding sites in this reporter plasmid (5'-GGGRNTTTCC-3', where R is a purine, Y is a pyrimidine and N is any nucleotide). To mutagenize the binding sequence, the "TTTC" sequence was mutated to "AAAA". The primers used to create the two mutant plasmids are:

Mutant-Fwd:

GGTACCTGAGCTCGCTAGCGGGAAAAAACGGGGACAAAACGGGAAAAAACGGGGA CAAAACGGGAAAAAACAGATCTGGCCTCGGCGGCCAAGCTTA

Mutant-Rev:

TAAGCTTGGCCGCCGAGGCCAGATCTGTTTTTTCCCGTTTTGTCCCCGTTTTTTCCCG TTTTGTCCCCGTTTTTTCCCGCTAGCGAGCTCAGGTACC

2.7.2.13 RT-PCR and quantitative RT-PCR (qRT-PCR)

Total RNA was isolated using RNeasy Mini kit (Qiagen). For cDNA synthesis, total RNA was reverse transcribed to cDNA with SuperScript VILO cDNA synthesis kit (Life Technologies). The resulting cDNA was diluted 10-20 fold prior to use. Quantitative RT-PCR (qRT-PCR) was performed in 10 μL reactions with either SYBR Green PCR Master Mix or Taqman Gene Expression Master Mix with AmpErase (ThermoFisher Scientific). All qRT-PCR analysis was performed on an Applied Biosystems QuantStudio 6 Flex Cycler (ThermoFisher Scientific). Relative gene expression levels were calculated using the comparative CT method.

Primers used in RT-PCR reactions were:

MAP3K7 (human) – Fwd: GATGGAATATGCTGAAGGGG, Rev: CACTCCTTGGGAACACTGTA

Map3k7 (mouse) – Fwd: GATGGAATATGCAGAGGGGG, Rev: CACTCCTTGGGAACACTGTA

CASP8 (human) – Fwd: GAACTTCAGACACCAGGC, Rev: CTTTGTCCAAAGTCTTTGCTG

Primers used in qRT-PCR reactions were:

CASP8 exclusion isoform (aberrant):

Fwd: GATGAATTTTCAAATGACTTTGGAC
Rev: TGATCAGACAGTATCCCCGAG

CASP8 inclusion isoform (canonical):
Fwd: TGATGAATTTTCAAATGGGGAGGA
Rev: ATCCTGTTCTCTTGGAGAGTCC

MAP3K7 mRNA half-life experiment (human):
Fwd (common): GCGTTTATTGTAGAGCTTCGG
Rev (canonical): GCACCATGCAGCACATTATATAAAG
Rev (aberrant): CATGCAGCACTGCGAAAGAAAG

Taqman probes were used for gene expression analysis of *TNF* (Hs00174128_m1), *IL-1β* (Hs01555410_m1), *GAPDH* (Hs02786624_g1), *Mpl* (Mm00440310_m1), *Pbx1* (Mm04207617_m1), *Itga2b* (Mm00439741_m1), and *Hprt* (Mm03024075_m1).

2.7.2.14 shRNA experiments

NALM-6 parental cells carrying the NF-κB luciferase reporter were transduced with a doxycycline-inducible lentiviral vector, T3G-dsRED-mirE-PGK-Neo-IRES-rtTA3 (Fellmann et al., 2013), expressing shRNAs for *MAP3K7* or a non-targeting renilla or firefly luciferase control. Transduced cells were selected with G418 (0.5 mg/mL; ThermoFisher Scientific), and the short hairpins were induced with the addition of doxycycline (2.0 µg/mL; Sigma Aldrich). All mouse shRNAs used in clonogenic assays were cloned into the retroviral pMSCV-LTR-mirE-PGK-SV40-IRES-GFP (MLS-E) backbone (Fellmann et al., 2013). All shRNAs were designed using the SplashRNA algorithm (Pelossof et al., 2017). The short hairpin sequences are:
sh.*MAP3K7*.748: TTAGGTAAATTTTTTATCAGTG
sh.*MAP3K7*.1041: TTTTCAACAATTTTGATTCTAA
sh.Luciferase control: TTAATCAGAGACTTCAGGCGGT
sh.*Ren*.713 control: CAGGAATTATAATGCTTATCTA
sh.*Mpl*.2121: TTATATAATAAACAGTGTCTAA
sh.*Mpl*.2368: TCAAATAAATAGATGACAGCAA
sh.*Pbx1*.824: TTCATCCAAACTCTGGTCTGTG

sh.*Pbx1*.1393: TCATTCAGAATTTCTGTGGCTT

sh.*Itga2b*.2279: TTCTCTTTCTTCTGAGTGCAGA

sh.*Itga2b*.3380: TTAGGAAAAGGGATGCACCCGG

sh.*UPF1*(TRCN0000022254): CCGGGCATCTTATTCTGGGTAATAACTCGAGTTATTACC CAGAATAAGATGCTTTTT

2.7.3 *QUANTIFICATION AND STATISTICAL ANALYSIS*

2.7.3.1 Statistical analyses

Statistical significance was determined by analysis of variance (ANOVA) after testing for normal distribution and equal variance, followed by Tukey's post-hoc test for multiple group comparisons. For non-normally distributed data, non-parametric test (Kruskal-Wallis) was used, followed by multiple group comparisons using false-discovery rate (FDR). For Kaplan Meier survival analysis, Mantel-Cox log-ranked test was used to determine statistical significance. For offspring frequency analysis, a Chi-Square test was performed to test the difference between observed and expected frequencies from different genotypes. Data were plotted using GraphPad Prism 7 software as mean values, and error bars represent standard deviation unless otherwise stated.

2.7.3.2 mRNA isolation, sequencing, and analysis

RNA was extracted from sorted mouse cell populations 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 101 bp reads per sample).

2.7.3.3 Genome annotations

Genome annotations for the human (NCBI GRCh37/UCSC hg19) and mouse (NCBI GRCm38/UCSC mm10) genomes were created as previously described (Dvinge et al., 2014). Genome annotations from the Ensembl (Flicek et al., 2013) and UCSC (Meyer et al., 2013) databases were merged with splicing event annotations from MISO v2.0 (Katz et al., 2010). An additional annotation of all possible combinations of annotated 5' and 3' splice sites found in the

merged annotation was created for read mapping. Constitutive introns were defined as those whose associated 5' and 3' splice sites were alternatively spliced in the UCSC annotation.

2.7.3.4 RNA-seq read mapping

RNA-seq reads were sequentially mapped to the transcriptome and genome as previously described (Dvinge et al., 2014). Reads were first mapped to the transcriptome using Bowtie v1.0.0 (Langmead et al., 2009) and RSEM v1.2.4 (Li and Dewey, 2011). The resulting read alignments were then filtered to require that reads spanning splice junctions overlapped the flanking exons by at least six nt. The remaining unaligned reads were then mapped to the genome and splice junctions using TopHat v2.0.8b(Trapnell et al., 2009), where reads were only allowed to align to the splice junctions present in the file of all possible combinations of annotated 5' and 3' splice sites described above. The resulting read alignments were then merged with the output of RSEM's alignment to create a final file of aligned reads.

2.7.3.5 Differential gene expression analysis

Gene expression analysis was performed using the gene expression estimates computed by RSEM in units of transcripts per million (TPM). Those estimates were then further normalized using the TMM method (Robinson and Oshlack, 2010), with a reference set of all protein-coding genes. Differentially expressed genes were defined as those with an associated Bayes factor ≥ 100 (computed using Wagenmakers's Bayesian framework (Wagenmakers et al., 2010)) and an associated fold-change ≥ 1.5.

2.7.3.6 Differential splicing analysis

Isoform ratios for annotated splicing events (cassette exons, competing 5' and 3' splice sites, and annotated retained introns) were calculated using MISO v2.0 (Katz et al., 2010). Splicing of constitutively spliced introns and junctions was quantified using only junction-spanning reads, as previously described (Hubert et al., 2013). Differential splicing in the murine data was identified by comparing samples from different genotypes for a single replicate in a pairwise fashion. The analysis was restricted to splicing events with at least 20 informative reads, where an informative read is defined as a read that distinguishes between isoforms. Differentially spliced events were defined as those with an associated Bayes factor was ≥ 5 (computed using Wagenmakers's Bayesian framework (Wagenmakers et al., 2010)) and absolute change in isoform ratio of $\geq 10\%$.

Differential splicing in the human patient cohorts was identified using a group statistical test to identify differences between patient samples with or without defined splicing factor mutations. Differentially spliced events were defined as those with an associated *p*-value ≤ 0.05 (computed using the Wilcoxon rank-sum test) and an absolute change in median per-group isoform ratio of ≥10%.

2.7.3.7 Gene Ontology (GO) enrichment analysis

GO enrichment analysis was performed using the GOseq method (Young et al., 2010) to correct for sequencing depth biases. The background set of genes was defined as all protein-coding genes. False discovery rates were calculated using the Wallenius method and corrected using the Benjamini-Hochberg method. We restricted reporting of enriched terms to those with at least two ancestors and fewer than 500 associated genes.

2.7.3.8 Motif enrichment analysis and sequence logos

The relative enrichment of different ESEs was computed by comparing motif occurrence within and adjacent to cassette exons that were promoted versus repressed in cells or samples with versus without defined splicing factor mutations. These analyses were performed using all cassette exons that were differentially spliced in at least one mouse replicate for a given genotype comparison. 95% confidence intervals were calculated by bootstrapping with 500 resampling steps. Sequence logos centered on intron-proximal or intron-distal 3' splice sites were created using all competing 3' splice sites that were differentially spliced in at least one mouse replicate for a given genotype comparison. The analysis was restricted to events with canonical GT and AG dinucleotides at the 5' and 3' splice sites. These analyses relied upon the GenomicRanges package in Bioconductor (Huber et al., 2015).

2.7.3.9 Analysis of expected and observed gene expression convergence

Expected gene expression programs for *Mx1*-Cre/*Srsf2*$^{P95H/+}$/*Sf3b1*$^{K700E/+}$ samples were calculated by computing the mean of individual gene expression levels between *Mx1*-Cre/*Srsf2*$^{P95H/+}$ and *Mx1*-Cre/*Sf3b1*$^{K700E/+}$ samples for each replicate. This method of computing the expected gene expression program was motivated by the assumption that mutations in *Srsf2* and *Sf3b1* have independent consequences for individual gene expression. Gene dysregulation for the expected

gene expression program was then calculated as described previously for the actual, observed gene expression values.

The numbers of dysregulated genes that were convergent between the expected gene expression program and the two *Mx1*-Cre/*Srsf2*$^{P95H/+}$ and *Mx1*-Cre/*Sf3b1*$^{K700E/+}$ single mutants was then determined for each replicate, again using a method identical to that for the actual, observed gene expression values (**Figure S4A**). A one-sided binomial proportion test was then used to test whether there were more dysregulated genes that were observed than expected (**Figure S4B**).

2.7.4 *DATA AND SOFTWARE AVAILABILITY*

2.7.4.1 Publicly available RNA-seq data

FASTQ files from published RNA-seq studies of patients with MDS (Dolatshad et al., 2015) and CLL (Darman et al., 2015) were downloaded from GEO series GSE63569 and GSE72790.

2.7.4.2 Accession codes

Gene Expression Omnibus: All newly generated RNA-seq data are deposited into the GEO database (accession number GSE97452).

2.8 ACKNOWLEDGEMENTS

This work was supported by the Leukemia and Lymphoma Society (S.C.-W.L., D.I., O.A.-W.), the NCI K99 CA218896 (S.C.-W.L.), Aplastic Anemia and MDS International Foundation (A.Y.), Lauri Strauss Leukemia Foundation (A.Y.), U.S. Dept. of Defense Bone Marrow Failure Research Program grant W81XWH-12-1-0041 (R.K.B., O.A.-W.), the Worldwide Cancer Research Fund (E.K.), the National Research Foundation of Korea (NRF) grant funded by the Korean government (Young Researcher Program, NRF-2017R1C1B2001991) (E.K,), the NRF grant (the Individual Research in Basic Science and Engineering program, NRF-2017R1D1A1B03034094) (E.J.), the American Society of Hematology (S.C.-W.L., J.T., B.H.D., O.A.-W.), the Edward P. Evans Foundation (R.K.B., O.A.-W.), the Taub Foundation (O.A.-W.), grant R01 HL128239 (R.K.B., O.A.-W.), the Ellison Medical Foundation grant AG-NS-1030-13 (R.K.B.), grant R01 DK103854 (R.K.B.), the Starr Foundation grant I8-A8-075 and I9-A9-059 (O.A.-W.), the Mr. William H. Goodwin and Mrs. Alice Goodwin Commonwealth Foundation for Cancer Research (O.A.-W.), and The Experimental Therapeutics Center of MSKCC (O.A.-W.). We like to acknowledge the

Molecular Cytology, Flow Cytometry, and Laboratory of Comparative Pathology Core Facilities for their technical assistance, and the MSK Cancer Center Support Grant P30 CA008748.

2.9 AUTHOR CONTRIBUTIONS

S.C.-W.L., K.N., E.K., R.K.B, and O.A.-W. designed the study. S.C.-W.L., K.N., E.K., E.J, S.X.L., B.L., D.I., M.K., M.Y., H.C., and Y.J.K. performed experiments. X.Z., M.K.K., Y.R.C. provided technical assistance. S.C.-W.L., K.N., E.K., E.J, S.X.L., B.L., D.I and A.P. analyzed data. B.H.D. and S.M. analyzed histology data. K.N., J.P., M.S., S.B., P.G.S. and R.K.B. performed RNA-seq analysis. A.Y. and J.T. provided critical feedback and discussion for this manuscript. E.O. and B.L.E. provided the *Sf3b1*K700E mutant mouse. S.C.-W.L., K.N., R.K.B, and O.A.-W. prepared the manuscript with help from all co-authors.

2.10 COMPETING FINANCIAL INTERESTS

J.P., M.S., S.B., and P.G.S. are employees of H3 Biomedicine, Inc.

2.11 REFERENCES

Ajibade, A. A., Wang, Q., Cui, J., Zou, J., Xia, X., Wang, M., Tong, Y., Hui, W., Liu, D., Su, B., *et al.* (2012). TAK1 negatively regulates NF-kappaB and p38 MAP kinase activation in Gr-1+CD11b+ neutrophils. Immunity *36*, 43-54.

Alexander, W. S., Roberts, A. W., Nicola, N. A., Li, R., and Metcalf, D. (1996). Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietic receptor c-Mpl. Blood *87*, 2162-2170.

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., *et al.* (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat Commun *7*, 10615.

Basiorka, A. A., McGraw, K. L., Eksioglu, E. A., Chen, X., Johnson, J., Zhang, L., Zhang, Q., Irvine, B. A., Cluzeau, T., Sallman, D. A., *et al.* (2016). The NLRP3 inflammasome functions as a driver of the myelodysplastic syndrome phenotype. Blood *128*, 2960-2975.

Bejar, R., Stevenson, K. E., Caughey, B. A., Abdel-Wahab, O., Steensma, D. P., Galili, N., Raza, A., Kantarjian, H., Levine, R. L., 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.

Bosman, M. C., Schepers, H., Jaques, J., Brouwers-Vos, A. Z., Quax, W. J., Schuringa, J. J., and Vellenga, E. (2014). The TAK1-NF-kappaB axis as therapeutic target for AML. Blood *124*, 3130-3140.
Carette, J. E., Raaben, M., Wong, A. C., Herbert, A. S., Obernosterer, G., Mulherkar, N., Kuehne, A. I., Kranzusch, P. J., Griffin, A. M., Ruthel, G.*, et al.* (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature *477*, 340-343.
Chaudhary, P. M., Eby, M. T., Jasmin, A., Kumar, A., Liu, L., and Hood, L. (2000). Activation of the NF-kappaB pathway by caspase 8 and its homologs. Oncogene *19*, 4451-4460.
Damm, F., Kosmider, O., Gelsi-Boyer, V., Renneville, A., Carbuccia, N., Hidalgo-Curtis, C., Della Valle, V., Couronné, L., Scourzic, L., Chesnais, V.*, et al.* (2012). Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. Blood *119*, 3211-3218.
Darman, R. B., Seiler, M., Agrawal, A. A., Lim, K. H., Peng, S., Aird, D., Bailey, S. L., Bhavsar, E. B., Chan, B., Colla, S.*, et al.* (2015). Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3? Splice Site Selection through Use of a Different Branch Point. Cell reports *13*, 1033-1045.
DeBoever, C., Ghia, E. M., Shepard, P. J., Rassenti, L., Barrett, C. L., Jepsen, K., Jamieson, C. H., Carson, D., Kipps, T. J., and Frazer, K. A. (2015). Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. PLoS computational biology *11*, e1004105.
DeDiego, M. L., Nieto-Torres, J. L., Regla-Nava, J. A., Jimenez-Guardeno, J. M., Fernandez-Delgado, R., Fett, C., Castano-Rodriguez, C., Perlman, S., and Enjuanes, L. (2014). Inhibition of NF-kappaB-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. J Virol *88*, 913-924.
Dolatshad, H., Pellagatti, A., Fernandez-Mercado, M., Yip, B. H., Malcovati, L., Attwood, M., Przychodzen, B., Sahgal, N., Kanapin, A. A., Lockstone, H.*, et al.* (2015). Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia *29*, 1092-1103.
Dvinge, H., Ries, R. E., Ilagan, J. O., Stirewalt, D. L., Meshinchi, S., and Bradley, R. K. (2014). Sample processing obscures cancer-specific alterations in leukemic transcriptomes. Proc Natl Acad Sci U S A *111*, 16802-16807.
Eftychi, C., Karagianni, N., Alexiou, M., Apostolaki, M., and Kollias, G. (2012). Myeloid TAKI [corrected] acts as a negative regulator of the LPS response and mediates resistance to endotoxemia. PloS one *7*, e31550.
Esplin, B. L., Shimazu, T., Welner, R. S., Garrett, K. P., Nie, L., Zhang, Q., Humphrey, M. B., Yang, Q., Borghesi, L. A., and Kincade, P. W. (2011). Chronic exposure to a TLR ligand injures hematopoietic stem cells. J Immunol *186*, 5367-5375.
Fang, J., Barker, B., Bolanos, L., Liu, X., Jerez, A., Makishima, H., Christie, S., Chen, X., Rao, D. S., Grimes, H. L.*, et al.* (2014). Myeloid malignancies with chromosome 5q deletions acquire a dependency on an intrachromosomal NF-kappaB gene network. Cell reports *8*, 1328-1338.
Fang, J., Bolanos, L. C., Choi, K., Liu, X., Christie, S., Akunuru, S., Kumar, R., Wang, D., Chen, X., Greis, K. D.*, et al.* (2017). Ubiquitination of hnRNPA1 by TRAF6 links chronic innate immune signaling with myelodysplasia. Nat Immunol *18*, 236-245.
Fei, D. L., Motowski, H., Chatrikhi, R., Prasad, S., Yu, J., Gao, S., Kielkopf, C. L., Bradley, R. K., and Varmus, H. (2016). Wild-Type U2AF1 Antagonizes the Splicing Program Characteristic of U2AF1-Mutant Tumors and Is Required for Cell Survival. PLoS Genet *12*, e1006384.

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D. Y., Barbosa, I. A., Kwon, J. S., Guan, Y.*, et al.* (2013). An optimized microRNA backbone for effective single-copy RNAi. Cell reports *5*, 1704-1713.
Ficara, F., Murphy, M. J., Lin, M., and Cleary, M. L. (2008). Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. Cell stem cell *2*, 484-496.
Flicek, P., Ahmed, I., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S.*, et al.* (2013). Ensembl 2013. Nucleic acids research *41*, D48-55.
Gekas, C., and Graf, T. (2013). CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age. Blood *121*, 4463-4472.
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.
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.
Harbour, J. W., Roberson, E. D., Anbunathan, H., Onken, M. D., Worley, L. A., and Bowcock, A. M. (2013). Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nature genetics *45*, 133-135.
Hernandez, L., Kim, M. K., Noonan, A. M., Sagher, E., Kohlhammer, H., Wright, G., Lyle, L. T., Steeg, P. S., Anver, M., Bowtell, D. D.*, et al.* (2015). A dual role for Caspase8 and NF-kappaB interactions in regulating apoptosis and necroptosis of ovarian cancer, with correlation to patient survival. Cell Death Discov *1*, 15053.
Hu, W. H., Johnson, H., and Shu, H. B. (2000). Activation of NF-kappaB by FADD, Casper, and caspase-8. The Journal of biological chemistry *275*, 10838-10844.
Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T.*, et al.* (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nature methods *12*, 115-121.
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.
Katz, Y., Wang, E. T., Airoldi, E. M., and Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods *7*, 1009-1015.
Kearney, C. J., and Martin, S. J. (2017). An Inflammatory Perspective on Necroptosis. Molecular cell *65*, 965-973.
Kilty, I., and Jones, L. H. (2015). TAK1 selective inhibition: state of the art and future opportunities. Future Med Chem *7*, 23-33.
Kim, E., Ilagan, J. O., Liang, Y., Daubner, G. M., Lee, S. C., Ramakrishnan, A., Li, Y., Chung, Y. R., Micol, J. B., Murphy, M. E.*, et al.* (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer cell *27*, 617-630.
Lamothe, B., Lai, Y., Hur, L., Orozco, N. M., Wang, J., Campos, A. D., Xie, M., Schneider, M. D., Lockworth, C. R., Jakacky, J.*, et al.* (2012). Deletion of TAK1 in the myeloid lineage results in the spontaneous development of myelomonocytic leukemia in mice. PloS one *7*, e51228.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.
Lasho, T. L., Jimma, T., Finke, C. M., Patnaik, M., Hanson, C. A., Ketterling, R. P., Pardanani, A., and Tefferi, A. (2012). SRSF2 mutations in primary myelofibrosis: significant clustering with IDH mutations and independent association with inferior overall and leukemia-free survival. Blood *120*, 4168-4171.
Lee, S. C., Dvinge, H., Kim, E., Cho, H., Micol, J. B., Chung, Y. R., Durham, B. H., Yoshimi, A., Kim, Y. J., Thomas, M.*, et al.* (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nature medicine *22*, 672-678.
Lee, S. T., Li, Z., Wu, Z., Aau, M., Guan, P., Karuturi, R. K., Liou, Y. C., and Yu, Q. (2011). Context-specific regulation of NF-kappaB target gene expression by EZH2 in breast cancers. Molecular cell *43*, 798-810.
Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics *12*, 323.
Makishima, H., Visconte, V., Sakaguchi, H., Jankowska, A. M., Abu Kar, S., Jerez, A., Przychodzen, B., Bupathi, M., Guinta, K., Afable, M. G.*, et al.* (2012). Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood *119*, 3203-3210.
Martin, M., Masshofer, L., Temming, P., Rahmann, S., Metz, C., Bornfeld, N., van de Nes, J., Klein-Hitpass, L., Hinnebusch, A. G., Horsthemke, B.*, et al.* (2013). Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. Nature genetics *45*, 933-936.
McNew, K. L., Whipple, W. J., Mehta, A. K., Grant, T. J., Ray, L., Kenny, C., and Singh, A. (2016). MEK and TAK1 Regulate Apoptosis in Colon Cancer Cells with KRAS-Dependent Activation of Proinflammatory Signaling. Molecular cancer research : MCR *14*, 1204-1216.
Meggendorfer, M., Roller, A., Haferlach, T., Eder, C., Dicker, F., Grossmann, V., Kohlmann, A., Alpermann, T., Yoshida, K., Ogawa, S.*, et al.* (2012). SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). Blood *120*, 3080-3088.
Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., Sloan, C. A., Rosenbloom, K. R., Roe, G., Rhead, B.*, et al.* (2013). The UCSC Genome Browser database: extensions and updates 2013. Nucleic acids research *41*, D64-69.
Mupo, A., Seiler, M., Sathiaseelan, V., Pance, A., Yang, Y., Agrawal, A. A., Iorio, F., Bautista, R., Pacharne, S., Tzelepis, K.*, et al.* (2016). Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. Leukemia.
Obeng, E. A., Chappell, R. J., Seiler, M., Chen, M. C., Campagna, D. R., Schmidt, P. J., Schneider, R. K., Lord, A. M., Wang, L., Gambe, R. G.*, et al.* (2016). Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. Cancer cell *30*, 404-417.
Okeyo-Owuor, T., White, B. S., Chatrikhi, R., Mohan, D. R., Kim, S., Griffith, M., Ding, L., Ketkar-Kulkarni, S., Hundal, J., Laird, K. M.*, et al.* (2015). U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. Leukemia *29*, 909-917.
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.
Patnaik, M. M., Lasho, T. L., Finke, C. M., Hanson, C. A., Hodnefield, J. M., Knudson, R. A., Ketterling, R. P., Pardanani, A., and Tefferi, A. (2013). Spliceosome mutations involving SRSF2, SF3B1, and U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, and prognostic relevance. Am J Hematol *88*, 201-206.
Pelossof, R., Fairchild, L., Huang, C. H., Widmer, C., Sreedharan, V. T., Sinha, N., Lai, D. Y., Guan, Y., Premsrirut, P. K., Tschaharganeh, D. F.*, et al.* (2017). Prediction of potent shRNAs with a sequential classification algorithm. Nature biotechnology *35*, 350-353.
Qian, H., Buza-Vidas, N., Hyland, C. D., Jensen, C. T., Antonchuk, J., Mansson, R., Thoren, L. A., Ekblom, M., Alexander, W. S., and Jacobsen, S. E. (2007). Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. Cell stem cell *1*, 671-684.
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.
Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology *11*, R25.
Sato, S., Sanjo, H., Takeda, K., Ninomiya-Tsuji, J., Yamamoto, M., Kawai, T., Matsumoto, K., Takeuchi, O., and Akira, S. (2005). Essential function for the kinase TAK1 in innate and adaptive immune responses. Nat Immunol *6*, 1087-1095.
Shikama, Y., Yamada, M., and Miyashita, T. (2003). Caspase-8 and caspase-10 activate NF-kappaB through RIP, NIK and IKKalpha kinases. Eur J Immunol *33*, 1998-2006.
Shirai, C. L., Ley, J. N., White, B. S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E. J., Okeyo-Owuor, T.*, et al.* (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer cell *27*, 631-643.
Shirai, C. L., White, B. S., Tripathi, M., Tapia, R., Ley, J. N., Ndonwi, M., Kim, S., Shao, J., Carver, A., Saez, B.*, et al.* (2017). Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. Nat Commun *8*, 14060.
Shu, H. B., Halpin, D. R., and Goeddel, D. V. (1997). Casper is a FADD- and caspase-related inducer of apoptosis. Immunity *6*, 751-763.
Simoes, A. E., Pereira, D. M., Gomes, S. E., Brito, H., Carvalho, T., French, A., Castro, R. E., Steer, C. J., Thibodeau, S. N., Rodrigues, C. M., and Borralho, P. M. (2015). Aberrant MEK5/ERK5 signalling contributes to human colon cancer progression via NF-kappaB activation. Cell Death Dis *6*, e1718.
Singh, A., Sweeney, M. F., Yu, M., Burger, A., Greninger, P., Benes, C., Haber, D. A., and Settleman, J. (2012). TAK1 inhibition promotes apoptosis in KRAS-dependent colon cancers. Cell *148*, 639-650.
Starczynowski, D. T. (2014). Errant innate immune signaling in del(5q) MDS. Blood *124*, 669-671.
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.

Tang, M., Wei, X., Guo, Y., Breslin, P., Zhang, S., Zhang, S., Wei, W., Xia, Z., Diaz, M., Akira, S., and Zhang, J. (2008). TAK1 is required for the survival of hematopoietic cells and hepatocytes in mice. The Journal of experimental medicine *205*, 1611-1619.
Thol, F., Kade, S., Schlarmann, C., Loffeld, P., Morgan, M., Krauter, J., Wlodarski, M. W., Kolking, B., Wichmann, M., Gorlich, K.*, et al.* (2012). Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood *119*, 3578-3584.
Thome, M., Schneider, P., Hofmann, K., Fickenscher, H., Meinl, E., Neipel, F., Mattmann, C., Burns, K., Bodmer, J. L., Schroter, M.*, et al.* (1997). Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. Nature *386*, 517-521.
Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Varney, M. E., Niederkorn, M., Konno, H., Matsumura, T., Gohda, J., Yoshida, N., Akiyama, T., Christie, S., Fang, J., Miller, D.*, et al.* (2015). Loss of Tifab, a del(5q) MDS gene, alters hematopoiesis through derepression of Toll-like receptor-TRAF6 signaling. The Journal of experimental medicine *212*, 1967-1985.
Vink, P. M., Smout, W. M., Driessen-Engels, L. J., de Bruin, A. M., Delsing, D., Krajnc-Franken, M. A., Jansen, A. J., Rovers, E. F., van Puijenbroek, A. A., Kaptein, A.*, et al.* (2013). In vivo knockdown of TAK1 accelerates bone marrow proliferation/differentiation and induces systemic inflammation. PloS one *8*, e57348.
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, L., Brooks, A. N., Fan, J., Wan, Y., Gambe, R., Li, S., Hergert, S., Yin, S., Freeman, S. S., Levin, J. Z.*, et al.* (2016). Transcriptomic Characterization of SF3B1 Mutation Reveals Its Pleiotropic Effects in Chronic Lymphocytic Leukemia. Cancer cell *30*, 750-763.
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.
Wei, Y., Dimicoli, S., Bueso-Ramos, C., Chen, R., Yang, H., Neuberg, D., Pierce, S., Jia, Y., Zheng, H., Wang, H.*, et al.* (2013). Toll-like receptor alterations in myelodysplastic syndrome. Leukemia *27*, 1832-1840.
Weinlich, R., Oberst, A., Beere, H. M., and Green, D. R. (2017). Necroptosis in development, inflammation and disease. Nat Rev Mol Cell Biol *18*, 127-136.
Xin, J., Breslin, P., Wei, W., Li, J., Gutierrez, R., Cannova, J., Ni, A., Ng, G., Schmidt, R., Chen, H.*, et al.* (2017). Necroptosis in spontaneously-mutated hematopoietic cells induces autoimmune bone marrow failure in mice. Haematologica *102*, 295-307.
Xu, Z., Tang, K., Wang, M., Rao, Q., Liu, B., and Wang, J. (2009). A new caspase-8 isoform caspase-8s increased sensitivity to apoptosis in Jurkat cells. J Biomed Biotechnol *2009*, 930462.
Yamagishi, M., Nakano, K., Miyake, A., Yamochi, T., Kagami, Y., Tsutsumi, A., Matsuda, Y., Sato-Otsubo, A., Muto, S., Utsunomiya, A.*, et al.* (2012). Polycomb-mediated loss of miR-31 activates NIK-dependent NF-kappaB pathway in adult T cell leukemia and other cancers. Cancer cell *21*, 121-135.
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.

Yoshihara, H., Arai, F., Hosokawa, K., Hagiwara, T., Takubo, K., Nakamura, Y., Gomei, Y., Iwasaki, H., Matsuoka, S., Miyamoto, K.*, et al.* (2007). Thrombopoietin/MPL signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic niche. Cell stem cell *1*, 685-697.

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome biology *11*, R14.

Yuan, R. T., Young, S., Liang, J., Schmid, M. C., Mielgo, A., and Stupack, D. G. (2012). Caspase-8 isoform 6 promotes death effector filament formation independent of microtubules. Apoptosis *17*, 229-235.

Zhang, J., Lieu, Y. K., Ali, A. M., Penson, A., Reggio, K. S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J. L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc Natl Acad Sci U S A *112*, E4726-4734.

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.

Zhao, Y., Ling, F., Wang, H. C., and Sun, X. H. (2013). Chronic TLR signaling impairs the long-term repopulating potential of hematopoietic stem cells of wild type but not Id1 deficient mice. PloS one *8*, e55552.

Zhou, Q., Derti, A., Ruddy, D., Rakiec, D., Kao, I., Lira, M., Gibaja, V., Chan, H., Yang, Y., Min, J.*, et al.* (2015). A chemical genetics approach for the functional assessment of novel cancer genes. Cancer research *75*, 1949-1958.

Chapter 3. SINGLE-CELL GENOMICS REVEALS THE GENETIC AND MOLECULAR BASES FOR ESCAPE FROM MUTATIONAL EPISTASIS IN MYELOID NEOPLASMS

This research was originally published in *Blood*. Justin Taylor*, Xiaoli Mi*, Khrystyna North*, Moritz Binder*, Alexander Penson, Terra Lasho, Katherine Knorr, Michael Haddadin, Bo Liu, Joseph Pangallo, Salima Benbarche, Daniel Wiseman, Ayalew Tefferi, Stephanie Halene, Yang Liang, Mrinal M. Patnaik, Robert K. Bradley, Omar Abdel-Wahab; Single-cell genomics reveals the genetic and molecular bases for escape from mutational epistasis in myeloid neoplasms. Blood 2020; 136 (13): 1477–1486.
*co-first authors

Contributions by Khrystyna North:

In this work I mapped, quantified and analyzed patient sample RNA-seq with different SF3B1 mutations.

The following figures display my contributions:
Figure 3:3 C, D, E
Figure 3:6 B
Figure 3:8 A

All figures shown highlight the most significant and relevant findings from the RNA-seq data I analyzed.

3.1 ABSTRACT

Large-scale sequencing studies of hematologic malignancies have revealed notable epistasis among high-frequency mutations. One of the most striking examples of epistasis occurs for mutations in RNA splicing factors. These lesions are amongst the most common alterations in myeloid neoplasms and generally occur in a mutually exclusive manner, a finding attributed to their synthetic lethal interactions and/or convergent effects. Curiously, however, patients with multiple concomitant splicing factor mutations have been observed, challenging our understanding of one of the most common examples of epistasis in hematologic malignancies. Here we performed bulk and single cell analyses of myeloid malignancy patients harboring ≥2 splicing factor mutations to understand the frequency and basis for the co-existence of these mutations. Although mutations in splicing factors were strongly mutually exclusive across 4,231 patients (q<0.001), 0.85% harbored two concomitant *bona fide* splicing factor mutations, ~50% of which were present in the same individual cells. However, the distribution of mutations in double mutants deviated from those in single mutants with selection against the most common alleles, SF3B1^{K700E} and SRSF2$^{P95H/L/R}$, and selection for less common alleles, such as SF3B1 non-K700E mutations, rare amino acid substitutions at SRSF2^{P95}, and combined U2AF1$^{S34/Q157}$ mutations. SF3B1 and SRSF2 alleles enriched in double mutants had reduced effects on RNA splicing and/or binding compared to the most common alleles. Moreover, dual U2AF1 mutations occurred in *cis* with preservation of the wild-type allele. These data highlight allele-specific differences as critical in regulating molecular effects of splicing factor mutations as well as their co-occurrences/exclusivities with one another.

3.2 INTRODUCTION

The RNA splicing machinery is frequently mutated in subjects with clonal hematopoiesis and patients with myelodysplastic syndromes (MDS), acute myeloid leukemia (AML), and myeloproliferative neoplasms (MPNs)[1-3]. Splicing factor mutations are concentrated in the core RNA splicing factor SF3B1, the accessary splicing factor SRSF2, the small subunit of the U2AF heterodimer U2AF1, and the minor spliceosome component ZRSR2, each of which has an essential role in recognition of splicing signals to ensure faithful gene expression. While

heterozygous point mutations occur at highly conserved amino acid residues of SF3B1, SRSF2, and U2AF1, conferring a change-of-function[4-6], a variety of mutations including frameshift, nonsense, and splice site changes occur throughout the X-chromosome encoded *ZRSR2*, presumably resulting in loss-of-function[7].

Intriguingly, although RNA splicing factor mutations occur in up to 45-85% of myeloid neoplasms, they consistently occur in a heterozygous state and mutually exclusive manner, such that it is rare to identify more than one RNA splicing factor mutation in an individual patient[1]. Since the discovery of these mutations, sequencing data from thousands of patients with myeloid neoplasms have reaffirmed mutual exclusivity of these mutations with great statistical significance compared to the likelihood of their co-existing by chance[8-11]. In parallel, efforts to identify the functional basis for epistasis of mutations in RNA splicing factors have shown that co-expression of the most common hotspot mutations in SF3B1 (*SF3B1*^{K700E}) and SRSF2 (*SRSF2*^{P95H}) *in vivo* in mice is intolerable to hematopoietic cells[12].

Despite these observations, several studies have noted rare patients with concurrent mutations in two RNA splicing factors in the same sample[9,10,13,14]. However, corresponding allelic frequencies of these mutations are absent from many series, precluding evaluation of whether both mutations are expressed within the same cell or not. Moreover, mutations in splicing factors may occur at multiple residues with potential distinct impacts on splicing[4,6,15], and prior studies have utilized different criteria for inclusion of variants as *bona fide* mutations. For these reasons, evaluation of known pathogenic mutant alleles (as opposed to mutant genes) where variant allele frequencies are available is required to clarify whether mutations in RNA splicing factors can actually co-exist. Furthermore, the advent of single-cell genomic technologies now provides an opportunity to definitively assess the presence of these mutations in the same cells.

Here we provide the clinical, genetic, and molecular characteristics of patients with dual splicing factor mutations in bulk malignant cell populations and at the single cell level. In so doing, we uncover enrichment of distinct alleles among patients with single versus dual splicing factor mutations and downstream effects on RNA splicing and binding, which provide a basis to explain the co-occurrence versus mutual exclusivity of mutant splicing factor alleles in patients with myeloid malignancies.

3.3 RESULTS

3.3.1 *Frequency of patients with two or more bona fide mutations in splicing factors*

We systematically annotated primary genomic DNA sequencing data for known hotspot mutations in *SF3B1*, *SRSF2*, and *U2AF1* as well as clearly deleterious mutations in *ZRSR2* across 4,231 patients with AML (n=2,822), MDS (n=1,050), or MPNs (n=359) using strict criteria for mutation calling (**Figure 3:1A,** and **Methods**). This included previously unpublished data from 1,242 patients at MSKCC as well as independent samples from six prior studies[8,10,16,17,24] (**Figure 3:4A**). Mutations in these four RNA splicing factors were seen in 22.9% of patients overall and exhibited strong mutual exclusivity (q<0.001). However, 0.85% of patients (n=36) had two concurrent splicing factor mutations (**Figure 3:4B**). No patient had mutations in three or more splicing factors. Of note, the subtypes of myeloid neoplasms present amongst dual mutant samples were reflective of the entire patient cohort (**Figure 3:4C**), suggesting that mutual exclusivity of splicing factor mutations is not related to any specific clinical subtype of myeloid neoplasm. The distribution of mutations in splicing factors across myeloid neoplasms was similar to prior studies[2,9,10,16,25] with an enrichment of SF3B1 K700 mutations in MDS over AML or MPNs and slightly higher frequency of SRSF2 mutations in AML patients over those with MPN or MDS (**Figure 3:4D**). There was also no significant difference in the tumor mutational burden in patients with one versus two mutations in splicing factors (**Figure 3:4E**).

3.3.2 *Allele-specific mutual exclusivity and co-occurrence of splicing factor mutations*

To address whether mutations in splicing factors co-exist within the same cells in patients with two mutations, we calculated the cancer cell fraction (CCF) of each splicing factor mutation based on its corresponding variant allele frequency (VAF). For this analysis, we included a cohort of 22 additional patients identified in clinical practice as having more than one splicing factor mutation for an overall cohort of 58 “double mutant” samples. Every possible combination of mutations in SF3B1, SRSF2, U2AF1, and ZRSR2 was observed and there were also patients who harbored two distinct concomitant mutations in either U2AF1 or SF3B1 in the same sample (**Figure 3:4F**).

Interestingly, in 63% of these patients, the two splicing factor mutations had a combined CCF greater than 1, suggesting that they co-exist within the same cell (**Figure 3:1B**). Overall, the CCF of mutations in *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2* within double-mutant samples mirrored that of single-mutant samples (**Figure 3:1C**). However, when we compared the frequency of mutations within each gene at the level of mutant alleles, we identified that *SF3B1*^{K700} and *SRSF2*$^{P95/P96}$ substitutions, which represent the two most common splicing factor mutations among all patients, were significantly less common in double mutants versus single mutants (**Figure 3:1D** and **Figure 3:4G-H**). For example, *SF3B1*^{K700} mutations accounted for 19% of all splicing factor mutations in single mutant patients but only 3.9% of all splicing factor mutations in dual mutant patients. Consistent with this result, the specific combination of *SF3B1*^{K700} and *SRSF2*$^{P95/P96}$ mutations occurred far less frequently than expected based on the frequencies of these mutations in the overall cohort (p=0.0013, Fisher's exact test; **Figure 3:1E**). These data demonstrate selection against co-occurrence of *SF3B1*^{K700} and *SRSF2*$^{P95/P96}$ mutations, the most frequent mutant alleles among splicing factors. At the same time, the fact that less common mutations in SF3B1 and SRSF2 can and do co-occur argues that the mutual exclusivity versus co-occurrence of these mutations occurs at the level of specific mutant alleles, rather than at a gene-specific level.

3.3.3 *Cellular-level mutual exclusivity and co-occurrence of splicing factor mutations*

Next, we sought to investigate whether mutual exclusivity of the most common splicing factor mutations and possible coexistence of less common alleles holds true at the single cell level. We performed high-throughput single-cell DNA sequencing (scDNA-seq) to quantitatively assess the clonal architecture of myeloid neoplasms harboring dual splicing factor mutations in the same sample (using the Tapestri® platform, Mission Bio, Inc.). We sequenced a total of 98,932 bone marrow mononuclear cells (BM MNCs) from 11 patients each of which bore two *bona fide* mutations in RNA splicing factors, using a custom panel of 36 amplicons across 17 genes intended to capture all mutations detected by bulk DNA sequencing in each double-mutant patient sample We sequenced a median of 7,643 BM MNCs per patient (interquartile range (IQR): 4,152-10,201) (**Figure 3:1F**). The median coverage was 192X coverage per amplicon per cell (IQR: 135-295X) (**Figure 3:1G**). The VAFs from bulk and scDNA-seq were significantly correlated (R^2 = 0.32, p=0.001), suggesting that despite variability in the data, the single cell sequencing results were

representative of bulk sequencing (**Figure 3:2A**). Target loci were genotyped in the vast majority of cells (median: 92%, IQR: 86-95%), with notable exception of *SRSF2*^{P95}, a locus prone to allele dropout (ADO)[26] (**Figure 3:5B** and **Methods**). We attempted to address ADO at *SRSF2* by jointly estimating the co-occurrence of *SRSF2, SF3B1,* and *U2AF1* mutations and ADO at both loci under the assumption that hotspot mutations in these genes are diploid and either WT or heterozygous in all cells. These are reasonable assumptions as <5% of patients had copy number alterations at regions of hotspot mutations in *SF3B1, SRSF2,* and *U2AF1* mutations in >500 myeloid neoplasm patients (**Figure 3:5C-D**) and bulk RNA-seq data consistently reveal simultaneous expression of both *SRSF2* WT and mutant alleles[8,24,27].

Our scDNA-seq analysis showed true co-occurrence of concomitant mutations in RNA splicing factors at the single cell level in multiple patients, including combined SF3B1^{K666T} and SRSF2^{P95H}, SF3B1^{H662Q} and SRSF2^{P95H}, SF3B1^{G740R} and U2AF1^{Q157R}, and U2AF1^{Q157R} and U2AF1^{S34F} mutations, amongst others (**Figures 3:1H, 3:2A-C; Figures 3:5E** and **3:6**). However, consistent with their exclusivity in bulk sequencing data, SF3B1^{K700E} and SRSF2^{P95H} mutations remained mutually exclusive at the level of individual cells (**Figures 3:1H** and **3:2A-B**). Moreover, in a male patient with two *ZRSR2* frameshift mutations (ZRSR2^{E118Dfs*28} and ZRSR2^{R290*}), the mutations existed in distinct cells- a result likely explained by the location of *ZRSR2* on chromosome X and the presumed convergent effects of loss-of-function mutations in the same gene (**Figure 3:2C**). In contrast, ZRSR2 mutations clearly co-existed within the same cells as hotspot mutations in other splicing factors (**Figure 3:2C**). We accounted for ADO under the aforementioned assumptions and found that adjusted estimates for mutant splicing factor clones closely resemble unadjusted estimates in each case (**Figure 3:7**). These data reaffirm the mutual exclusivity of SF3B1^{K700E} and SRSF2^{P95H} mutations while highlighting the potential for co-occurrence of other splicing mutant alleles at the single cell level.

We next evaluated the clonal architecture of the mutations in splicing factors and other driver genes implicated in myeloid malignancies. In samples with true co-occurrence of splicing factor mutations, we most frequently detected a dominant single splicing factor mutant clone followed by acquisition of a second splicing factor mutation within the same cells (**Figure 3:1H** and **Figure 3:5E**). This most commonly consisted of an SF3B1$^{non\text{-}K700}$ or U2AF1 mutant clone

followed by acquisition of an SRSF2 mutation. Quantitative analysis of the precise clonal architecture of the dual splicing factor mutant patients including additional driver genes showed that, in each case, pathogenic mutations in epigenetic modifiers such as *TET2, ASXL1, KDM6A,* or *DNMT3A* nearly always preceded acquisition of an RNA splicing factor mutation (**Figure 3:2** and **Figure 3:6**).

3.3.4 *Functional basis for co-occurrence of splicing factor mutations*

In order to understand the allele-specific basis for rare co-occurrence of splicing factor mutations, we next evaluated the functional impact of splicing factor mutations selected in double mutants in more detail. The most common mutational co-occurrence amongst splicing factors in dual mutant samples was the combination of SRSF2$^{P95/P96}$ and SF3B1 mutations in residues other than K700 (**Figure 3:1E**). But even in these cases, the specific mutant alleles in SRSF2 observed deviated from the most common SRSF2 mutant allele substitutions. For example, we identified one case of an *SRSF2*^{P95A} mutation in a patient with an *SF3B1*^{K666N} mutation (**Figure 3:3A**). Previous work has identified that mutations at Proline 95 in SRSF2 alter the ability of SRSF2 to physically interact with exonic splicing enhancer sequences[6,28]. While WT SRSF2 recognizes C- and G-rich sequences equally well, SRSF2$^{P95H/L/R}$ mutants have enhanced binding avidity to C-rich RNA sequences- a biochemical hallmark of pathologic substitutions at SRSF2^{P95}. However, the effect of SRSF2 P95A mutations (which are overall extremely rare) on RNA binding has not been tested in a similar assay as SRSF2$^{P95H/L/R}$ mutants. We therefore purified SRSF2's RNA recognition motif domain (RRM; as previously described[6,29] and in **Methods**) with or without P95H, L, R, or A substitutions and performed isothermal titration calorimetry (ITC) with the RNA ligand 5'-uCCAGu-3', a previously demonstrated optimal SRSF2 target according to the SSNG consensus sequence[29]. In contrast to the nearly 5-fold increase in RNA binding affinity seen with SRSF2 P95H, P95L, and P95R substitutions relative to WT SRSF2 (**Figure 3:3B**), P95A substitutions did not influence RNA binding affinity (**Figure 3:3B**). These data indicate the importance of evaluating the exact allelic substitution in splicing factor mutants and suggest that the specific splicing factor mutations present in double mutant cases may allow for escape from epistasis due to mitigated effects on RNA binding and/or splicing.

SF3B1^{K700E} exhibits strong mutual exclusivity with other splicing factor mutations, a phenomenon not recapitulated by other hotspot mutations in SF3B1. We therefore characterized differences in RNA splicing between the most common mutant alleles in SF3B1 found in myeloid neoplasms, SF3B1^{K700E} and SF3B1^{K666} substitutions, by performing RNA-seq on blood or bone marrow MNCs from a new cohort of ten SF3B1 WT, six SF3B1^{K666} mutant, and eight SF3B1^{K700E} mutant patients with MDS or AML. To differentiate whether the two groups of mutants have distinct consequences on splicing or whether the SF3B1^{K666} mutants have a less potent response, we compared the quantitative extent of mis-splicing of the events that responded most strongly to each type of SF3B1 mutation. Interestingly, this analysis revealed that although both SF3B1^{K700E} and SF3B1^{K666} mutations cause mis-splicing of a similar sets of genes, SF3B1^{K700E} mutations result in quantitatively more dramatic changes in splicing than do SF3B1^{K666} mutations. SF3B1^{K666} mutations drive only modest changes in mis-splicing of the top SF3B1^{K700E} mutation-responsive events; conversely, SF3B1^{K700E} mutations are just as capable as SF3B1^{K666} mutations of driving strong mis-splicing of the top SF3B1^{K666}-responsive events (**Figure 3:3C-E; Figure 3:8A**). These data demonstrate important allele-specific differences in splicing amongst the distinct SF3B1 mutations seen in myeloid neoplasm which may underlie the co-occurrence and exclusivity of SF3B1 mutations with mutations in other splicing factors.

In contrast to significant exclusivity of the most common mutant alleles in SRSF2 and SF3B1, U2AF1^{S34} and U2AF1^{Q157} hotspot mutations co-occurred at a higher rate than expected by chance alone (**Figure 3:1E**). Moreover, scDNA-seq analysis of >6,000 cells from a patient harboring both U2AF1^{S34F} and U2AF1^{Q157} mutations clearly revealed that both mutations occurred in the same cells. While we hypothesize that the U2AF1 mutations occurred around the same time, as virtually all of the U2AF1 mutant cells carried both mutations (**Figure 3:3F**), it is also conceivable that dual mutants may have a significant proliferative advantage over single mutants such that they became the only detectable clone over time. Evaluation of the clonal architecture of this patient based on sequencing of the five mutations identified in bulk DNA identified that initiating mutations in *KDM6B* followed by *DNMT3A* were followed by acquisition of the dual U2AF1 mutant clone (**Figure 3:3G-H**).

Co-occurrence of multiple U2AF1 mutations is surprising given prior data suggesting the requirement for expression of a WT allele in the presence of a hotspot mutation in U2AF1 for cell survival[30]. We therefore evaluated whether co-occurring U2AF1^{S34F} and U2AF1^{Q157R} mutations were present on the same allele or different alleles using long-range RT-PCR spanning exons 2-7 of *U2AF1* (which encode the S34 and Q157 residues of U2AF1) followed by subcloning and sequence analysis of individual clones. All clones that emerged from this analysis were either entirely WT for *U2AF1* or contained both *U2AF1*^{S34F} and *U2AF1*^{Q157} mutations (**Figure 3:3I**). In addition, there are distinct mRNA isoforms for U2AF1, one of which includes a premature termination codon that targets the resulting U2AF1 mRNA for nonsense-mediated mRNA decay.[31] However, *U2AF1*$^{S34F/Q157}$ double-mutant clones did not differ from U2AF1 WT clones in their usage of U2AF1 isoforms (**Figure 3:8B**). Consistent with these findings, *U2AF1*$^{S34F/Q157}$ double-mutant patients harbored differential cassette exon splicing events characteristic of both the *U2AF1*^{S34F} and *U2AF1*^{Q157} single-mutant state (**Figure 3:3J**). While these data indicate that dual U2AF1 hotspot mutations occurred in *cis* with preservation of one WT allele in this sample, further evaluation of additional samples will be important to understand if co-existing *U2AF1*$^{S34F/Q157}$ hotspot mutations are tolerated in *trans*.

3.4 Discussion

Genomic analyses of patients with myeloid malignancies have successfully identified not only individual genes which are recurrently mutated, but also groups of genes which have important genetic interactions with one another. Notable groups of mutations which display statistically significant mutual exclusivity in patients with MDS and AML include mutations in TET2 and IDH1/2 enzymes[32], cohesin subunits, and RNA splicing factors (SF3B1, SRSF2, U2AF1 and ZRSR2)[1]. The most common explanation for such epistatic interactions is that mutations are mutually exclusive due to either redundant effects of each mutation and/or intolerability of co-expression of these mutations within the same cell. Studying the basis for these genetic interactions has elucidated novel disease biology. For example, discovery of the mutual exclusivity of TET2 and IDH1/2 mutations in AML led in part to the identification of the convergent effects of these mutations on DNA cytosine modification[32]. We focused here on evaluating the clinical and genetic characteristics of patients with ≥2 simultaneous mutations in RNA splicing factors to understand

the basis for rare coexistence of these mutations. Such patients have been anecdotally noted in a number of studies but the frequency, characteristics, and basis for the existence of such cases have not been systematically studied.

The existence of double splicing factor mutant patients appears counter to the mutual exclusivity of these mutations in the vast majority of patients[1], prior functional data demonstrating intolerability of the most common splicing factor mutations[12], and the known requirement of WT splicing in cells carrying hotspot mutations in splicing factors[12,30,33,34]. However, our data uncover that mutual exclusivity of mutations in RNA splicing factors occurs at the level of specific mutant alleles, dependent on the precise amino acid substitutions in SF3B1, SRSF2, and U2AF1, rather than at the gene level. In other words, mutations in RNA splicing factors can and do coexist in the same individuals and even in the same cells, but only combinations involving less common mutant alleles are identified in such individuals.

Allele-specific differences in RNA splicing factor mutations make sense in light of the fact that distinct hotspot mutations in RNA splicing factors are known to have unique disease manifestations and, in some cases, are known to have unique effects on RNA splicing and gene expression. For example, mutations at the S34 and Q157 residues of U2AF1 have distinct sequence-specific effects on cassette exon usage and mis-spliced target genes[4]. Consistent with these non-overlapping effects of U2AF1 S34 and Q157 mutations, we uncovered significant co-enrichment of U2AF1 S34 and Q157 mutations in rare patients and this co-mutation was enriched even within individual cells. This finding suggests potential cooperative interaction of U2AF1 S34 and Q157 mutations, a hypothesis interesting to examine in future studies. Nonetheless, $U2AF1^{S34/Q157}$ dual mutations occurred in *cis* with preservation of the WT allele, in line with prior work demonstrating requirement of the WT allele in U2AF1 mutant cells[30].

In accordance with the concept of allele-specific regulation of mutational co-occurrences, specific combinations of splicing factor mutations exhibit significant mutual exclusivity. In fact, the most common mutation substitutions in SF3B1 ($SF3B1^{K700E}$) and SRSF2 ($SRSF2^{P95H/L/R}$) were substantially selected against at the bulk cellular level as well as within individual cells. In one particularly illustrative example, we found that one patient who ostensibly appeared to have co-

existing SRSF2^{P95} and SF3B1 mutations, actually had selection for an extremely rare mutation in SRSF2 (SRSF2^{P95A}), which although located at the commonly mutated residue in SRSF2, did not alter the ability of SRSF2 to bind RNA. In contrast, SF3B1 hotspot mutations outside of K700 frequently co-existed with other splicing factor mutations. This is likely due to allele-specific effects of distinct SF3B1 mutations on RNA splicing and gene expression, as has been demonstrated for mutant U2AF1 noted above.

SF3B1 mutations have been suspected to have allele-specific differences given the enrichment of specific SF3B1 mutations in distinct subtypes of cancer. For example, SF3B1^{R625} substitutions are common in melanomas but rare in other cancers[35] and SF3B1^{G742} mutations are most common in CLL[36]. While these findings suggest allele-specific differences across distinct SF3B1 mutant hotspots, these differences have not been elucidated. In fact, one recent study identified adverse outcome amongst MDS patients with SF3B1^{K666} mutations specifically[37], however global analyses of splicing were absent from this study (and the few splicing changes suggested as specific to SF3B1^{K666N} genotype are actually evident in SF3B1^{K700E} samples). Here, we found that K700E and K666 mutations in SF3B1, which represent the two most commonly mutated SF3B1 residues in myeloid neoplasms, have significant differences in their global effects on splicing. Compared to SF3B1^{K700E}, SF3B1^{K666} mutations have less dramatic effects on splicing, which may account for intolerability of the former, and tolerability of the latter, with other splicing factor mutations. Although both K666 and K700 residues are located in the HEAT domain of SF3B1, recent structural analyses of human SF3B1 highlighted key functional differences between these locations[38]. The K666 residue is involved in intramolecular hydrogen bonds within SF3B1 to maintain its tertiary structure. In contrast, the K700 residue is exposed on the surface of SF3B1 and appears to be involved in the interaction of SF3B1 with other U2 snRNP components[38]. Thus, mutations in K666 and K700 are expected to create distinct structural disturbances in SF3B1 which likely account for their subsequent different effects on RNA splicing. Strong data on the clinical and morphologic associations of individual mutations in RNA splicing factors in myeloid malignancies necessitates a closer future evaluation of the potential effects of combined mutations of RNA splicing factors on disease characteristics such as morphologic features, blood counts, and outcome.

The discovery that cells containing hotspot mutations in RNA splicing factors are intolerant to additional genetic perturbations in splicing led to a therapeutic effort to modulate splicing as a novel form of therapy for splicing factor mutant leukemias.[31] However, the allele-specific effects of splicing factor mutations on vulnerability to additional genetic perturbations to splicing identified here suggest the possibility that response to such therapies may vary based on the exact mutant allele present. For example, cells bearing $SF3B1^{K700E}$ and $SRSF2^{P95H/L/R}$ substitutions might be more sensitive to splicing modulatory drugs than $U2AF1^{S34}$ or $U2AF1^{Q157}$ mutations or rarer mutant alleles in SRSF2 or SF3B1 which have less prominent effects on RNA splicing than K700E or P95H/L/R substitutions. It is also possible that ZRSR2 mutant cells may be less responsive to such therapies given that ZRSR2 differs from other leukemia-associated mutant splicing factors in that it functions in the minor spliceosome. These hypotheses should be considered in future preclinical as well as clinical studies testing drugs which globally perturb RNA splicing.

One frequent reason for co-occurrence of mutations that are normally mutually exclusive is distribution of such mutations to distinct subclones. Studying the clonal architecture of such cases has recently been instrumental in understanding novel mechanisms of resistance to mutationally targeted therapies. For example, while mutations in IDH1 and IDH2 typically do not co-occur, resistance to IDH2 inhibitors can develop due to clonal outgrowth of IDH1 subclones[39]. Similarly, FLT3 inhibitor resistance has been shown in some cases to arise from clonal outgrowth of RAS mutant cells. One can envision then that prospective therapies targeting specific RNA splicing factor mutations could potentially select for outgrowth of cells bearing other splicing factor mutations that have convergent effects on cell survival.

3.5 FIGURES AND FIGURE LEGENDS

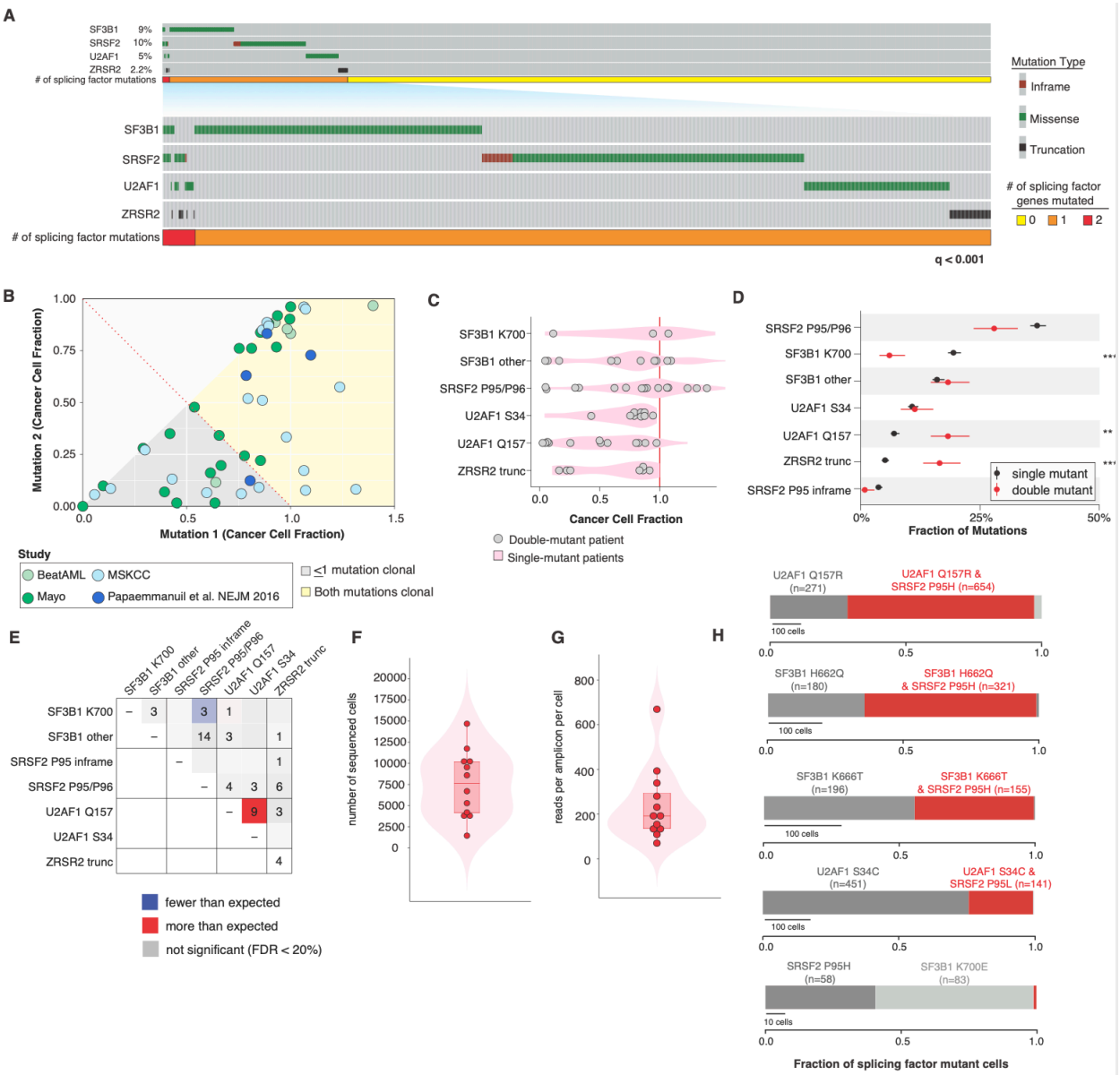


Figure 3:1 Genetic features of patients harboring two concomitant mutations in RNA splicing factors at the bulk and single cell level.

Figure 3:1. Genetic features of patients harboring two concomitant mutations in RNA splicing factors at the bulk and single cell level. (A) Oncoprint of hotspot mutations in *SF3B1, SRSF2*, an *U2AF1* as well as clearly deleterious mutations (nonsense or frameshift mutations) in *ZRSR2* across 4,231 patients. Each column represents a single individual patient. The number of patients with zero, one, or two splicing factor mutations is shown in yellow, orange, and red,

respectively. Overall, mutations in each gene exhibited strong mutual exclusivity ($q<0.001$; Fisher's exact test). (B) Cancer cell fraction (CCF) of each mutant splicing factor from genomic DNA sequencing of a cohort of 58 dual splicing factor mutant samples. (C) CCF of mutations at $SF3B1^{K700}$, other residues of SF3B1 ("SF3B1 other"), $SRSF2^{P95/P96}$, $U2AF1^{S34}$, and $U2AF1^{Q157}$ as well as ZRSR2 truncation mutations ("ZRSR2 trunc"). (D) Percentage of single and double splicing factor mutant patients (in black and red respectively) with mutations in SRSF2, SF3B1, U2AF1, and ZRSR2. Error bars show one standard deviation, based on a binomial distribution. **$p<0.005$, ***$p<0.0005$ (Fisher's exact test). (E) Plot describing number of patients with co-existing mutant alleles in splicing factors. The expected number was based on the fraction of samples with exactly two mutations under the assumption of no mutual exclusivity and using a Poisson distribution. Distribution of the number of (F) total sequenced cells per patient and (G) reads per amplicon per cell from single cell genomic DNA sequencing. Each point represents a sample from a unique patient. (H) Fraction of mutated cells with one or two mutations in RNA splicing factors within each dual splicing factor mutant patient. Red bar denotes fraction of individual cells where two splicing factor mutations were identified within the same cell. The number of cells containing each mutation is indicated.

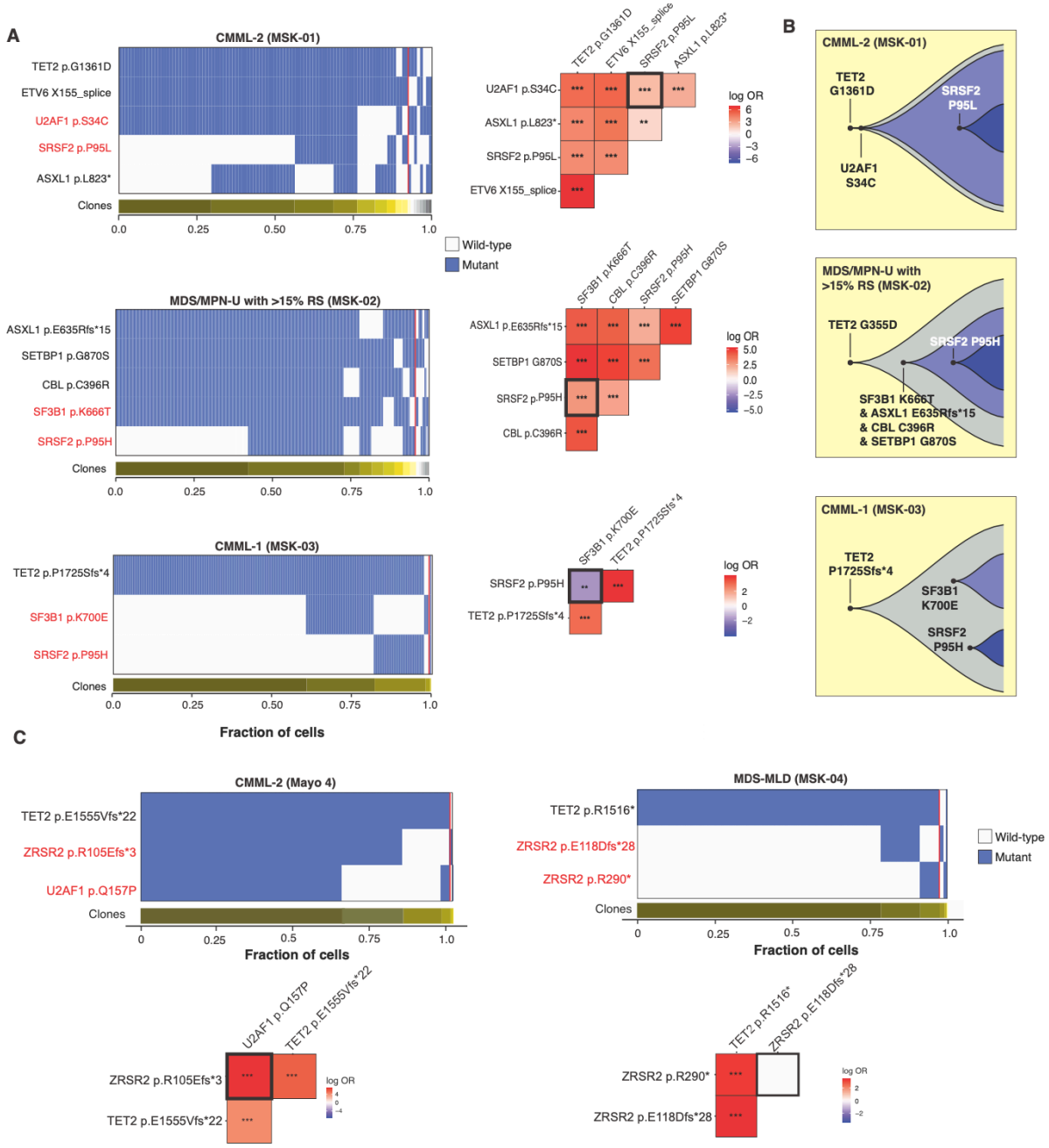
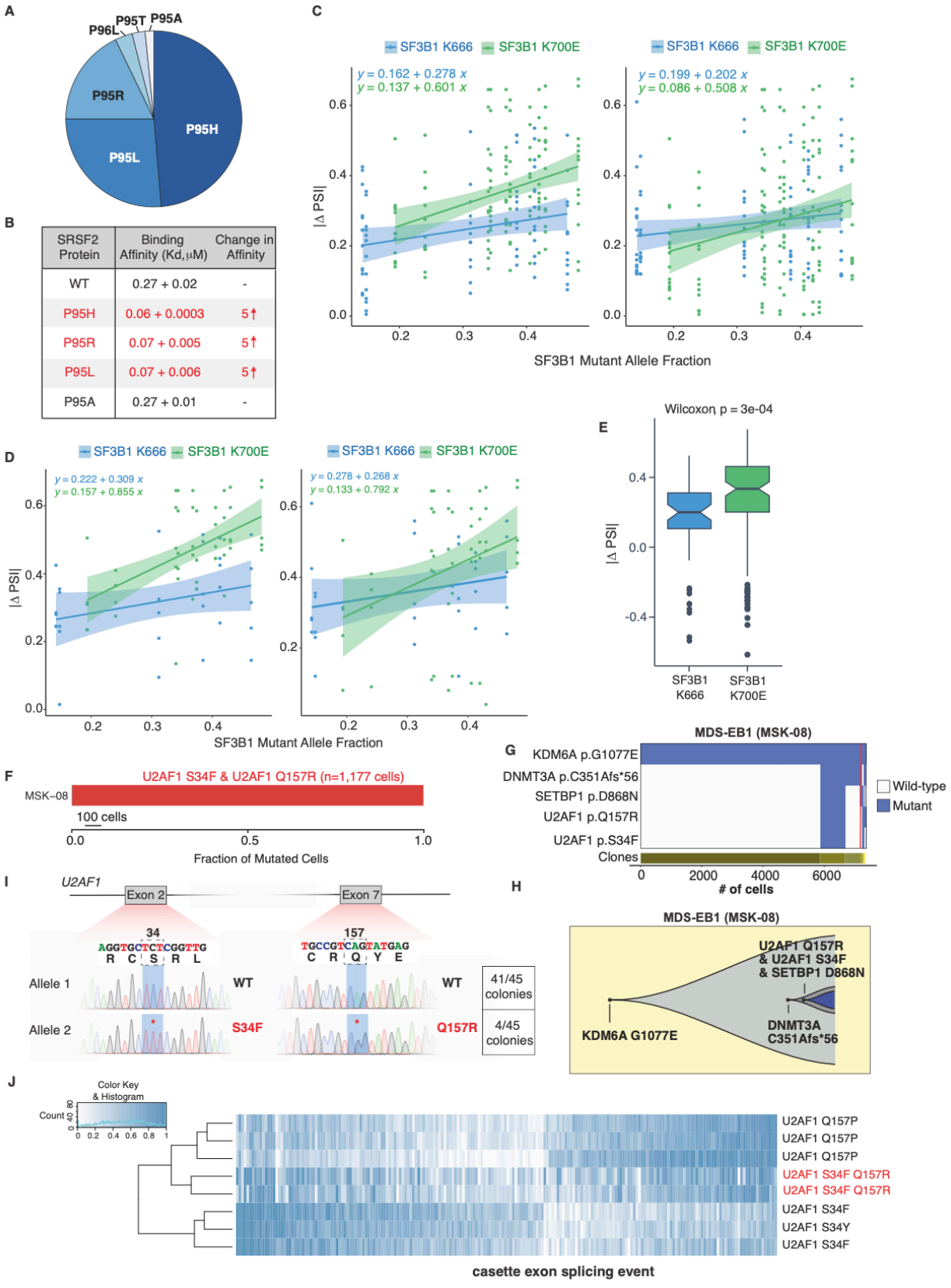


Figure 3:2 Allele- specific mutational co-occurrences in RNA splicing factor mutations

Figure 3:2. Allele-specific mutational co-occurrences in RNA splicing factor mutations. (A) Oncoprint indicating cellular-level co-occurrence of mutations in select double-splicing factor mutant patient cases (the clinical diagnosis and sample ID is listed above each Oncoprint). Each column in the heat map (left) depicts an individual cell with the genotype of each sequenced cell for each variant. Clustering is based on the genotypes of driver mutations and a horizontal bar depicts the detected clones in each case. Mutant and wild-type cells are indicated in blue and white,

respectively. The subclones located to the right of the red line comprised <1% of the total sequence cells, since such small subclones can represent false positive or negative genotypes as a result of allele dropout or multiplets. The figures on the right show the pairwise association of mutations. The color and size of each panel represent the degree of the logarithmic odds ratio (log OR). The vertical bar indicates the association of the colors with the log OR. Co-occurrence and mutual exclusivity are indicated by red and blue, respectively. The statistical significance of the associations based on the false discovery rate (FDR) is indicated by the asterisks (*FDR < 0.1, **FDR < 0.05, ***FDR < 0.001). (B) Fish plots showing the inferred clonal hierarchy based on the single-cell genotype data for the 3 patients in (A). (C) Oncoprint as in (A) but evaluating cellular co-occurrence or mutual exclusivity of deleterious ZRSR2 mutations with mutations in other splicing factors (left) or with one another (right). Abbreviations: CMML: chronic myelomonocytic leukemia; MDS-MLD: myelodysplastic syndrome with multilineage dysplasia; MDS/MPN-U with >15% RS: MDS/myeloproliferative neoplasm unclassified with >15% ring sideroblasts.



SRSF2 Protein	Binding Affinity (Kd, μM)	Change in Affinity
WT	0.27 + 0.02	-
P95H	0.06 + 0.0003	5↑
P95R	0.07 + 0.005	5↑
P95L	0.07 + 0.006	5↑
P95A	0.27 + 0.01	-

Figure 3:3 Allele-specific effects on RNA binding and splicing in splicing factor mutations seen in patients harboring two concomitant mutations in splicing factors

Figure 3:3. Allele-specific effects on RNA binding and splicing in splicing factor mutations seen in patients harboring two concomitant mutations in splicing factors. (A) Pie chart of SRSF2 P95 amino acid substitutions across the entire cohort. (B) Binding affinities of WT versus P95H/L/R/A mutant SRSF2 peptides to UCCAGU RNA oligonucleotides as absolute Kd values. The column labeled "change in affinity" provides the Kd ratio of mutant:WT peptide. (C) Comparison of the quantitative effects of $SF3B1^{K700E}$ and $SF3B1^{K666N}$ mutations on splicing, stratified by mutant allele fraction. Each point illustrates the absolute change in isoform usage (ΔPSI) for one of the top 20 most mis-spliced events associated with each mutation. For each panel, the top 20 most mis-spliced events were computed using only samples with $SF3B1^{K700E}$ (left) or $SF3B1^{K666N}$ (right) mutations. Mis-splicing of those 20 events was then computed for all samples, irrespective of mutation, and plotted as illustrated. $SF3B1^{K700E}$ and $SF3B1^{K666N}$ mutations cause mis-splicing of similar sets of genes, but $SF3B1^{K700E}$ mutations cause more dramatic changes. Lines, shading, and equations indicate the best-fit linear regressions and corresponding 95% confidence intervals. (D) As (C), but computed using the top 5 most mis-spliced events for each mutation ($SF3B1^{K700E}$, left; $SF3B1^{K666N}$, right). (E) Box plot illustrating the data from (C) and associated *p*-value, computed using a two-sided Wilcoxon rank-sum test. (F) Fraction of $U2AF1^{S34F}$, $U2AF1^{Q157R}$, or dual $U2AF1^{S34F/Q157R}$ mutated cells from a patient harboring both $U2AF1^{S34F/WT}$ and $U2AF1^{Q157R/WT}$ mutations. Red bar indicates fraction of $U2AF1^{S34F/Q157R}$ dual mutant cells. (G) Clonal hierarchy of mutations in the patient from (E). Each column represents a cell at the indicated scale as in (A). Cells with mutations and wild-type cells are indicated in blue and white, respectively. (H) Fish plots showing the inferred clonal hierarchy based on the single-cell genotype data from (G). (I) Sanger sequencing electropherograms from representative single cell clones from the patient in (F). As enumerated on right, all colonies were either *U2AF1* dual WT or $U2AF1^{S34F/Q157R}$ dual mutant indicating that these mutations always occur in *cis* with preservation of the WT allele. (J) Heatmap of percent spliced in values of cassette exons in U2AF1S34, Q157, and U2AF1S34/Q157 dual mutant patients displaying cassette exon splicing events specific to U2AF1S34 or Q157 single mutant state (standard deviation of <0.2 amongst single mutants and mean(U2AF1 S34) - mean(U2AF1 Q157) < 0.32). Each row is a unique patient and each column a single splicing event.

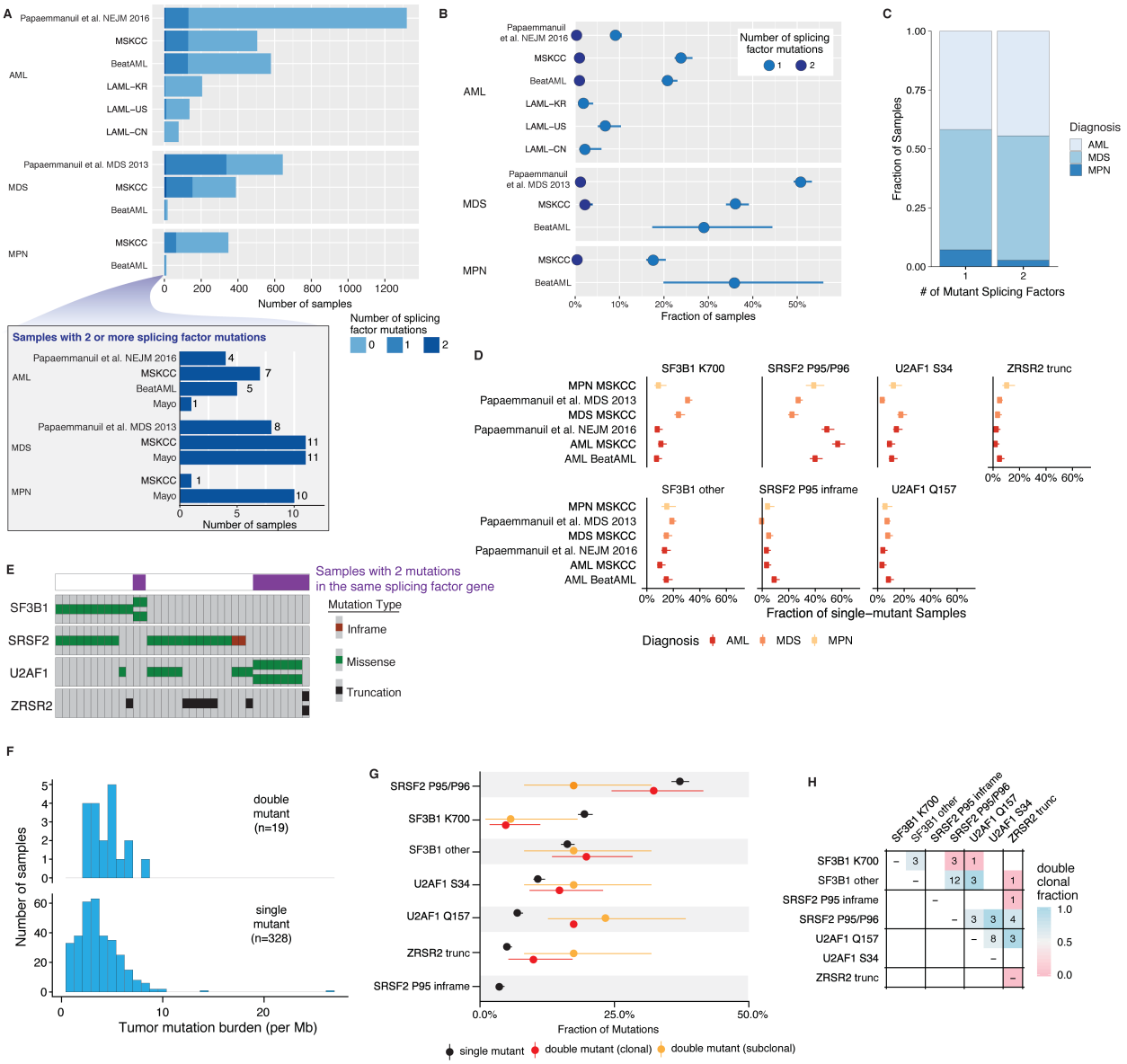


Figure 3:4 (Supplementary) : Characteristics of study cohort and genetic alterations in splicing factors analyzed

Figure 3:4. Characteristics of study cohort and genetic alterations in splicing factors analyzed. (A) Histogram of patients with acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), and myeloproliferative neoplasms (MPNs) with zero, one, or two mutations in RNA splicing factors studied here. Inset shows exact numbers of patients with two mutant splicing factors. **(B)** Percentage of samples within each study that contain one or two mutations in RNA splicing factors. **(C)** Fraction of samples with diagnosis of AML, MDS, or MPN based on presence of one versus two RNA splicing factor mutations. **(D)** Percentage of samples amongst single

mutants that contained mutations at the SF3B1^{K700}, SRSF2$^{P95/P96}$, U2AF1^{S34}, or U2AF1^{Q157} residues, other residues of SF3B1 ("SF3B1 other"), inframe deletions or insertions around the SRSF2^{P95} residue ("SRSF2 P95 inframe"), and ZRSR2 truncating mutations ("ZRSR2 trunc"). **(E)** Oncoprint of patients with two splicing factor mutations with annotation to describe patients with >1 mutation within the same gene. **(F)** Histogram of tumor mutational burden in patients with two or more (top; "double mutant") mutant splicing factors versus a single mutant splicing factor (bottom). **(G)** Percentage of patients with mutations in SRSF2, SF3B1, U2AF1, and ZRSR2 as a single mutation (in black) or dual mutation where ≥1 mutation was subclonal (orange) or both were clonal (red). Error bars show one standard deviation, based on a binomial distribution. **(H)** Plot describing number of patients with co-existing mutant alleles in splicing factors restricted to double-mutant patients where both mutations were clonal. The expected number was based on the fraction of samples with exactly two mutations under the assumption of no mutual exclusivity and using a Poisson distribution.

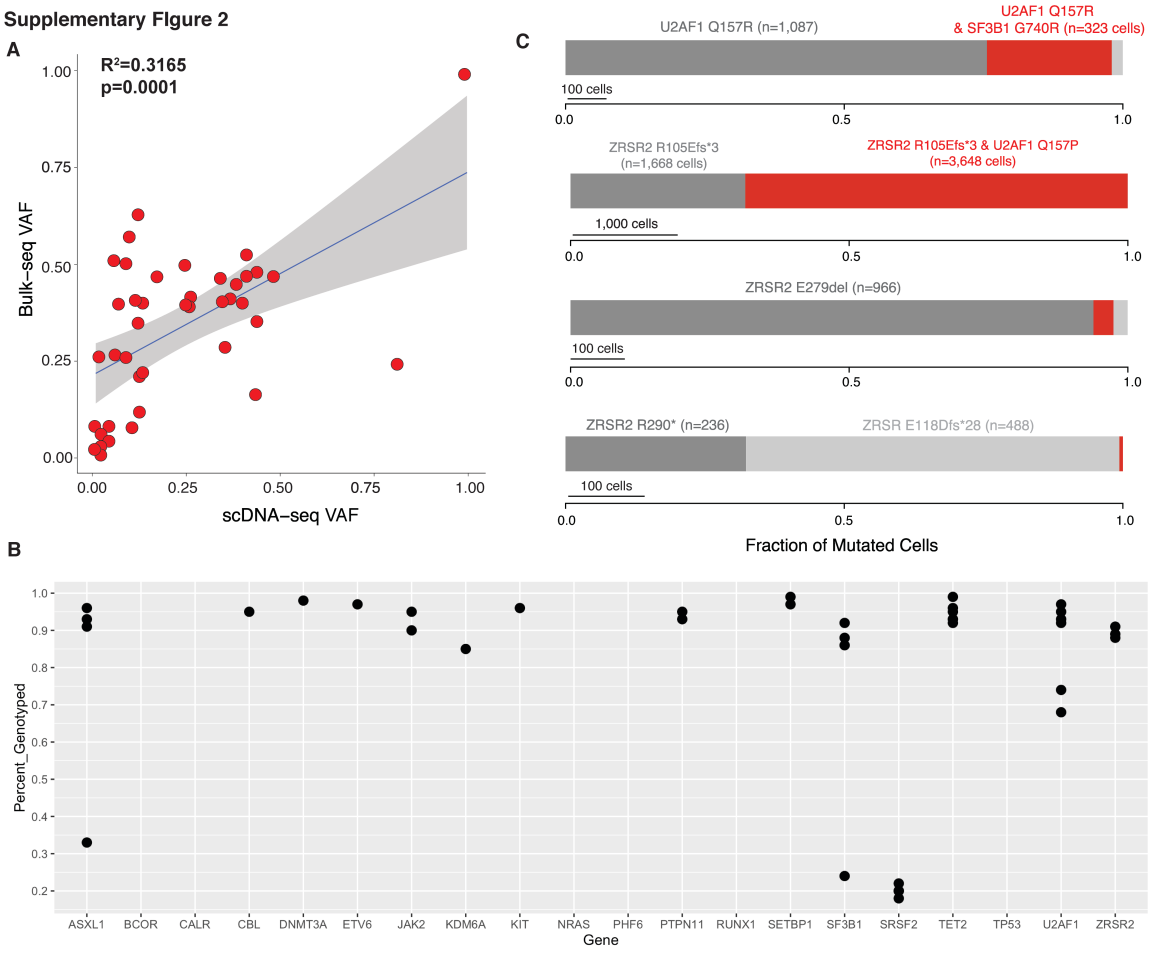


Figure 3:5 (Supplementary) Single cell genomic analysis of myeloid neoplasm patients bearing two mutations in RNA splicing factors.

Figure 3:5. Single cell genomic analysis of myeloid neoplasm patients bearing two mutations in RNA splicing factors. (A) Correlation of the variant allele fraction (VAF) from bulk sequencing and single cell DNA sequencing. The X-axis shows the VAF from the single-cell genotype data (scDNA-seq VAF). The Y-axis shows the VAF from the bulk next-generation sequencing (bulk VAF). Each dot represents a detected variant. The linear trendline was added to best fit the distribution of the dots. The shaded area around the trendline represents the 95% confidence intervals. (B) Genotyping efficiencies from scDNA-seq for each gene sequenced. Each dot represents a detected variant. (C) Fraction of mutated cells with one or two mutations in RNA splicing factors within each unique dual splicing factor mutant patient. Red bar denotes fraction of individual cells where two mutations were identified within the same cell.

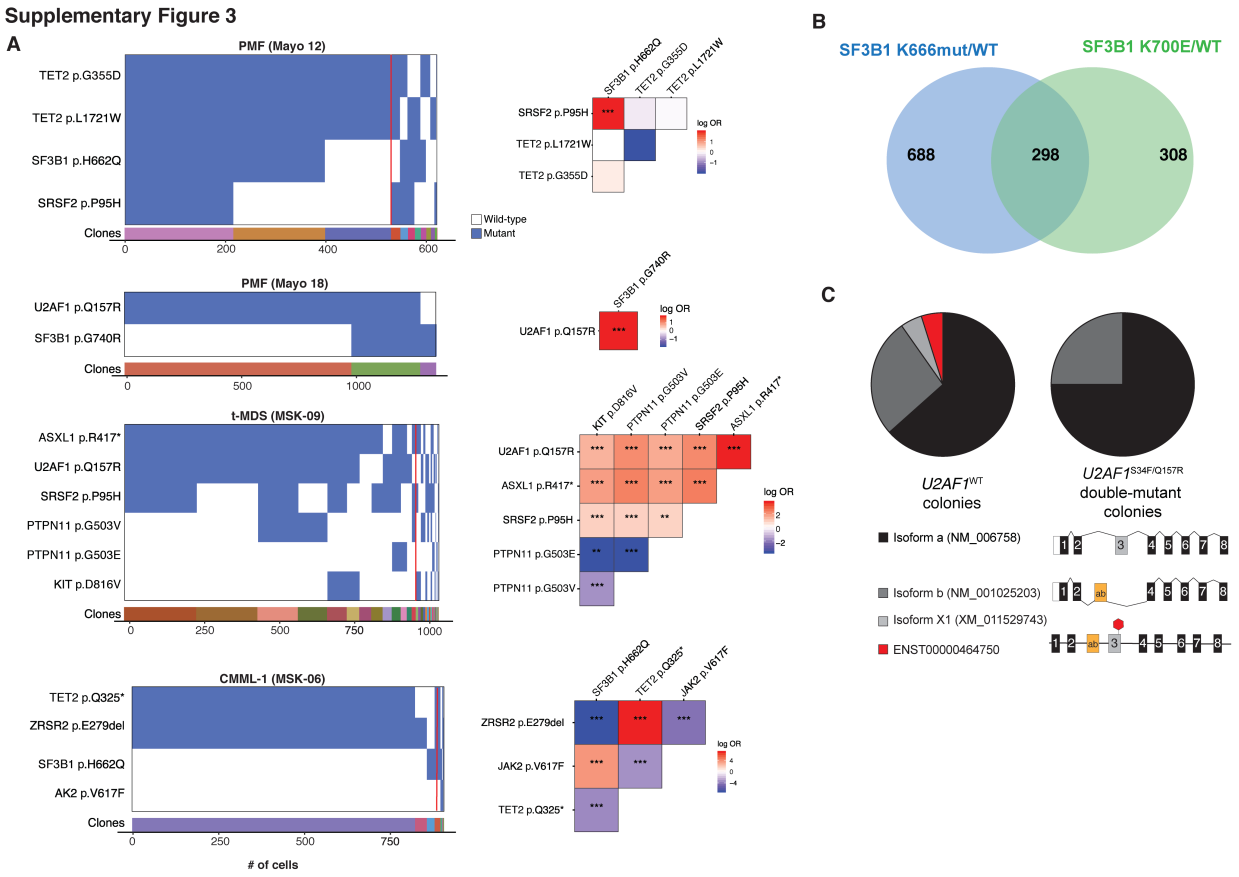


Figure 3:6 (Supplementary) Oncoprints of mutations in individual cells of dual splicing factor mutant myeloid neoplasm patients

Figure 3:6. Oncoprints of mutations in individual cells of dual splicing factor mutant myeloid neoplasm patients. (A) Cellular-level co-occurrence of mutations in select double-splicing factor mutant patient cases (clinical diagnosis and sample ID are listed above each oncoprint). Oncoprint (left) shows the genotype of each sequenced cell for each variant, with clustering based on the genotypes of driver mutations. Each column represents a cell at the indicated scale. Mutant and wild-type cells are indicated in blue and white, respectively. The subclones located to the right of the red line comprised <1% of the total sequence cells, since such small subclones can represent false positive or negative genotypes as a result of allele-drop out or multiplets. The figures on the right show the pairwise association of mutations. The color and size of each panel represent the degree of the logarithmic odds ratio (log OR). The vertical bar is a key indicating the association of the colors with the log OR. Co-occurrence and mutual exclusivity are indicated by red and blue, respectively. The statistical significance of the associations based on the false discovery rate (FDR) is indicated by the asterisks (*FDR < 0.1, **FDR < 0.05, ***FDR < 0.001). (B) Venn diagram

illustrating the overlap between differentially spliced events in SF3B1^{K700E} and SF3B1^{K666} mutant versus WT samples. (C) Pie-chart distribution of *U2AF1* mRNA isoforms in U2AF1^{WT} and U2AF1$^{S34F/Q157}$ double-mutant clones. Abbreviations: CMML: chronic myelomonocytic leukemia; PMF: primary myelofibrosis; t-MDS: therapy-related myelodysplastic syndrome.

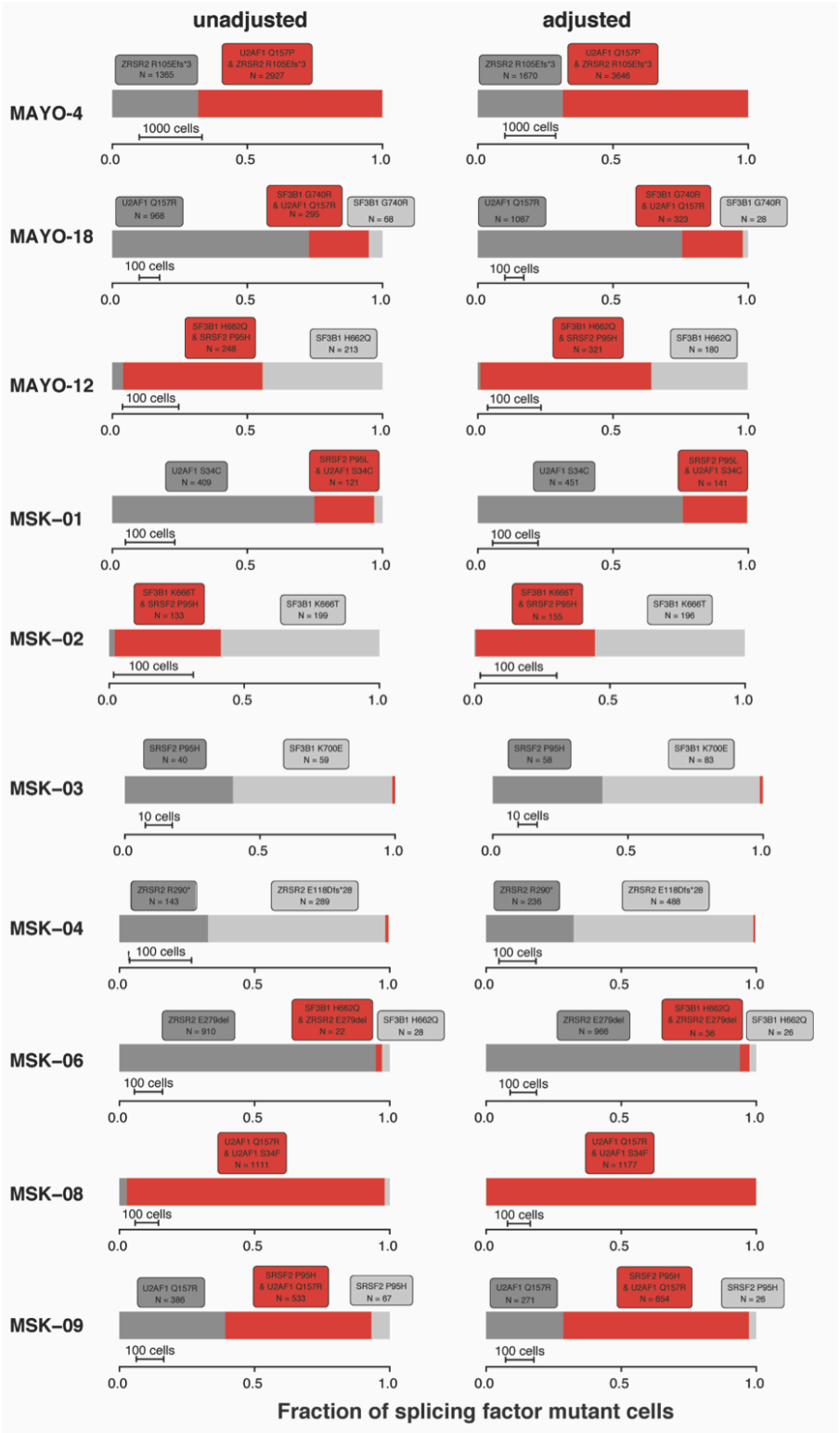


Figure 3:7 (Supplementary) Clonal structure of RNA splicing factor mutations in double mutant patients before and after adjustments for allele dropout (ADO)

Figure 3:7 (Supplemental) Clonal structure of RNA splicing factor mutations in double mutant patients before and after adjustments for allele dropout (ADO). ADO was adjusted under the assumptions that i) mutant and wild-type alleles are equally likely to be amplified and ii) mutations are consistently heterozygous in diploid cells. The left panel shows unadjusted estimates of mutant subclones within each patient. The right panel shows adjusted estimates in corresponding patients. Red bar denotes fraction of cells with both splicing factor mutations. Gray bars denote fraction of cells with one of the two splicing factor mutations.

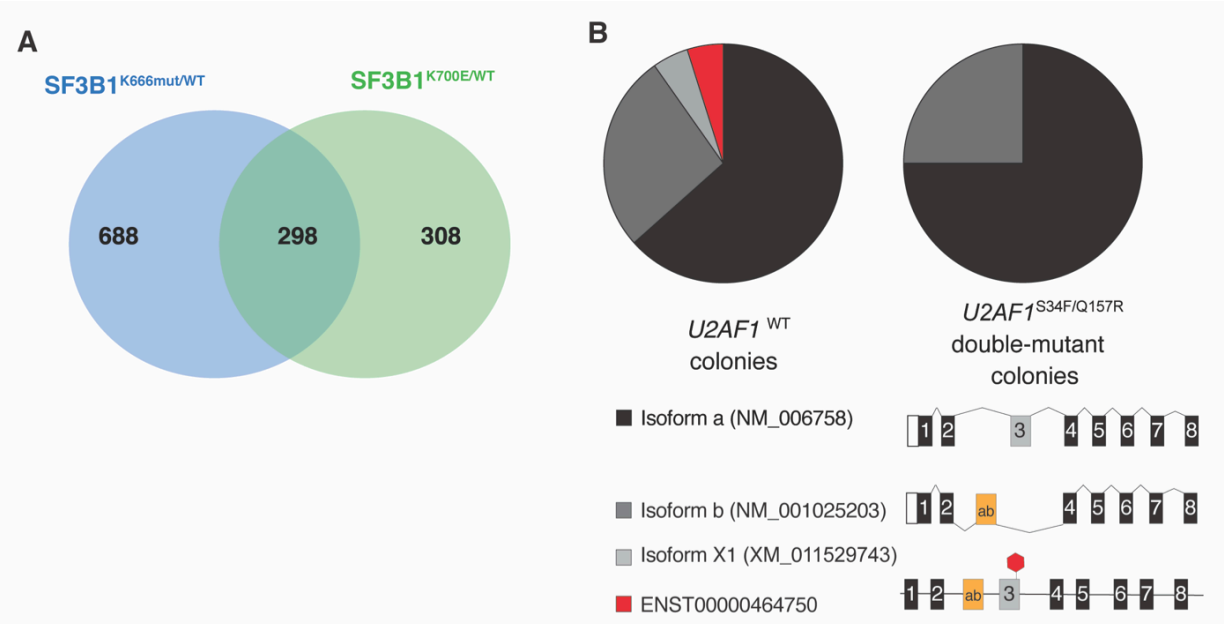


Figure 3:8 (Supplementary) Allele-specific differences in splicing based on SF3B1 mutational hotspot and *U2AF1* isoform usage in *U2AF1*^{WT} versus *U2AF1*$^{S34F/Q157R}$ double mutant cells.

Figure 3:8. Allele-specific differences in splicing based on SF3B1 mutational hotspot and U2AF1 isoform usage in U2AF1WT versus U2AF1S34F/Q157R double mutant cells. (A) Venn diagram illustrating the overlap between differentially spliced events in SF3B1K700E and SF3B1K666 mutant versus WT samples. (B) Pie-chart distribution of U2AF1 mRNA isoforms in U2AF1WT and U2AF1S34F/Q157 double-mutant clones. Abbreviations: CMML: chronic myelomonocytic leukemia; PMF: primary myelofibrosis; t-MDS: therapy-related myelodysplastic syndrome.

3.6 METHODS

3.6.1 *Patient Samples*

Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center, the University of Manchester, and the Mayo Clinic and conducted in accordance to the Declaration of Helsinki protocol. Patients with myeloid malignancies, including AML, MDS, and MPN, were identified from four public studies[8,10,16,17] and in patients at clinical practice at MSK and the Mayo clinic. The overall frequency of splicing factor mutations and comparisons between single and double mutants were evaluated in this cohort of 4,231 patients (1,319 from the German-Austrian AML Study Group[14], 1,242 from MSK, 644 from the United Kingdom, 608 from the Beat AML program[4], and 418 from the Internal Cancer Genome Consortium (ICGC)[16] (**Figure 4:4A**)). These genomic DNA sequencing data will be publicly available at https://cbioportal.mskcc.org/study/summary?id=mds_mskcc_2020. For analysis specific to double mutants, 36 samples identified from this cohort were combined with an additional 22 samples contributed by Mayo Clinic (unpublished) for a total of 58 samples.

3.6.2 *Bulk DNA Mutational Analysis*

Mutation data were systemically re-annotated for known hotspot mutations in *SF3B1*, *SRSF2*, and *U2AF1*, as well as truncating mutations in *ZRSR2* using the Ensembl Variant Effect Predictor (VEP)[18]. For samples sequences by MSK IMPACT we used a variant allele fraction (VAF) threshold of 1% for hotspot mutations in *SF3B1, SRSF2,* and *U2AF1*, and 5% for variants of unknown significance, consistent with normal clinical practice for MSK IMPACT data.[19,20] However, for the varied techniques used for mutational calling from non-MSK data (including whole exome sequencing and targeted DNA sequencing of varying depth), the lowest VAF reported was 5%. Several functional alleles specific to myeloid diseases were added to a list derived primarily from solid tumors[21] The frequencies of patients with one versus two splicing factor mutations were determined and Fisher's exact test was performed to assess for mutual exclusivity. Single and double mutants were compared in terms of distribution of myeloid neoplasm subtypes and tumor mutation burden. Mutation burden was calculated by dividing the number of non-synonymous mutations by the expected breadth of sequencing in the IMPACT Heme panel[22] (1.21 Mb). For samples with double splicing factor mutations, the cancer cell

fraction (CCF) of each mutation was summed to gauge the likelihood that they co-occur in the same cells as opposed to different cells of a patient sample. Although 58 patient samples with double splicing factor mutations were identified as described above, the CCF analysis could only be performed on 46 samples where VAFs of both splicing factor mutations were available. The CCF of hotspot mutations in *SF3B1*, *SRSF2*, and *U2AF1* was assumed to be two times the VAF. For nonsense and frameshift mutations on *ZRSR2*, the ploidy was assumed to be 1 for males and 2 for females when this information was available. When it was not, VAFs above 0.5 were taken to imply monoploidy. Co-occurrence of the splicing factor mutations in the same cells was substantiated by a combined CCF equal to or greater than 1. Copy number analysis was performed on 600 patients in the MSK Cohort sequenced with MSK IMPACT (using previously published methods[19,20]).

3.6.3 *Single Cell DNA sequencing Analysis*

Targeted single cell DNA sequencing of cryopreserved bone marrow mononuclear cells (BM MNCs) from patients with double splicing factor mutations was performed on a microfluidic, droplet-based platform developed by Mission Bio as previously described[23]. Barcoded samples proceeded to targeted PCR amplification using a custom panel of 36 amplicons targeting 47 mutations across all four splicing factors and 13 additional driver genes involved in myeloid malignancies. The identity and mutational profile of each cell were preserved through the process as each amplicon was tagged with a unique cell barcode.

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

Pooled single cell DNA libraries were sequenced on Illumina MiSeq with paired-end multiplex runs (2 x 150 bp). Raw sequencing reads in FASTQ files were processed using the Tapestri Pipeline, which uses Bluebee's High Performance Genomics Platform. The pipeline trims adapter information, aligns sequences to the reference genome (hg19), assigns sequence reads to individual cell barcodes, calls genotypes using GATK, and generates annotated loom files. Resulting data that met the following filtering criteria were included for downstream analysis (**Supplementary Table 5**): read depth > 10, genotype quality (i.e., difference between the likelihoods of the two

most likely genotypes on a scale of 0-99) > 30, alternate allele frequency (i.e., number of reads that support each of the reported alleles) > 20, variants genotyped in > 10% of cells, variants mutated in >1% of cells, and cells with >50% of genotypes present. VAFs were determined by aggregated reads from filtered cells, i.e. number of reads with variant of interest / number of total reads. The co-occurrence or mutual exclusivity of the dual splicing factor alterations was calculated at the same time as the rate of allele dropout (ADO) using the Bayesian statistical package STAN[1]. Since ADO of the relevant splicing factors can strongly influence the observed degree of co-occurrence, it was modeled under the assumption that splicing factor alterations were diploid and either wild-type or heterozygous in all cells and also that mutant and wild-type alleles are equally likely to be amplified. The estimated median ADO rate was 22% overall (IQR: 14% - 26%) but 73.9% (ICR: 68 - 77%) at the locus encoding SRSF2 P95. We attempted to address this by jointly estimating the co-occurrence of *SF3B1, SRSF2,* and *U2AF1* mutations and ADO at both loci under the assumption that hotspot mutations in these genes were diploid and either wild-type or heterozygous in all cells.

Subclones from each sample were identified using Tapestri Insights and annotated NGS data were subsequently analyzed in R. Oncoprints were generated where each column represents an individual cell and each row represents a mutation. Cells were clustered by subclone and then arranged by size of subclones from large to small. Pairwise associations between mutations were evaluated by Fisher's exact test with correction for multiple testing using the Benjamini-Hochberg method. The order of mutations in each sample was inferred based on the principle of maximum parsimony and depicted on fishplots in accordance with the proportions of the subclones.

3.6.4 *U2AF1 allele-specific sequencing*

Cryopreserved BM MNCs from a patient with co-occurring U2AF1^{S34F} and U2AF1^{Q157R} mutations by bulk and single cell DNA sequencing were thawed and lysed in TRI Reagent®. Total RNA was extracted using the Direct-zol™ RNA MiniPrep. Next, RT-PCR was performed using the QIAGEN One-Step RT-PCR Kit with the following primers: forward, 5'-GGCACCGAGAAAGACAAAGT-3'; reverse, 5'-AGCTCTCTGGAAATGGGCTT-3'. The PCR product was confirmed by gel electrophoresis and purified using the QIAquick PCR Purification Kit. Next, the U2AF1 PCR product was cloned into the pGEM®-T vector at a 3:1

insert to vector molar ratio. The ligated product was transformed into NEB® Turbo Competent E. coli (High Efficiency). Transformed bacteria were then plated onto LB agar plates with 100 μg/ml ampicillin, 0.1 mM IPTG, and 80 μg/ml X-Gal and incubated for overnight hours at 37°C. The resulting white colonies generally contain inserts whereas the blue colonies contain empty vectors. White colonies were thus selected for Sanger sequencing using the T7 primer which binds upstream of the insert on the vector backbone. U2AF1^{S34F} results from a C-to-T mutation that changes the codon from TCT to TTT. U2AF1^{Q157R} results from an A-to-G mutation that changes the codon from CAG to CGG.

3.6.5 *Genome annotations, RNA-seq read mapping, and Differential splicing analysis*

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

3.6.6 *Differential RNA splicing analysis*

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

3.6.7 *Isothermal titration calorimetry (ITC)*

SRSF2 peptides were purified with the same protocol as previously described10 and ITC was performed as previously described.11 Briefly, cDNAs of SRSF2 WT and P95A, P95H, P95HL, and P95R mutants encoding amino acids 1-101 of SRSF2 (which encompasses the RNA recognition motif (RRM) of SRSF2 were cloned into the pet28a plasmid using the restriction sites BamH1/Xho1. SRSF2 proteins were overexpressed at 37°C for 3 h in Escherichia coli BL21 (DE3) codon plus cells in minimal M9 medium (1 g l−1 15N-NH4Cl, 2 g l−1 13C-glucose) using 0.1 mM IPTG. Protein was then purified by two successive nickel affinity chromatography steps and dialysed against an NMR buffer (50 mM L-Glu, 50 mM L-Arg and 20 mM Na2HPO4/NaH2PO4 at pH 5.5). A last purification step by size exclusion chromatography with a Superdex75 column (GE Healthcare) was necessary to remove residual RNases in the solution. The protein could be concentrated to over 2 mM with a 10-kDa molecular mass cutoff membrane.

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

3.7 ACKNOWLEDGEMENTS

We are grateful for the support of the Mt. Sinai School of Medicine Technology Development Lab for help with the MissionBio Tapestri platform and sequencing as well as the Manchester Cancer Research Centre Biobank for help with samples. We also thank Mission Bio for providing assistance with data generation, analysis, and interpretation. J.T. is supported by the Conquer

Cancer Foundation of the American Society of Clinical Oncology, the American Association for Cancer Research, the American Society of Hematology (ASH), the Robert Wood Johnson Foundation, and the NIH/NCI (1K08CA230319-01). O.A.-W. and R.K.B are supported by the Evans MDS Foundation, the US National Institutes of Health (R01 HL128239), the Dept. of Defense Bone Marrow Failure Research Program (W81XWH-12-1-0041), and the Leukemia & Lymphoma Society. R.K.B. is supported in part by the US National Institutes of Health (R01 DK103854). O.A.-W. is supported by the Henry & Marilyn Taub Foundation.

3.8 AUTHOR CONTRIBUTIONS

J.T., X.M., K.N., M.B., A.P., M.M.P., R.K.B., and O.A.-W. designed the study. J.T., M.B., T.L., K.N., M.H., D.W., A.T., M.M.P., and O.A.-W. provided patient samples. J.T., X.M., M.B., A.P., J.P., S.B., D.W., M.M.P., and O.A.-W. performed computational analyses of mutational data. K.N., A.P., and R.K.B. performed computational analyses of RNA-seq data. X.M. and B.L. performed cloning and colony assays. S.H. and Y.L. performed ITC assays. J.T., K.N., X.M., A.P., R.K.B, and O.A.-W. wrote the manuscript with approval from all co-authors.

3.9 DISCLOSURE OF CONFLICTS OF INTEREST

O.A.-W. has served as a consultant for H3B Biomedicine, Foundation Medicine Inc, Merck, and Janssen, and is on the Scientific Advisory Board of Envisagenics Inc; O.A.-W. has received prior research funding from H3B Biomedicine unrelated to the current manuscript.

3.10 REFERENCES

1. Carpenter B, Gelman A, Hoffman MD, et al. Stan: A Probabilistic Programming Language. 2017. 2017;76(1):32.
2. Dvinge H, Ries RE, Ilagan JO, Stirewalt DL, Meshinchi S, Bradley RK. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. Proc Natl Acad Sci U S A. 2014;111(47):16802-16807.
3. Flicek P, Ahmed I, Amode MR, et al. Ensembl 2013. Nucleic Acids Res. 2013;41(Database issue):D48-55.
4. Meyer LR, Zweig AS, Hinrichs AS, et al. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. 2013;41(Database issue):D64-69.
5. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7(12):1009-1015.

6. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
7. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
8. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105-1111.
9. Wagenmakers EJ, Lodewyckx T, Kuriyal H, Grasman R. Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. Cogn Psychol. 2010;60(3):158-189.

Chapter 4. SYNTHETIC INTRONS ENABLE MUTATION-DEPENDENT TARGETING OF CANCER CELLS

Khrystyna North*, Salima Benbarche*, Joseph Pangallo, Caroline Erickson, Hana Cho, Jose Mario Bello Pineda, James D. Thomas, Jacob T. Polaski, Omar Abdel-Wahab, Robert K. Bradley

*Equal contribution

Contributions by Khrystyna North:

In this work I designed and tested synthetic introns in both single intron and screen contexts in vitro. I also analyzed and visualized all screen results.

The following figures display my contributions to this project:

Figure 4:1 a, b, c and e, f, g: I selected the best endogenous introns to test, cloned, and designed the synthetic intron fluorescent reporter and performed/analyzed the flow cytometry data shown.
Figure 4:2 a, b, c, d, e: I designed and cloned the reporter shown. I made the lentivirus and performed the pilot screen as well as the IC50 curve and analyzed the data from the pilot screen.
Figure 4:3 a, b, c, d, e, f, g, h: I cloned the intron library and performed the HSV-TK, GCV screen, and analyzed/visualized the data shown.
Figure 4:6 a: I cloned and designed the reporter shown.
Figure 4:7 a, b, c, d, e, f: I cloned the intron library and performed the HSV-TK, GCV screen, and analyzed/visualized the data shown.

4.1 ABSTRACT

Many cancers carry recurrent change-of-function mutations in RNA splicing factor genes, which induce sequence-specific changes in RNA splicing. Here, we describe a method to harness this change in RNA splicing activity to drive spliceosomal mutation-dependent gene expression in cancers and selectively eliminate these tumors. We engineered synthetic introns which were efficiently spliced in leukemia cells bearing the most common *SF3B1* mutations, but unspliced in wild-type cells—and vice versa—to yield mutation-dependent protein production. A massively parallel screen of 8,881 distinct introns delineated ideal intronic size, mapped essential sequence elements, and revealed the basis of mutation-dependent splicing. Key synthetic introns from the screen enabled intron-dependent delivery of thymidine kinase and subsequent ganciclovir-mediated elimination of leukemia cells bearing *SF3B1* mutations *in vitro* and *in vivo*, while leaving wild-type cells unaffected. This approach significantly decreased the growth of otherwise lethal leukemia xenografts. The modular, compact, and specific nature of synthetic introns thereby provide a means to exploit cancer-specific changes in RNA splicing for genotype-dependent gene expression and gene therapy.

4.2 INTRODUCTION

Recurrent mutations affecting an RNA splicing factor occur in many cancer types, with frequencies ranging from 65-83% in myelodysplastic syndromes with ring sideroblasts (MDS-RS) and 14-29% in uveal melanoma to 15-35% in acute myeloid leukemia (AML) and 2-3% in breast adenocarcinoma. [1-6] These lesions are attractive targets for therapeutic development thanks to their pan-cancer nature, frequent presence in the dominant clone, and particular enrichment in diseases with few effective therapies. Accordingly, several studies have demonstrated that cancer cells bearing spliceosomal mutations are preferentially sensitive to further splicing perturbation, including treatment with compounds that inhibit normal spliceosome assembly or function. [5, 6] However, the therapeutic index of drugs that inhibit global splicing activity is not yet clear. Moreover, therapeutic approaches that target the function of the mutant splicing machinery itself have not yet been identified.

Spliceosomal mutations alter splice site and exon recognition to cause dramatic mis-splicing of a restricted set of genes, while leaving most genes unaffected. Although these splicing changes promote aberrant self-renewal, transformation, and other pro-tumorigenic phenotypes, we hypothesized that we could exploit this splicing dysregulation for therapeutic development. [7-10] We therefore sought to develop synthetic constructs that were differentially spliced in cells with or without recurrent mutations in *SF3B1*, the most commonly mutated spliceosomal gene in cancer, to allow for cancer cell-specific protein production.[11]

4.3 RESULTS

We first identified endogenous genes that responded most strongly and consistently to *SF3B1* mutations, which are near-universally present as heterozygous, missense changes affecting a few residues. We queried the transcriptomes of 35 cancer cohorts to identify 20 distinct cancer types with more than one *SF3B1*-mutant sample, with a total of 271 samples from patients carrying *SF3B1* mutations. 1,608 splicing events were significantly differentially spliced between samples bearing no spliceosomal mutations (wild-type; WT) and *SF3B1*-mutant samples in at least one cohort, with a subset exhibiting highly consistent differential splicing (**Figure 4:1a-b**). [12,13] *SF3B1* mutations were associated with diverse splicing changes, including differential 3' splice site (3'ss) selection and exon recognition and differential intron retention.

We selected six introns representing two major classes of splicing events for further study. SF3B1 mutations activate intron-proximal cryptic 3'ss in *MAP3K7*, *ORAI2*, and *TMEM14C*, and promote efficient intron removal in *MTERFD3*, *MYO15B*, and *SYTL1* (**Figure 4:1c**). These were among the strongest and most consistent mis-splicing events, and preferentially caused either open reading frame disruption (*MAP3K7*, *ORAI2*, and *TMEM14C*) or preservation (*MTERFD3*, *MYO15B*, and *SYTL1*) in *SF3B1*-mutant samples. We confirmed that the mis-splicing observed in primary patient samples was recapitulated in isogenic K562 (erythroleukemic) and NALM-6 (B-cell acute lymphoblastic leukemia) cells, as well as MEL202 and MEL270 (uveal melanoma) cells, with or without a recurrent *SF3B1* mutation (**Figure 4:1d, Figure 4:5a**).

These six endogenous introns served as starting points for the development of synthetic introns that functioned as compact and modular molecular switches. We reduced each intron to

250 nt in length by taking the first 100 and last 150 nt and inserted each intron into the mEmerald coding sequence in a location that preserved the 5'ss and 3'ss strengths of the endogenous genes as well as generated exons of roughly comparable sizes. These choices were guided by the increased complexity of the 3'ss versus 5'ss, SF3B1's functional role in 3'ss recognition, and the importance of exon length in splicing. We cloned each split mEmerald sequence into a vector with constitutive expression of the non-overlapping fluorophore mCardinal (**Figure 4:1e**). The resulting vectors permitted quantitative assessment of mutation-dependent protein production by measuring the ratio of mEmerald to mCardinal in cells with or without an SF3B1 mutation via flow cytometry.

We transfected each construct into isogenic WT or *SF3B1*-mutant K562 cells and measured mutation-dependent splicing and protein production using flow cytometry. Of the six initial synthetic introns, three exhibited mutation-dependent specificity of ≥2-fold (synMAP3K7i4-250, synTMEM14Ci1-250, and synMTERFD3i1-250) and two others drove modestly mutation-dependent protein production (synORAI2i1-250 and synMYO15Bi4-250; **Figure 4:1f-g**). Mutation-dependent protein production arose from mutation-dependent splicing changes, as designed (**Figure 4:5b-c**). These proof-of-principle studies confirmed the feasibility of using synthetic introns for mutation-dependent gene expression.

We next tested the therapeutic potential of using synthetic introns to achieve mutation-dependent toxin delivery to cancer cells. We selected the herpes simplex virus thymidine kinase (HSV-TK) system, in which treatment of HSV-TK-expressing cells with the prodrug ganciclovir (GCV) causes cytotoxic metabolite production. [14,15] As GCV is an FDA-approved antiviral therapy with low toxicity for cells lacking HSV-TK, HSV-TK is an attractive system for cancer gene therapy.

Following the same approach used for fluorescent protein expression, we inserted the *MTERFD3*-derived synthetic intron, which was more efficiently excised in *SF3B1*-mutant cells, into the HSV-TK coding sequence (**Figure 4:2a**, **4:6a**). We cloned this split HSV-TK sequence or an intronless HSV-TK into a lentiviral expression vector, infected isogenic WT or *SF3B1*-mutant K562 cells, selected positive integrants, and treated with GCV (**Figure 4:2b**). Untransduced cells exhibited minimal loss of viability, while cells transduced with an intronless

HSV-TK construct died rapidly, independent of *SF3B1* mutational status. *SF3B1*-mutant cells expressing synthetic intron-containing HSV-TK exhibited a rapid and dose-dependent loss of viability, indistinguishable to that caused by intronless HSV-TK; in contrast, WT cells expressing synthetic intron-containing HSV-TK exhibited no significant differences in viability from untransduced cells (**Figure 4:2c**).

We next identified key sequence features that conferred mutation responsiveness to our synthetic intron with sequence analysis, cDNA cloning, and RNA lariat sequencing. The intron has a simple 5'ss region, with a near-consensus ss followed by a pyrimidine-rich region of unknown function. In contrast, its 3'ss region is very complex. It has two cryptic 3'ss at positions -11 and -22 relative to the canonical (frame-preserving) 3'ss, with a highly unusual TG dinucleotide at the most intron-proximal site; four adenine branchpoints at positions -43, -48, -55 and -61; a thymine-rich region that resembles a polypyrimidine tract interrupted by branchpoints; and a long, upstream purine-rich region of unknown function (**Fig. 4:2e**). Because of the intron's complexity, the sequence features that govern mutation responsiveness were not a priori obvious. We therefore developed a massively parallel assay to map and functionally interrogate key sequence features. We first synthesized a pilot mini-library of eight introns, each with one or more perturbations to potentially critical features, and cloned this mini-library into HSV-TK. We introduced the corresponding mini-library into WT or *SF3B1*-mutant K562 cells with a lentiviral vector at a low multiplicity of infection, treated with GCV, and measured relative depletion of each construct by high-throughput sequencing of the entire introns from genomic DNA at days 0 and 6.

This pilot experiment demonstrated the utility of parallel screening for functional interrogation (**Figure 4:2f**). The original 250 nt synthetic intron was markedly depleted in *SF3B1*-mutant, but not WT, cells. Shortening the intron to 150 or 100 nt resulted in robust or modest mutation responsiveness, respectively. Mutation responsiveness was maintained even after mutating all four known branchpoints (A>G, which is a poor branchpoint) or inserting a single consensus branchpoint within the pyrimidine-rich region of the 3'ss. Removing either the 5'ss or canonical (frame-preserving) 3'ss prevented introns from becoming depleted even in *SF3B1*-mutant cells, as expected for abolition of splicing, while removing the cryptic 3'ss at position -11 resulted in strong depletion irrespective of genotype. We validated mini-screen results by

introducing each individual construct in our mini-library into WT or *SF3B1*-mutant K562 cells and measuring relative depletion, confirming that our parallelized functional screen yielded accurate estimates of fitness costs following GCV treatment (**Figure 4:2f**). Additionally we tested few control introns in MCF10A, a breast cancer cells line with (SF3B1*K700E*) or without an endogenous *SF3B1* mutation. Expression of synthetic intron-containing HSV-TK in isogenic MCF10A breast cancer cells and subsequent GCV administration similarly resulted in strongly mutation-dependent cell death (**Figure 4:2g**), confirming that our synthetic intron enables efficient targeting of multiple cancer types with distinct *SF3B1* mutations. The synthetic intron was efficiently excised in *SF3B1*-mutant K562 and MCF10A cells, but not in *SF3B1*-wild-type K562 or MCF10A cells, confirming that GCV sensitivity arose from mutation-dependent splicing of the synthetic intron (**Figure 4:2f-g**).

We next expanded our approach to a massively parallel assay. We designed 8,881 distinct introns to test the functional consequences of perturbing diverse features, including intron length, 5'ss and canonical 3'ss strengths, cryptic 3'ss position and multiplicity, pyrimidine and purine contents, branchpoint position and multiplicity, and nucleotide and dinucleotide identity. We synthesized these 8,881 introns as an oligonucleotide array, cloned this array into HSV-TK, infected K562 cells, and then sequenced all entire introns from genomic DNA to estimate how each affected cell viability upon GCV administration in triplicate. The resulting data illuminated global features governing mutation responsiveness. Shortening the original 250 nt synthetic intron to 150 nt, while maintaining mutation responsiveness, required preserving the first 25 and last 125 nt (**Figure 4:3a**). Extreme shortening to 75 nt was possible, but reduced mutation responsiveness (**Figure 4:3b**). Significant perturbations adjacent to splice sites were not tolerated. On average, constructs with single-nucleotide mutations affecting the 5'ss were rarely depleted in either genotype; those with mutations affecting either 3'ss had reduced mutation responsiveness; and constructs with mutations affecting neither splice site frequently maintained responsiveness (**Figure 4:3c**). Mutation responsiveness required keeping the canonical 3'ss modestly stronger than the cryptic 3'ss; splice site strengths could be shifted as long as this imbalance was maintained, but not exaggerated (**Figure 4:3d**).

Our massively parallel assay also enabled us to gain high-resolution insight (**Figure 4:3e-f**). Deletion scanning revealed that loss of either cryptic 3'ss caused genotype-independent depletion, while perturbation of the pyrimidine-rich, branchpoint-containing region abolished depletion for both genotypes. In contrast, the immediately upstream sequence was dispensable. Sliding creation of an additional cryptic 3'ss or conversion of pyrimidine-rich sequence to purines generally maintained mutation responsiveness, as long as the critical ~30 nt upstream of the canonical 3'ss were preserved. In contrast, inserting a consensus branchpoint sequence typically reduced mutation responsiveness, unless this insertion was performed concordantly with ablation of all four endogenous branchpoints. In that context, branchpoint insertion maintained mutation responsiveness, notably even when the new branchpoint was located unusually deep within the intron.

Saturation mutagenesis revealed that the intron is remarkably robust to single-nucleotide mutations, with almost all constructs maintaining excellent mutation responsiveness (**Figure 4:3g**). The cryptic 3'ss at position -11 was a notable exception, where C/T mutations of the AG dinucleotide were notably depleted in WT cells. Several positions proved unexpectedly important in *SF3B1*-mutant cells, including +6 of the 5'ss and -3, -6, and -10 of the canonical 3'ss. Mutation of many positions to adenine within the pyrimidine-rich, branchpoint-containing region was associated with strong depletion, while mutations that ablated branchpoints within this region preserved mutation responsiveness, confirming that variable branchpoint multiplicity is tolerated.

These observations were highly similar when the same modifications were applied to synMTERFD3i1-100, indicating that the critical sequence elements are independent of intron length (**Figure 4:7**). While the parent synMTERFD3i1-100 intron splicing was less extreme than the synMTERFD3i1-150 described above it appeared to be more responsive to inserted components and nucleotide modifications. Inserting a consensus branch point sequence, for example, improves genotype dependent splicing. This shorter intron is also, much more sensitive to the ablation of the “TG” splice site indicating that this splice site may play in important role in splice site selection.

Finally, we searched for possible epistatic interactions within the critical 4 and 20 nt of the 5'ss and 3'ss regions. We performed saturation mutagenesis of all 54 and 1,710 nucleotide pairs that did not disrupt the GT or AG of the 5'ss and canonical 3'ss and searched for enrichment or depletion exceeding that expected based on single-nucleotide mutagenesis by ≥2-fold (**Figure 4:3h**). For WT cells, only a single interaction at the 3'ss met this threshold: the G and following nucleotide of the most intron-proximal cryptic 3'ss. *SF3B1*-mutant cells exhibited more complex interactions, particularly for positions between the cryptic 3'ss, reinforcing the complex and essential nature of this region. No epistatic interactions met our threshold at the 5'ss for either genotype.

Having identified the key features governing mutation responsiveness, we next tested whether synthetic introns permitted mutation-dependent cancer cell killing *in vivo*. We introduced Luciferase-GFP constructs into WT or *SF3B1*-mutant K562 cells expressing HSV-TK interrupted by our original 250 nt synthetic intron, performed tail vein injections of these cells into NSG mice, and monitored leukemia burdens with live imaging. While both genotypes formed aggressive leukemias, GCV administration resulted in immediate and sustained reductions in *SF3B1*-mutant leukemic burden, with no effects on WT leukemias, and correspondingly significantly increased survival (**Figure 4a-b**).

4.4 Discussion

Our study demonstrates the feasibility and therapeutic potential of harnessing pro-tumorigenic splicing alterations to engineer new molecular therapeutics. As *SF3B1* mutations are common across diverse cancer types, synthetic introns may facilitate the development of pan-cancer gene therapies; they were equally efficacious for mutation-dependent targeting of leukemias and melanomas both *in vitro* and *in vivo* (**Figure 4:2, 4:4**). We expect synthetic introns to be similarly effective for targeting other recurrent, cancer-associated spliceosomal gene mutations—such as mutations in *U2AF1*, *SRSF2*, and *RBM10*, amongst others—most of which induce specific alterations in splicing mechanisms. Finally, our study illustrates the power of massively parallel assays for functional interrogation, including the derivation of rational rules governing mutation-dependent splicing that will facilitate the future design and improvement of these and other synthetic introns.

4.5 FIGURES AND LEGENDS

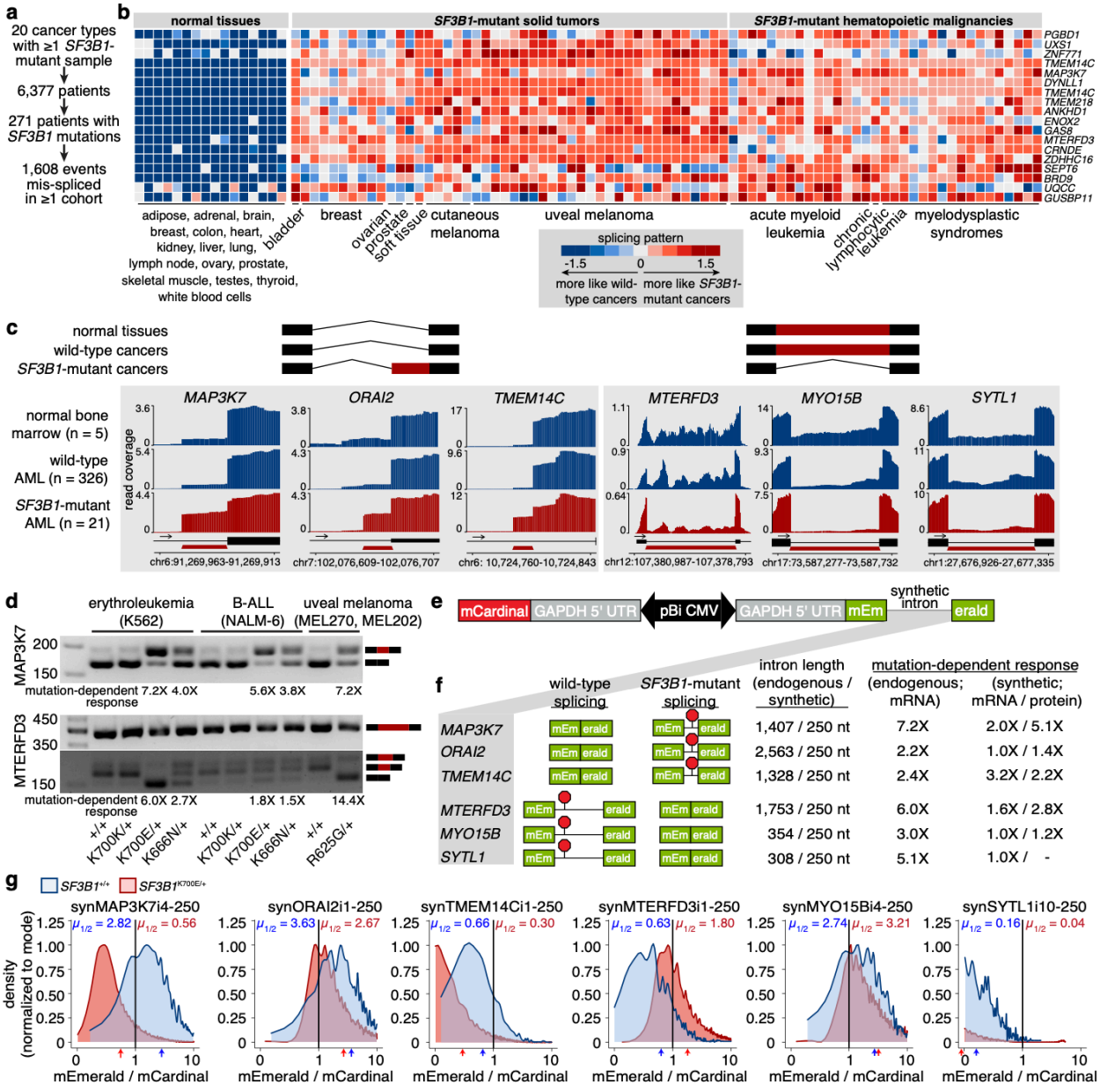


Figure 4:1 Synthetic introns can mimic *SF3B1* mutation-dependent mis-splicing in cancers.

Figure 4:1. Synthetic introns can mimic *SF3B1* mutation-dependent mis-splicing in cancers. **(a)** Workflow to identify differentially spliced events in *SF3B1*-mutant patient samples. **(b)** Heatmap illustrating z score-normalized expression of the top-ranked, mis-spliced isoforms. Top-ranked isoforms were defined as those with |delta psi| >= 0.1 and s.d. (psi) <= 0.15 across all *SF3B1*-mutant samples. Plot restricted to samples bearing the most common *SF3B1*K700 and R625 mutations with mutant allele expression >= 25%. Sample origins described in Methods. **(c)**

RNA-seq read coverage plots, averaged over the indicated samples, for the six introns selected for follow-up studies. Samples from Beat AML cohort. **(d)** RT-PCR analysis of endogenous *MAP3K7* and *MTERFD3* splicing in cells engineered to bear the indicated mutations in endogenous *SF3B1* (K562, NALM-6) or carrying them endogenously (MEL270, MEL202). *MAP3K7* isoforms arise from cryptic 3' splice site; *MTERFD3* isoforms arise from both cryptic 3' splice site usage and intron retention. n = 4 (K562, NALM-6) and 2 (MEL270, MEL202) biologically independent cell lines. **(e)** Schematic of the fluorescent reporter created to test synthetic intron function. **(f)** Expected splicing outcomes, intron lengths, and mutation-dependent response for each tested intron. Mutation-dependent response defined as the ratio of the indicated isoforms in *SF3B1*-mutant:WT cells (mRNA) and median mEmerald:mCardinal signal (protein). **(g)** Histograms of mEmerald:mCardinal signal, measured by flow cytometry. Arrows indicate medians ($\mu_{1/2}$) for each genotype. Representative images from n = 2 biologically independent experiments.

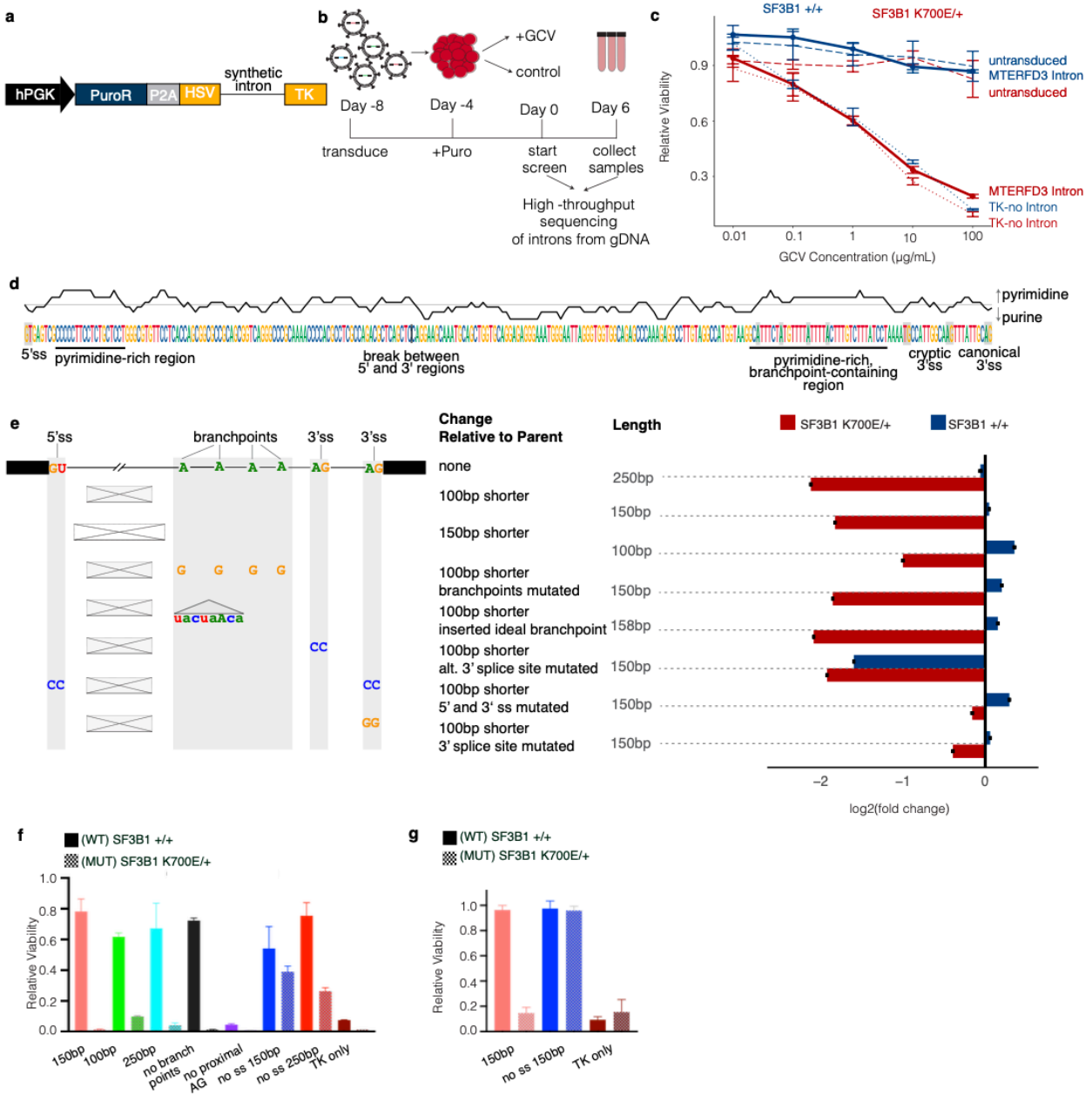


Figure 4:2 Synthetic introns enable mutation-dependent cancer cell killing.

Figure 4:2. Synthetic introns enable mutation-dependent cancer cell killing. (a) Schematic of HSV-TK-expressing construct**. (b)** Diagram of the experimental timecourse. **(c)** Relative viability of K562 cells expressing the indicated constructs, normalized to untreated samples. Viability measured by ATP after 3 days of treatment. Experimental schema in (b). Vector is hPGK-HSV-TK-P2A-mCherry. Data represented as mean ± s.d. n = 3 biologically independent experiments. **(d)** Diagram of 250 nt synthetic intron, with notable features highlighted. **(e)** Left, diagram of modifications in each intron relative to the original 250 nt intron. Right, depletion/enrichment at

day 6 relative to day 0, estimated by full-length intron sequencing from genomic DNA. GCV concentration, 100 μg/mL. **(f)** Relative viability measured with respect to untreated transduced controls of K562 cells expressing the indicated constructs demonstrated in (e). Viability was measured after 11 days of treatment with GCV at a concentration of 100μg/mL. Vector is hPGK-PuroR-P2A-HSV-TK (**g)** Relative viability measured with respect to untreated transduced controls of MCF10A cells expressing the indicated constructs. Viability was measured after 7 days of treatment with GCV at a concentration of 1μg/mL. Vector is hPGK-PuroR-P2A-HSV-TK

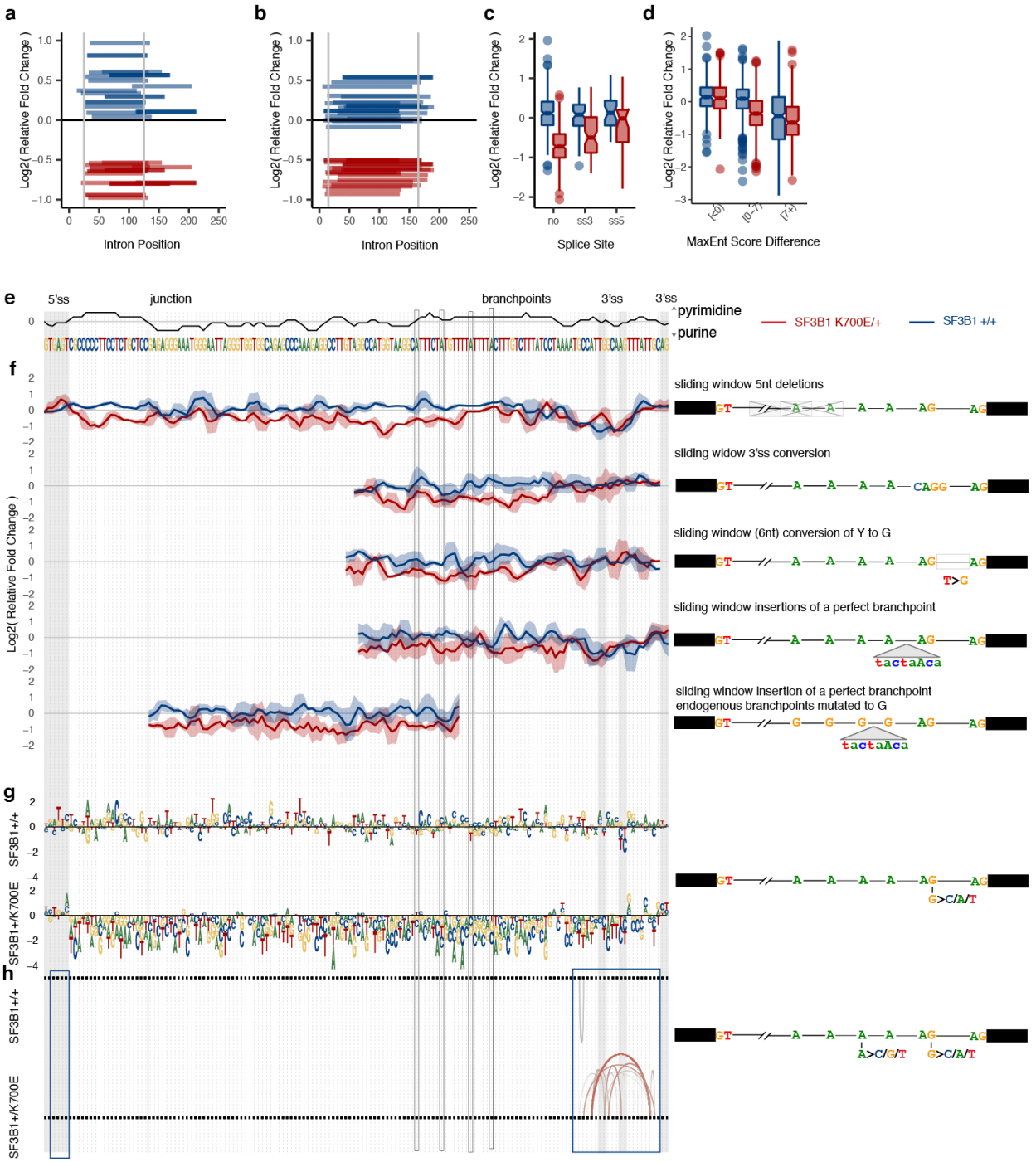


Figure 4:3 Massively parallel screening reveals critical elements governing synthetic intron function.

Figure 4:3. Massively parallel screening reveals critical elements governing synthetic intron function. (a) Relative fold-change of synthetic introns derived from synMTERFD3i1-250 by deletion of 100 consecutive nt. Each horizontal line indicates the nucleotides deleted in a single such variant, with the vertical position indicating the fold-change in *SF3B1*-mutant (red) or WT (blue) cells. Plot restricted to introns with $\log_2$ (fold-change) < -0.5 or > -0.1 in *SF3B1*-mutant and WT cells, respectively. Grey lines, deletion used to create synMTERFD3i1-150 (first 25 nt and last 125 nt preserved). **(b)** As (a), but for deletions resulting in introns of lengths 125, 100, 85 or 75 nt. Deletions were made to synMTERFD3i1-150 or synMTERFD3i1-100. Grey lines, deletion used to create synMTERFD3i1-100 (first 15 nt and last 85 nt preserved). **(c)** Box plots illustrating relative fold-changes for all introns derived from synMTERFD3i1-150 by single-nucleotide mutations, where the mutations affected the 10 5'-most nucleotides (5'ss), 26 3'-most nucleotides (3'ss), or neither (no). **(d)** Box plots illustrating relative fold-changes for all introns derived from synMTERFD3i1-150 by one or two single-nucleotide mutations, grouped by the difference in splice site strengths (MaxEnt) between the most intron-distal and most intron-proximal 3' splice sites. **(e)** Top, purine:pyrimidine ratio calculated with a 5 nt sliding window. Bottom, schematic of synMTERFD3i1-150 intron. **(f)** Line plots illustrating relative fold-changes for introns with the indicated perturbations at the indicated positions to synMTERFD3i1-150 for *SF3B1*-mutant (red) or WT (blue) cells. Data presented as geometric mean of fold-changes for the three closest introns (line) ± geometric s.d. from the mean (shading). From top to bottom, perturbations are: sliding deletions (5 nt); sliding conversion of 4 nt to 3'ss (CAGG), where position corresponds to placement of the AG; sliding conversion of all Ys within 6 nt to G, where position corresponds to center of the 6 nt window; sliding insertion of consensus branchpoint context (tactaAca), where position represents point of insertion; simultaneous ablation of branchpoints (A>G) and sliding insertion of consensus branchpoint context (tactaAca), where position represents point of insertion. **(g)** Sequence logo plots illustrating relative fold-change for single-nucleotide mutations to synMTERFD3i1-150 in WT (top) or *SF3B1*-mutant (bottom) cells. Height of each nucleotide indicates the fold-change for that mutation. **(h)** Arc plot illustrating the directions and magnitudes of synergistic fold-changes for each pair of mutations at the 5'ss or 3'ss to synMTERFD3i1-150 in WT (top) or *SF3B1*-mutant (bottom) cells. Vertical height indicates fold-change from combinatorial mutation relative to that expected from the two underlying single-nucleotide

mutations. Plot restricted to $|\log_2(\text{synergy})| > 1$. Arc above the x-axis indicates positive relative fold change and below the x axis indicates negative relative fold change.

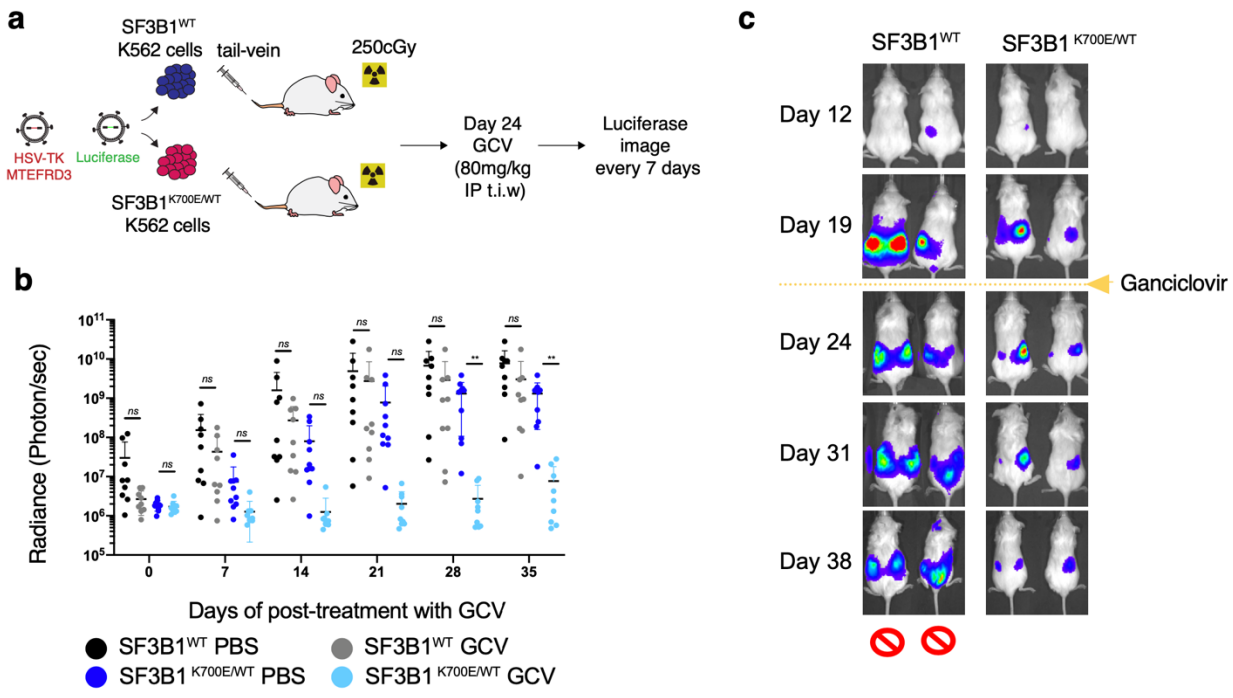


Figure 4:4 Synthetic introns enable mutation-dependent cancer cell targeting *in vivo*.

Figure 4:4. Synthetic introns enable mutation-dependent cancer cell targeting *in vivo*.(a) Schematic of xenograft experiments with K562 cells expressing HSV-TK with synMTERFD3i1-150. **(b)** Quantification of tumor burden, estimated by whole-body bioluminescent signal. K562 cells with and without an SF3B1 mutation expressing HSV-TK with synMTERFD3-150 and co-expressing Luciferase. Error bars, mean ± SD. P value calculated by unpaired two-tailed t-test. ** p-value Day 28 (0.0047) and ** p-value Day 35 (0.0035) **(c)** Representative bioluminescence images from n=5 mice per genotype.

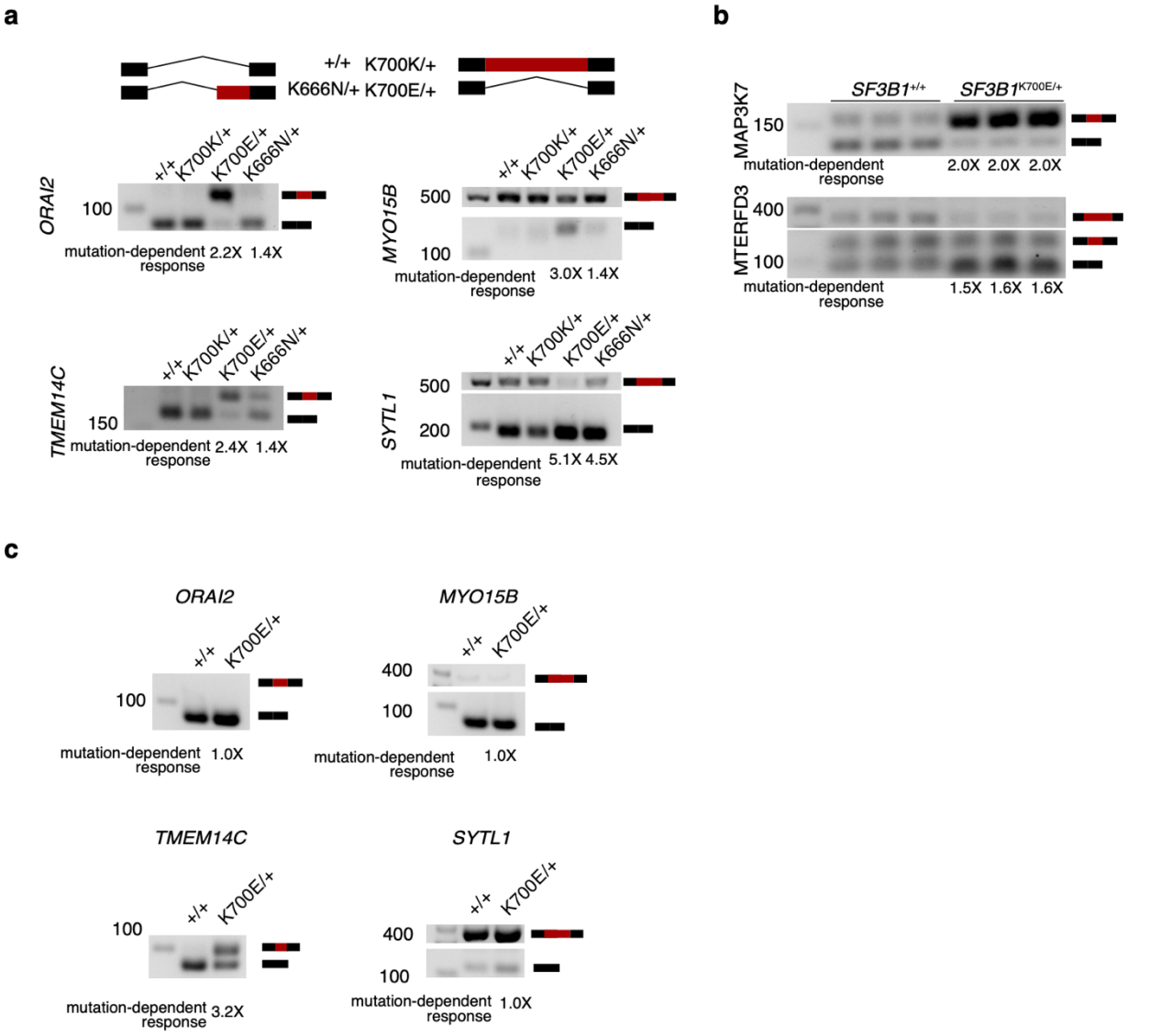


Figure 4:5 (Supplementary) Validation of *SF3B1* mutation-dependent differential splicing for endogenous and synthetic introns.

Figure 4:5. Validation of *SF3B1* mutation-dependent differential splicing for endogenous and synthetic introns. **(a)** RT-PCR analysis of splicing of the indicated endogenous introns in K562 cells engineered to bear the indicated mutations in endogenous *SF3B1*. n = 4 biologically independent cell lines. **(b)** RT-PCR analysis of synthetic intron splicing for the indicated introns following transfection of the fluorescent reporter construct into isogenic K562 cells with the indicated *SF3B1* genotypes. n = 2 biologically independent cell lines and n = 3 biologically independent experiments. **(c)** As (b), but for n = 2 biologically independent cell lines.

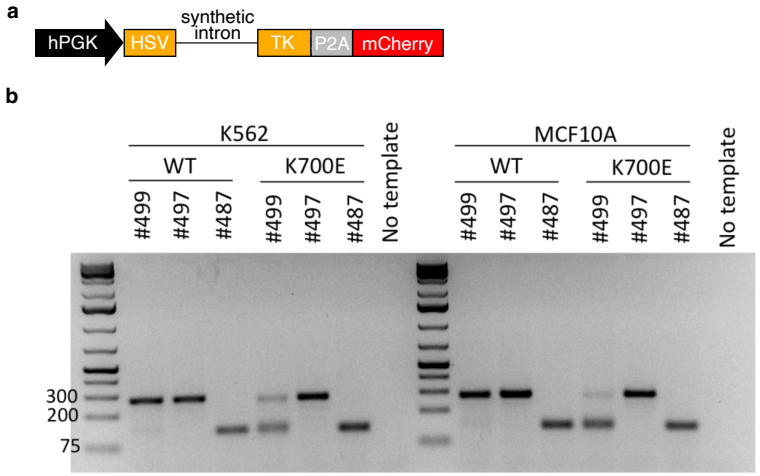


Figure 4:6 (Supplementary) PCR validation of the pilot screen and alternate construct design

Figure 4:6 PCR validation of the pilot screen and alternate construct design (a) Schematic of split HSV-TK construct with mCherry. **(b)** RT-PCR demonstrating mutation-dependent excision of the synthetic intron. Vector is hPGK-PuroR-P2A-HSV-TK.

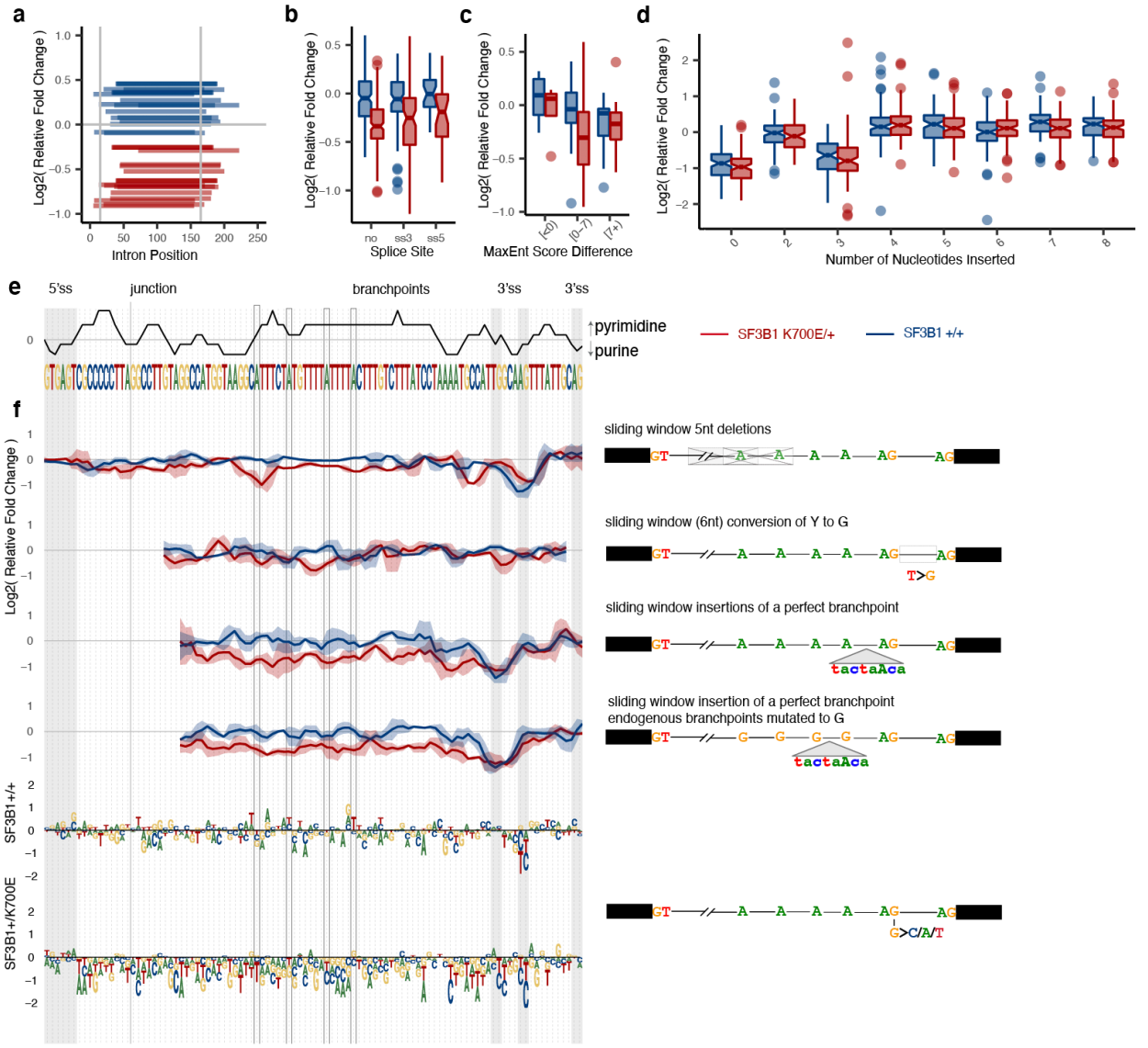


Figure 4:7 (Supplementary) Massively parallel screening reveals critical elements governing the function of very short synthetic introns

Figure 4:7. Massively parallel screening reveals critical elements governing the function of very short synthetic introns.

(a) Relative fold-change of synthetic introns derived from synMTERFD3i1-250 by deletion of 150 consecutive nt. Each horizontal line indicates the nucleotides deleted in a single such variant, with the vertical position indicating the fold-change in *SF3B1*-mutant (red) or WT (blue) cells. Plot restricted to introns with $\log_2$ (fold-change) < -0.5 or > -0.1 in *SF3B1*-mutant and WT cells, respectively. Grey lines, deletion used to create synMTERFD3i1-100 (first 15 nt and last 85 nt preserved). **(b)** Box plots illustrating relative fold-changes for all introns derived from

synMTERFD3i1-100 by single-nucleotide mutations, where the mutations affected the 10 5'-most nucleotides (5'ss), 26 3'-most nucleotides (3'ss), or neither (no). **(c)** Box plots illustrating relative fold-changes for all introns derived from synMTERFD3i1-100 by one or two single-nucleotide mutations, grouped by the difference in splice site strengths (MaxEnt) between the most intron-distal and most intron-proximal 3' splice sites. **(d)** Box plot showing the fold-change of introns with 0-8 nucleotides between canonical 3'ss and the AG of an inserted P(Y) with a perfect 3'ss derived from synMTERFD3i1-100. **(e)** Top, purine:pyrimidine ratio calculated with a 5 nt sliding window. Bottom, schematic of synMTERFD3i1-100 intron. **(f)** Line plots illustrating relative fold-changes for introns with the indicated perturbations at the indicated positions to synMTERFD3i1-100 for *SF3B1*-mutant (red) or WT (blue) cells. Data presented as geometric mean of fold-changes for the three closest introns (line) ± geometric s.d. from the mean (shading). From top to bottom, perturbations are: sliding deletions (5 nt); sliding conversion of all Ys within 6 nt to G, where position corresponds to center of the 6 nt window; sliding insertion of consensus branchpoint context (tactaAca), where position represents point of insertion; simultaneous ablation of branchpoints (A>G) and sliding insertion of consensus branchpoint context (tactaAca), where position represents point of insertion. **(g)** Sequence logo plots illustrating relative fold-change for single-nucleotide mutations to synMTERFD3i1-100 in WT (top) or *SF3B1*-mutant (bottom) cells. Height of each nucleotide indicates the fold-change for that mutation.

4.6 METHODS

4.6.1 *Expression vector cloning*

mCardinal-pBiCMV-mEmerald: Oligonucleotides containing the endogenous GAPDH 5' UTR (from genomic DNA), mEmerald coding sequence (from mEmerald-N1; Addgene Plasmid 53976), and mCardinal coding sequence (from mCardinal-N1; Addgene Plasmid 54590) were synthesized and cloned into the pRRLSIN.cPPT.PGK-GFP.WPRE (Addgene Plasmid 12252) backbone, along with the pBiCMV promoter (from pBi-CMV1 (Clontech); Addgene Vector 6166), to replace the PGK-GFP sequence. The GAPDH 5' UTR sequence is: gcctcaagaccttgggctgggactggctgagcctggcgggaggcggggtccgagtcaccgcctgccgccgcgcccccggtttctataaa ttgagcccgcagcctcccgcttcgctctctgctcctcctgttcgacagtcagccgcatcttcttttgcgtcgccagccgagccacatcgctcag acacc. Orientation of fragments was as illustrated in **Fig. 1e**. hPGK-HSV-TK-P2A-mCherry: The

HSV-TK coding sequence (from pAL119-TK; Addgene Plasmid 21911) was cloned into the pRRLSIN.cPPT.PGK-SF3B1 WT-FLAG-P2A-mCherry.WPRE [PMID 31961934] backbone to replace the SF3B1 WT-FLAG sequence. The HPGK-HSV-TK-P2A-mCherry sequence was flipped using XhoI and SalI enzymes so that the intron is not spliced out during lentivirus production. hPGK-PuroR-P2A-HSV-TK: The puromycin resistance coding sequence (from pLenti CMV GFP Puro; Addgene Plasmid 17448) with P2A was cloned into the hPGK-HSV-TK-P2A-mCherry backbone after excising the P2A-mCherry sequence. All pieces were amplified using Phusion or Q5 polymerase from NEB. Assembly was done using NEBuilder HiFi (New England Biolabs) according to manufacturer's instructions. All truncated intron sequences were initially ordered as gBlocks from IDT.

4.6.2 *Transfection and Flow Cytometry*

K562 Cells were transfected with fluorescent reporters using Lonza Cell Line Nucleofector V Kit as described in the kit protocol. Flow was done 72 hours after transfection using GFP and APC wavelengths. The cells were spun down and resuspended in PBS for this. Shown cells were filtered to express the mCherry control and then for each cell the mEmerald expression was divided by the mCardinal expression. Gates were first set to capture all live cells then we gated to only analyze mCardinal positive cells.

4.6.3 *Lentivirus production*

Expression vector plasmids were co-transfected with psPAX2 (Addgene plasmid 12260) and envelope vector pMD2.G (Addgene plasmid 12259) into 293T cells. Lentivirus was collected from the supernatant 48 hours after transfection. Stable cell lines were made by transducing K562, MEL202, or MEL270 cells with lentivirus at multiplicities of infection of 1 (Fig. 4:2c), (Fig. 4:2f,g), 0.3 (mini-library), and 0.1 (full library). Positive integrants were selected by treating with puromycin (for hPGK-PuroR-P2A-HSV-TK) or flow sorting for mCherry (hPGK-HSV-TK-P2A-mCherry).

4.6.4 *Cell viability measurements*

Cell viability was measured for single-construct experiments in cell culture by the CellTiter-Glo Luminescent Cell Viability Assay (Promega). **Figure 4:2c** K562 cells expressing HSV-TK with the indicated synthetic introns were seeded at a density of 5000 cells/100 uL/well in 96-well plate in triplicate, and then treated with 0-100 ug/mL GCV or PBS (control). **Figure 4:2g**, K562 cells expressing HSV-TK with the indicated synthetic introns were seeded at a density of 5000 cells/100 uL/well in 96-well plate in triplicate, and then treated with 100 ug/mL GCV or PBS (control). Viability was measured after 11 days of treatment. For **Figure 4:2f** cells were seeded at a density of 10,000cells/100µL/well in a 96 well plate in triplicate, and then treated with 1 µg/mL GCV. Viability was measured after 7 days.

4.6.5 *RT-PCR of Endogenous and Reporter RNA*

Total RNA was extracted using Direct-zol RNAMini Prep (Zymo Research.) CDNA was synthesized using Superscript IV Reverse Transcriptase (Thermo Fisher Scientific) suing the manufacturer's protocol. Gene Specific primers used for amplifications are in Supplementary Table X . Amplicons were analyzed using agarose gel electrophoresis and quantified using ImageJ.

4.6.6 *Cell Culture*

K562 Cells were grown at 37 °C and 5% atmospheric CO2 in Iscove's Modified Dulbecco's Medium (IMDM; Gibco) supplemented with 10% fetal bovine serum (Gibco).

4.6.7 *K562 xenografts*

Luciferase-expressing K562 cells were established by infecting cells with lentivirus created from pMSCV-Luciferase-PGK-GFP (Addgene plasmid 18782; HygR replaced by GFP) at MOIs of 0.9 (*SF3B1*$^{+/+}$) and 0.5 (*SF3B1*$^{+/K700E}$). GFP^{+} cells were isolated by flow sorting 7 days after infection. K562 cells expressing Luciferase and HSV-TK with synMTERFD3i1-250 were intravenously injected into sub-lethally irradiated (250 cGy) NOD-*scid* IL2Rgnull (NSG) mice (2 million cells/mouse, n=10 pr 9). Leukemic cells were allowed to grow for 24 days before mice were treated with Ganciclovir (80 mg/kg) via intraperitoneal injection three times per week. Bioluminescence imaging was carried out weekly with 150 mg/kg of D-Luciferin.

4.6.8 *Mini-library construction, screen, and analysis*

Introns in the mini library were ordered individually as gBlocks from IDT and mixed in equal proportions before being cloned into hPGK-PuroR-P2A-HSV-TK. (The cloning protocol here was identical to the pgFARM paper) The Intron mix was amplified using NEBNext High Fidelity Rady Mix and purified using 1.8X Ampure XP SPRI beads (Beckman Coulter.) The backbone for the library was amplified using Q5 polymerase from NEB. The library was transformed and amplified using Endura ElectroCompetent Cells (Lucigen 60242-2) and large LB plates. The library was maxiprepped using Macherey-Nagel MaxiPrep kit (Cat # 740414.10.)

WT or *SF3B1*-mutant K562 cells were infected with lentivirus encoding the mini-library at an MOI of 0.3 and treated with GCV (100 ug/mL) for 6 days. Genomic DNA was collected at day 0 and day 6 and the resulting Illumina libraries were sequenced with 2x150 bp reads (Illumina MiSeq). Depletion/enrichment of each construct was estimated as follows. For each sample, reads were normalized to the total reads mapped. The relative fraction of reads mapping to each intron was then estimated by dividing the numbers of normalized reads mapped to an intron by the total reads mapped in the sample. The standard deviation was calculated using the sample proportion P for each intron ($\sigma = \text{sqrt}\ [\ P(1 - P) / n\ ]$). A fold-change was calculated for each intron by dividing the proportion of the intron in the treated GCV sample by the fraction of the intron in the untreated samples. Error propagation was used to estimate the standard deviation. The final relative fold-changes were computed by normalizing fold-changes such that the fold-change of synMTERFD3i1-250 in the pilot screen was identical to the experimentally measured fold-change in a single-construct experiment (**Figure 4:2e**, 100 µg/mL of GCV).

4.6.9 *Full library construction, screen, and analysis*

WT or *SF3B1*-mutant K562 cells were infected with lentivirus encoding the full library at an MOI of 0.1 and treated with GCV (100 ug/mL) for 8 days. Genomic DNA was collected at day 0 and day 8 and the resulting Illumina libraries (triplicates) were sequenced with both 2x150 bp and 2x250 bp reads (Illumina MiSeq).

Relationships between differences in 3' splice site strengths and mutation-dependent responses (**Figure 4:3**) were analyzed as follows. All AG dinucleotides that occurred in each intron

sequence were identified. The corresponding 23 nt of context (20 nt before the intron-exon junction and 3 nt after) that define each such candidate 3'ss were extracted and used to compute a 3'ss strength with MaxEnt. After computing these MaxEnt scores for all candidate 3'ss in an intron, the difference in strength between the two most intron-distal 3'ss was computed as (score for most intron-distal 3'ss - score for next most intron-distal 3'ss). This is equivalent to comparing the difference in splice site strength between the most intron-distal cryptic 3' splice site with the canonical 3' splice site, unless the canonical 3' splice site is ablated by the mutation, in which case two cryptic 3' splice sites will be compared. This analysis was restricted to introns derived by introducing one or two single-nucleotide mutations to synMTERFD3i1-150.

4.6.10 *gDNA PCR*

gDNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. Intronic region of interest was amplified using primers listed in Supplementary Table X and analyzed using agarose gel electrophoresis.

4.6.11 *RNAseq Analysis*

RNAseq data anslsyis reference used previously: Dvinge, H. et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. Proc. Natl Acad. Sci. USA 111, 16802–16807 (2014).

4.6.12 *Library Mapping*

Reads were first trimmed using cutadapt 2.1 to remove sequenced Illumina adapters and most of TK reads. Adapters were defined as "acgcggcggtggtaatgacaa" and "ggccgatatctcaccctggtc" for the 3' end and 5' respectively. Trimmed reads were then combined into single reads using FLASH-1.2.11 with a minimum length sequence of 70. Merged reads are mapped using bowtie2 –very-sensistive parameters. Mapped reads are then filtered for a minimum MAPQ score of 1 and counted for each intron.

4.7 ACKNOWLEDGEMENTS

RKB was supported by the Leukemia & Lymphoma Society (1344-18) and NIH/NIDDK (R01 DK103854). OA-W and RKB are supported by the Edward P. Evans Foundation, NIH/NCI (R01 CA251138-01), and the NIH/NHLBI (R01 HL128239). OA-W is supported by NIH/NCI R01CA242020) and the Henry and Marilyn Taub Foundation for MDS Research. The results published here are based in part upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

4.8 COMPETING INTERESTS

O.A.-W. has served as a consultant for H3B Biomedicine, Foundation Medicine Inc, Prelude Therapeutics, Merck, and Janssen, and is on the Scientific Advisory Board of Envisagenics Inc and AIChemy; O.A.-W. has received prior research funding from LOXO Oncology and H3B Biomedicine unrelated to the current manuscript

4.9 REFERENCES

1. Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet. 2016;17(1):19-32.
2. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011;365(15):1384–1395.
3. Visconte V, Makishima H, Jankowska A, et al. SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia*. 2011
4. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365(26):2497–2506
5. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer*. 2016;16(7):413–430.
6. M. Seiler, et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nat. Med., 24 (2018), pp. 497-504
7. Kim E, Ilagan JO, Liang Y, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell.* 2015;27(5):617–630.
8. Ilagan JO, Ramakrishnan A, Hayes B, Murphy ME, Zebari AS, Bradley P, Bradley RK. U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res. 2015 Jan;25(1):14-26. doi: 10.1101/gr.181016.114. Epub 2014 Sep 29.
9. Lee SC, Dvinge H.,et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat Med. 2016 Jun;22(6):672-8. doi: 10.1038/nm.4097. Epub 2016 May 2. Erratum in: Nat Med. 2016 Jun 7;22(6):692.

10. Lee SC, North K., et al. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. Cancer Cell. 2018 Aug 13;34(2):225-241.e8. doi: 10.1016/j.ccell.2018.07.003.
11. Seiler M, Peng S, Agrawal AA, et al. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep*. 2018;23(1):282–296.e4.
12. KI, Boultwood J. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia. 2015 May;29(5):1092-103. doi: 10.1038/leu.2014.331. Epub 2014 Nov 27. Erratum in: Leukemia. 2015 Aug;29(8):1798. Erratum in: Leukemia. 2015 Aug;29(8):1798
13. Tyner JW, et al. Functional genomic landscape of acute myeloid leukaemia. Nature. 2018 Oct;562(7728):526-531.
14. Moolten et al. Tumor Chemosensitivity Conferred by Inserted Herpes Thymidine Kinase Genes: Paradigm for a Prospective Cancer Control Strategy. Cancer Res October 1 1986 (46) (10) 5276-5281;
15. Zarogoulidis P, Darwiche K, Sakkas A, et al. Suicide Gene Therapy for Cancer - Current Strategies. *J Genet Syndr Gene Ther*. 2013;4:16849. doi:10.4172/2157-7412.1000139

Chapter 5. DISCUSSION

My graduate research focused on three distinct but related questions exploring the molecular drivers of disease and sequence characteristics that influence mutant SF3B1 promoted alternative splicing. SF3B1 is one of the most mutated splicing factors among the most diverse set of cancers. This makes SF3B1 an essential gene of focus in cancer biology. Understanding how mutations in SF3B1 drive disease and why specific mRNA sequences are susceptible to mutant SF3B1 miss splicing contributes to downstream treatment development in many patients.

In Chapter 2, I show that co-expressed *SF3B1* and *SRSF2* mutations are lethal to the cell, and when expressed individually, both mutations result in hyperactive NF-kB signaling. Before this work, there were two unanswered questions in the field: 1.) Are splicing factor mutations mutually exclusive, or have we just not screened enough patients? 2.) How do two very different splicing factors, when mutated, lead to the same disease's development? Mutual exclusivity in cancer is an important topic that can guide our understanding of cancer development and treatment decisions. Using an isogenic mouse model with either a single mutation in *Sf3b1* or *Srsf2* or a double mutation, we show that the two splicing factors have very distinct consequences on splicing and, when expressed together, block hematopoiesis *in vivo*. Analysis of the disrupted splicing in Srsf2P95H/+ Sf3b1K700E/+ cells showed, surprisingly, that the impaired differentiation and survival of hematopoietic stem/progenitor cells was not due to widespread disruption in splicing but as a result of several key alternatively spliced genes such as the thrombopoietin receptor, *Mpl.* This finding confirmed that splicing factors do not co-occur in patients in part due to their synthetic lethality. In the same study, we also show that mis-splicing in two different genes, *MAP3K7* and *CASP8* by mutant SF3B1 and SRSF2, respectively, results in hyperactive NF-kB signaling. Recent research has shown that NF-kB activity in cancer can promote cellular proliferation, suppress apoptosis, and in some cases, even facilitate metastasis.

The study in Chapter 2 did not address the anecdotal evidence we had, at the time, of increased tolerance of SF3B1 K666 mutations in the double mutant contexts. In Chapter 3, we next asked whether some specific and minor splicing factor mutants could escape epistasis and, if so, how. Increased patient data and single-cell sequencing confirmed that SF3B1 K700E/+ and

SRSF2P95H/+ mutations are not tolerated in the same cell as previously showed but revealed that SF3B1 K666N/+ and SRSF2P95H/+ are. This raised a question about whether the two different SF3B1 mutants result in different differential splicing consequences. Here we show that SF3B1 with a K666 mutation has a significantly less dramatic effect on splicing than a K700E mutation. This finding indicates that some therapies developed primarily for patients with SF3B1 K700E mutations may not fit if an alternative mutation in the protein is present.

Chapter 2 and Chapter 3 showed us that patient data was not enough to elucidate the sequence characteristics that make genes most vulnerable to SF3B1 miss splicing. We also saw the importance of having a therapy that would be effective no matter what the specific SF3B1 mutation is. Chapter 4 addresses both of these concerns by introducing a massively-parallel intron screen, which allowed us to identify important intronic sequence components and fine-tune an intron dependent delivery system of thymidine-kinase for ganciclovir-mediated elimination of leukemia cells. From this screen, we learned that the most critical sequence components that govern SF3B1 genotype-specific splicing are in the first 25 and last 125 nucleotides of an intron. We also show that SF3B1 genotype-dependent splicing can be maintained even when the intron is cut down to 100bp. Although less responsive, the smaller introns can be made more genotype-specific with modifications to the branchpoint region, among others. I am excited to see the future of this technology. This work will open opportunities for not only new and more targeted treatment opportunities for patients with SF3B1 mutations, but it also sets guidelines for engineering other intron-switches responding to SRSF2, U2AF1, and ZRSR2.