



# VCU

Virginia Commonwealth University  
VCU Scholars Compass

---

Theses and Dissertations

Graduate School


---

2022

## Alternatively Spliced CSF3R in Human Health and Disease

Borwyn A. Wang  
*Virginia Commonwealth University*

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>

 Part of the [Cancer Biology Commons](#), [Developmental Biology Commons](#), and the [Molecular Genetics Commons](#)

© Borwyn A. Wang

---

Downloaded from

<https://scholarscompass.vcu.edu/etd/6896>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact [libcompass@vcu.edu](mailto:libcompass@vcu.edu).

© Borwyn A. Wang 2022

All Rights Reserved

**Alternatively Spliced *CSF3R* in Human Health and Disease**

A dissertation submitted in partial fulfillment of the requirement for the degree of Doctor of  
Philosophy at Virginia Commonwealth University.

by  
Borwyn A. Wang  
B.S. University of Mary Washington  
M.S. Virginia Commonwealth University

Thesis Advisor: Seth J. Corey, MD, MPH  
Professor, Pediatric and Molecular Medicine, Department of Pediatrics and Cancer Biology,  
Cleveland Clinic

Virginia Commonwealth University  
Richmond, Virginia  
April 4, 2022

## Acknowledgements

My lab mates and colleagues from VCU, LRI, and CWRU have been instrumental in my success throughout the last five and half years. First and foremost, I would like to acknowledge Dr. Seth Corey for his mentorship throughout this project. He has taught me how to become a better writer, presenter, and critical thinker. I especially want to thank Dr. Hrishikesh (Rishi) Mehta for his continuous support and guidance since my joining the lab. His patience and scientific rigor are qualities that I will carry forward into the future. I also wish to thank Dr. Usua Oyarbide, Dr. Bhavuk Garg, and Dr. Rebecca Anderson for their thoughtful discussions and constant encouragement to keep going when times became challenging.

I would like to thank my committee members Dr. Rita Shiang, Dr. Joseph Landry, Dr. Larissa Litovchick, and Dr. Mikhail Dozmorov for their advice and input on this project. I thank Dr. Shiang for being my primary advisor after I moved to Cleveland.

I would also especially like to thank Dr. Srinivasa Penumutchu (Vas) for helping me with the ITC experiment and teaching me how to analyze the data. Dr. Juan Sabin helped us curve fit a particularly unique case and I thank him for his expertise.

The mouse studies would not have been possible without help from the Flow Core at LRI. I thank Jena Korecky for performing my flow cytometry sorting experiments.

I would like to thank the VCU CCTR program for supporting my scientific journey and pushing me to achieve excellence.

Finally, I would like to thank my mother and father for their unconditional love and support. My father encouraged me to pursue higher education and I wish he were still alive to see the completion of this work. My mother has been with me every step of the way and for that, I am forever grateful.

## Table of Contents

Acknowledgements.....	ii
List of Tables .....	vii
List of Figures .....	viii
List of Abbreviations.....	x
Abstract.....	xii
Alternative <i>CSF3R</i> Splice Isoforms in Human Health and Disease .....	xii
Overview.....	14
Hypotheses.....	16
Objectives.....	17
Chapter 1: Background.....	19
Colony Stimulating Factor 3 Receptor .....	19
Alternative Splicing of <i>CSF3R</i> in Neutrophil Development and Diseases.....	24
Overview of mRNA Splicing.....	27
Mechanism of RNA Splicing .....	29
Regulation of Alternative Splicing .....	32
Chapter 2: Comparative Analysis of Neutrophils and <i>CSF3R</i> Isoforms Across Species and Non-Human Primates .....	41
2.1 Rationale .....	41
2.2 Materials and Methods .....	41
Comparison of <i>CSF3R</i> Between Humans and Non-Human Primates .....	41
Isolation of Rhesus Macaque Mononuclear Cells and Neutrophils .....	42

cDNA Synthesis.....	42
PCR.....	43
2.3 Results .....	43
Mouse Lemur CSF3R is Most Evolutionary Divergent from Humans.....	43
CSF3R Class IV is Present in Rhesus Macaques .....	50
2.4 Discussion.....	51
Chapter 3: The Role of MDS-Associated Splicing Factor Mutations on CSF3R Splicing .....	52
3.1 Rationale.....	52
3.2 Materials and Methods .....	53
Cell Culture and Reagents .....	53
Isolation, Purification, and Differentiation of Human CD34+ Cells .....	53
Construction of CSF3R Minigene ESE Deletional Mutants.....	54
SRSF2 RRM Protein Purification .....	54
Isothermal Titration Calorimetry .....	55
Mouse Lin <sup>-</sup> Sca-1 <sup>+</sup> c-Kit <sup>-</sup> (LSK) Purification and Colony Assays .....	55
Csf3r <sup>-/-</sup> Mice Genotyping .....	56
RNA-Sequencing and Analysis of Patient Data .....	57
PCR and qPCR Primer Design .....	57
cDNA Synthesis.....	58
PCR and Quantitative Real-Time PCR.....	58
Lentivirus Production .....	59

Mutagenesis of SRSF2 P95H .....	59
Generation of LUC7L2 Knockout Cells.....	60
Cell Transfection and Transduction.....	60
Western Blotting.....	61
Cloning of CSF3R Class III into CSII-CMV-IRES2 Lentiviral Backbone.....	61
Statistical Analysis .....	62
Study Approval .....	62
3.3 Results .....	63
MDS-Associated Splicing Factor Mutations Lead to Altered Splicing of <i>CSF3R</i> .....	63
SRSF2 P95 Mutations Increase <i>CSF3R</i> Class IV Splicing .....	74
Co-Mutations in Splicing Factor Mutated Myeloid Neoplasms May Enhance <i>CSF3R</i> Splicing .....	80
ESEs Contribute to <i>CSF3R</i> Class I and Class IV Splicing.....	84
Overexpression of SRSF2 Lead to Impaired Granulopoiesis in CD34+ Cells.....	96
Alternative <i>CSF3R</i> Isoforms Do Not Rescue Granulopoiesis in <i>Csf3r</i> -Null Progenitor Cells in Response to Granulocytic Differentiating Factors.....	99
3.4 Discussion .....	105
Chapter 4: Perspectives and Future Directions .....	112
Cell Line Models of SRSF2 P95 Mutations .....	112
<i>In Vivo</i> Models of Class III and Class IV on Granulopoiesis.....	113
The Role of LUC7L2 on <i>CSF3R</i> Splicing .....	113
Implications for Splicing Modulators as Therapeutics .....	113

Chapter 5: Concluding Remarks .....	120
Development of MDS May Depend on Presence of Alternative Splice Isoforms .....	120
<i>CSF3R</i> Splicing is affected by MDS-Associated Splicing Factor Mutations .....	121
SRSF2 Regulates <i>CSF3R</i> Class IV Splicing Through ESE Recognition .....	122
Conclusion .....	123
Appendix A .....	125
Appendix B .....	142
Appendix C .....	143
References .....	148



## List of Tables

<b>Table 2.1 Rhesus Macaques Sample Characteristics .....</b>	<b>42</b>
<b>Table 2.2. Standard PCR Mix.....</b>	<b>43</b>
<b>Table 3.1. ESE Motif Outputs from ESEfinder 3.0.....</b>	<b>87</b>
<b>Table 3.2. Dissociation Values for SRSF2 RRMs Titrated with ESE1 or ESE2. ....</b>	<b>91</b>

## List of Figures

Figure 1.1. Schema of Hematopoiesis. ....	20
Figure 1.2. Schematic of CSF3R Protein Isoforms. ....	22
Figure 1.3. Schema of CSF3R Signaling. ....	24
Figure 1.4. Alternative Splicing of <i>CSF3R</i> Pre-mRNA. ....	26
Figure 1.5. Mechanism of Pre-mRNA Splicing. ....	31
Figure 1.6. Schematic Representation of Alternative Splicing. ....	34
Figure 1.7. Regulation of Alternative Splicing by SR Proteins and hnRNPs. ....	36
Figure 1.8. Splicing Factor Mutations on Alternative Splicing. ....	40
Figure 2.1. CSF3R Phylogenetic Tree of Humans and Non-Human Primates. .....	47
Figure 2.3. Comparison of CSF3R Class III Nucleotide Sequences between Humans and Non-Human Primates.. ....	48
Figure 2.4 Comparison of CSF3R Class IV Nucleotide Sequences between Humans and Non-Human Primates.....	50
Figure 3.1. LUC7L2 CRISPR-Cas9 Knockout Clones in K562 Cells.....	67
Figure 3.2. <i>CSF3R</i> Minigene Splicing Assay with MDS-Related Splicing Factors. ....	68
Figure 3.3. PCR Gels from <i>CSF3R</i> Minigene Splicing Assays with MDS-Related Splicing Factors. ....	69
Figure 3.4. Proliferation and Viability of LUC7L2 KO in HL60 Cells. ....	70
Figure 3.5. Cytospins of LUC7L2 KO in HL60 Cells Differentiated with ATRA. ....	71
Figure 3.6. Cytospins of LUC7L2 KO in K562 Cells Differentiated with PMA. ....	72
Figure 3.7. Correlation of <i>CSF3R</i> Class III and IV Expression with Recurrent Mutations of Splicing Factors Found in Myeloid Neoplasia. ....	73
Figure 3.8. Class III and Class IV <i>CSF3R</i> Splicing in K562 Cells with SRSF2 Expression. ....	76
Figure 3.9. PCR Gels of Class I, Class III, and Class IV <i>CSF3R</i> in K562 Cells Expressing SRSF2... ..	77
Figure 3.10. Distribution of Patients with Splicing Factor Mutations Stratified by Disease.. ....	78
Figure 3.11. <i>CSF3R</i> Splicing by SRSF2 P95 Mutations in Patients with Myeloid Neoplasms. ....	79

Figure 3.12. Frequency of Genes Co-Mutated with Splicing Factors in Patients with Myeloid Neoplasms. ....	81
Figure 3.13. Association of Co-Mutations with High and Low Class III <i>CSF3R</i> Expression. ....	82
Figure 3.14. Association of Co-mutations with High and Low Class IV <i>CSF3R</i> Expression. ....	83
Figure 3.15. SRSF2 WT and SRSF2 P95H RRM Structure and Purity.....	90
Figure 3.16. Global Curve Fit for SRSF2 RRMs Titrated with ESE1 with Residual Plots.....	94
Figure 3.19. Class IV:I Ratios in ESE-Deleted K562 Cells.....	95
Figure 3.20. Expression of Exogenous SRSF2 in CD34+ Cells.....	97
Figure 3.21. Differentiation of CD34+ Cells with Expression of WT or Mutant SRSF2. ....	98
Figure 3.22. <i>CSF3R</i> is Developmentally Regulated in Human CD34+ Cells. ....	101
Figure 3.23. Colony Scores from <i>Csf3r</i> WT LSK Cells. ....	102
Figure 3.24. Comparison of <i>CSF3R</i> Isoforms and their Contributions to Colony Growth. ....	103
Figure 3.25. Flow Cytometric Analysis for CD11b <sup>+</sup> CD11c <sup>-</sup> Ly6C <sup>low</sup> Ly6G <sup>+</sup> (Neutrophils) and CD11b <sup>+</sup> CD11c <sup>-</sup> Ly6C <sup>+</sup> Ly6G <sup>-</sup> (Monocytes). ....	104
Figure 3.26. Proposed Model of SRSF2 WT and SRSF2 P95H Binding Competition to ESEs. ....	111
Figure 4.1. Class IV:I <i>CSF3R</i> in K562 Cells Treated with Dasatinib and Imatinib.....	117
Figure 4.2. Class IV:I <i>CSF3R</i> in K562 Cells Treated with Risdiplam.....	118

## List of Abbreviations

(Abbreviations appear in alphabetical order)

5'SS	5' Splice Site
3'SS	3' Splice Site
AML	Acute Myeloid Leukemia
ATRA	All Trans Retinoic Acid
CMML	Chronic Myelomonocytic Leukemia
CSF3R	Colony Stimulating Factor 3 Receptor
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FACS	Fluorescence-Activated Cell Sorting
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
G-CSF	Granulocyte Colony Stimulating Factor
HSC	Hematopoietic Stem Cell
LUC7L2	LUC7 Like 2
MDS	Myelodysplastic Syndromes
MFI	Mean Fluorescence Intensity
mRNA	Messenger RNA
mRNP	Messenger Ribonucleoprotein
NMD	Nonsense Mediated Decay
ITC	Isothermal Titration Calorimetry
qPCR	Quantitative (real-time) Polymerase Chain Reaction
RRM	RNA Recognition Motif

RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SF3B1	Spliceosome Factor 3B Complex
snRNA	Small Nuclear RNA
snRNP	Small Nuclear Ribonucleoprotein
SR	Serine/Arginine
SRSF2	Serine/Arginine-Rich Splicing Factor 2
U2AF35/U2AF1	U2 Small Nuclear RNA Auxiliary Factor 1
ZRSR2	Zinc Finger CCCH-Type, RNA Binding Motif and Serine/Arginine Rich 2

## Abstract

### Alternative *CSF3R* Splice Isoforms in Human Health and Disease

A dissertation submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by  
Borwyn A. Wang  
B.S. University of Mary Washington  
M.S. Virginia Commonwealth University

Thesis Advisor: Seth J. Corey, MD, MPH  
Professor, Pediatric and Molecular Medicine, Department of Pediatrics and Cancer Biology,  
Cleveland Clinic

Myelodysplastic syndromes (MDS) constitute the most common group of bone marrow failure disorders, characterized by ineffective hematopoiesis and a significant risk of transformation to AML. Efforts to understand the molecular basis of MDS led to the identification of acquired somatic mutations in the RNA splicing factors *SF3B1*, *U2AF1*, *SRSF2*, *PRPF8*, and *DDX41*. Haploinsufficiency of *ZRSR2* and *LUC7L2* are also recurrently found in MDS. Altogether, these abnormalities are seen in 45-85% of adult MDS patients. Despite evidence that these mutated genes are founder lesions, how they perturb hematopoiesis and promote expansion of abnormal clones remain unknown. Elevated alternative colony stimulating factor 3 receptor (*CSF3R*) isoforms, called classes, are observed in MDS and other myeloid neoplasms and may drive disease progression. *CSF3R* and its cognate ligand, G-CSF, are critical components that regulate neutrophil development. Our previous work show that Class IV *CSF3R* confers a differentiation-defective phenotype, but the mechanisms describing its regulation by splicing factors were unclear. I show that a small group of MDS-associated splicing factor mutations have differential effects on *CSF3R* splicing into Class III and Class IV. Specifically, I demonstrate that splicing

factor SRSF2 P95 mutations promote Class IV splicing in minigene splicing assays and validated these findings in patients with myeloid neoplasms. Since SRSF2 regulates splicing in part by recognizing exonic splicing enhancer (ESE) motifs on target pre-mRNA, deletion of two critical motifs (ESE1 and ESE2) within *CSF3R* exon 17 decreased Class IV transcript expression. Binding affinities measured by isothermal titration calorimetry (ITC) revealed that SRSF2 WT and SRSF2 P95H have opposing dissociation constants where binding of SRSF2 P95H to the stronger ESE motif may favor Class IV splicing. Human CD34+ cells overexpressing SRSF2 P95H mutations showed impaired neutrophil differentiation accompanied by an increase in Class IV expression in response to G-CSF. I also describe the phenotypes of Class III and Class IV *CSF3R* in isolation by adding back individual isoforms in mice lacking *Csf3r*. Colony assays of Class IV add-back experiments showed impaired neutrophil differentiation with increased colony forming unit granulocytes (CFU-G) while Class III add-back showed greater dysgranulopoiesis with dysmorphic neutrophils and an overall decrease in colony numbers. Having established that splicing factor mutations can favor unproductive *CSF3R* isoforms to drive disease progression, treating cells with splicing modulators to correct the isoform imbalance is an approach gaining attention. K562 cells treated with Risdiplam, a drug used to modify the splicing of SMN2 which is responsible for spinal muscular atrophy, reduced the expression of Class IV while having no effect on Class III. We demonstrate for the first time a mechanism of Class IV *CSF3R* splicing through SRSF2 and show how increased levels of alternate isoforms can contribute to disease progression. Our findings shed insights into how aberrant splicing of *CSF3R* drives MDS and provides a new model of dysgranulopoiesis.

## Overview

Mutations in the RNA splicing machinery are understood to be the most common abnormalities in MDS (1-3). Within the last decade, much progress has been made in understanding its pathogenesis. MDS is a highly heterogeneous disease that generally affects the aging population. The incidence rate remains low in individuals <40 years, at ~0.1 per 100,000 people per year. The incidence rises to 26.9 per 100,000 people per year for individuals 70-79 years and further to 55.4 per 100,000 people per year for those aged >80 years (4, 5). It is characterized by ineffective hematopoiesis with peripheral blood cytopenias, dysplastic changes in  $\geq 10\%$  of cells of one or more myeloid lineages, and a significant risk of transformation into AML (~30-40%). Advances in sequencing techniques have revealed that MDS patients carry a median of 9 somatic mutations within the coding sequences or approximately 15,000 mutations across the entire genome (6, 7). The numbers and types of mutations vary depending on MDS subtype, but typical patients have a median of 2 or 3 driver mutations out of >30 driver genes identified (6, 8). Patients with high risk MDS have a median survival of <2 years and correlate with the number of driver mutations. Aside from splicing factors, other mutated genes include RUNX1, TP53, STAG2, NRAS, and EZH2 (6). Currently, the only curative treatment is hematopoietic stem cell transplant, which is only a viable option for a small percentage of patients. For those who are ineligible for stem cell transplant, they undergo maintenance therapy that aim to improve blood counts and quality of life. Therapies for this subgroup include hypomethylating agents, induction chemotherapy, or best supportive care (9, 10).

Splicing factor mutations represent a major class of driver mutations occurring as founder lesions in early HSC clones (11, 12). A few of these splicing factors show discrete hotspot mutations with the absence of frameshift or nonsense alterations, which suggest that they confer neomorphic functions and result in aberrant regulation of target genes. The precise mechanisms of how splicing factor mutations drive MDS pathogenesis are still largely unclear, but several



putative genes have been identified as targets of missplicing. *CSF3R* is one such target that may explain common MDS phenotypes because aberrant splicing results in usage of alternative isoforms that have impaired function. *CSF3R* represents one gene out of thousands that collectively drive disease progression with compounded effects. With limited FDA approved therapies for MDS, splicing modulators to correct isoform imbalance has become a promising approach in recent years.

## Hypotheses

We sought to identify which splicing factors and their mutations regulated *CSF3R* alternative splicing into Class III and Class IV. We also wanted to determine if Class III or the differentiation-defective Class IV isoforms could partially rescue the phenotype of mice lacking *Csf3r*.

1. Alternative splicing of *CSF3R* is a normal biological phenomenon in developing neutrophils. Exon 17 contains many regulatory sequences and alternative splice sites that promote alternative splicing depending on the cellular context. Given its susceptibility to undergo alternative splicing, which involve tight regulation by splicing factors, I hypothesized that aberrant splicing activity of MDS-associated splicing factors and low expression of LUC7L2 promote the expression of Class III and Class IV *CSF3R*.
2. Normal granulopoiesis relies on proper signaling from cell surface receptors and their ability to bind to their cognate ligands. Class III and Class IV are elevated in AML patients, which suggest that they confer improper maturation signals that enhance the growth of immature neutrophils. I further hypothesized that increased expression of Class III and Class IV *CSF3R* will perturb signaling resulting in abnormal granulopoiesis.
3. Since Class IV *CSF3R* is elevated in MDS patients with SRSF2 mutations, it was suspected that *CSF3R* is a target of SRSF2 but no direct relationship has ever been reported. Therefore, I hypothesized that mutant SRSF2 P95H facilitates Class IV splicing by competing with wildtype SRSF2 in recognizing exonic splicing enhancer motifs.

## Objectives

**1. Compare *CSF3R* isoforms between humans and non-human primates.** The evolution of splicing, conservation of splice sites, and the emergence of new exons is an intriguing question as to why certain species such as mice do not phenocopy diseases found in humans. One striking observation is that mice do not have the Class III or Class IV *CSF3R* isoforms found in humans and their predominate leukocyte is the lymphocyte rather than the neutrophil. Neutrophils play an important role in innate immunity, whereas lymphocytes play major roles in adaptive immunity. As vertebrates evolved, so did the complexity of their gene structures and splicing patterns. Conservation of splice sites in addition to gene duplication and exon insertion lead to more splicing variation and generation of isoforms. Some isoforms have cell type or condition-specific functions. Others are induced by stress or enriched in disease states. The abundance and splice site choice under different conditions are tightly regulated and disruptions in the splicing network can lead to production of alternate isoforms that do not function properly. Production of these isoforms result in reduced capacity of the canonical isoforms to preserve homeostasis, which may render some cells susceptible to diseases. Our goal was to briefly compare *CSF3R* isoforms between humans and non-human primates to determine when they could have emerged during evolution.

**2. Determine which MDS-associated splicing factor mutations affect *CSF3R* splicing into Class I, III and IV.** MDS-related splicing factor mutations arise in hematopoietic stem and progenitor cells. However, splicing mechanisms of specific genes involved in the pathogenesis of MDS remain unclear. Previous RNA-sequencing studies on patients with myeloid neoplasms with splicing factor mutations have shown that genes display frequently dysregulated splicing events with a pattern of reduced levels of intron retention across disease samples. My approach to determine relevant splicing factors and their mutations on *CSF3R* splicing is by using a minigene

splicing assay to screen for potential candidates. Prior to this study, there were no reports on which splicing factors directly regulated *CSF3R* splicing.

**3. Assess the effect of SRSF2 P95 mutations on Class III and Class IV splicing and its contribution to granulopoiesis.** Studies in heterozygous SRSF2 P95H knock-in mice showed downregulation of *Csf3r* and impaired granulopoiesis. Patients with SRSF2-mutated MDS and AML have increased levels of Class III and Class IV. My goal was to investigate if SRSF2 P95L and SRSF2 P95R also promoted Class IV splicing. It is presumed that these mutations function similarly, but studies to date have only focused on SRSF2 P95H. We also wanted to establish a link between SRSF2 P95 mutations and dysgranulopoiesis, with Class IV being a driver of a disease phenotype.

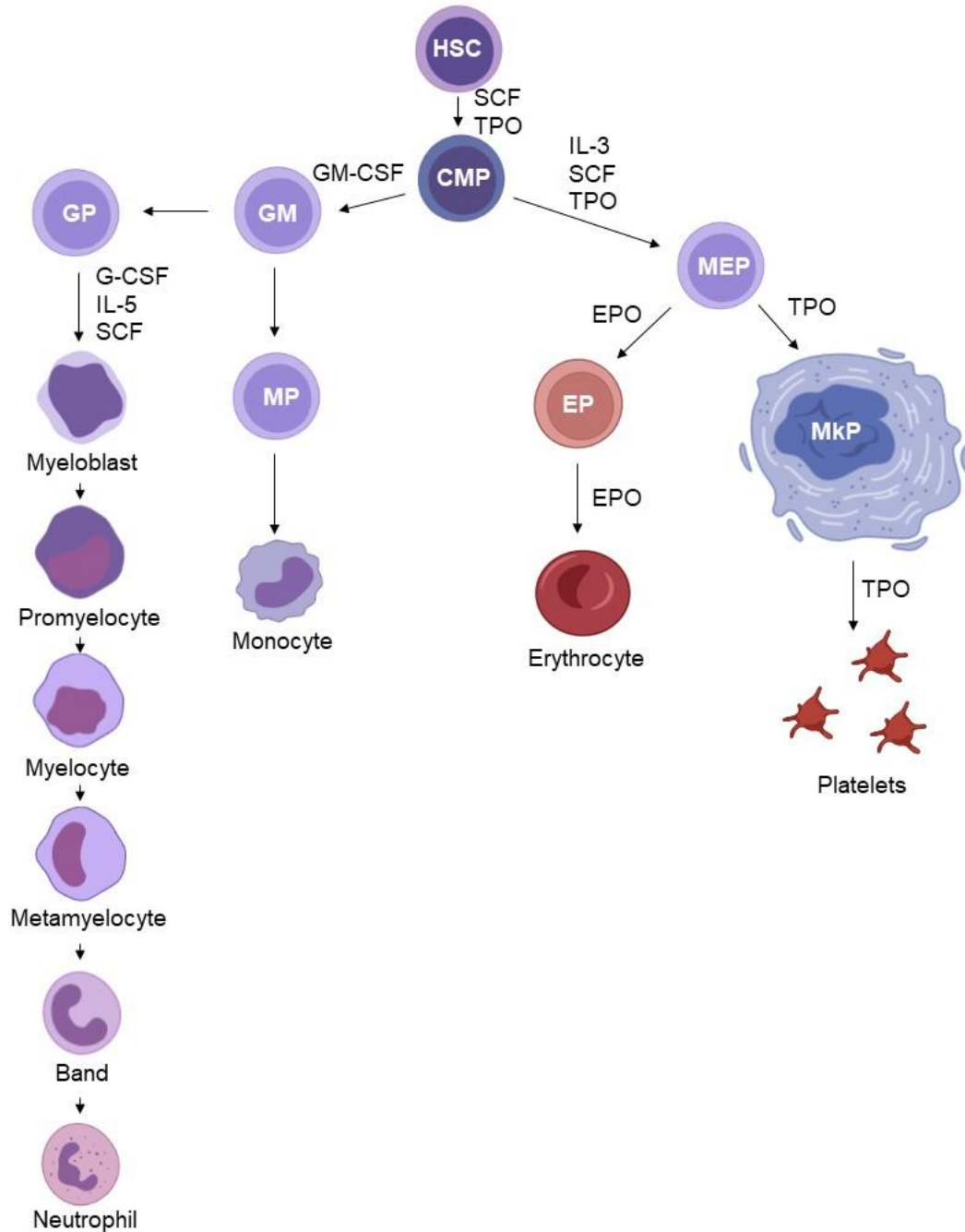
**4. Characterize the dysgranulopoiesis in the context of Class III and Class IV *CSF3R*.** Elevated levels of alternatively-spliced *CSF3R* interferes with neutrophil development. Normally, these isoforms exist as heterodimers with low percentages of Class IV homodimers. However, the contribution of individual isoforms and its effects on granulopoiesis has not been assessed in an *ex vivo* model. My goal was to determine how add-back of Class I, Class III, and Class IV affected neutrophil development in response to differentiating factors.

Overall, the work presented in my dissertation highlights the importance of specific isoform expression in the context of splicing factor mutations to drive hematopoietic diseases. I found that Class III results in severe dysgranulopoiesis when expressed alone in *Csf3r*-null mice. I also found that SRSF2 P95H promotes Class IV splicing through recognition of two key ESE sequences in exon 17. The impact of this work can be extended to other types of diseases and gene targets where correction of alternative isoform usage can be a potential therapeutic intervention.

## Chapter 1: Background

### Colony Stimulating Factor 3 Receptor

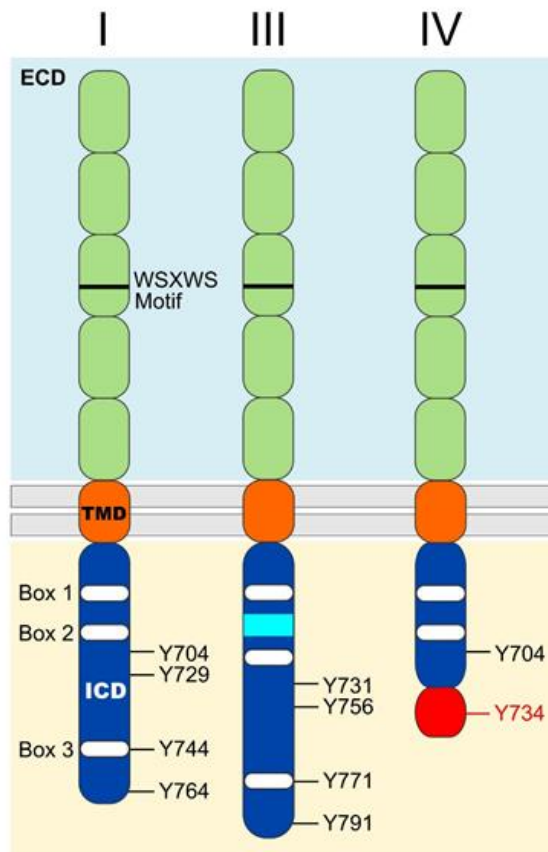
CSF3R is the receptor for G-CSF, a critical growth factor for neutrophil development. It is the major regulator of granulopoiesis and promotes proliferation, differentiation, and survival of granulocytic progenitor cells (13) (**Figure 1.1**). G-CSF is essential for steady state granulopoiesis, mobilizing neutrophil reserves, and bone marrow granulopoiesis. CSF3R is expressed on neutrophilic progenitors and is increased as the cells differentiate (14). Knockout of *Csf3r* or *Csf3* results in neutropenia in mice and zebrafish, validating its significance in neutrophil production (15).



**Figure 1.1. Schema of Hematopoiesis.** A simplified diagram of myeloid hematopoiesis is depicted. Blood cell components arise from an HSC and gradually become more committed in response to differentiating cytokines. The CMP gives rise to granulocyte/monocyte progenitors and megakaryocyte/erythrocyte progenitors that further differentiate into cells of the myeloid lineage: neutrophils, monocytes, erythrocytes, and platelets. Adapted from Lyman GH, Dale DC. Introduction to the Hematopoietic Growth Factors. In: Lyman GH, Dale DC, editors. Hematopoietic Growth Factors in Oncology. Boston, MA: Springer US; 2011.

It belongs to the type I cytokine receptor superfamily that play important roles in lineage commitment and development of hematopoietic cells. The extracellular domain contains an N-terminal immunoglobulin-like domain and five fibronectin III-like domain repeats, a highly conserved WSXSW (Trp-Ser-X-Trp-Ser) motif, four conserved cysteines, a transmembrane domain, and a cytoplasmic domain containing box 1 and 2 (shared with all members of type I cytokine receptor family) and box 3 (shared with the gp130 signal transduction chain) along with four tyrosine residues (Y704, Y729, Y744, Y764) (16-18) (**Figure 1.2**). These residues are essential in transducing differentiation signals and truncation of this region results in abrogation of maturation signaling. They mediate and promote enhanced STAT3 activation and Ras/Raf/ERK1/2 signaling. It also promotes negative regulation of CSF3R signaling by serving as a binding site for JAK/STAT inhibitor protein SOCS3 and by receptor internalization and targeting to lysosomes for degradation mediated by the box 3 subdomain. Early studies that investigated the role of each tyrosine residue on its ability to activate proliferation and differentiation signaling were conducted by mutating each of the tyrosine residues to phenylalanine and transfecting them into mouse myeloid cells. In the murine *Csf3r*, Y703 and Y728 (Y704 and Y729 in humans) were able to transmit differentiation signals in LGM-1 cells in response to G-CSF. However, transfectants that lacked Y703 retained blast morphologies and showed a corresponding decrease in the neutrophil differentiation marker myeloperoxidase (MPO). Furthermore, transfectants that lacked Y728 became macrophage-like in morphology and did not express MPO in response to G-CSF (19). In contrast, a study that used CSF3R mutants to assess macrophage differentiation responses in mouse M1 cells demonstrated that Y704, Y729, Y764 all responded to G-CSF in a similar manner as wildtype CSF3R. However, mutation of Y744 (located in box 3) to phenylalanine showed decreased differentiation of M1 cells (20). In mouse 32D cells, Y704 had the greatest effect on neutrophil differentiation and cell survival in response to low concentrations of G-CSF while Y764 supported a strong proliferation response. Mutants that lacked Y764 were still able to differentiate into neutrophils but did not proliferate. In

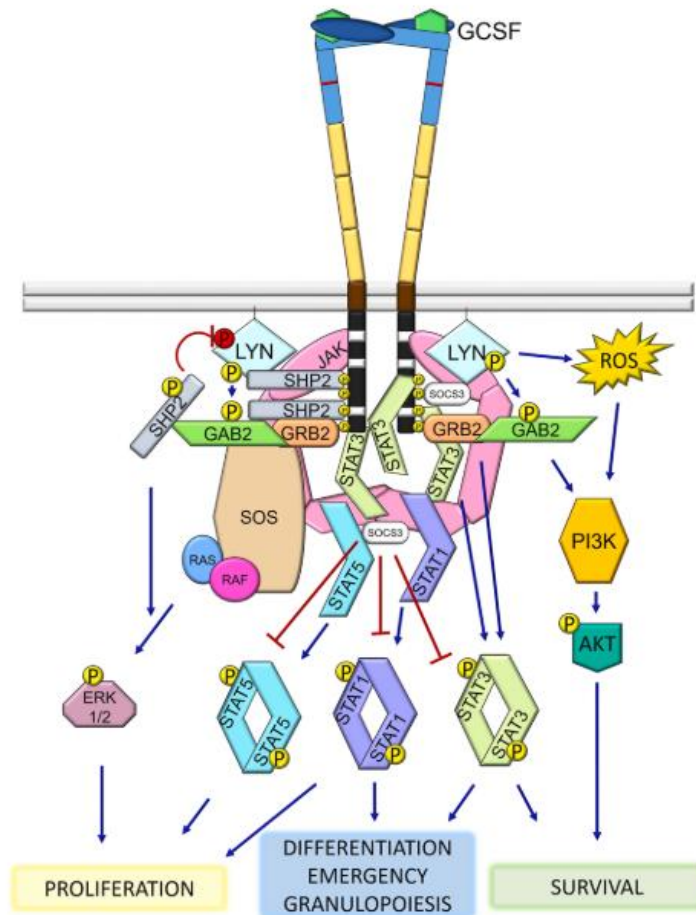
addition, deletion of all four tyrosine residues was most deleterious to cell survival, with the loss of Y704, Y729, and Y744 having less impact on survival and the loss of Y764 having intermediate effects on survival. Even though the involvement of the cytoplasmic tyrosine residues appears to be dependent on cell line, studies have generally shown that Y704, Y729, and Y744 are involved in activating signals for cell differentiation while Y764 activates signals for cell proliferation (21-24).



**Figure 1.2. Schematic of CSF3R Protein Isoforms.** Class I, Class III, and Class IV protein isoforms depicting the extracellular domain (ECD), transmembrane domain (TMD) and intracellular domain (ICD). Differences between the isoforms arise from insertions and deletions within the ICD. Class I contains box 1-3 and the four tyrosine residues important for transducing intact maturational signaling. Class III contains an insertion of 27 amino acids between box 1 and box 2 and Class IV contains a Y734 from a novel 34 amino acid sequence.



Binding of G-CSF to CSF3R results in homodimerization of receptor subunits that lead to the rapid phosphorylation and activation of JAK1 and JAK2 tyrosine kinases. In turn, JAK activation leads to phosphorylation of four tyrosine residues (Y704, Y729, Y744, and Y764) located on the cytoplasmic tail of CSF3R (25). Since CSF3R does not have intrinsic kinase activity, the phosphorylation of these tyrosine residues is critical in activating downstream effectors such as members of the JAK/STAT, PI3K/AKT, and MAPK/ERK pathways (26, 27). They also serve as docking sites for Src homology 2 (SH2) domain-containing proteins that undergo phosphorylation to recruit additional molecules to transmit growth signals. For example, the STAT proteins bind to the phosphorylated tyrosine residues through their SH2 domains and translocate to the nucleus where they bind to their specific sequences to regulate target genes. SOCS proteins (suppressors of cytokine signaling) also bind to the cytoplasmic domains and/or JAKs to inhibit activation of STAT proteins via their SH2 domains. Y764 phosphorylation activates MAPK/ERK, which is mediated through adapter protein Shc and Grb2 as well as members of the p21/Ras pathway. Loss of Y764 leads to a decrease in p21/Ras activation and a subsequent increase in neutrophil differentiation. Other MAP kinases controlled by Y764 include p38 and JNK. Even though a clear role for these proteins remains to be elucidated, there is evidence to suggest that G-CSF induced proliferation is mediated by the MAPK/ERK pathway (**Figure 1.3**) (28).



**Figure 1.3. Schema of CSF3R Signaling.** G-CSF binds to CSF3R which triggers receptor conformational changes that activate multiple tyrosine kinases such as JAK1/2 and LYN in the cytoplasmic domain. Important signaling pathways include STAT3 and STAT5, PI3K/AKT, and RAS/MAPK. Signaling is regulated by members of the SOCs family and tyrosine phosphatases SHP1 and SHP2. Blue arrows depict phosphorylation/activation signaling and red arrows depict dephosphorylation/inhibitory signaling. Figure reprinted with permission from Mehta HM, Corey SJ. G-CSF, the guardian of granulopoiesis. *Seminars in Immunology*. 2021:101515.

### Alternative Splicing of *CSF3R* in Neutrophil Development and Diseases

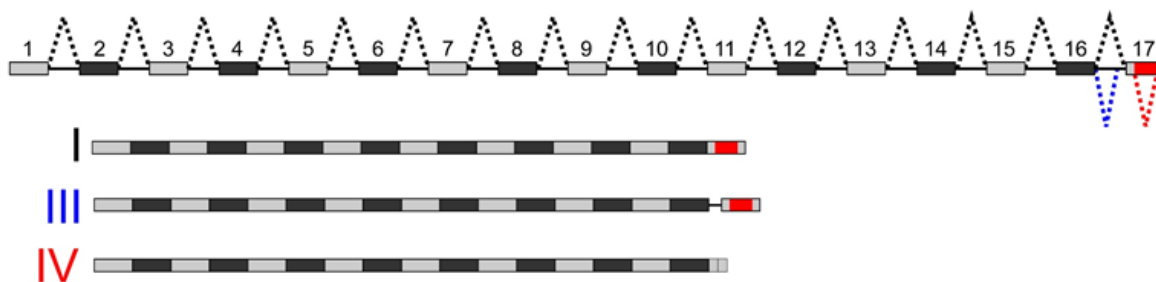
*CSF3R* comprises of 16 introns and 17 exons where all of the donor and acceptor splice sites conform to the 5'-GT/AG-3' consensus sequences. Exons 1 and 2 and the first 20 nucleotides of exon 3 comprise the noncoding region, with the remaining nucleotides of exon 3 encoding the signal sequence. The cytokine receptor homology domain is encoded by exons 5-8, the IgG-like domain by exon 4, and the fibronectin-like domains by exons 9-14. Most of exon 15 (the first 69

nucleotides) encodes the transmembrane domain and the remaining nucleotides of exon 15-17 encode the intracellular domain, which includes a 3' noncoding region (16).

Alternative splicing of *CSF3R* results in eight isoforms (Class I-VIII), of which three have been described in humans (Class I-III) (29-32) (**Figure 1.4**). Constitutive splicing results in the canonical, wildtype Class I isoform, which contains 836 amino acids and is the dominant isoform. It was cloned from placental cells and U937 cDNA in the early 1990s, when many functional studies of *CSF3R* were conducted (29, 33). Class I retains all of the functional domains described above. Class II results from an 88 base pair deletion of nucleotides 2034-2121 that leads to loss of the transmembrane domain. Isolated from U937 cells, it produces a soluble receptor with an altered C-terminal 150 amino acids with unknown function (34). Class III results from an insertion of 81 base pair (27 amino acids) between nucleotides 2209 and 2210 at the boundary between exon 16 and 17 due to alternate usage of a 3' splice site acceptor site in intron 16 (16, 30). Class IV is a truncated, differentiation-defective, receptor resulting from excision of 420 base pairs within exon 17 (34). Interestingly, the 5' and 3' boundary of this region contains the canonical GU/AG consensus sequences that may act as weak splice sites. The resulting isoform has a deletion of the C-terminal 87 amino acids, which ablate the tyrosine residues Y729, Y744, and Y764, which are replaced with a 34 amino acid novel sequence.

Class V-VIII are lesser studied *CSF3R* isoforms. Class V was isolated in circulating neutrophils and shares the same splice site acceptor as well as the novel 34 amino acid sequence as Class IV but lacks 130 amino acids of the membrane proximal region containing box 3. Removal of this region impairs both proliferative and differentiation signaling (34, 35). Class VI was isolated from the bone marrow of healthy individuals and has a deletion of 82 base pairs (nucleotides 2128-2209) that results in complete deletion of exon 16 encoding the membrane proximal region of the cytoplasmic tail, creating a frameshift that alters the entire cytoplasmic domain. While this isoform is still capable of binding G-CSF, it may not be clinically relevant because its levels were not found

at significantly increased levels in AML patients (36). Class VII was also isolated from normal bone marrow and contains a 60 base pair insertion between nucleotides 2033-2034. The insertion takes place immediately upstream of exon 15 from the use of an alternative donor splice site in the intron (36). A possible eighth isoform was discovered in the bone marrow of seven out of thirteen MDS patients with a deletion of a CAG trinucleotide at positions 2128-2130 at the boundary between the transmembrane and cytoplasmic domain. The trinucleotide is located at an acceptor splice site in the intron that is adjacent to another CAG at the 5' end of exon 16 where the deletion can result from use of an alternative 3' splice site. It is an activating isoform that provides a growth advantage to abnormal MDS clones from prolonged STAT3 activation and increased colony growth in response to G-CSF. It is currently unknown if this isoform exists in normal bone marrow or is associated with other myeloid malignancies (31).



**Figure 1.4. Alternative Splicing of *CSF3R* Pre-mRNA.** The splice isoforms relevant to humans arise from splice events in the terminal intron and exon. The retained intron of Class I is depicted in red and the usage of an alternate 3'SS in intron 16 is depicted in blue.

Granulopoiesis relies on well-regulated splicing mechanisms for normal physiological development and any dysregulation of the network can lead to diseases. In granulocyte maturation, intron retention is seen at different stages in specific genes involved in normal neutrophil development. This observation was supported by comparing the transcriptomes of promyelocytes, myelocytes, and neutrophils where it was shown that a high level of non-intron

retaining transcripts were specific to particular stages of neutrophil differentiation. Intron-retaining transcripts were increased with progressive differentiation and increasing cellular maturity of neutrophils. TCGA analysis has also shown that intron retention levels are globally increased even in the absence of splicing factor mutations, which suggests that intron retention is regulated by a combination of *cis* and *trans*-acting elements (37). The excision of 420 base pairs from *CSF3R* exon 17 can be considered an intron-retaining event resulting in Class IV because the GU/AG sequences marking the intron-exon boundary may act as weak splice sites. Weaker splice sites contribute to intron retention by preventing the spliceosome from recognizing introns that would otherwise be excised (38). *CSF3R* is susceptible to intron retention rather than exon skipping, partly due to the location of the GC defining intron-exon junctions. Intron retention is the only form of alternative splicing that is regulated in a coordinated manner that increases as neutrophils mature (39).

While splicing factors are frequently mutated in MDS and secondary AML, aberrant splicing of downstream targets such as *CSF3R*, *EPOR*, and *MPL* can lead to improper isoform ratios, which contributes to disease development (40). RNA-seq analysis from MDS and AML patients revealed increased levels of Class III and Class IV *CSF3R* relative to Class I (34, 41). G-CSF has also been found to preferentially stimulate the proliferation of monosomy 7 cells bearing the Class IV isoform (42). Since different isoforms of *CSF3R* have specific roles in normal neutrophil development, altered ratios due to dysregulated splicing events may act as weak drivers of leukemogenesis.

### **Overview of mRNA Splicing**

Splicing is a complex mechanism in which multicellular organisms process pre-mRNA into mature RNA by removing intronic sequences from the immature transcript. The physical gene structure in bacteria was established by the late 1970s in which a 1:1 relationship between genes, RNA, and protein expression was assumed to be a universal phenomenon in all organisms.

However, several observations about the biochemical properties of genetic material and its expression in cells questioned the gene structure in higher eukaryotes. First, RNAs transcribed in the nucleus were physically separated from the translational machinery in the cytoplasm. Second, comparison of mRNA with genomic sequences suggested that gene expression mechanisms in bacteria may not be universal. The DNA content of eukaryotic germ cells varied without significant changes to the number of genes. Third, studies involving hnRNAs suggested that long RNAs were transcribed from diverse nuclear sequences. These observations led to efforts comparing RNA length in the nucleus versus the cytoplasm in late-stage adenovirus infection with viral DNA that resulted in the discovery of “split genes” (43-45). The landmark discovery of RNA splicing in 1977 showed electron micrographs of DNA and RNA hybrids (R-loops) and the subsequent joining of remaining RNA after the removal of viral sequences where the cap structure at the 5' end and polyA tail at the 3' end was conserved. A comparison of the corresponding DNA sequences revealed that the shorter mature RNA sequences that accounted for the differences in length between the nucleus and the cytoplasm were due to removal of intervening sequences (introns) and joining of the remaining sequences (exons). On average, the pre-mRNA is four to ten times longer than the mature RNA.

Shortly after this discovery, other genes were also shown to have intervening intronic sequences, such as the globin and immunoglobulin genes. In addition, conserved sequences were found within each intron-exon boundary that were common to yeast, plants, and vertebrates—suggesting that splicing is an evolutionarily general and conserved event. The question remained as to why introns are evolutionarily conserved. The intron-exon structure has clearly been important in the generation of new genes during evolution; therefore, the placement of introns is perhaps not entirely random. Some have speculated that genes originally evolved as exons that encode sequences for stable protein-folding domains. Assembly of several exon sequences by splicing would produce a protein composed of stable folding domains that have a

high probability of producing a functional protein. For example, fibronectin has three types of exons that correspond to the three types of protein domains, which are the same extracellular domains found on genes encoding type I cytokine receptors, like *CSF3R* (46). Gene duplication of the fibronectin gene provides a particularly salient example of a common mechanism for generating genes encoding many cell surface receptors and other types of proteins in higher organisms. Approximately 95-100% of human pre-mRNAs that have more than one exon undergoes splicing (47).

### **Mechanism of RNA Splicing**

The excision of introns and resulting joining of exons is directed by sequences at the intron-exon junction called splice sites and catalyzed by a macromolecular protein complex called the spliceosome. The 5'SS is the intron-exon junction at the 5' end of an intron that includes a GU dinucleotide. The 3' end of the intron has a branch point sequence (BPS), a polypyrimidine tract ((Py)<sub>n</sub>), and a terminal AG dinucleotide representing the 3'SS that are all conserved elements. The spliceosome, comprising of at least 170 proteins and five snRNPs, assembles onto the splice site sequences and catalyzes two transesterification steps. First, the 2'-hydroxyl group of a special A residue at the BPS attacks the phosphate at the 5' splice site. The 5' exon is cleaved from the intron and the 5' end of the intron is ligated to the branch point 2'-hydroxyl, producing a detached 5' exon and an intron/3' –exon fragment in a lariat configuration containing an A nucleotide at the branch point. Second, the phosphate at the 3' end of the intron is attacked by the 3'-hydroxyl of the detached exon. The two exons are ligated and the intron is released in a lariat form (48, 49).

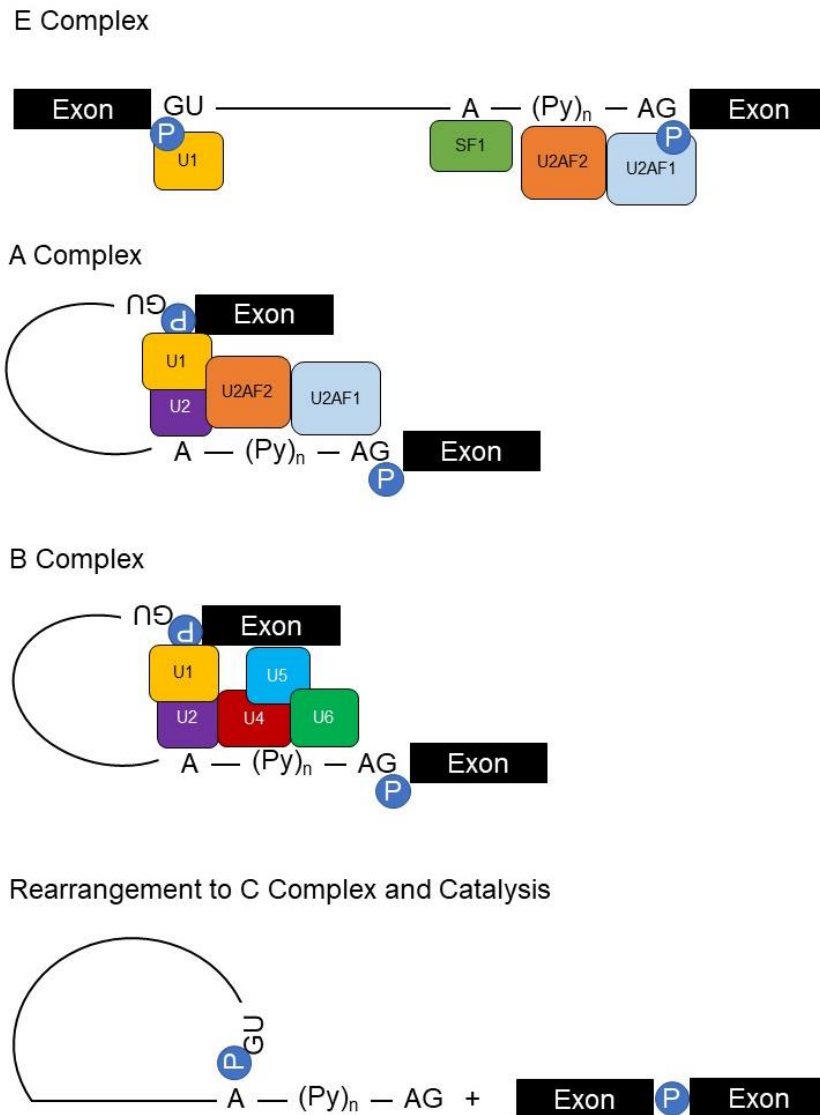
The U1, U2, U4/U6, and U5 snRNPs comprise the main components for the major spliceosome, which is responsible for removing >99% of pre-mRNA introns (50). Some metazoan and plants have another type of spliceosome, the minor spliceosome, which is composed of functionally analogous U11/U12 and U4atac snRNPs with the U5 snRNP shared between the major and minor spliceosome. The mechanism of RNA splicing via the major spliceosome will be

discussed here. In the first part of spliceosome assembly, the U1 snRNP binds to the 5'SS of the intron through base pairing interactions with the 5' end of the U1 snRNA. These interactions are stabilized by members of the SR protein family and proteins of the U1 snRNP. In general, these RNA-RNA interactions are weak and require stabilization by auxiliary proteins. On the 3' end, SF1/BBP protein and U2AF bind to the branch point sequence and polypyrimidine tract. SF1/BBP interacts with the 65 kDa subunit of U2AF through its C-terminal RRM and the 35 kDa subunit of U2AF (U2AF35) is tightly bound to U2AF65 in the U2AF heterodimer, which binds the AG dinucleotide of the 3'SS at the intron-exon junction. These combined interactions form the ATP-independent early (E) complex in spliceosome assembly and is important in the initial recognition of the 3'SS and 5'SS of the intron. Following the E complex formation, it joins with the U2 snRNP in an ATP-dependent manner that forms the A complex. The base-pairing interaction is stabilized by heteromeric protein complexes of the U2 snRNP, mainly SF3a, SF3b, and the SR rich domain of U2AF65. The joining of the U2 snRNP results in the displacement of SF1/BBP from the branch point sequence and is replaced by the association of SF3b14a/p14 with the BPS adenosine. Next, it joins with the U4/U5/U6 tri-snRNP to form the B complex. The spliceosome is still catalytically inactive at this stage and requires major conformational changes to catalyze the splicing reaction. During spliceosome activation, the U1 snRNP and U4 snRNP are released, forming the B\* complex. Once activated, it goes through the first catalytic step, generating the catalytic (C) complex. Additional rearrangements occur prior to the second catalytic step. Afterwards, the spliceosome dissociates and releases the mRNA in the form of a messenger RNP. The U2, U5, and U6 snRNPs are also released and recycled for additional rounds of splicing (48) (**Figure 1.5**).

Splicing is a highly dynamic process. As described above, RNA and protein interactions on the pre-mRNA are recognized multiple times to ensure reaction precision. Many of these reactions are generally weak and are often enhanced by auxiliary proteins. Indeed, the splice site consensus sequences are usually not sufficient information to determine whether it will assemble



a spliceosome and catalyze the splicing reaction. Additional *cis* and *trans*-regulatory elements are needed to activate pathways that lead to alternative splicing. Alterations in splice site choice are regulated by other processes such as post translational modifications of SR proteins, *cis/trans* exonic regulatory elements, and splice site mutations.

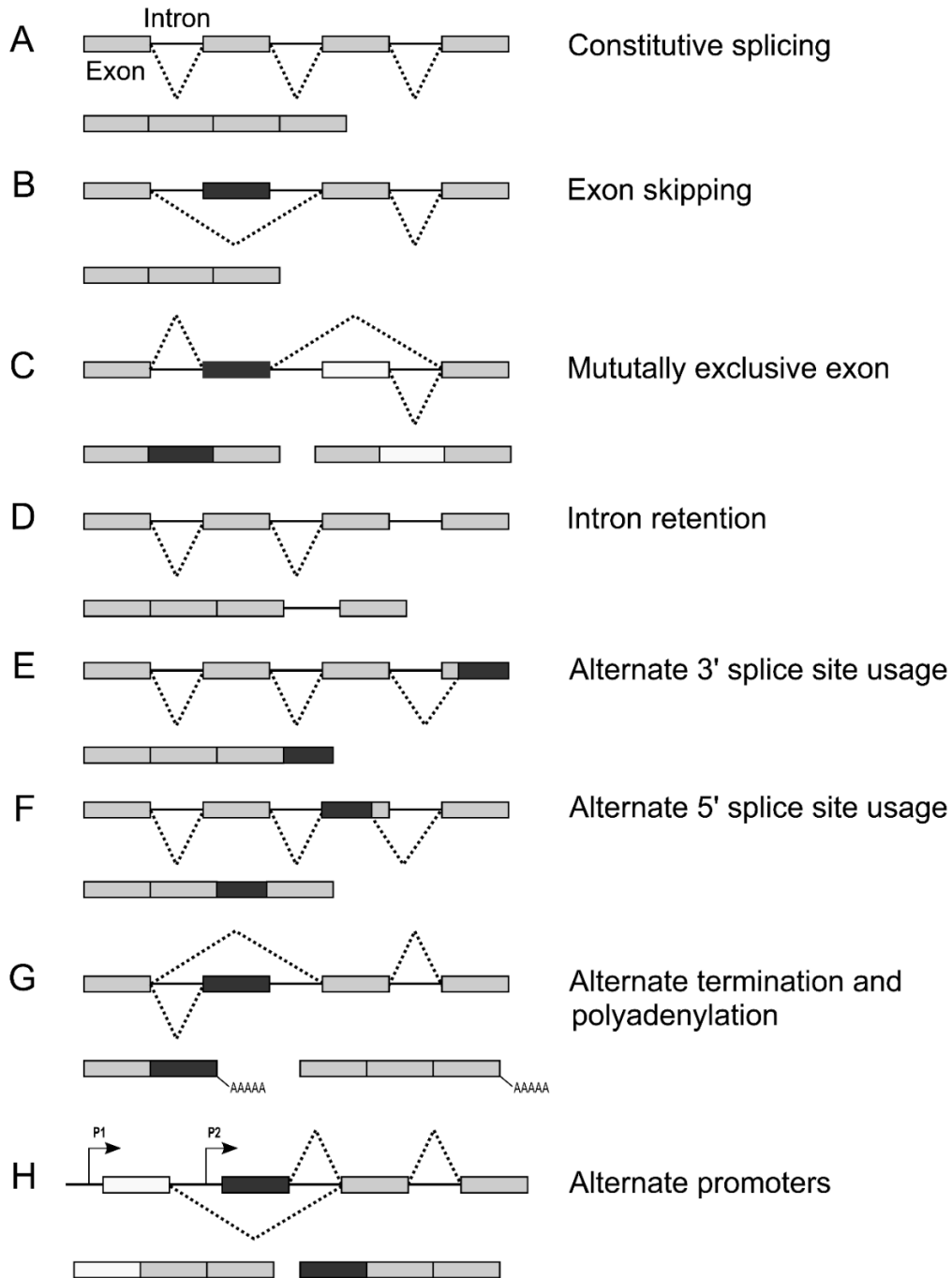


**Figure 1.5. Mechanism of Pre-mRNA Splicing.** Splicing takes place in two transesterification steps. The spliceosome contains five snRNPs that assemble onto the intron: U1/2/4/5/6. The early (E) complex is formed when the U1 snRNP binds to the 5'SS, the branch point by SF1, the polypyrimidine tract by U2AF2, and the AG dinucleotide by U2AF1. The A complex is formed when the U2 snRNP binds to the branch point. The accumulation of the U4/5/6 tri-snRNP forms the B complex, which is then rearranged to the form the C complex.

## Regulation of Alternative Splicing

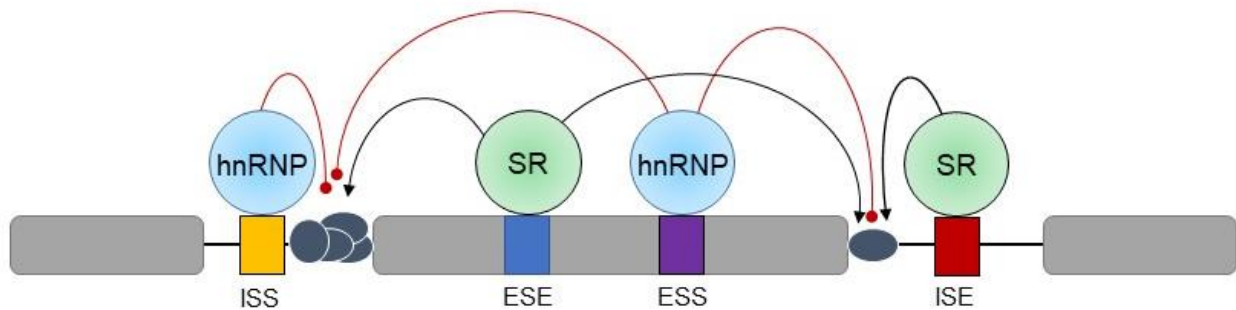
Alternative splicing is how eukaryotic organisms generate multiple transcripts from a single gene, generating diversity to its transcriptome and proteome without increasing the size of their genomes. Nearly all multi-exon organisms undergo alternative splicing and is the basis for the discrepancy between the 24,000 protein-coding genes in humans and the 100,000 proteins thought to be synthesized (47). In alternative splicing, the spliceosome uses different splice sites to generate other transcripts besides the constitutive mRNA. Distinct splicing isoforms can produce functionally equivalent, divergent, or inactive proteins. The types of alternative splicing include exon skipping (cassette exon), mutually exclusive exons, intron retention, and alternate 3' or 5' splice site usage. Other non-splicing mechanisms to generate isoforms are alternate termination and polyadenylation and alternate promoters (47) (**Figure 1.6**). Exon skipping occurs most often where cassette exons can either be included or excluded. Mutually exclusive exons occur when two or more adjacent cassette exons are spliced in a way that only one exon in the group is included at a time. Intron retention occurs when the intron remains in the final mRNA transcript. Intron retention is a widespread mechanism that regulates neutrophil development and diversifies most cancer transcriptomes (51). Alternate 5' and 3' splice site selection lengthens or shortens the exon length. Multiple promoters and transcription start sites varies the level of transcription initiation and translation efficiency of mRNA with different leader exons as well as generating new protein isoforms that differ at the amino terminus. Alternate polyadenylation generates diversity at the 3' end of transcripts and controls the 3' UTR. These examples emphasize that exon definition requires extensive crosstalk between several *cis*-acting elements and *trans*-acting splicing factors. Mutations that disrupt only one of the few critical elements can significantly affect the splicing pathway that can lead to diseases. Mutations in these core elements can lead to improper expression of protein isoforms which can drive disease progression.

Alternative splicing typically occurs at sites where the consensus sequences are less well conserved and can depend on the cell context and/or condition. For example, the GC content in introns or exons influence how the spliceosome defines introns and exons. GC-rich introns are more susceptible to intron definition resulting in intron retention as opposed to GC-rich exons which result in exon skipping. This feature allows *cis*-intronic sequences and *trans*-acting splicing factors to dictate which sequences to include or exclude at the intron-exon junction (52, 53). This phenomenon rarely occurs in the minor class introns in part because they are defined by highly conserved sequence elements and the lack of U12 introns found in the major class of introns available for alternative selection within the same gene. Since spliceosome assembly onto intron-exon boundaries by itself is not sufficient information to catalyze the splicing reaction, there is still a high degree of selectivity despite the rather poorly conserved nature of splice site sequences in higher eukaryotes. It is estimated that only ~50% of the information that defines a splice site is at the splice site itself (54). It remains unclear how much of alternative splicing is constitutive. In other words, the extent to which multiple isoforms are produced at the same ratios in all cell types is unknown due to studies that focused on splicing patterns in different cellular environments. Splicing regulation can be tissue specific, developmental, differentiation-specific, or can be a response to external stimuli (55). The biochemical mechanisms of alternative splicing are still largely unknown, but it is unlikely that specific and distinct factors dedicated to each of the more than 100,000 alternative splicing decisions that occur in human cells exist. Several genomes worth of regulatory proteins would be required if this were not the case. Presumably, splicing regulation would likely depend on a small number of factors that are specifically dedicated to one or few alternative splicing events. Indeed, a small number of proteins have been repeatedly found to be responsible in part for regulating a large number of alternative splicing events. So far, the most studied splicing regulators are members of the SR-protein and hnRNP families.



**Figure 1.6. Schematic Representation of Alternative Splicing.** (A) Constitutive splicing and alternative splicing types include (B) exon skipping, (C) mutually exclusive exons, (D) intron retention, (E) alternate 3'SS usage, and (F) alternate 5'SS usage. Other types of mechanisms that produce alternative isoforms are (G) alternate termination and polyadenylation and (H) alternate promoters. Exons are represented by the rectangles and introns are represented by lines. Dashed lines represent splicing events. The mature mRNA transcripts are depicted below each pre-mRNA transcript.

The decision to include or exclude certain exons depends in part on *cis* regulatory elements and proteins. *Cis* regulatory elements are divided into four categories: ESEs, ESSs, ISEs, and ISSs. ESEs are usually bound by SR proteins while ISSs and ESSs are commonly bound by hnRNPs (54). Members of the SR family of proteins play a role in both constitutive and alternative splicing. They are involved in the steps of splicing regulation by binding to ESEs through their RRM and RS domains to mediate protein-RNA and protein-protein interactions respectively (56-58). The RRMs determine the sequence specificity of RNA binding while the RS domain mediates protein-protein interactions that are important in the recruitment of splicing machinery elements (59). When bound, these elements can promote exon definition by directly recruiting the spliceosome through their RS domain and/or by antagonizing the action of nearby splicing silencers (**Figure 1.7**). All canonical SR proteins have common characteristics and are highly conserved, which include a similar structure with one or two RNP type RNA binding domains at their amino termini and a variable length RS domain at their C-terminal end. They are extensively phosphorylated and usually function as activators (60, 61). When an SR protein binds to an ESE, it recruits U2AF1 to the polypyrimidine tract and activates the adjacent 3'SS. It has been found that mammalian ESEs are usually purine-rich sequences and associate with specific SR proteins to promote the usage of adjacent splice sites. The RS domain strengthens protein-protein interactions between the U2AF1 bound to 3' end, the U1 snRNP bound at the 5' end, and the SR protein bound to the ESE (62). There is also an RS domain-independent mechanism where the main function of the SR protein that is bound to an ESE is to antagonize the inhibitory effect of an hnRNP bound to an ESS. Nonsense, missense, and translationally silent mutations can interrupt ESEs to cause exon skipping that can have serious consequences on the resulting protein.



**Figure 1.7. Regulation of Alternative Splicing by SR Proteins and hnRNPs.** SR proteins and hnRNPs are two groups of well-studied splicing regulators. SR proteins regulate constitutive and alternative splicing by binding to ESE elements on pre-mRNA to promote spliceosome assembly at adjacent splice sites. hnRNPs bind to ESS and ISS elements that repress the splicing of some alternatively-spliced exons.

### Aberrant RNA Splicing in Hematologic Malignancies

A landmark study published in 2011 by Yoshida *et al* discovered that a handful of spliceosomal genes are recurrently mutated in adult MDS, which are clonal disorders involving HSCs. These somatic mutations occur in splicing factors U2AF1, SRSF2, ZRSR2, U2AF65, SF3B1, SF3A1, PRPF40B, SF1, and LUC7L2 (**Figure 1.8**). They are collectively mutated in up to 85% of MDS patients, suggesting that these genetic contributions are important to the pathophysiology of MDS. Notably, these mutations are usually heterozygous point mutations, occur at hotspots, are mutually exclusive, and are associated with different diseases and MDS subtypes that correlate with patient outcomes. According to the Human Gene Mutation Database, mutations that disrupt normal splicing are estimated to account for up to a third of all disease-causing mutations. Many of the hallmarks of cancer can be attributed to abnormal gene regulation that is a consequence of alternative splicing during cancer progression. Transcriptomic studies have revealed multiple splicing changes in genes between normal and abnormal HSCs (11, 63, 64). Investigations in the

whole transcriptome of HSCs have been underway to understand how distinct genes often associated with myeloid neoplasms are regulated by alternative splicing in normal conditions.

SF3B1 mutations are associated with cryptic 3'SS recognition and alternative branch point usage of genes related to heme biosynthesis and iron metabolism as well as cancer specific genes (65, 66). Almost all of its mutations lie within the HEAT repeat region which provides scaffolding for the U2 snRNP to support interactions with other SF3B subunits. SF3B1 K700E mutations account for more than 50% of all observed variants and strongly associate with erythroid defects like refractory anemia and ringed sideroblasts (12, 67). SF3B1 mutations have a positive predictive value of 97.7% for disease phenotype with ringed sideroblasts. Mutations result in neomorphic activity that cause hundreds of aberrant splicing alterations and dysregulated gene expression in common alternative splicing signatures in different types of cancers. Approximately 50% of the aberrant mRNAs undergo NMD resulting in downregulation of canonical transcripts leading to downregulation of gene and protein expression (68).

U2AF1 mutations occur in 11% of MDS patients and affect the S34 and Q157 residues within the first and second CCCH zinc finger domains respectively (69). It preferentially recognizes the RNA sequence motif yAG|r which matches the genomic consensus 3'SS splice site and intron-exon boundary that crosslinks with U2AF1. U2AF1 mutations cause allele specific alterations in recognition of the 3'SS consensus sequence. U2AF1 S34 mutations are associated with nearly identical alterations at the -3 position and Q157 are associated with alterations at the +1 position. S34F/Y mutations have a preference for C instead of T at the -3 position while Q157P/R mutations have a preference for G instead of A at the +1 position (70, 71). Proper splicing relies on the accurate recognition of the 3'SS. Inappropriate binding of U2AF1 to other sites can result in expression of alternate protein isoforms that can contribute to disease. Even though U2AF1 mutations are associated with altered 3'SS, only a small fraction of cassette exons is affected by expression of mutant proteins (72). KEGG pathway analysis have shown missplicing in genes

related to RNA processing, RNA localization/transport, RNA binding, and ribosomal pathways (73). Mice expressing S34F showed altered hematopoiesis and changes in RNA splicing in hematopoietic progenitor cells by RNA-sequencing analysis (69). Knockdown of U2AF1 in erythroid cells derived from human CD34+ cells led to apoptosis and delayed erythroid differentiation via alternative splicing of apoptosis-associated gene transcripts (74).

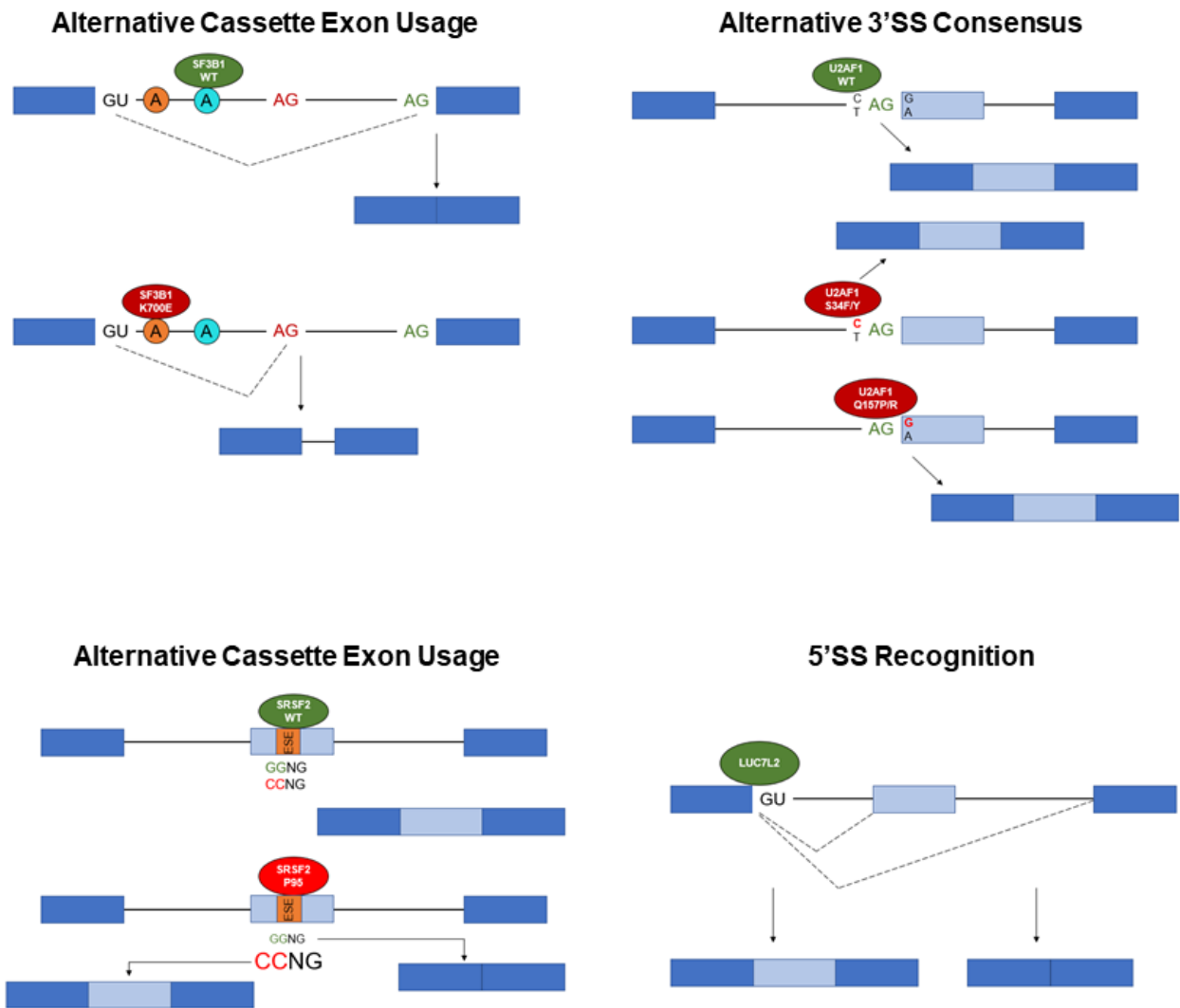
SRSF2 is a prototypical SR protein that is involved in constitutive and alternative splicing by binding to ESE sequences in pre-mRNA through its RRM (75). SRSF2 is involved in pathways that require tight control and regulation of genes and its expression is regulated by a negative feedback loop. SRSF2 mutations occur in 20-30% of MDS and ~50% of CMML patients where mutations are consistently associated with poor patient outcomes (1, 76, 77). The most common mutations occur on the P95 residue within the C-terminal end of the RRM. SRSF2 P95H conditional knock-in mice showed impaired hematopoietic differentiation, increased HSPC subsets, and morphologic dysplasia in neutrophils (78, 79). SRSF2 P95 mutations alter binding specificity to ESEs on mRNA by changing the conformation of both RRM termini, altering its preferred binding sequence to CCNG over GGNG motifs (78). In contrast, SRSF2 WT can recognize both sequences with equal affinity (80). SRSF2 is essential for hematopoiesis since ablation of the gene causes embryonic lethality with enhanced apoptosis and decreased levels of hematopoietic stem/progenitor cells in mice (81, 82). Intact *SRSF2* is therefore essential for the functional integrity of the hematopoietic system and provides another example in which mutations are not simply loss of function, but rather a change of function.

*LUC7L2* has been identified as a candidate gene in the development of myeloid malignancies (83). It resides on chromosome 7q34, a region that is frequently altered in patients with acute AML and MDS. Loss of chromosome 7 or the deletion of its long arm (-7/del7q) is one of the most common cytogenetic abnormalities found in both childhood and adult MDS and is associated with poor clinical outcomes and higher transformation to AML (84, 85). Expression and deletion



analyses have identified decreased expression of LUC7L2 on chromosome 7q34 and may play a role in clonal evolution in MDS/AML pathogenesis. Despite these clinical observations, the molecular functions of LUC7L2 remain largely uncharacterized and its role in driving myeloid malignancies unknown. CLIP-seq studies have shown that LUC7L2 have high enrichment over input in exons near the 5'SS, implicating its role in the regulation of splicing through the 5'SS and not the 3'SS like the other splicing factors associated with MDS (86). One study reported decreased hematopoietic differentiation of del(7q)-MDS induced pluripotent stem cells that was rescued by 7q dosage correction (87). Hematopoietic defects are mediated by combined haploinsufficiency of EZH2 in cooperation with other candidate genes residing on the same region as LUC7L2 (88).

Splicing factor mutations are considered to be founder lesions that contribute to the growth advantage of an early clone, but how they affect gene regulation that results in dysplasia in individual myeloid lineages remains unclear. Impaired granulopoiesis may be an effect of mutations in gene networks involved in specific myeloid lineages or on those responsible for granulocytic differentiation. Gaining a mechanistic understanding of this is challenging because it will require transcriptome analysis of specific populations of blood cells and lineages rather than a mixed population of hematopoietic cells. Since dysregulated splicing can produce abnormal levels of isoforms that affect normal hematopoiesis, investigating the splicing patterns of specific genes involved in this process will be important in providing new models for dysgranulopoiesis and methods for modulating splicing.



**Figure 1.8. Splicing Factor Mutations on Alternative Splicing.** MDS-associated splicing factor mutations are change of function mutations that result in aberrant protein isoforms that can drive disease progression. SF3B1 K700E recognizes an alternate BP sequence resulting in the use of an alternate cassette exon. U2AF1 S34F and Q157P recognize an alternative 3'SS consensus sequence. SRSF2 P95H recognizes alternate ESE motifs to include or exclude exons. LUC7L2 deficiency affects its binding to the 5'SS that functions to retain or excise introns. Adapted from Rahman MA, Krainer AR, Abdel-Wahab O. SnapShot: Splicing Alterations in Cancer. Cell. 2020;180(1):208-e1.

## **Chapter 2: Comparative Analysis of Neutrophils and *CSF3R* Isoforms Across Species and Non-Human Primates**

### **2.1 Rationale**

The abundance of *CSF3R* isoforms and its splicing patterns seen throughout different species may explain why vertebrates like cats, dogs, non-human primates, and humans are able to develop bone marrow failures like MDS and other species like mice and jawed/jawless fish do not develop these diseases (89-91). Studies have indicated that the emergence of G-CSF and *CSF3R* led to a reorganization of granulocyte distribution between amphibians and reptiles that played key roles in the evolution of the modern-day mammalian neutrophil. Mammalian neutrophils do share developmental and functional properties as other more primitive counterparts but a notable species-specific difference comes from the distribution of neutrophils in the peripheral blood, which might have resulted from evolutionary pressures (91). As different populations of white blood cells predominate different species and intraspecies, due to environmental pressures on the immune system to favor certain populations at the expense of other white blood cells, these same pressures would have subjected *CSF3R* isoforms to exist in some species at the expense of the canonical isoform. Thus, the evolution of *CSF3R* has been a driving force in neutrophil development in mammals.

### **2.2 Materials and Methods**

#### **Comparison of *CSF3R* Between Humans and Non-Human Primates**

Gene and nucleotide alignments between humans and select non-human primates based on the Nonhuman Primate Reference Transcriptome Resource were conducted in Ensembl Genome Browser 105. The phylogenetic tree of *CSF3R* was constructed using the entire gene and the nucleotide alignments were made using *CSF3R* exon 17.

## Isolation of Rhesus Macaque Mononuclear Cells and Neutrophils

Mononuclear cells and neutrophils were isolated from bone marrow aspirates and peripheral blood of three rhesus macaques (**Table 2.1**) using Ficoll-Paque PREMIUM (Cytiva). Bone marrow aspirates (5 mL) and peripheral blood (5 ml) were layered onto 5 mL of Ficoll-Paque solution and spun at 400xg for 30 min at 20°C. The interphase cells containing the mononuclear cells were collected, spun at 300xg for 10 min, and resuspended in autoMACS Running Buffer (Miltenyi Biotec). To remove the platelets, cells were centrifuged at 200xg for 15 min. The plasma and Ficoll-Paque layer were aspirated, leaving the red blood cells (RBCs) and neutrophils at the bottom of the tube. RBCs were lysed with ACK lysis buffer (Gibco) and spun at 300xg for 10 min. The supernatant containing the lysed RBCs were aspirated and the remaining neutrophil pellet was resuspended in 10 mL autoMACS Running Buffer. The cell suspension was spun at 300xg for 10 min and the supernatant was aspirated. Bone marrow aspirates and peripheral blood samples were received from Dr. Robert Donahue (NIH/NHLBI).

**Table 2.1 Rhesus Macaques Sample Characteristics**

ID#	Gender	Wt. (Kg)	DOB
ZK39	Male	14	01/01/2013
ZL08	Male	11.1	03/23/2013
ZL12	Male	9	04/04/2013

## cDNA Synthesis

Cell pellets were lysed with 1 mL TRIzol (Invitrogen). Total RNA was purified by the addition of 200  $\mu$ L chloroform, precipitated with 500  $\mu$ L isopropanol, rehydrated in 50  $\mu$ L nuclease-free water, and heated at 55°C for 12 min. RNA was subsequently cleaned up using Monarch RNA Cleanup Kit (NEB) and concentrations were measured using a Nanodrop One spectrophotometer (Thermo Fisher) at an absorbance of 260 nm. 1  $\mu$ g of RNA was used to synthesize 1  $\mu$ g cDNA by RT-PCR using iScript cDNA Synthesis Kit (Bio-Rad).

## PCR

To determine if Class I and Class IV isoforms exist in rhesus macaques, the forward PCR primer was designed at the start of rhesus macaque *CSF3R* exon 17 (5'-GAGGCCTTCCAGCTGCCT-3') and the reverse primer was designed near the end of exon 17 (5'-ACAAAAGGCCATTGGGTGGG-3'). The PCR components were added according to the recommendations of the Phusion High-Fidelity DNA Polymerase (NEB) (**Table 2.2**). Standard PCR cycling conditions were performed with an annealing temperature of 60°C and extension time of 7 min. PCR reactions were subsequently visualized on a 1% ethidium bromide agarose gel.

**Table 2.2. Standard PCR Mix**

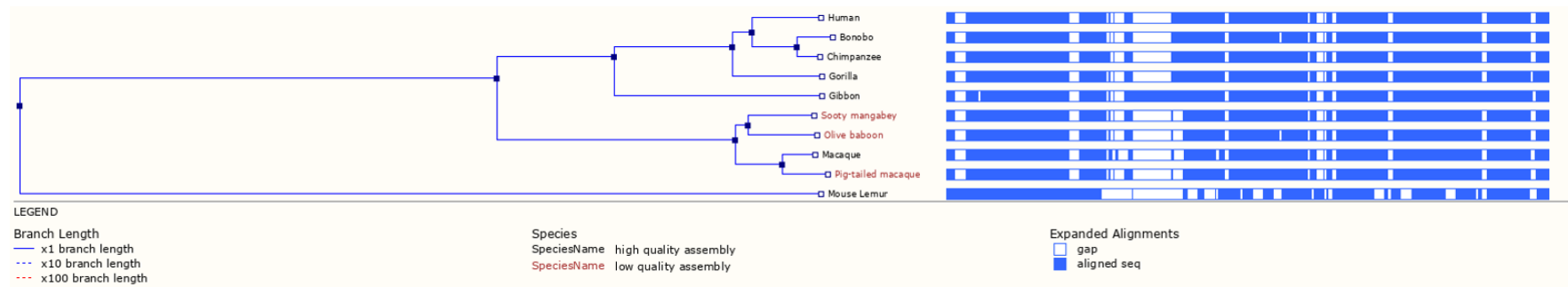
Component	20 $\mu$ L Reaction	Final Concentration
Nuclease-free water	to 20 $\mu$ L	--
5X Phusion HF Buffer	4 $\mu$ L	1X
10 mM dNTPs	0.4 $\mu$ L	200 $\mu$ M
10 $\mu$ M Forward Primer	1 $\mu$ L	0.5 $\mu$ M
10 $\mu$ M Reverse Primer	1 $\mu$ L	0.5 $\mu$ M
Template DNA	1 $\mu$ L	50 ng
Phusion DNA Polymerase	0.2 $\mu$ L	1.0 units/50 $\mu$ L PCR

## 2.3 Results

### Mouse Lemur *CSF3R* is Most Evolutionary Divergent from Humans

Gene alignments of the selected non-human primates to the human *CSF3R* show that there are high sequence similarities between bonobos (the closest human relatives), chimpanzees, and gorillas. The gene evolutionarily diverges beginning with gibbons and continues to do so where the human and mouse lemur *CSF3R*, the most distant non-human primate in our list, show the

most differences in alignment (**Figure 2.1 and Figure 2.2**). The underlined nucleotide sequence in **Figure 2.3** depicts the 81 bp of intron 16 that is inserted into the Class III isoform. This region is fairly conserved, aside from a single base pair change at the +23 position and a change from a G>A at the -1 position from the Class III alternate 3'SS between humans and non-human primates. Comparison of the Class IV region show that all species have the consensus GT/AG at the 5' and 3' splice sites (**Figure 2.4**), which implies that Class IV is evolutionarily conserved as long as the mechanism by which the major spliceosome recognizes these splice sites.



**Figure 2.1. CSF3R Phylogenetic Tree of Humans and Non-Human Primates.** Gene alignments were created on Ensembl comparing humans and select non-human primates. The list of non-human primates was curated from the Nonhuman Primate Reference Transcriptome Resource.



# CSF3R

Exon 17

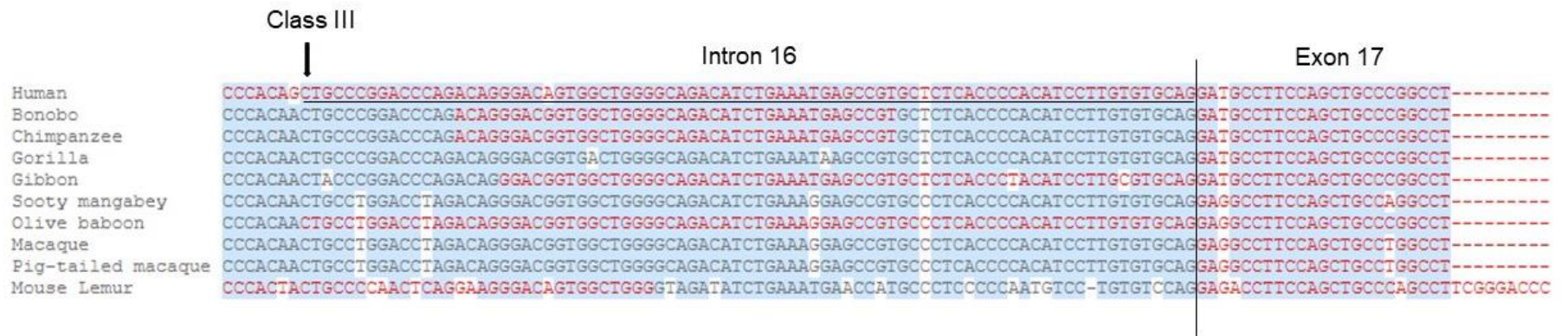
Human	CCCACAGCTGCCCGGACCCAGACAGGGACAGTGGCTGGGGCAGACATCTGAAATGAGCCGTGCTCTCACCCCACATCCTTGTGTGCAGGATGCCTTCCAGCTGCCCGGCC	-----
Bonobo	CCCACAACCTGCCCGGACCCAGACAGGGACGGTGGCTGGGGCAGACATCTGAAATGAGCCGTGCTCTCACCCCACATCCTTGTGTGCAGGATGCCTTCCAGCTGCCCGGCC	-----
Chimpanzee	CCCACAACCTGCCCGGACCCAGACAGGGACGGTGGCTGGGGCAGACATCTGAAATGAGCCGTGCTCTCACCCCACATCCTTGTGTGCAGGATGCCTTCCAGCTGCCCGGCC	-----
Gorilla	CCCACAACCTGCCCGGACCCAGACAGGGACGGTGGCTGGGGCAGACATCTGAAATGAGCCGTGCTCTCACCCCACATCCTTGTGTGCAGGATGCCTTCCAGCTGCCCGGCC	-----
Gibbon	CCCACAACCTACCCGGACCCAGACAGGGACGGTGGCTGGGGCAGACATCTGAAATGAGCCGTGCTCTCACCCCTACATCCTTGCCTGCAGGATGCCTTCCAGCTGCCCGGCC	-----
Sooty mangabey	CCCACAACCTGCCCTGGACCTAGACAGGGACGGTGGCTGGGGCAGACATCTGAAAGGAGCCGTGCCCTCACCCCACATCCTTGTGTGCAGGAGGCCCTTCCAGCTGCCAGGCC	-----
Olive baboon	CCCACAACCTGCCCTGGACCTAGACAGGGACGGTGGCTGGGGCAGACATCTGAAAGGAGCCGTGCCCTCACCCCACATCCTTGTGTGCAGGAGGCCCTTCCAGCTGCCCGGCC	-----
Macaque	CCCACAACCTGCCCTGGACCTAGACAGGGACGGTGGCTGGGGCAGACATCTGAAAGGAGCCGTGCCCTCACCCCACATCCTTGTGTGCAGGAGGCCCTTCCAGCTGCCCGGCC	-----
Pig-tailed macaque	CCCACAACCTGCCCTGGACCTAGACAGGGACGGTGGCTGGGGCAGACATCTGAAAGGAGCCGTGCCCTCACCCCACATCCTTGTGTGCAGGAGGCCCTTCCAGCTGCCCGGCC	-----
Mouse Lemur	CCCACACTGCCCCAACTCAGGAAGGGACAGTGGCTGGGGTAGATATCTGAAATGAACCATGCCCTCCCCAATGTCC-TGTGTCCAGGAGACCTTCCAGCTGCCAGCCTTCGGGACC	-----

TGGCACGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGATGAAAGAAAGCCGGTGGCCCTGGGAGTCCCATAACAGCTCAGAGACCTGTGGCCTCCCCACTCTGGTCCAGACCTATGT	TGGCACGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGATGAGAGAAAGCCGGTGGCCCTGGGAGTCCCATAACAGCTCAGAGACCTGTGGCCTCCCTACTCTGGTCCAGACCTATGT	TGGCACGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGATGAGAAGAAAGCCGGTGGCCCTGGGAGTCCCATAACAGCTCAGAGACCTGTGGCCTCCCTACTCTGGTCCAGACCTATGT	TGGCACGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGACGAGAAGAAAGCCAGTGGCCCTGGGAGTCCCATAACAGCTCAGAGACCTGTGGCCTCCCCACTCTGGTCCAGACCTATGT	CGGCACGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGACGAGAAGAAAGCCGGTGGCCCTGGGAGTCCCATAACAGCTCAGAGACCTGTGGCCTCCCCACTCTGGTCCAGACCTATGT	GGGCATGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGACGAGAAGAAAGCCACTGGCCCTGGGAGTCCCATGACAGCTCAGAGACCTGTGGCCTCCCCACTCTGGTCCAGACCTATGT	GGGCATGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGACGAGAAGAAAGCCACTGGCCCTGGGAGTCCCATGACAGCTCAGAGACCTGTGGCCTCCCCACTCTGGTCCAGACCTACGT	GGGCATGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGACGAGAAGAAAGCCACTGGCCCTGGGAGTCCCATAACAGCTCAGAGACCTGTGGCCTCCCCACTCTGGTCCAGACCTATGT	GGGCATGCCACCCATCACCAAGCTCACAGTGTGGAGGAGGATGAGAAGAAAGCCACTGGCCCTGGGAGTCCCATAACAGCTCAGAGACCTGTGGCCTCCCCACTCTGGTCCAGACCTATGT	TGGCATGCCACCCATCACCAAGATCACAGTGTAGAGGAGGAAAGAAAGAAAGCTGGGCCCTGGGAGTCCAGTACAGCTCAGAGACCTG---CCTCCCCACCCTGGTCCAGACCTATGT
GGACGACTGTGTCTTTGGGCCACTGCTCAACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGGGAGCTTCTAGGGCTTCCTGGGGTTCCTTCTTGGGCCTGCCT	GGACGACTGTGTCTTTGGGCCACTGCTCAACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGGGAGCTTCTAGGGCTTCCTGAGGTTCCCTTCTTGGGCCTGCCT	GGACGACTGTGTCTTTGGGCCACTGCTCAACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGGGAGCTTCTAGGGCTTCCTGAGGTTCCCTTCTTGGGCCTGCCT	GGACGACTGTGTCTTTGGGCCACTGCTCAACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGGGAGCTTCTAGGGCTTCCTGAGGTTCCCTTCTTGGGCCTGCCT	GGACGACTGTGTCTTTGGGCCACTGCTCGACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGGGAGCTTCTAGGGCTTCCTGGGGTTCCTTCTTGGGCCTGCCT	GGACGACTGTGTCTTTGGGCCACTGCTCGACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGAAAGCTTCTAGGGCTTCCTGGGGTTCCTTCTTGGGCCTGCCT	GGACGACTGTGTCTTTGGGCCACTGCTCGACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGAAAGCTTCTAGGGCTTCCTGGGGTTCCTTCTTGGGCCTGCCT	GGACGACTGTGTCTTTGGGCCACTGCTCGACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGAAAGCTTCTAGGGCTTCCTGGGGTTCCTTCTTGGGCCTGCCT	GGATGACTGTGTCTTTGGGCCACTGCTAGACTTCCCCCTCCTGCAGGGGATCCAGGTCCATGGGATGAGGGGACTAGGGGGCTTCTAGGACTTCCTGGGACTCCCCACTGGGCCTGCCT	
CTT-AAAGGCCTGAGCTAGCTGGAGAAGAGGGGAGGGTCCATGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAAGGCCTGAGCTAAGCTGGAGAAGAGGGGAGGGTCCATGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAAGGCCTGAGCTAAGCTGGAGAAGAGGGGAGGGTCCATGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAAGGCCTGAGCTAGCTGGAGAAGAGGGGAGGGTCCATGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAAGGCCTGAGCTAGCTGGAGAAGAGGGGAGGGTCCATGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAATGCCTGAGCTAGCTGGAGGAGAGGGGAGGGTCCAGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAAGGCCTGAGCTAGCTGGAGGAGAGGGGAGGGTCCAGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAAGGCCTGAGCTAGCTGGAGGAGAGGGGAGGGTCCAGAGCCCATGACTAAAACTACCCAGC-----	CTT-AAAGGCCTGAGCTAGCTGGAGGAGAGGGGAGGGTCCAGAGCCCATGACTAAAACTACCCAGC-----	CTTAAAGGCCTGAGCTATCCAGAGGAGAGGGGAGGGTCCAGTAAAGCCCATCACTGAAAACCTACCCAGCTACAGGAAAGGCAGAGGTTCCCTGTCTCTCCCAATCTCCAGTCCCAGTAT







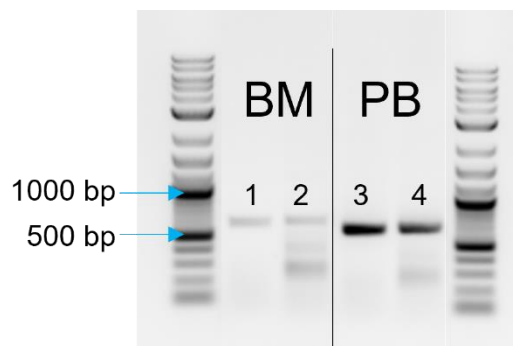
**Figure 2.3. Comparison of CSF3R Class III Nucleotide Sequences between Humans and Non-Human Primates.** Sequence alignments were made using Ensembl. The underlined nucleotide sequence represents the 81 bp of intron 16 that are inserted into the Class III CSF3R isoform as a result of an alternate 3'SS usage. The vertical line depicts the boundary of intron 16 and exon 17.





### **CSF3R Class IV is Present in Rhesus Macaques**

Rhesus macaque is the most widely used non-human primate model organism. RNA-sequencing studies revealed that this species is approximately three times as diverse as humans (92) and believed to have shared a common ancestor with humans 25 million years ago (93). Based on the nucleotide sequence alignment between humans and rhesus macaques, the Class IV region is mostly conserved between the two species. Therefore, we expected the primers to have amplified both Class I and Class IV. The agarose gel showed the Class I amplicon (702 bp) present in mononuclear cells (lanes 1 and 3) and neutrophils (lanes 2 and 4) in both the bone marrow and peripheral blood. However, the Class IV amplicon (283 bp) was only detected in bone marrow and peripheral blood neutrophils (lanes 2 and 4) (**Figure 2.5**). These results represent samples from one rhesus macaque (DF6F).



**Figure 2.5. PCR Amplification of Class I and Class IV CSF3R in Rhesus Macaques.** The agarose gel represents a cropped image of the primer set that specifically amplified Class I and Class IV in rhesus macaques (sample ID ZK39). Mononuclear cells and neutrophils isolated from the bone marrow (BM) and peripheral blood (PB) were used as input templates for the PCR reactions. Lanes 1 and 3 depict mononuclear cells and lanes 2 and 4 depict neutrophils. The expected amplicon sizes for Class I and Class IV are 702 bp and 283 bp respectively.

## 2.4 Discussion

Rhesus macaque genes share a high degree of gene similarities with humans and critical genes like *CSF3R* that are important in hematopoiesis, which is similar across non-human primate species and humans, are expected to have conserved splicing patterns and mechanisms. Our PCR results indicated that Class I is expressed in mononuclear cells from the bone marrow and peripheral blood. Mononuclear cells from bone marrow include lymphocytes, monocytes, stem cells, and progenitor cells and mesenchymal stromal cells, whereas mononuclear cells in peripheral blood contain lymphocytes and monocytes. Since lymphocytes do not express *CSF3R*, most of the Class I expression in bone marrow mononuclear cells stems from monocytes and progenitor cells. Class I expression in the peripheral blood mononuclear cells is mostly seen in monocytes or contaminating neutrophils. Class IV was seen in both bone marrow and peripheral blood neutrophils, which is the expected cell population in which this isoform would be present. It is also possible that Class IV is present in mononuclear cells as well, but its expression may be diluted from the presence of other cell types.

The amplification of Class I and Class IV was expected, but sequencing of the bands will provide confirmation of these observations. The presence of Class IV implies that it may also be important in regulating hematopoiesis and future studies will include determining if *CSF3R* is regulated in rhesus macaques as they are in humans. It is likely that the rhesus macaques use the same splicing mechanism to produce Class I and Class IV because the splice site sequences at the 5' and 3' boundary of the Class IV region is highly conserved. The existence of Class III remains to be answered, but if present in rhesus macaques, would suggest that the recognition of the alternate 3'SS by the spliceosome may not entirely rely on sequence alone.

## Chapter 3: The Role of MDS-Associated Splicing Factor Mutations on *CSF3R* Splicing

### 3.1 Rationale

The high prevalence of mutations in splicing factors, which are not typically found in solid tumors, suggests that disruption of RNA splicing might be a common disease mechanism in MDS (3, 94-97). Advances in sequencing have revealed a complex genomic landscape in MDS patients that largely varies depending on disease subtype (98-101). Since the discovery of recurrent somatic mutations in spliceosome genes more than a decade ago (1), the molecular mechanisms that describe MDS pathogenesis remain unclear. A vast number of studies have focused on global splicing alterations that have resulted in a comprehensive catalogue of missplicing events caused by aberrant regulation from mutated splicing factors (94, 99, 102-104). While these top-down approaches have highlighted the importance of targeting splicing factor mutations for potential therapeutics (105-113), there is still little overlap between the genes that are misspliced between different patient cohorts (94). Bulk RNA-sequencing often dilutes expression of important gene isoforms and varying degrees of mutational burden within patient cohorts can affect differential splicing patterns (100). Splicing patterns between different species are also usually not conserved (114). Gene expression analysis from mice used as models for hematopoiesis (115, 116) may not entirely translate to humans because of different predominating leukocyte populations and splicing regulatory networks (117). For these reasons, splicing of genes like *CSF3R* in which non-canonical isoforms with altered functions are elevated in myeloid neoplasms are often overlooked because they are enriched in specific cell populations.

Global gene expression and aberrant splicing analyses are useful markers for disease prognoses (7, 9). To understand the molecular mechanisms driving MDS pathogenesis, I proposed a bottom's up approach using *CSF3R* splicing as a model to study the effects of splicing on hematopoiesis. Investigating the splice isoforms of *CSF3R* is particularly relevant in normal and abnormal granulopoiesis because of their importance in regulating neutrophil development

(40). Neutrophils represent a well-documented example of the importance of maintaining proper splicing regulation in normal physiological processes (37, 39, 118, 119). Studying alternate *CSF3R* isoform production in the presence of splicing factor mutations will add new insights into a complex disease with an additional goal of highlighting the subtle, but impactful ways alternative isoforms can drive diseases.

## **3.2 Materials and Methods**

### **Cell Culture and Reagents**

All cells were cultured in a humidified atmosphere at 37°C with 5% CO<sub>2</sub>. K562 and HL-60 cells were obtained from ATCC (Manassas, VA). K562 cells were maintained in IMDM (Gibco) supplemented with final concentrations of 10% FBS (R&D Systems) and 1% penicillin/streptomycin (Gibco). HL-60 were maintained in IMDM supplemented with final concentrations of 20% FBS and 1% penicillin/streptomycin. Cells were cultured at a density between 1x10<sup>5</sup>-1x10<sup>6</sup> cells/mL with medium renewal every 2-3 days. Cells were maintained in low passage states by cryopreserving in Bambanker (Fujifilm Wako Chemicals) freezing medium in liquid nitrogen.

### **Isolation, Purification, and Differentiation of Human CD34+ Cells**

Mononuclear cells were isolated from human umbilical cord blood (Cleveland Clinic Cord Blood Center) by density gradient centrifugation using Ficoll-Paque PREMIUM. Cord blood was layered on the Ficoll-Paque solution and spun at 400xg for 30 min at 20°C. The interphase cells were collected, spun at 300xg for 10 min, and resuspended in autoMACS Running Buffer. Cells were centrifuged again at 200xg for 15 min to remove platelets and then magnetically labeled with the CD34 MicroBead Kit (Miltenyi Biotec). The magnetically labeled CD34+ cell fraction was eluted from the column and cultured in StemSpan SFEM II (STEMCELL Technologies) supplemented with StemSpan CD34+ Expansion Supplement (STEMCELL Technologies) for 7 days. Cells were cultured for 7 days in StemSpan SFEM II with 10% FBS, 1% penicillin/streptomycin supplemented

with 100 ng/mL human SCF, 10 ng/mL human IL-3, 10 ng/mL human TPO, and 10 ng/mL human G-CSF at 37°C, 5% CO<sub>2</sub>. Cells were then cultured in StemSpan SFEM II with 10% FBS, 1% penicillin/streptomycin, and 10 ng/mL human G-CSF for an additional 7 days.

### **Construction of *CSF3R* Minigene ESE Deletional Mutants**

Putative SRSF2 ESE motifs were found using ESEfinder 3.0 with *CSF3R* exon 17 as the input sequence. Top scoring motifs above the default threshold (2.383) were identified as ESE1 (5.44) and ESE2 (5.97) with respective sequences 5'-GGGGACCCUGGU-3' and 5'-UGGGCCACUGCU-3'. ESE1 and ESE2 were deleted using Quikchange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacture's protocol using forward primer 5'-GCCAGCCCCTTGGGTAACCCAGCCC-3' and reverse primer 5'-GGGCTGGGGTTACCCAAGGGGCTGGC-3' for ESE1 and forward primer 5'-GACGACTGTGTCTTTGCTCAACTTCCCCCTCC-3' and reverse primer 5'-GGAGGGGGAAGTTGAGCAAAGACACAGTCGTC-3' for ESE2. Bacterial transformations were grown on LB agar plates with 0.1 mg/mL ampicillin for 16 h. Individual colonies were plucked and grown in LB broth with 0.1 mg/mL ampicillin for 16 h at 37°C with shaking at 225 rpm. Successful ESE deletional mutants were confirmed by sequencing after sample preparation using QIAprep Spin Miniprep Kit (QIAGEN).

### **SRSF2 RRM Protein Purification**

Gene fragments (Integrated DNA Technologies) of SRSF2 WT and SRSF2 P95H RRM (amino acids 1-101) were cloned into pMCSG7 plasmid using restriction cut sites NdeI and EcoRI with T4 ligase (Promega Corporation). Protein was expressed in BL21(DE3) competent cells (New England Biolabs), grown in M9 minimal medium supplemented with 13.3 mM glucose, 9.3 mM NH<sub>4</sub>Cl, 30 mg/mL thiamine/HCl, 2.5 M MgCl<sub>2</sub>, 1 M CaCl<sub>2</sub>, and 0.1 mg/mL ampicillin. Bacterial cultures were grown at 37°C until the optical density reached 0.8-1 at which it was moved to 20°C and protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside for an



additional 16 hours. Cells were centrifuged at 6,000 rpm for 30 min at 4°C and lysed with buffer containing 20 mM Tris/NaCl and 20 mM imidazole, pH 8 with the addition of 1X PIC and 1 mM PMSF. Following sonication of the pellet, the cells were centrifuged again at 12,000 rpm for 25 min at 4°C. Protein was purified by nickel affinity chromatography using ÄKTA start and the bound protein was eluted in buffer solution containing 20 mM Tris/NaCl, 250 mM imidazole, and 10 mM  $\beta$ -mercaptoethanol. Briefly, following nickel affinity chromatography, the 6-histidine tag at the N-terminus of SRSF2 WT and P95H proteins were cleaved in lysis buffer with TEV enzyme at 25°C for 8-12 h. The cleaved SRSF2 proteins were further purified as a monomer by size exclusion chromatography using a Superdex 75 column (1.6x60 cm; Pharmacia) that was equilibrated with buffer containing 50 mM L-Glu, 50 mM L-Arg, and 20 mM  $\text{Na}_2\text{HPO}_4/\text{NaHPO}_4$ , 4 mM TCEP at pH 5.5. The purity of the proteins was further analyzed by SDS page analysis and folding of the proteins were checked by two-dimensional NMR HSQC experiments. HPLC purified RNA oligonucleotides were purchased from Integrated DNA Technologies and resuspended in ITC buffer (50 mM L-Glu, 50 mM L-Arg, and 20 mM  $\text{Na}_2\text{HPO}_4/\text{NaHPO}_4$  at pH 5.5).

### **Isothermal Titration Calorimetry**

The binding affinity measurements of SRSF2 RRM1s with RNA oligos were characterized by using VP-ITC (MicroCal) with 8  $\mu\text{M}$  (2 mL) of protein titrated with 80  $\mu\text{M}$  (500  $\mu\text{L}$ ) of RNA in ITC buffer. All experiments were performed at 25°C with 29 consecutive injections of 10  $\mu\text{L}$  of RNA. Curves were fitted by Affinimeter software using a stoichiometric equilibria model described by Free species  $\rightleftharpoons$  MA  $\rightleftharpoons$  M<sub>2</sub>A  $\rightleftharpoons$  M<sub>3</sub>A for ESE1 and Free species  $\rightleftharpoons$  MA for ESE2. Dissociation constants ( $K_d$ ) were calculated based on global fitting of two independent measurements for both SRSF2 WT and SRSF2 P95H using the KinITC routines supplied with the Affinimeter software.

### **Mouse Lin<sup>-</sup>Sca-1<sup>+</sup>c-Kit<sup>-</sup> (LSK) Purification and Colony Assays**

Lineage-negative cells were isolated from the bone marrows of the femurs and tibias from 8-12 week-old male and female wildtype *Csf3r* or *Csf3r*<sup>-/-</sup> mice using Direct Lineage Cell Depletion

Kit (Miltenyi Biotec). LSK cells were purified using anti-CD117 (BD Pharmingen) and anti-Sca-1 (Cedarlane) and cultured in StemSpan SFEM II supplemented with 1% penicillin/streptomycin, 100 ng/mL murine SCF, 100 ng/mL human IL-11, and 20 ng/mL murine IL-3 at 37°C for one day. Cells were transduced the following day with the cDNA for Class I, Class III, Class IV, or empty vector control at MOI 10 and cultured for an additional three days. Cells were sorted for GFP and 100 cells were plated onto MethoCult 3534 (STEMCELL Technologies) supplemented with 100 ng/mL G-CSF in triplicate plates. Colonies were scored on day 8 along with cytopins, differential counts, and flow cytometry. Cytopins of colonies were prepared by cytocentrifugation at 500 rpm for 5 min. Cells were fixed and stained using the PROTOCOL Hema 3 staining kit (Fisher Scientific). Flow cytometry measured the populations of neutrophils and monocytes using the antibodies CD11b (Miltenyi Biotec), CD11c (BioLegend), Ly6G (BioLegend), and Ly6C (BD Pharmingen). Animal studies were approved by the Cleveland Clinic Institutional Animal Care and Use Committee. *Csf3r*<sup>-/-</sup> mice were obtained from Daniel Link (Washington University Medical School). A targeting vector was designed to replace exons 3-8 that encode the endogenous *Csf3r* with a neomycin resistance cassette that abolishes gene function. The vector was electroporated into 129X1/SvJ-derived RW-4 embryonic stem cells. Correctly targeted cells were injected into C57BL/6 blastocytes and the resulting chimeric males were bred into Black Swiss females. Mice were backcrossed for at least ten generations onto a C57BL/6J background. These mice exhibit fewer circulating neutrophils as well as 50% decrease of mature myeloid cells, metamyelocytes, HSCs, and granulocytes in the bone marrow. Neutrophils from these mice display impaired chemotaxis in response to G-CSF and IL-8. Mononuclear cells derived from the bone marrow also show increased apoptosis.

### ***Csf3r*<sup>-/-</sup> Mice Genotyping**

DNA from ear clippings of mice at least five weeks old were isolated by incubating in a base solution containing 25 mM NaOH and 0.2 mM EDTA at 95°C for 30 min, then neutralized in a

solution containing 40 mM Tris-HCl. Touchdown PCR was conducted on samples using a common forward primer 5'-ACATAAGCCTGTGGGAAGGA-3', wildtype *Csf3r* reverse primer 5'-GCTGGTTCTCCACTCATTG-3', and *Csf3r* knockout reverse primer 5'-CTCCAGACTGCCTTGGGAAAA-3'. PCR products were visualized on a 2% agarose gel with ethidium bromide. Wildtype *Csf3r* PCR bands are 136 bp, *Csf3r*<sup>-/-</sup> PCR bands are 78 bp, and heterozygotes are both 78 bp and 136 bp.

### RNA-Sequencing and Analysis of Patient Data

Sequencing data of alternative splicing events from MDS, AML, and MDS/MPN patients were analyzed from Hershberger *et al.* Briefly, bone marrow samples from patients were collected between 09/2005 and 04/2017. Paired-end reads of 100 bp were sequenced with a median depth of 50 million reads per sample. Sequences were aligned using STAR (STAR 2.5.0) to human reference genome hg19. Annotated splice junctions were quantified using rMATS 4.0.1 and filtered for coverage and variance. Skipped and included junction counts on *CSF3R* exon 17 corresponding to Class I, Class III, or Class IV were tabulated and isoform ratios were calculated.

*CSF3R* isoform proportions were calculated using the equation  $\frac{1}{(1+\text{Class IV}:\text{Class I}+\text{Class III}:\text{Class I})}$ .

### PCR and qPCR Primer Design

Primers for RT-PCR were designed to amplify either both Class I and Class IV or both Class I and Class III in the *CSF3R* minigene. The primers are as follows: Class I and Class III: F=5'-TGTGTGGGGAGCGTGGATTC-3', R=5'-CGGCTTCTTTTCATCCTCCTCC-3'; Class I and Class IV: F=5'-TGTGTGGGGAGCGTGGATTC-3', R=5'-CCCAAGAGTGTCTATAACAACAACAAAACACTGC-3'. Specific qPCR primers to detect Class I, Class III, and Class IV in the *CSF3R* minigene are as follows: Class I: F=5'-GAGGACCCACAAGGATGCCTT-3', R=5'-CAGCAGCTGCCATAAAGGAC-3'; Class III: F=5'-CTTCTACACACCCATGTCCCGC-3', R=5'-AGCACGGCTCATTTCAGATGTC-3'; Class IV: F=5'-ATGCCTTCCAGCTGCCC-3', R=5'-GCCCAGCCTGATCGCT-3'. Specific qPCR primers to

detect endogenous Class I, Class III, Class IV, and Actin are as follows: Class I: F=5'-CCACAATCATGGAGGAGGATGC-3', R=5'-CAGCAGCTGCCCATAAAGGAC-3', Class III: F=5'-CAGGAAGAATCCCCTCTGGC-3', R=5'-GTCTGCCCCAGCCACTG-3'; Class IV: F=5'-GTCTGCCCCAGCCACTG-3', R=5'-GCCCAGCCTGATCGCT-3'; Actin: F=5'-GAATCAATGCAAGTTCGGTTCC-3'; R=5'-TCATCTCCGCTATTAGCTCCG-3'.

### **cDNA Synthesis**

Total RNA was collected from cells by lysing in 1 mL TRIzol (Invitrogen). RNA purification was performed by chloroform-based aqueous extraction (200  $\mu$ L), isopropanol precipitation (500  $\mu$ L), and rehydration in nuclease-free water. RNA were treated with TURBO DNA-free Kit (Invitrogen) and cleaned up using Monarch RNA Cleanup Kit (NEB). RNA (1  $\mu$ L) was measured using a Nanodrop One spectrophotometer (Thermo Fisher) at an absorbance of 260 nm. One  $\mu$ g of total RNA was used to generate 1  $\mu$ g cDNA by RT-PCR (iScript cDNA Synthesis Kit, BioRad). Samples without reverse transcriptase were included as negative controls along with a no template control.

### **PCR and Quantitative Real-Time PCR**

Isoform specific mRNA levels were measured by SYBR Green fluorescence-based qPCR (PowerUp SYBR Green Master Mix, Applied Biosystems) using primers specific for Class I, Class III, or Class IV to a final concentration of 500 nM. Primer efficiencies were validated using 5-fold serial dilutions of reference cDNA (minigene only transfected cells). Reactions were performed using 40 cycles of the standard two step PCR on the QuantStudio 3 (Applied Biosystems) with 40 ng of cDNA. Amplicon specificities were also validated by melt curve analysis and visualization on agarose gels. qPCR data was analyzed in Microsoft Excel by relative quantification using the  $2^{-\Delta\Delta Ct}$  method with respect to human beta-actin and the equation  $2^{-Ct(\text{Class III or IV})}/2^{-Ct(\text{Class I})}$  for calculation of Class III:I and Class IV:I mRNA ratios. These methods are outlined in [Harvey and Cheng]. Three biological replicates consisting of three technical replicates were used for gene expression and *CSF3R* isoform ratio calculations.

## Lentivirus Production

293FT cells (Invitrogen) were plated in T-150 culture flasks until 90% confluency with high glucose DMEM (Gibco) supplemented with 10% FBS and 1% penicillin/streptomycin with 0.1 mM MEM Non Essential Amino Acids, and 2 mM L-glutamine co-transfected with 20 µg of the pCMV-VSV-G-RSV-REV envelope vector, 20 µg of the pCAG-HIVgp packaging vector, and 34 µg of the transfer vector (CSII-CMV-MCS-IRES2-Venus, CSII-CMV-MCS-IRES2-Venus-HA-CSF3RI, CSII-CMV-MCS-IRES-Venus-HA-CSF3RIII, CSII-CMV-MCS-IRES-Venus-HA-CSF3RIV, pRRL\_SRSF2\_WT\_mCherry, pRRL\_SRSF2\_P95H\_mCherry, pRRL\_SRSF2\_P95L\_mCherry, or pRRL\_SRSF2\_P95R\_mCherry) using Lipofectamine 2000 (Invitrogen) in 9 mL of Opti-MEM I (Gibco) and 20 mL of complete medium according to the manufacturer's instructions. The medium was changed twice with the addition of 5 µM Forskolin. The media containing the virus was collected, filtered through a 0.45-micron strainer, and ultracentrifuged (Beckman 358126) at 25,000 rpm for 90 min at 4°C. Pellets were resuspended in 10 µL of RPMI and stored at -80°C. Transduction efficiency was determined by flow cytometry of +Venus 293FT cells infected with serially diluted lentivirus. Titers were calculated using the equation  $\frac{\text{cell number} \cdot \text{transduction efficiency}}{\text{lentivirus volume} \cdot \text{dilution factor}}$ .

## Mutagenesis of SRSF2 P95H

The Q5 Site-Directed Mutagenesis Kit (NEB) was used to generate the SRSF2 P95H mutation. Forward (5'-ACCCGGACTCACACCACAGC-3') and reverse (5'-GGCGGCCGTAGCGGCCCAT-3') primers were used to target c.284C>A on the pcDNA3.1-SC35-cMyc plasmid (a gift from Kathleen Scotto, Rutgers University, Addgene plasmid #44721). The PCR contained the hot start master mix, 10 µM of the forward and reverse primers, and 25 ng of the wildtype SRSF2 plasmid. PCR cycling conditions were followed according to the manufacturer's protocol with a  $T_m = 60^\circ\text{C}$ . A kinase, ligase, and DpnI treatment was performed on the PCR product and the reaction was subsequently transformed into One Shot Stbl3 Chemically Competent Cells (Invitrogen) following

the manufacturer's instructions. The transformation mixture (20-75  $\mu$ L) was spread onto ampicillin plates and incubated at 37°C for 16 h. Individual colonies were picked and sequenced using T7 primers to confirm the nucleotide substitution. QIAprep Spin Miniprep and Plasmid Plus Maxi Kits (QIAGEN) were used to isolate and purify plasmid DNA for sequencing and downstream cell transfections.

### **Generation of LUC7L2 Knockout Cells**

LUC7L2 CRISPR/Cas9 knockout in K562 and HL-60 cells were generated using primers that targeted exon 2 (guide target GTTCCAGAAAGGACATCATG) and exon 4 (guide target TTGTGATCGTAGAACAGAAG) respectively. Knockout pools were obtained from Synthego and underwent single cell sorting by flow cytometry. Single cells representing knockout clones were grown in 96-well plates and screened for LUC7L2 protein expression by western blot. Two K562 knockout clones (C7 and C11) and four HL-60 knockout clones (C4, C31, C34, C41) were isolated.

### **Cell Transfection and Transduction**

K562 cells were co-transfected with 1  $\mu$ g each of a construct including the *CSF3R* minigene  $\pm$  wildtype SF3B1 or SF3B1 K700E, *CSF3R* minigene  $\pm$  wildtype U2AF1, U2AF1 S34F, or U2AF1 Q157P, or *CSF3R* minigene  $\pm$  wildtype SRSF2 or SRSF2 P95H using Cell Line Nucleofector Kit V (Lonza) according to the manufacturer's instructions. Cells transfected with the *CSF3R* minigene only were used as a control. Two million cells were transfected and cultured in complete media conditions for 16 h before total RNA was collected. Stable exogenous expression of wildtype SRSF2 or mutant P95H, P95L, or P95R were introduced into K562 cells by lentiviral transduction. Cells were transduced at an MOI of 5 for two days, followed by sorting for equivalent mean fluorescence intensity (MFI) for mCherry expression via flow cytometry. CD34+ were transduced at MOI 10 for three days, followed by sorting for equivalent MFI +mCherry. The *CSF3R* minigene was constructed with a pET backbone with a CAT primer anchoring gene and

a partial intron 16 fused to exon 17 and was a gift from Dr. Chonghui Chen. Plasmids for U2AF1 WT and U2AF1 S34F were gifts from Dr. Dennis Liang Fei (NIH/NHGRI) and U2AF1 Q157P was a gift from Dr. Babal Ja (Cleveland Clinic LRI). pRRL\_SRSF2\_WT\_mCherry, pRRL\_SRSF2\_P95H\_mCherry, pRRL\_SRSF2\_P95L\_mCherry, and pRRL\_SRSF2\_P95R\_mCherry were gifts from Dr. Robert Bradley (Fred Hutchinson Cancer Research Center, Addgene plasmids #84020, #84023, #84022, and #84021).

### **Western Blotting**

Protein immunoblots were performed on lysates made from cells lysed with RIPA buffer with 1 mM sodium orthovanadate, 1 mM phenylmethylsulfonyl fluoride, and 1X protease inhibitor cocktail. Protein concentrations were measured using Pierce BCA Assay (Thermo Fisher) and denatured in 1X Laemmli buffer (Bio-Rad) containing  $\beta$ -mercaptoethanol for 10 min at 95°C. Protein (20  $\mu$ g) were loaded onto 1.5 mm mini-PROTEAN polyacrylamide gels (Bio-Rad), electrophoresed, and transferred onto nitrocellulose membranes (Bio-Rad). Membranes were blocked with 5% non-fat dry milk in tris-buffered saline with 0.1% Tween (MLK-TBST) for 1 h at RT, washed with TBST, and subsequently incubated with anti-FLAG (Invitrogen), anti-SRSF2 (Invitrogen), anti-LUC7L2 (Invitrogen) anti-actin (Santa Cruz), or anti-tubulin (Cell Signaling) antibodies in 5% MLK-TBST overnight at 4°C. Membranes were incubated with secondary antibodies conjugated to horseradish peroxidase (Rockland Immunochemicals) in 5% MLK-TBST for 1 h at RT. Proteins were detected by the addition of ECL Start Western Blotting detection reagent (GE Healthcare), incubated for 1 min, and developed on photographic film (Agfa Healthcare) using an SRX-101a developer (Konica Medical Corp) or imaged using a ChemiDoc MP Imaging System (Bio-Rad).

### **Cloning of CSF3R Class III into CSII-CMV-IRES2 Lentiviral Backbone**

gBlock Gene Fragment of Class III specific sequence was synthesized by Integrated DNA Technologies. The fragment was PCR amplified using “cold fusion” primers that had a 15 bp

homology to both ends of the Class I lentiviral vector. The Class III PCR product and Class I lentiviral vector were both double digested with BlnI and AgeI at 37°C overnight. Both reactions were cleaned up using Monarch PCR & DNA Cleanup Kit (NEB) and loaded onto a 0.6% agarose gel. A band 11.06 kb vector and a 1.4 kb band for the Class III insert were excised from the gel and purified using QIAquick PCR & Gel Cleanup Kit (QIAGEN). Molar ratios of vector:insert of 1:1, 3:1, and 8:1 were combined with T4 ligase (Promega Corporation) and the reaction ran overnight at 4°C. The ligation mixture was transformed into XL10-Gold Ultracompetent Cells (Agilent) and plated onto LB agar plates with 0.1 mg/mL ampicillin and incubated at 37°C for 16 h. Colony PCR was performed on all colonies using primers that span the Class III specific sequence in intron 16. To screen positive clones, an amplicon size of 146 bp was detected. Positive clones were selected for restriction enzyme digestion with BamHI and banding pattern was compared with the Class I lentiviral vector. Ones with the correct banding pattern were further sent for sequencing to confirm the insertion of Class III.

### **Statistical Analysis**

Statistical analysis and graphing were performed on GraphPad Prism 9 (GraphPad Software). One-way ANOVA was used to compare Class III:I and Class IV:I ratios in the minigene splicing assays, in patients with splicing factor mutations compared to control samples, and in neutrophil differentiation status between Class I, Class III, and Class IV.

### **Study Approval**

Animal studies were performed and approved by the Cleveland Clinic Institutional Animal Care and Use Committee. All patients gave written informed consent, and the study was approved by the Internal Review Board in accordance with the Declaration of Helsinki.



### 3.3 Results

#### MDS-Associated Splicing Factor Mutations Lead to Altered Splicing of *CSF3R*

Splicing factor mutations represent a new class of driver lesions that are seen in up to 85% of adult MDS. More than 17,000 transcripts undergo dysregulated splicing in patients with spliceosomal gene mutations. Since aberrant alternative splicing events in genes related to hematopoiesis have already been well characterized in numerous studies, our motivation was to study the effect of *CSF3R* splicing, a gene that is important in regulating granulopoiesis, in the context of MDS-associated splicing factor mutations. Since RNA-sequencing analyses have already identified *Csf3r* as a target of Srsf2 P95H in mouse models, we further investigated if *CSF3R* was a target of other splicing factor mutations to promote the production of alternative splice isoforms that may be weak drivers of dysgranulopoiesis, contributing to MDS pathogenesis.

To determine which splicing factors and their mutations affect *CSF3R* splicing into Class III and Class IV, we designed a *CSF3R* minigene comprising of the terminal exon 17 with a partial intron 16 and a CAT primer anchoring gene (**Appendix A, Figure 1**). We screened for MDS-associated splicing factors of interest by transiently co-expressing the *CSF3R* minigene with either SF3B1 WT or SF3B1 K700E, U2AF1 WT or U2AF1 S34F/Q157P, SRSF2 WT or SRSF2 P95H in 293FT and K562 cells. Since MDS patients often present with haploinsufficiency of LUC7L2 due to -7/del(7q), we knocked out LUC7L2 in K562 and HL60 cells (**Figure 3.1; Appendix A, Figure 4A**). In these cells, only the *CSF3R* minigene was transiently expressed along with a LUC7L2 WT control. When we analyzed the ratios of Class III:I and Class IV:I after 16 h of transfection, we found that cells with SF3B1 K700E expression did not change Class III:I or Class IV:I levels (**Figure 3.2A; Figure 3.3A**). Even though there was no difference in expression in Class IV:I by qPCR, PCR showed a smaller ~200 bp band that resulted from a probable cryptic 3'SS usage (**Figure 3.3A**). In contrast, U2AF1 S34F and U2AF1 Q157P mutations do not affect Class III splicing but does affect Class IV. qPCR analysis show that the

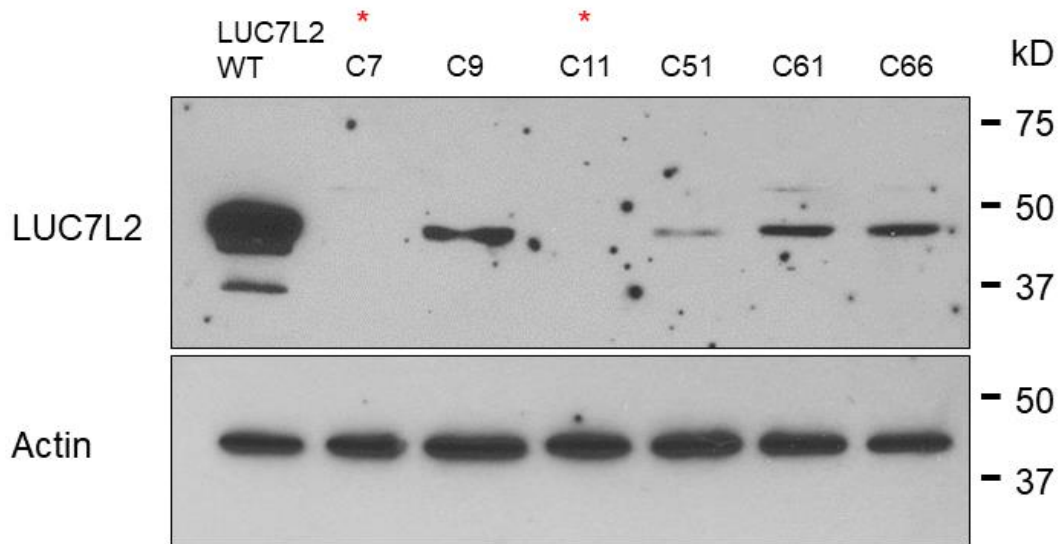
U2AF1 S34F and U2AF1 Q157P mutations have opposing effects on Class IV splicing, with U2AF1 S34F promoting a very significant decrease in Class IV:I and U2AF1 Q157P promoting an increase compared to a minigene-only transfected control (**Figure 3.2B, Figure 3.3B**). To validate the finding that U2AF1 S34F promotes intron 17 excision, we created K562, U937, and HL60 cells lines with stable expression of FLAG-tagged U2AF1 WT and U2AF1 S34F. qPCR analysis of Class IV:I ratios in these cell lines confirm that U2AF1 S34F decreases Class IV splicing compared to the parental cells and U2AF1 WT (**Appendix A, Figure 3**). K562 cells with overexpression of SRSF2 P95H (**Appendix A, Figure 2**) significantly increased Class IV:I levels compared to the minigene only transfected control with no difference in Class III:I (**Figure 3.2C, Figure 3.3C**). Similarly, the related SR protein LUC7L2 also showed significantly increased Class IV:I ratios in both K562 and HL60 knockout cells compared to the wildtype control with no differences in Class III:I (**Figure 3.2D, Figure 3.3D**). Class IV:I findings were validated in LUC7L2 knockout in HL60 cells with increased Class IV splicing in three clones (**Appendix A, Figure 4**).

LUC7L2 haploinsufficiency represents a unique case in its role in alternative splicing. While most other commonly mutated splicing factors in MDS are involved in 3'SS recognition, LUC7L2 is involved in 5'SS recognition and is responsible for both intron excision and retention. It resides on 7q34, a region that is frequently altered in patients with MDS. One study reported decreased hematopoietic differentiation of del(7q)-MDS induced pluripotent stem cells that was rescued by 7q dosage correction. Since our minigene splicing assay showed increased Class IV in LUC7L2 knockout K562 and HL60 cells, we expected to see impaired ATRA-induced granulocytic differentiation in these cells due to abundance of the differentiation-defective Class IV isoform. Cell proliferation of knockout clones 4, 31, and 34 treated with 0.1% DMSO did not show any difference in growth compared to their wildtype counterparts. Viability also remained consistently above 90% in unstimulated cells (**Figure 3.4**). We expected to observe decreased viability as the cells differentiated upon ATRA stimulation, which was seen in cells with wildtype LUC7L2.

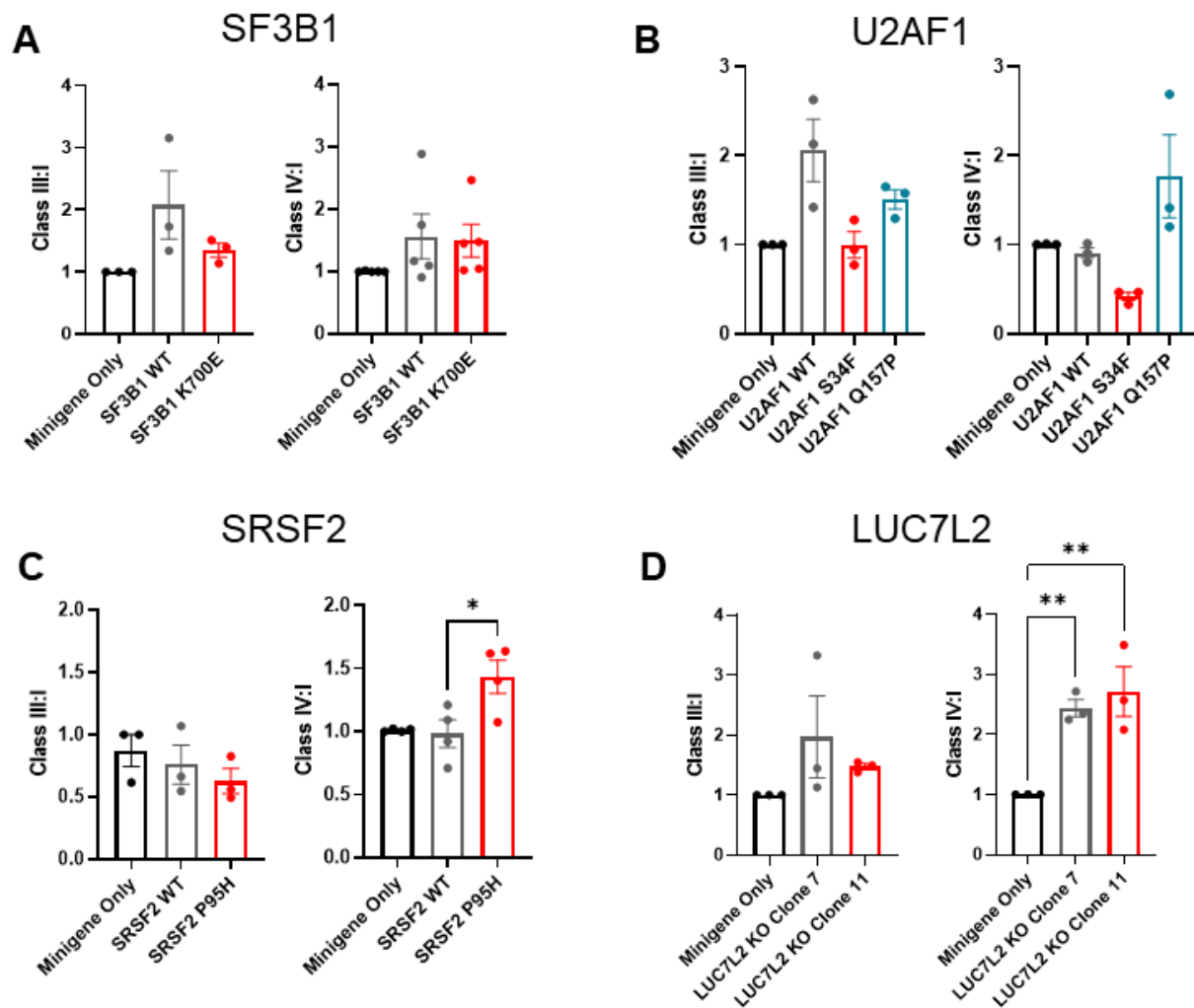
However, an interesting and unexpected observation was the increased viability in LUC7L2 knockout clones treated with ATRA compared to ATRA treated HL60 cells with intact LUC7L2, which would suggest that the cells remained in an undifferentiated state. Cytospin assessment of cell morphology confirmed that the LUC7L2 knockout cells showed impaired granulocytic differentiation compared to the wildtype cells upon ATRA treatment (**Figure 3.5**). To determine if LUC7L2 deficiency impaired differentiation in other hematopoietic lineages, LUC7L2 KO in K562 cells were induced to differentiate into megakaryocytes through PMA treatment. After 6 days, megakaryocyte-like cell morphologies were seen in cytopins with no visible differences between 0.1% DMSO treated cells compared to PMA treated cells (**Figure 3.6**). These results suggest that LUC7L2 deficiency specifically affects granulocyte maturation.

To correlate our *in vitro* minigene splicing assays with patient data, we analyzed a cohort of 1,258 patients diagnosed with MDS, AML, and CMML and assessed the proportions of Class I, Class III, and Class IV *CSF3R* in patients with splicing factor mutations compared to those without splicing factor mutations. We extracted junction counts of *CSF3R* from processed RNA-sequencing data from the Munich Leukemia Laboratory (MLL) cohort as described by Hershberger, *et al.* Junction counts specific to Class I-IV *CSF3R* were tabulated and proportions of each splice isoform relative to another were correlated with disease types and gene mutations, which included splicing factors. Since Class IV transcript levels are increased in the presence of mutated splicing factors in our minigene assay, we expected that production of these alternate splice isoforms was produced at the expense of Class I. Except for U2AF1 mutations, proportions of Class I were significantly decreased in patients with SF3B1 and SRSF2 mutations and LUC7L2 haploinsufficiency compared to patients with wildtype splicing factors (**Figure 3.7A**). Class III was significantly elevated in patients with SF3B1 and SRSF2 mutations as well as LUC7L2 haploinsufficiency (**Figure 3.7B**). In contrast, Class IV was significantly elevated in patients with SRSF2 and U2AF1 mutations and LUC7L2 haploinsufficiency (**Figure 3.7C**). Since the minigene

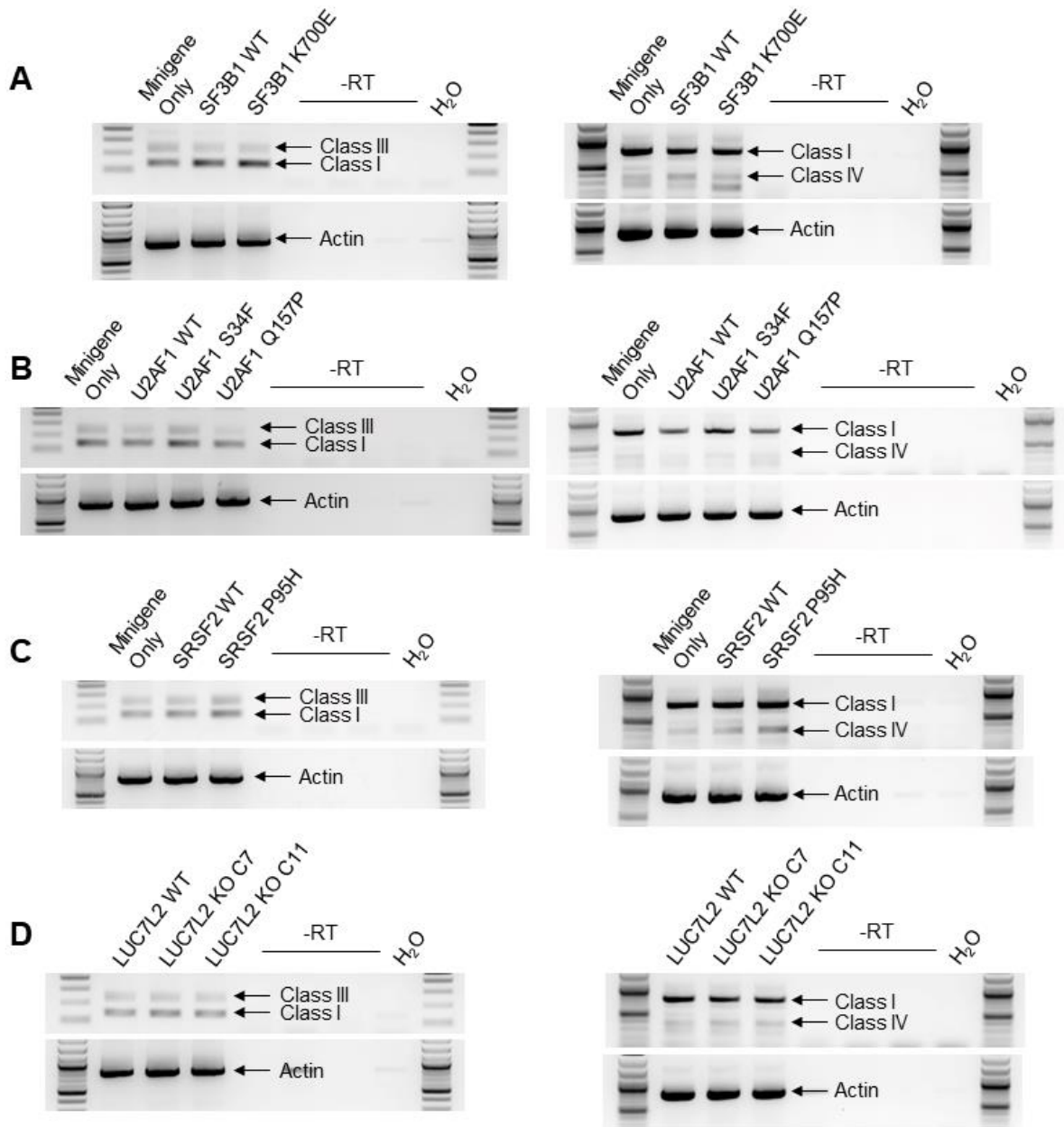
splicing assay showed opposing splicing effects in U2AF1 S34F and U2AF1 Q157P expressing cells, we then stratified the patients based on type of U2AF1 mutation. These groups included S34F, Q157P, Q157R, and R156H. Mutations that did not fit under either of these categories and had <5 patients were grouped into the “other” category. In general, patients with U2AF1 mutations were associated with low Class III (**Figure 3.7D**) but high Class IV expression (**Figure 3.7E**). In addition, U2AF1 Q157P mutations were associated with low Class III and high Class IV expression. Conversely, U2AF1 S34F mutations were associated with high Class III but low Class IV expression. Even though the SF3B1 K700E mutation did not increase Class III or Class IV splicing in the minigene assay, we found that SF3B1 K700E, R625C, K666N, and H662Q mutations all consistently showed patients with high Class III expression (**Appendix A, Figure 5A**). SF3B1 R265C and K666N were the only mutations that were associated with both high Class III and Class IV expression (**Appendix A, Figure 5B**).



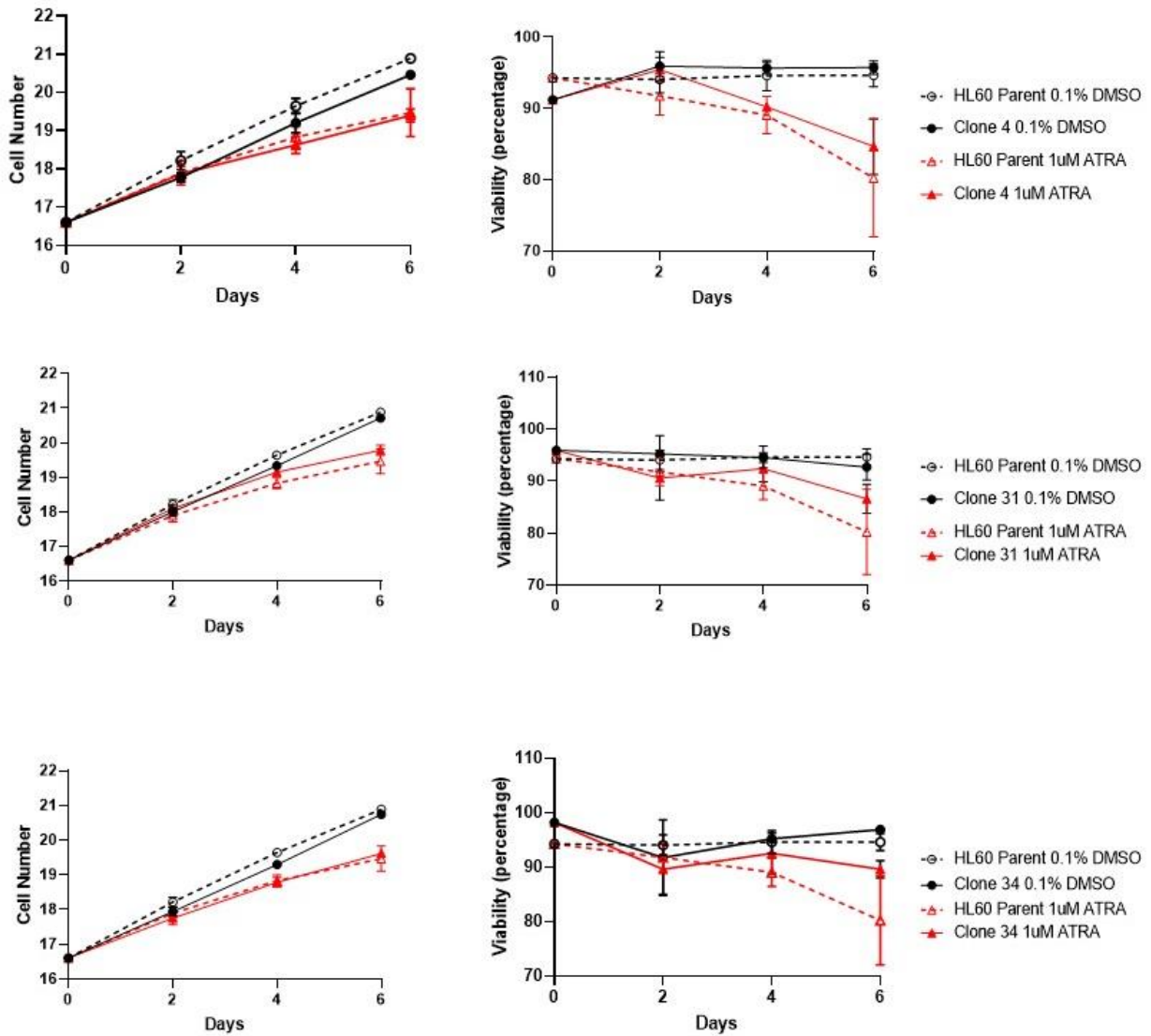
**Figure 3.1. LUC7L2 CRISPR-Cas9 Knockout Clones in K562 Cells.** Western blot of LUC7L2 (47 kDa) in K562 cells show absence of protein. Single cells of a knockout pool were sorted by flow cytometry and subcultured representing individual clones. Two clones (C7 and C11) were isolated from K562 cells as denoted by the red asterisks.



**Figure 3.2. *CSF3R* Minigene Splicing Assay with MDS-Related Splicing Factors.** *CSF3R* Class III:I and Class IV:I ratios in K562 cells with exogenous expression of (A) SF3B1 WT or SF3B1 K700E, (B) U2AF1 WT, U2AF1 S34F, or U2AF1 Q157P, (C) SRSF2 WT or SRSF2 P95H, (D) LUC7L2 WT or knockout by CRISPR editing. The *CSF3R* minigene was transiently expressed in K562 cells, and total RNA was extracted after 18 h. Class I, III, and IV transcript levels were measured using qPCR and ratios of Class III:I and IV:I were calculated relative to cells expressing WT splicing factor. Error bars represent the mean with standard error of the mean from three independent experiments corrected for multiple comparisons by controlling the False Discovery Rate. \* $p \leq 0.0332$ , \*\* $p \leq 0.0021$ .

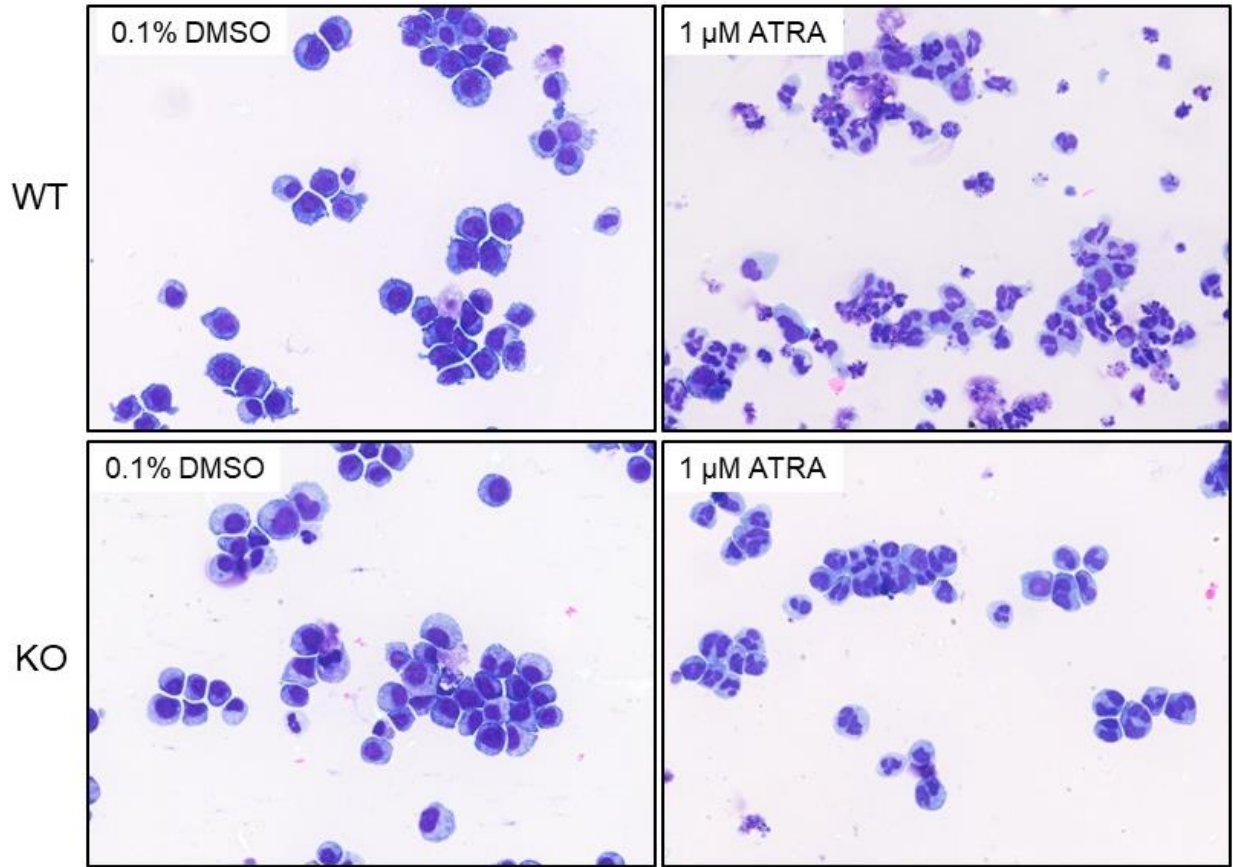


**Figure 3.3. PCR Gels from *CSF3R* Minigene Splicing Assays with MDS-Related Splicing Factors.** In addition to qPCR analyses of Class III and Class IV splicing, bands were visualized on a 1% or 2% agarose gel following PCR reactions. Gels show *CSF3R* Class I, Class III, and Class IV amplicons in K562 cells with exogenous expression of (A) SF3B1 WT or SF3B1 K700E, (B) U2AF1 WT, U2AF1 S34F, or U2AF1 Q157P, (C) SRSF2 WT or SRSF2 P95H, (D) LUC7L2 WT or knockout by CRISPR editing.

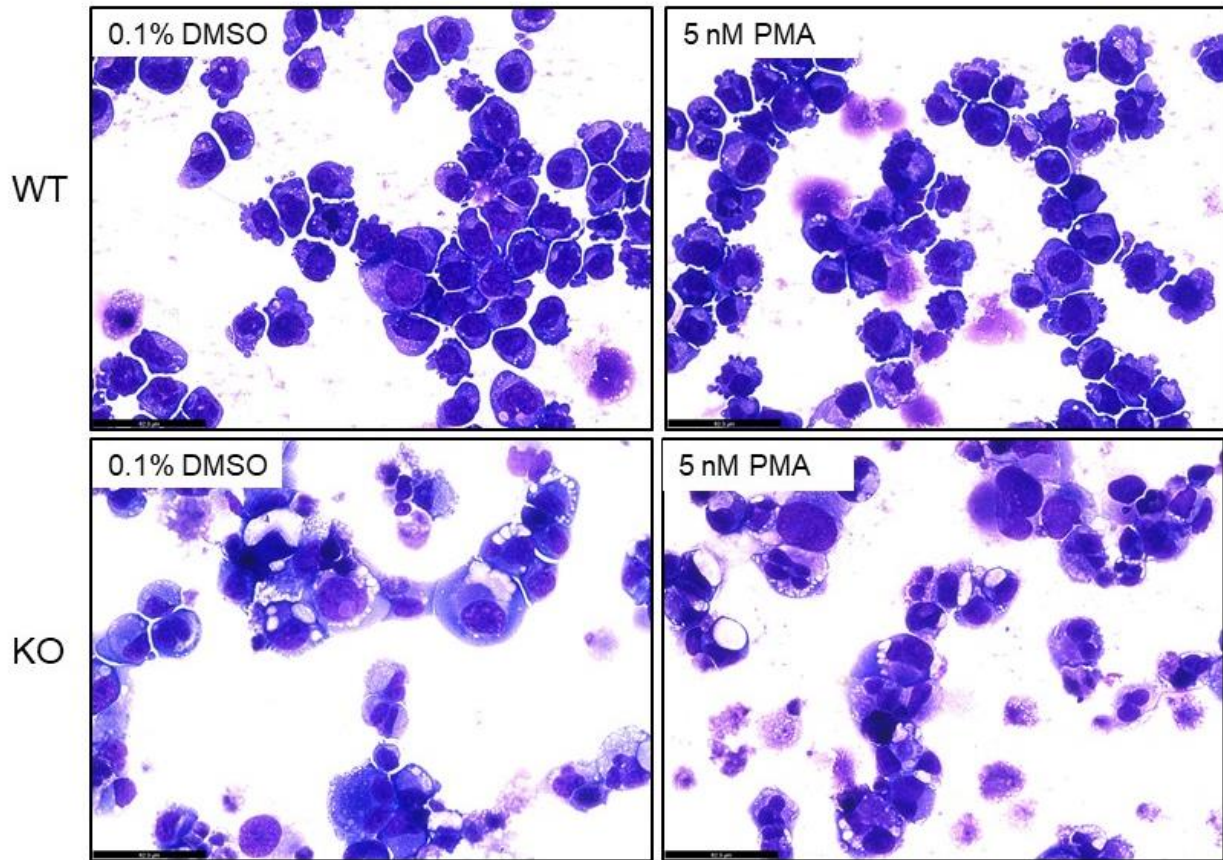


**Figure 3.4. Proliferation and Viability of LUC7L2 KO in HL60 Cells.** Three knockout clones were differentiated with 1  $\mu$ M ATRA over 6 days. Cell counts were performed every 2 days and viabilities were assessed using trypan blue.

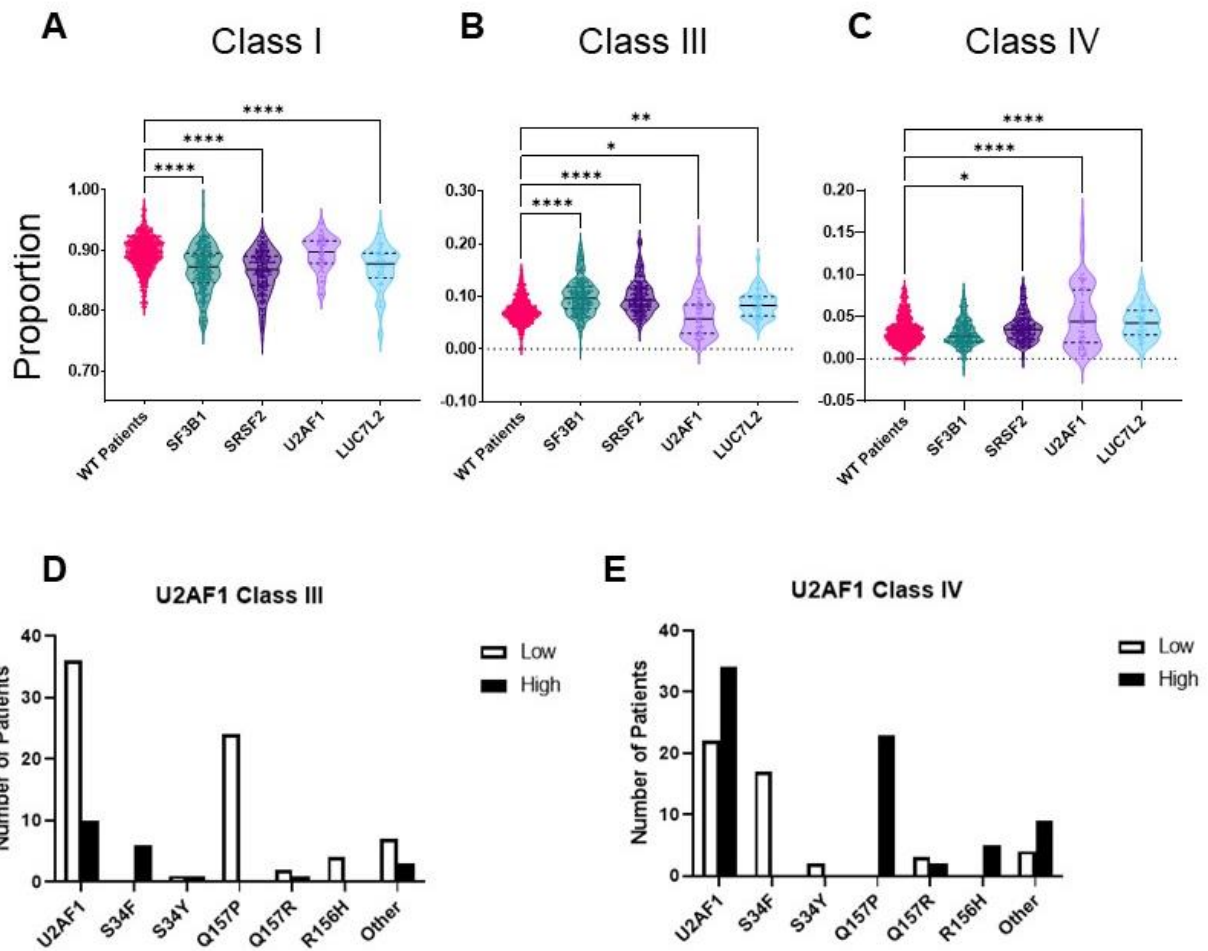




**Figure 3.5. Cytospins of LUC7L2 KO in HL60 Cells Differentiated with ATRA.** Morphologies of HL60 cells treated with 1  $\mu$ M ATRA were assessed by cytopins after 6 days. The top row of images depict HL60 cells with wildtype LUC7L2 treated with 0.1% DMSO and 1  $\mu$ M ATRA. The bottom row of images depict HL60 cells with LUC7L2 deficiency treated with 0.1% DMSO and 1  $\mu$ M ATRA.



**Figure 3.6. Cytospins of LUC7L2 KO in K562 Cells Differentiated with PMA.** Morphologies of K562 cells treated with 5 nM PMA were assessed by cytopins after 6 days. The top row of images depict K562 cells with wildtype LUC7L2 treated with 0.1% DMSO and 5 nM PMA. The bottom row of images depict K562 cells with LUC7L2 deficiency treated with 0.1% DMSO and 5 nM PMA.



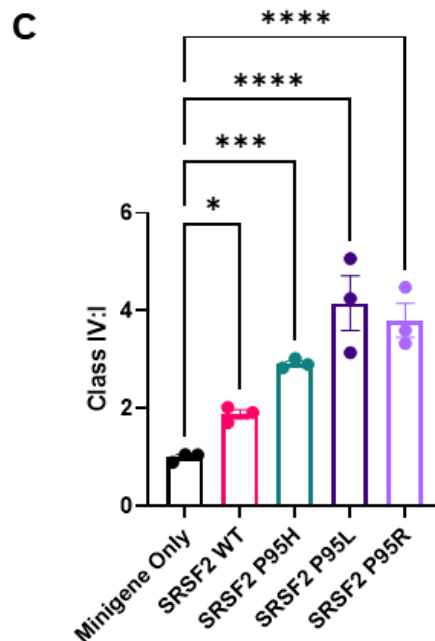
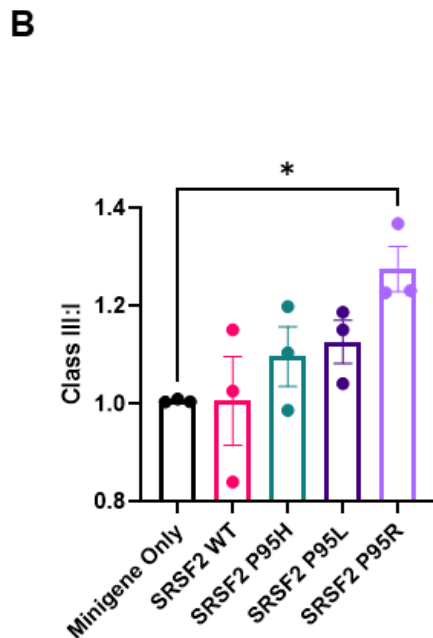
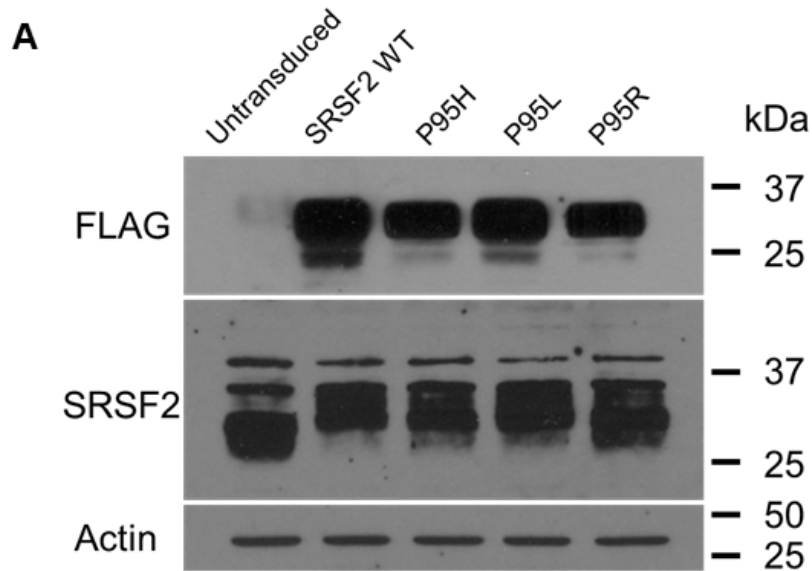
**Figure 3.7. Correlation of *CSF3R* Class III and IV Expression with Recurrent Mutations of Splicing Factors Found in Myeloid Neoplasia.** (A) Comparison of *CSF3R* Class I expression in patients with splicing factor mutations (SF3B1 K666R, E622D, K700E, H662Q, T663I, R625C, D781G, G742D, N626D; SRSF2 P95H, P95L, P95R, P95T, Y44H; U2AF1 S34F, S34Y, Q157R, Q157P, Q157H, R156H; LUC7L2 R271\* and haploinsufficiency) compared to patients without splicing factor mutations across all myeloid malignancies. (B) Comparison of *CSF3R* Class III expression in patients with splicing factor mutations compared to patients without splicing factor mutations across all myeloid malignancies. (C) Comparison of *CSF3R* Class IV expression in patients with splicing factor mutations compared to patients without splicing factor mutations across all myeloid malignancies. Correlation of U2AF1 mutations with low or high (D) Class III and (E) Class IV expression on *CSF3R* splicing. \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ , \*\*\*\* $p \leq 0.0001$ .

### **SRSF2 P95 Mutations Increase *CSF3R* Class IV Splicing**

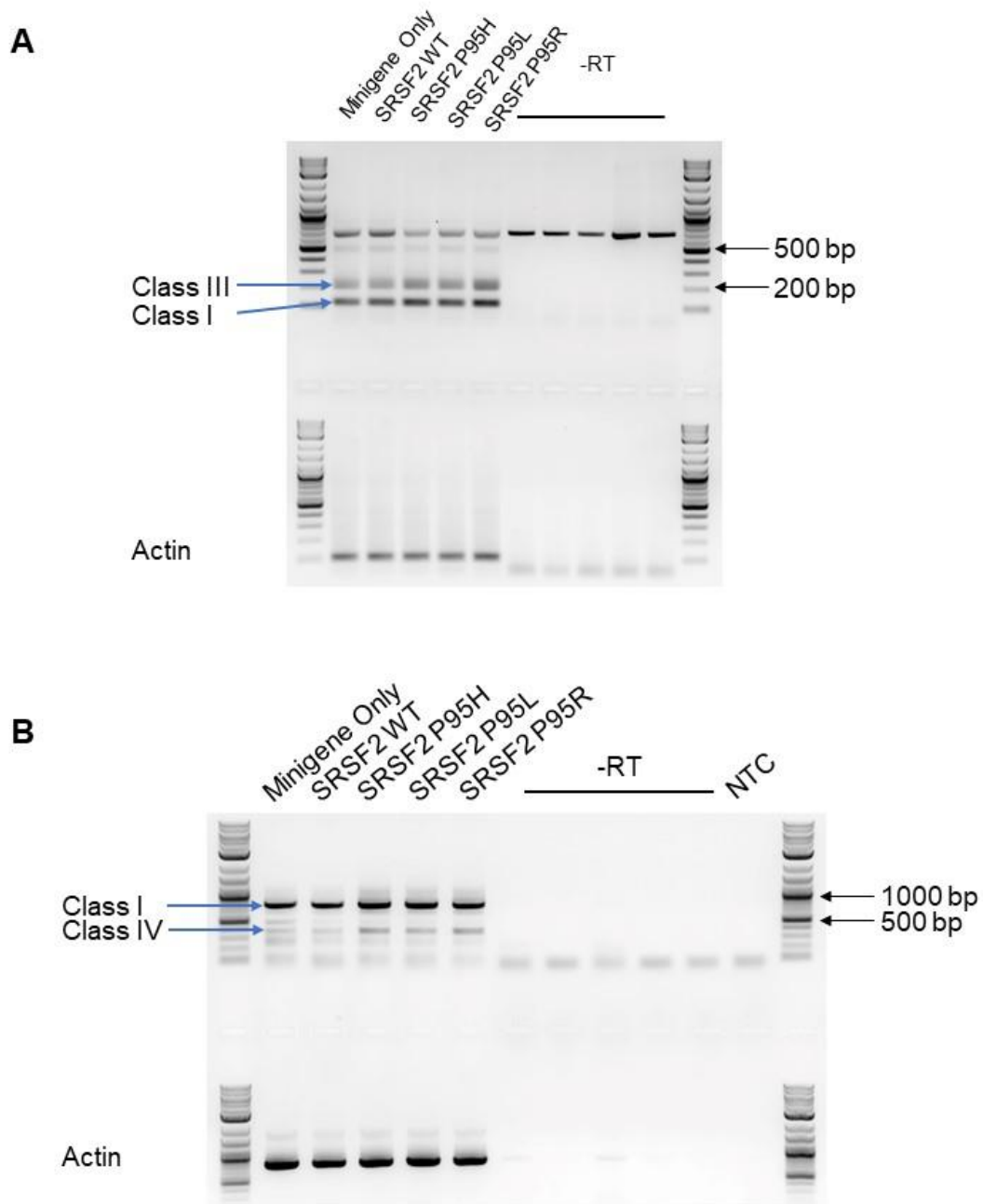
We chose to further investigate the impact of SRSF2 P95H mutations on *CSF3R* splicing into Class I, Class III, and Class IV because of supporting literature that report the downregulation of *Csf3r* in *Srsf2* P95H knock-in mice that contributed to impaired hematopoiesis and myeloid dysplasia. To study the role of SRSF2 P95 mutations on *CSF3R* splicing, we created K562 cells with stable expression of SRSF2 WT, SRSF2 P95H, SRSF2 P95L, and SRSF2 P95R (**Figure 3.8A**). Sequencing confirmed that the point mutations at position 95 encoding a proline all corresponded to the change in amino acid substitutions to either a histidine, leucine, or arginine (**Appendix A, Figure 6**). Cells were transduced with FLAG-tagged SRSF2 lentivirus and sorted for +mCherry with equal MFIs by FACS. Sorting for equal MFIs was used as a surrogate measure of selecting cells that had approximately equal expression of SRSF2. Since transient expression relies in part on transfection efficiency, consistent expression of SRSF2 is an improved model to study mutant SRSF2 effects on *CSF3R* splicing. Since the *CSF3R* isoform ratio calculations are derived from the same transcript, this method bypasses the need for normalizing to a housekeeping gene in qPCR experiments. Class III:I was increased in all SRSF2 P95 mutants with P95R showing significance (**Figure 3.8B**). Class IV:I was also significantly increased in all SRSF2 P95 mutants (**Figure 3.8C**). These data were demonstrated by both qPCR and PCR with higher intensity bands depicting increased splicing. For Class III:I PCR (**Figure 3.9A**), band sizes are Class I = 141 bp and Class III = 222 bp. For Class IV:I, band sizes are Class I = 794 bp and Class IV = 374 bp (**Figure 3.9B**).

AML patients with SRSF2 P95H mutations are reported to have increased Class III:I and Class IV:I *CSF3R* (34, 41). In addition, SRSF2 P95 mutations are often associated with monocyte skewing in mouse knock-in models and in patients with CMML (76, 120, 121). To validate these findings in our patient cohort, we stratified patients based on their diagnoses (AML, MDS, or CMML) and compared *CSF3R* isoform proportions in those with and without splicing factor

mutations. Patients with mutated *SF3B1* or *SRSF2*, or *LUC7L2* haploinsufficiency were observed to have significantly high proportions of Class III while only patients with *LUC7L2* haploinsufficiency showed significantly increased Class IV proportions in AML (**Figure 3.10A**). Patients with mutated *SF3B1* or *SRSF2* have increased Class III proportions compared to patients with mutated *SRSF2* or *LUC7L2* that show increased Class IV proportions in MDS (**Figure 3.10B**). *SRSF2* mutations were observed in patients with MDS, AML, and CMML. However, *SRSF2* mutations were almost exclusive to CMML, in which very few (0.03-0.07%) other splicing factor mutations were seen (**Figure 3.10C**). Studies to date all suggest similar functions between P95 mutations and have not investigated if differences in amino acid substitutions impose differential effects on gene splicing. To study if *CSF3R* splicing is influenced by different *SRSF2* P95 mutations, we stratified patients across all myeloid neoplasms by *SRSF2* P95H, *SRSF2* P95L, and *SRSF2* P95R mutations and observed the percentage of patients with either high or low Class III and Class IV expression. We found that differences in amino acid substitutions did not affect Class III splicing. In contrast, high Class IV expression was mostly driven by the *SRSF2* P95R mutation over *SRSF2* P95H and *SRSF2* P95L mutations (**Figure 3.11**).

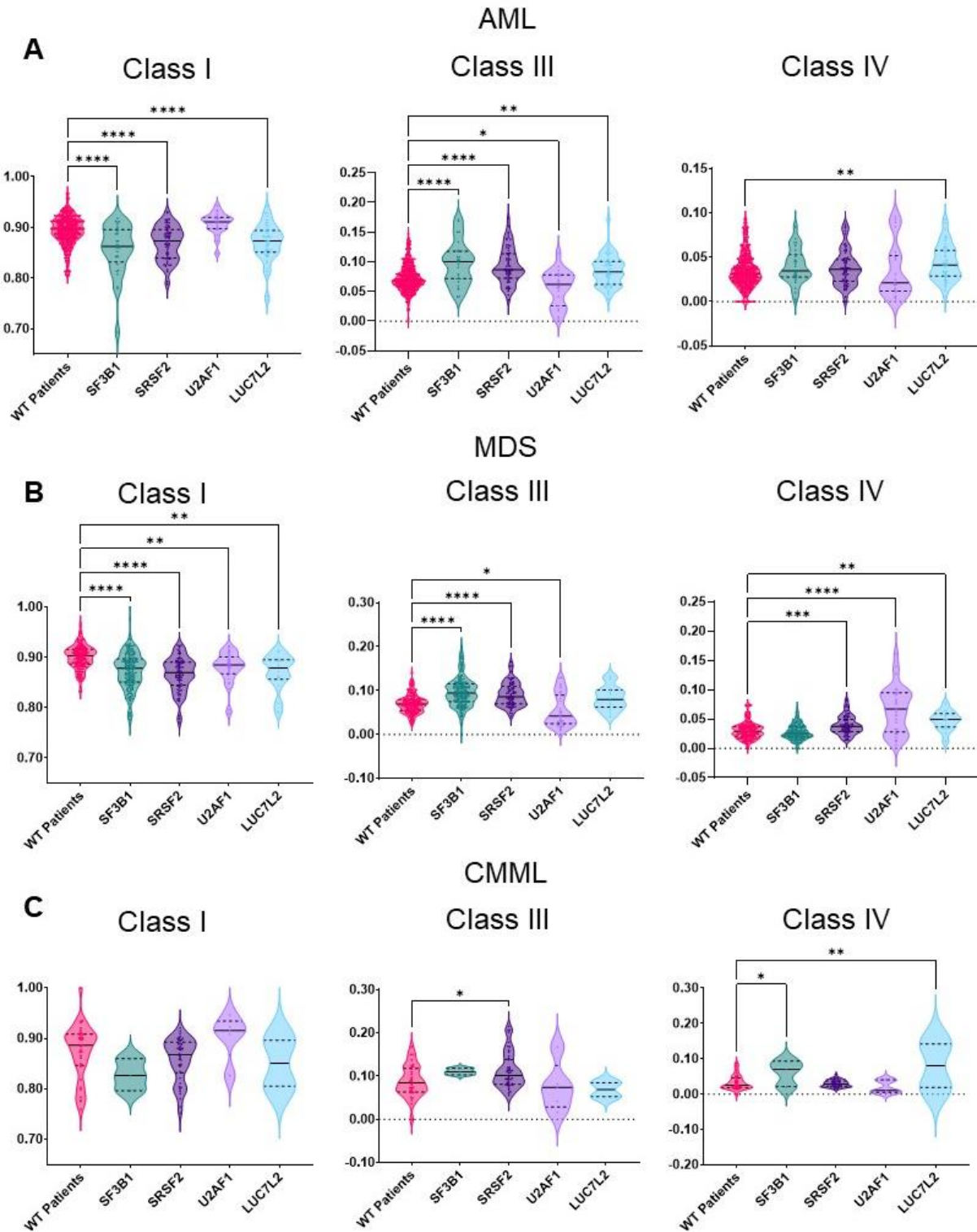


**Figure 3.8. Class III and Class IV *CSF3R* Splicing in K562 Cells with SRSF2 Expression.** (A) Western blot showing FLAG-tagged SRSF2 in K562 cells. (B) Class III:I ratio measured by qPCR. (C) Class IV:I ratio measured by qPCR. Error bars represent the standard error of means from three independent experiments corrected for multiple comparisons by controlling the False Discovery Rate. \* $p \leq 0.0332$ , \*\* $p \leq 0.0021$ , \*\*\* $p \leq 0.0002$ , \*\*\*\* $p \leq 0.0001$ .



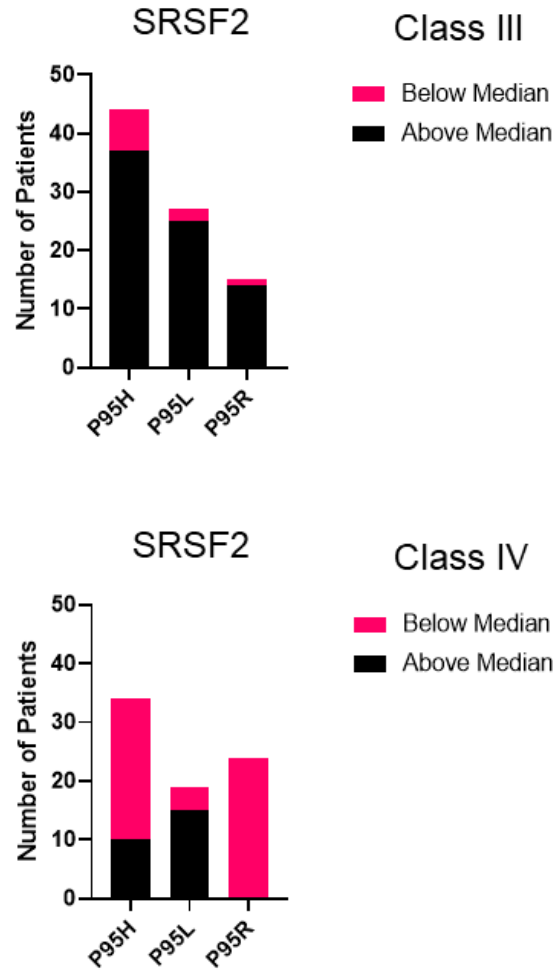
**Figure 3.9. PCR Gels of Class I, Class III, and Class IV *CSF3R* in K562 Cells Expressing *SRSF2*.** (A) Class I (141 bp) and Class III (222 bp) bands. PCR products were run on a 2% agarose gel and is a representative image of three independent experiments. (B) Class I (794 bp) and Class IV (374 bp) bands. PCR products were run on a 1% agarose gel and is a representative image of three independent experiments.





**Figure 3.10. Distribution of Patients with Splicing Factor Mutations Stratified by Disease.** Expression of *CSF3R* Class I, Class III, and Class IV in patients with (A) AML, (B) MDS, and (C) CMML. \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ , \*\*\*\* $p \leq 0.0001$ .



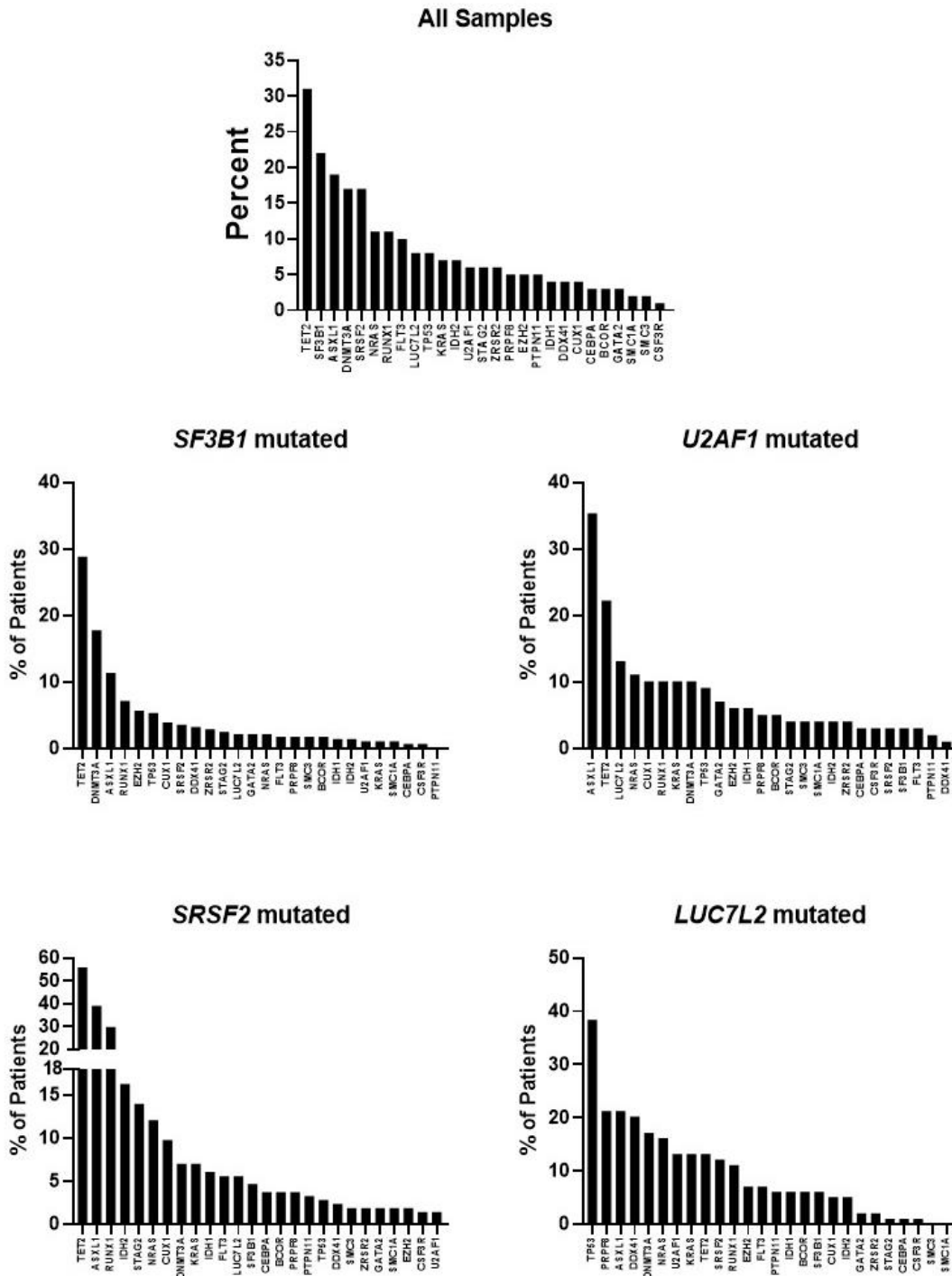


**Figure 3.11. *CSF3R* Splicing by SRSF2 P95 Mutations in Patients with Myeloid Neoplasms.** Patients with high and low Class III and Class IV expression were stratified by SRSF2 P95H (Class III: high n=37, low n=7; Class IV: high n=10, low n=24), SRSF2 P95L (Class III: high n=25, low n=2; Class IV: high n=15, low n=4), or SRSF2 P95R mutations (Class III: high n=14, low n=1; Class IV: high n=0, low n=24) (x-axis). The y-axis reflects the percentage of patients with high and low Class III and Class IV expression.

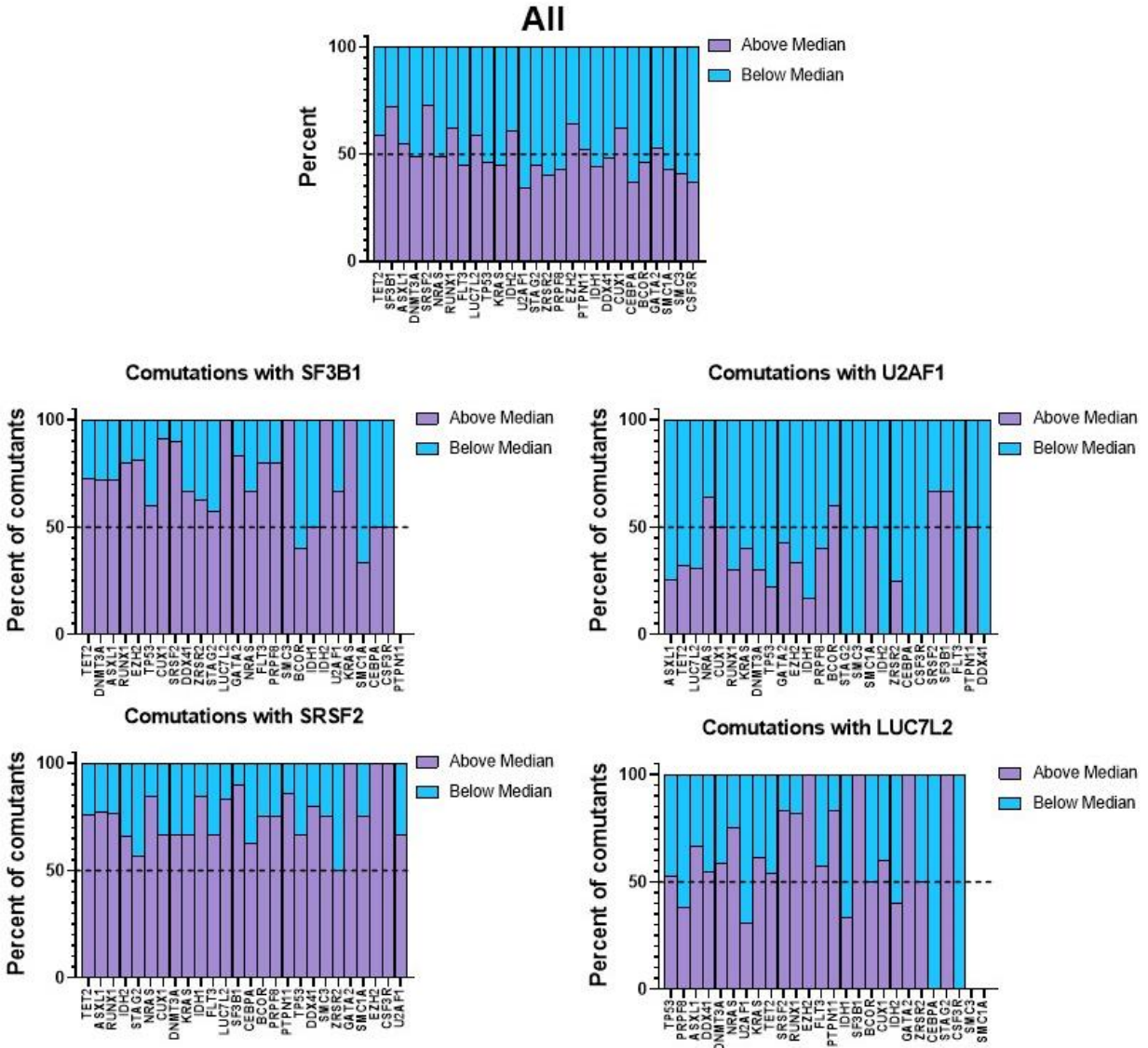
### **Co-Mutations in Splicing Factor Mutated Myeloid Neoplasms May Enhance CSF3R Splicing**

Splicing factor mutations are known to be founder lesions that occur in an HSC clone that gains competitive growth advantages. In some instances of low risk MDS, a clone will undergo somatic reversion with functional recovery of subsequent cell populations that results in HSCs with higher fitness, leading to improved blood counts. However, some HSC clones retain high-risk somatic mutations that expand without functional recovery and acquire secondary mutations that lead to high risk MDS and transformation into AML. In our patient cohort, TET2 mutations (31%) were most commonly observed, followed by SF3B1 (22%), ASXL1 (19%), and DNMT3A (17.5%). TET2 mutations were also the most common mutations to co-occur with SF3B1 (28%) and SRSF2 (56%) mutations. ASXL1 mutations occurred in 36% of patients with U2AF1 mutations and TP53 were found in 38% of patients with LUC7L2 haploinsufficiency (**Figure 3.12**). In general, mutations that were associated with increased Class III splicing were in SF3B1, SRSF2, RUNX1, LUC7L2, IDH2, EZH2, and CUX1. However, the genes that are overall mutated across all myeloid neoplasms do not reflect the same ones when grouped by individual splicing factor mutations. SF3B1 mutations are most commonly observed with TET2, CUX1, SRSF2, GATA2, FLT3, and PRPF8 mutations. U2AF1 mutations are associated with NRAS, BCOR, SRSF2, and SF3B1 mutations. SRSF2 has the highest frequency of co-mutations across all genes assessed. With the exception of ZRSR2, mutations in all genes were correlated with increased Class III. Notable genes co-mutated with LUC7L2 are SRSF2, RUNX1, PTPN11, ASXL1, and NRAS (**Figure 3.13**). Overall, the gene mutations that were strongly associated with increased Class IV were LUC7L2, DDX41, CUX1, GATA2, and SMC1A. As we observed earlier, some mutations are more strongly associated with certain splicing factors. Genes that were co-mutated with SF3B1 included IDH2, ASXL1, RUNX1, CUX1, SRSF2, LUC7L2, GATA2, NRAS, U2AF1. Genes that were co-mutated with U2AF1 included ASXL1, TP53, GATA2, SMC1A, IDH2. In contrast to genes related to SRSF2-associated Class III splicing, those that are seen in high Class IV expression represent the fewest. These genes include CSF3R, U2AF1, and DDX41. In LUC7L2, Genes that strongly

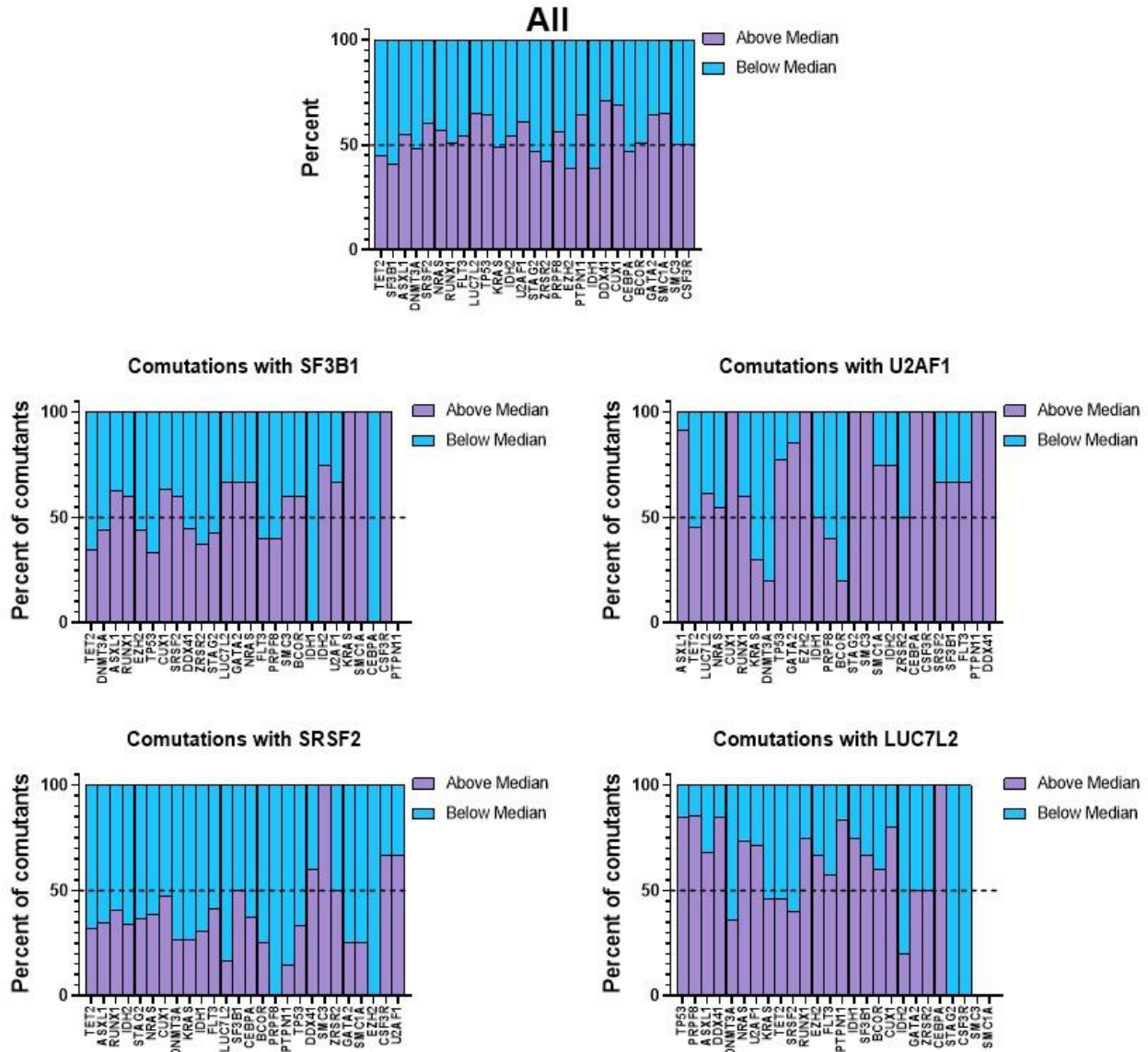
correlated with LUC7L2 haploinsufficiency were TP53, PRPF8, ASXL1, DDX41, NRAS, RUNX1, PTPN11, IDH1, and CUX1 (Figure 3.14).



**Figure 3.12. Frequency of Genes Co-Mutated with Splicing Factors in Patients with Myeloid Neoplasms.** Percentage of mutated genes in patients with myeloid neoplasms are plotted on the y-axis in order from most to less frequently observed.



**Figure 3.13. Association of Co-Mutations with High and Low Class III *CSF3R* Expression.** Genes that were co-mutated in each patient with a single MDS-related splicing factor mutation were counted and grouped into high and low Class III expression.



**Figure 3.14. Association of Co-mutations with High and Low Class IV *CSF3R* Expression.** Genes that were co-mutated in each patient with a single MDS-related splicing factor mutation were counted and grouped into high and low Class IV expression.

## ESEs Contribute to *CSF3R* Class I and Class IV Splicing

Since SRSF2 functions in part by binding to ESEs to regulate exon inclusion or exclusion, we sought to identify these motifs within *CSF3R* exon 17, hypothesizing that SRSF2 P95H would bind to these motifs with higher affinity than SRSF2 WT to promote Class IV splicing. SRSF2 P95H binds to pyrimidine-rich sequences (CCNG) with higher affinity than purine-rich sequences (GGNG). In contrast, WT SRSF2 can bind to both sequences with equal affinity. Using ESEfinder 3.0, we input the nucleotide sequence of exon 17 to identify putative motifs. ESEfinder is a web-based resource to identify exonic splicing enhancers that are responsive to SR proteins SRSF1, SRSF2, SRSF5, and SRSF6. ESEfinder searches for putative motifs by using weight matrices corresponding to the motifs for SRSF1, SRSF2, SRSF5, and SRSF6. The matrices are based on frequency values derived from the alignment of winner sequences obtained by functional SELEX experiments. Only 'hits' or 'high-score motifs' are shown following the query and outputs the position of the first nucleotide, sequence of the motif, and the calculated score. A motif is considered a high score when the number is greater than the threshold value defined in the input. In total, we found 57 putative motifs (**Table 3.1**) with 36 of them residing in the Class IV intron. In general, these putative motifs were observed to be pyrimidine-rich sequences that would indicate stronger binding to SRSF2 P95H in comparison to SRSF2 WT. Two particularly high-scoring motifs were found in our search query: one at position 353 (GGACCCTG, score 5.44) and the other at position 401 (GGCCACTG, score 5.97), named ESE1 and ESE2 respectively. Both motifs are located within intron 17. To determine if SRSF2 could recognize these motifs, we conducted binding studies using ITC. First, we purified the RRM (amino acids 1-101, **Appendix A, Figures 7 and 8**) of SRSF2 WT and SRSF2 P95H.  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR was performed on both SRSF2 WT and SRSF2 P95H titrated with a control sequence 5'-UCCAGU-3' to confirm proper protein folding. We overlaid our results with the HSQC spectra reported in Daubner, *et al* (80) to validate our structure of SRSF2 WT RRM (**Appendix A, Figure 9**). SRSF2 WT and SRSF2 P95H RRM

HSQC spectra were then overlaid to confirm proper folding (**Figure 3.15A**). Purity was assessed by Coomassie Blue staining on a 20% SDS-PAGE. In standard ITC runs, RNA is titrated with the protein. Since SRSF2 purification resulted in low yields, we performed a reverse titration, loading 2 mL of 8  $\mu$ M protein into the chamber and 500  $\mu$ L of 80  $\mu$ M of either the control sequence (5'-UCCAGU-3'), ESE1, or ESE2 on the VP-ITC. We first tested binding of SRSF2 WT with the control sequence, which is an established motif that binds SRSF2 in a 1:1 curve. As expected, our results showed that the control sequence and SRSF2 WT RRM bound 1:1 with  $N=1.09$  fitted to a one site model with a  $K_d$  of 18.1 nM using Origin Software (**Appendix A, Figure 10**). We then proceeded to assess the binding affinities of SRSF2 WT and SRSF2 P95H RRMs with ESE1 and ESE2. An immediate observation about the binding curve for SRSF2 RRM with ESE1 was that it did not conform to a one site model. Instead, the curve implied that two or three proteins interacted with the RNA at the beginning of the titrations (proteins in excess) and eventually tapered a 1:1 interaction with the RNA in excess (**Appendix A, Figure 11A**). A binding model of Free Species  $\rightleftharpoons MA \rightleftharpoons M_2A \rightleftharpoons M_3A$  was used to describe the protein:RNA interactions of SRSF2 RRMs to ESE1 (Figure . Since the additional minigene nucleotides generates two potential ESE motifs on ESE1 (GGGGACCCUGGU, GGGGACCCUGGU), a stronger binding site might influence the initial protein:RNA interaction over the weaker binding site. To test this model, we titrated both SRSF2 WT and SRSF2 P95H RRMs with ESE1 without the additional *CSF3R* minigene nucleotides at the 5' and 3' ends (3'-GGACCCUG-5') and observed that the 1:1 interaction grows, but the 2:1 isotherm still prevails (**Appendix A, Figure 11B**). Using this model and a global curve fitting (**Figure 3.16**), the  $K_d$  values were calculated to be  $0.104 \pm 0.038$ ,  $1.10 \pm 0.505$ ,  $0.339 \pm 0.011$  for SRSF2 WT and  $0.011 \pm 0.002$ ,  $0.262 \pm 0.427$ , and  $0.052 \pm 0.102$  for SRSF2 P95H (**Table 3.2**). The curve fitting for SRSF2 RRMs titrated with ESE2 resulted in a typical 1:1 isotherm (Free species  $\rightleftharpoons MA$ ) (**Figure 3.17**) with  $K_d$  values of  $0.87 \pm 0.13$  for SRSF2 WT and  $0.180 \pm 0.01$  for SRSF2 P95H (**Table 3.2**). The curves indicate that SRSF2 WT has a stronger binding affinity to

ESE1 compared to SRSF2 P95H whereas SRSF2 P95H has a stronger binding affinity to ESE2 compared to SRSF2 WT.

Next, we assessed the contribution of ESE1 and ESE2 in splicing of *CSF3R* into Class I and Class IV by deleting each motif in the *CSF3R* minigene and transiently expressing the construct in K562 cells with exogenous expression of SRSF2 WT or SRSF2 P95H. When either ESE motifs were deleted, Class IV splicing became undetectable by RT-PCR compared to the controls (**Figure 3.18 A and B**). When we measured Class IV:I ratios in these cells by qPCR, the ratios in K562 cells with SRSF2 P95H and the deletional mutants were comparable to cells with SRSF2 P95H and wildtype *CSF3R* minigene (**Figure 3.19 A and B**). Deletion of ESE1, but not ESE2, strongly affected Class IV:I in the presence of SRSF2 WT.

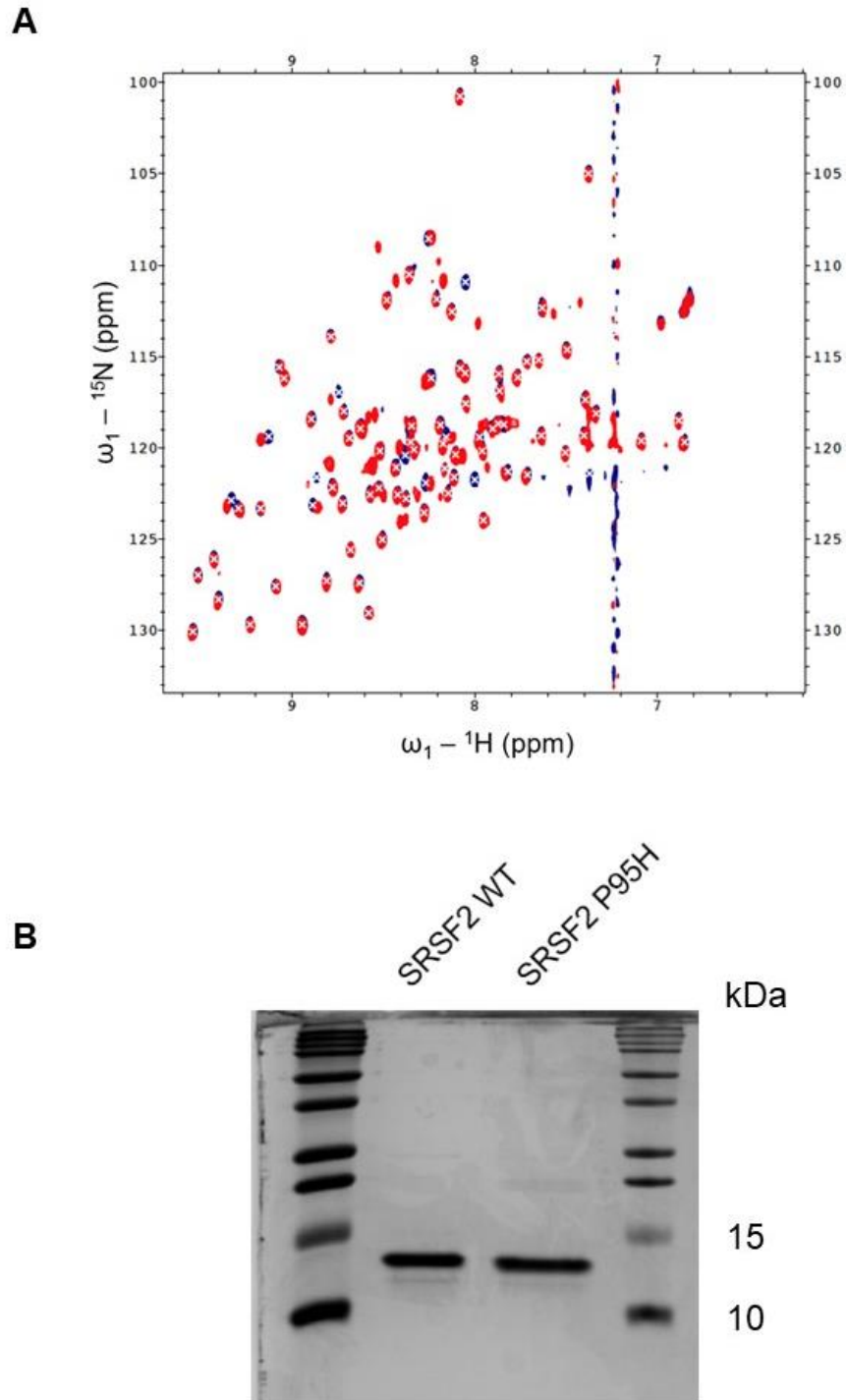


**Table 3.1. ESE Motif Outputs from ESEfinder 3.0.** The input sequence of *CSF3R* exon 17 has 57 putative ESE motifs. The list of motifs represents all scores above the threshold score. The asterisks denote ESE1 and ESE2, the two highest-scoring motifs.

Putative ESE motifs	Position (5' end)	Score
AGCCGGTG	74	2.88
GTGCCCTG	79	3.49
GAGTCCCA	88	2.45
GACCTGTG	108	4.41
GGCCTCCC	115	2.78
GACCTATG	135	4.05
TGCTCCAG	143	3.14
GGACCCAA	153	3.96
AGCCCCAA	176	3.22
AATCCCAG	182	2.74
GTCTGGCA	189	2.66
GGCACCAG	193	2.73
GTCCTTTA	208	2.86
GGCAGCTG	221	2.85
AGCTGCTG	248	3.57
GGCCAGGG	253	2.74
GGGCACTA	262	3.22
CTCCGCTG	271	2.87
GACTCCAC	281	2.57
AGCCCCTC	262	2.43

GGCCTCAC	295	2.40
CACCCCA	300	2.96
AGCCCAA	307	3.22
GTCCTATG	315	3.76
AACCTCTG	325	3.57
GGTTCCAG	332	4.62
AGCCCCTT	343	2.43
GCCCCTTG	344	2.68
GGACCCTG	353	5.44* (ESE1)
AACCCAG	363	3.31
AGCCCAA	369	3.22
GACGACTG	385	2.92
GGCCACTG	401	5.97* (ESE2)
GGATCCGG	431	2.89
GGTCCATG	438	4.60
AGCTTCTA	463	3.25
GGCTTCCT	472	2.55
GCTTCCTG	473	3.61
GGTTCCT	482	2.74
GTTCCCTT	483	2.97
GGCCTGAG	510	3.75
GGTCCATA	538	4.15
AGCCCATG	546	3.42
GACTAAAA	553	2.63
TACCCAG	563	3.01

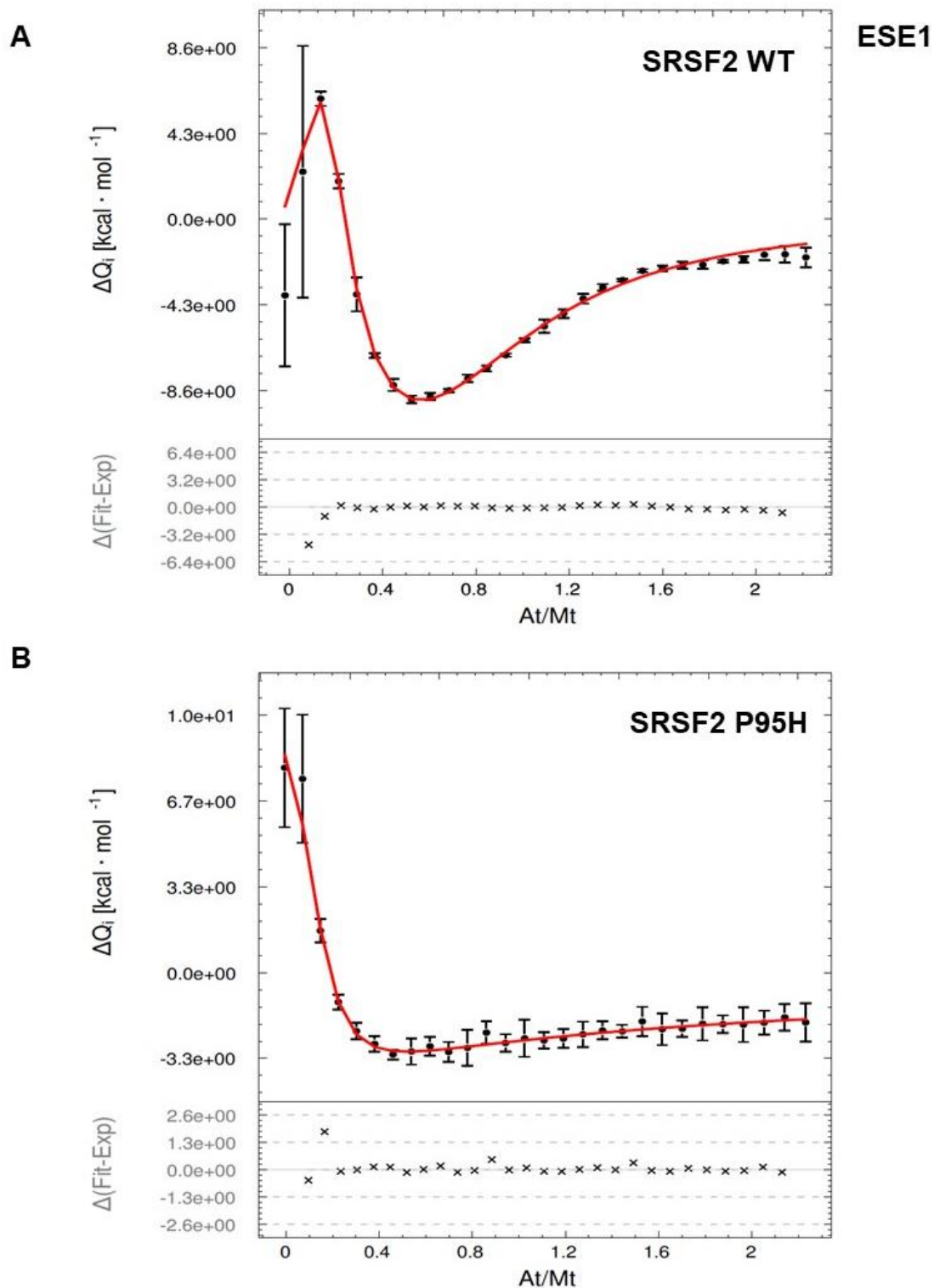
GGCTCTCA	575	3.39
ATCTCCAG	585	2.79
AGTCACCA	591	2.56
GGCCTCCC	632	2.78
GCCTCCCA	633	3.09
GATCTGCA	644	2.75
GTTTCCTA	708	4.55
GTATTGTA	724	2.43
GGTTTGCA	743	2.88
GTTTGCAG	744	3.08



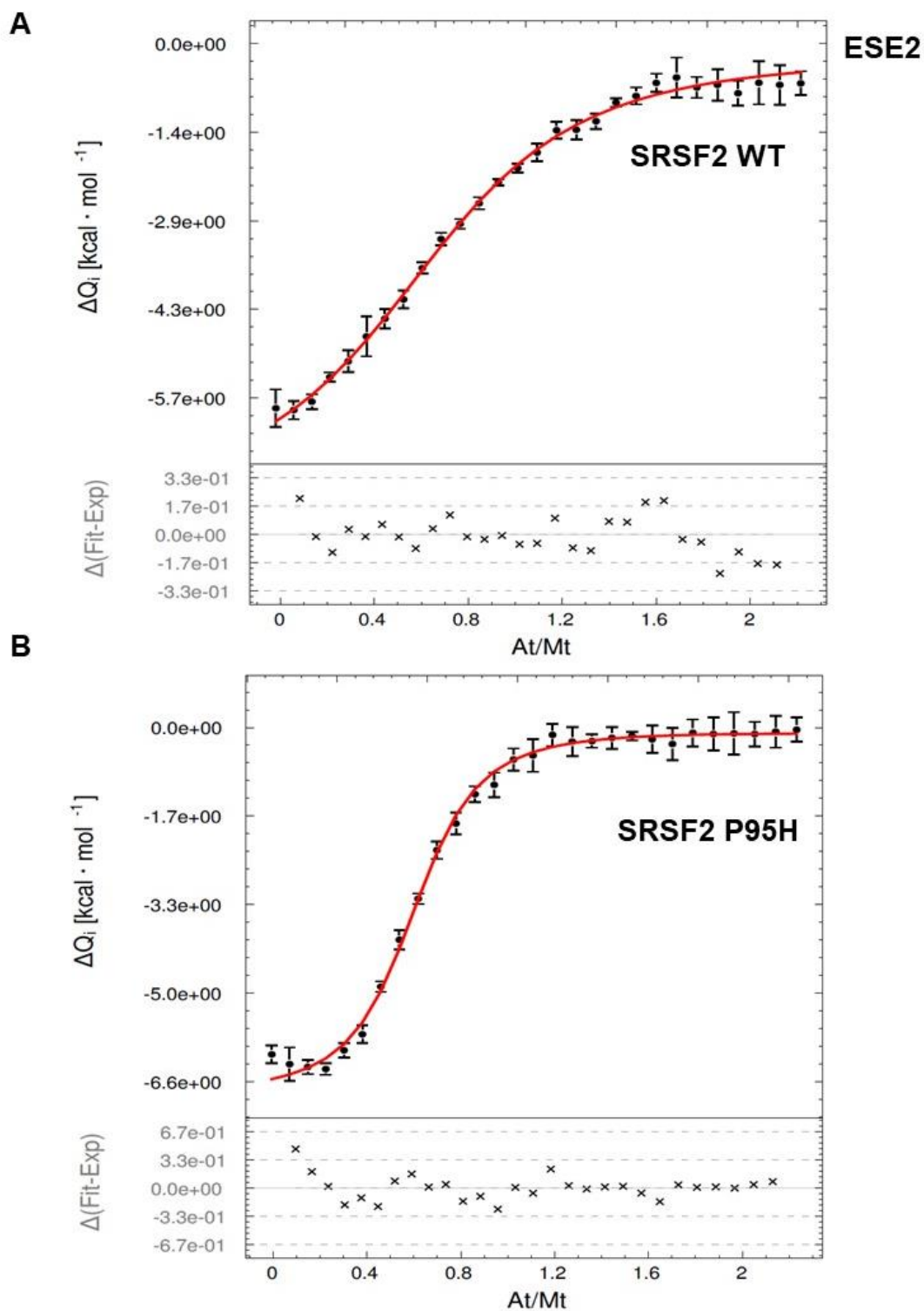
**Figure 3.15. SRSF2 WT and SRSF2 P95H RRM Structure and Purity.** (A)  ${}^1\text{H}$ - ${}^{15}\text{N}$  HSQC spectra with NMR titrations (100  $\mu\text{M}$ ) of  ${}^{15}\text{N}$ -labelled SRSF2 WT (blue) and SRSF2 P95H (red) RRM with unlabelled 5'-UCCAGU-3'. (B) Coomassie Blue staining of purified SRSF2 WT and SRSF2 P95H RRM (10  $\mu\text{g}$ ).

**Table 3.2. Dissociation Values for SRSF2 RRMs Titrated with ESE1 or ESE2.** Dissociation values ( $K_d$ ) from SRSF2 WT and SRSF2 P95H RRMs titrated with ESE1 and ESE2 were calculated from the association values ( $K$ ) given by the Affinimeter software. Values represent the standard deviation from two independent experiments.

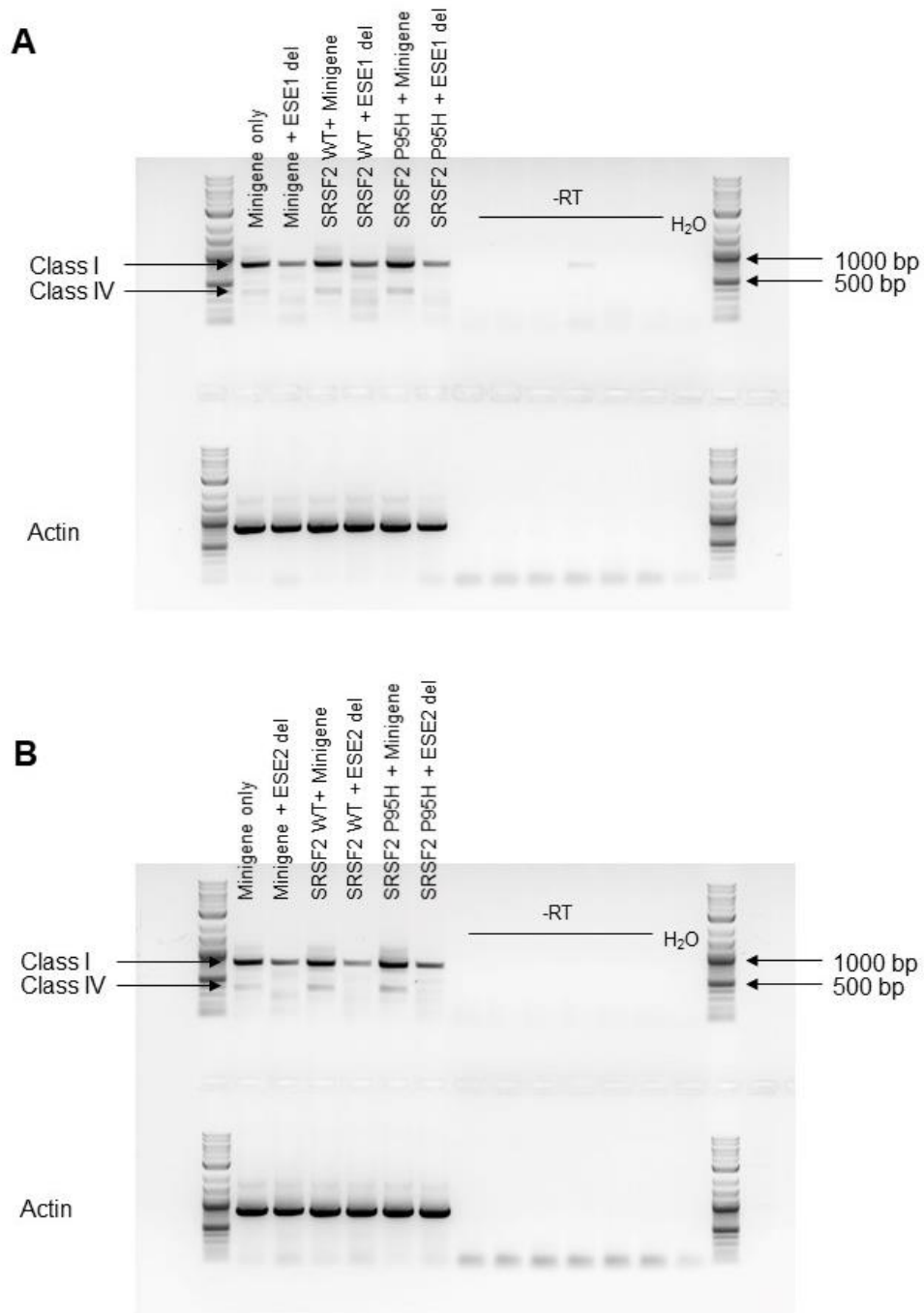
	<b><math>K_d</math> (<math>\mu\text{M}</math>)</b>	<b><math>K_d</math> (<math>\mu\text{M}</math>)</b>
	<b>ESE1 (SRSF2 WT)</b>	<b>ESE1 (SRSF2 P95H)</b>
[1] Free species $\rightleftharpoons M_1A_1$	0.104 $\pm$ 0.038	0.011 $\pm$ 0.002
[2] $M_1A_1 + M_1 \rightleftharpoons M_2A_1$	1.10 $\pm$ 0.505	0.262 $\pm$ 0.427
[3] $M_2A_1 + M_1 \rightleftharpoons M_3A_1$	0.339 $\pm$ 0.011	0.052 $\pm$ 0.102
	<b><math>K_d</math> (<math>\mu\text{M}</math>)</b>	<b><math>K_d</math> (<math>\mu\text{M}</math>)</b>
	<b>ESE2 (SRSF2 WT)</b>	<b>ESE2 (SRSF2 P95H)</b>
[1] Free species $\rightleftharpoons MA$	0.87 $\pm$ 0.13	0.180 $\pm$ 0.01



**Figure 3.16. Global Curve Fit for SRSF2 RRMs Titrated with ESE1 with Residual Plots.** (A) Curve for SRSF2 WT RRM titrated with ESE1. (B) Curve for SRSF2 P95H RRM titrated with ESE1. Graphs are representative of two independent ITC runs with error bars showing standard deviation.

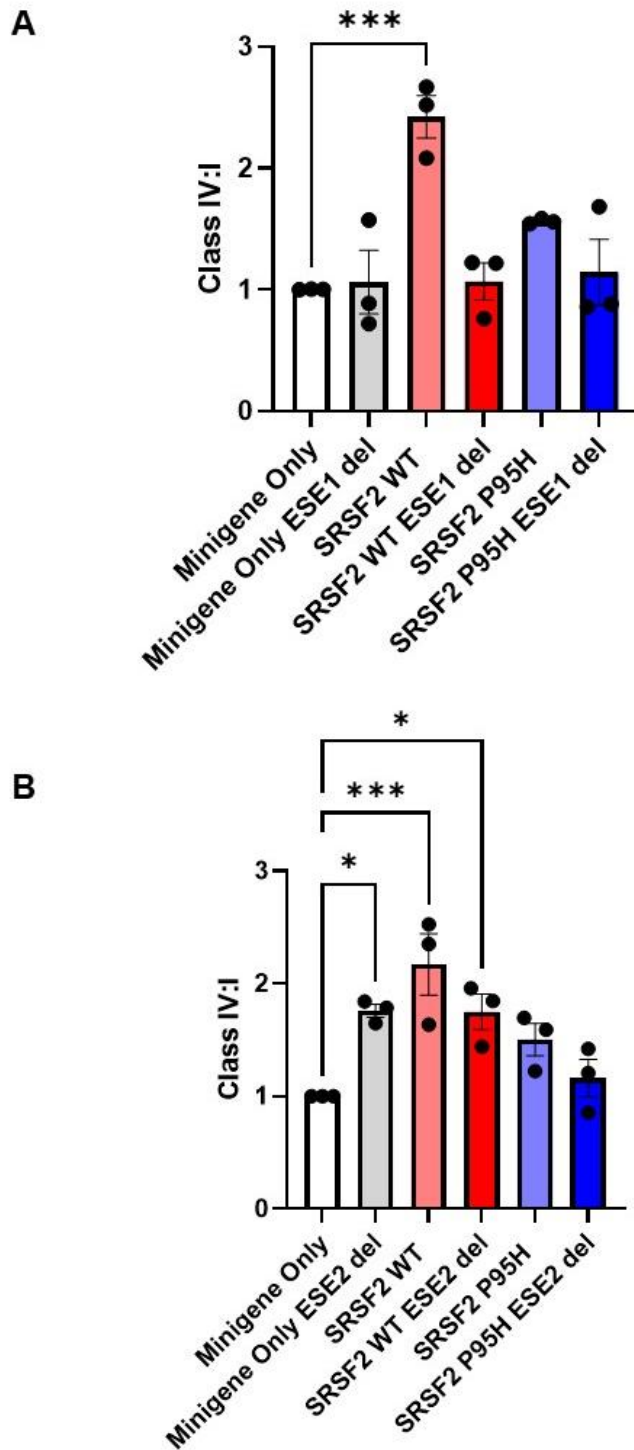


**Figure 3.17. Global Curve Fit for SRSF2 RRMs Titrated with ESE2 with Residual Plots.** (A) Curve for SRSF2 WT RRM titrated with ESE2. (B) Curve for SRSF2 P95H RRM titrated with ESE2. Graphs are representative of two independent ITC runs with error bars showing standard deviation.



**Figure 3.18. ESE1 and ESE2 Deletion in the *CSF3R* Minigene Contribute to Class IV Splicing.** PCR 1% agarose gels of (A) ESE1 and (B) ESE2 deleted motifs in the *CSF3R* minigene showing Class I and Class IV amplicons.

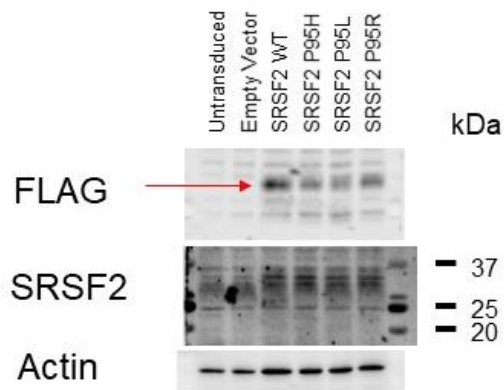




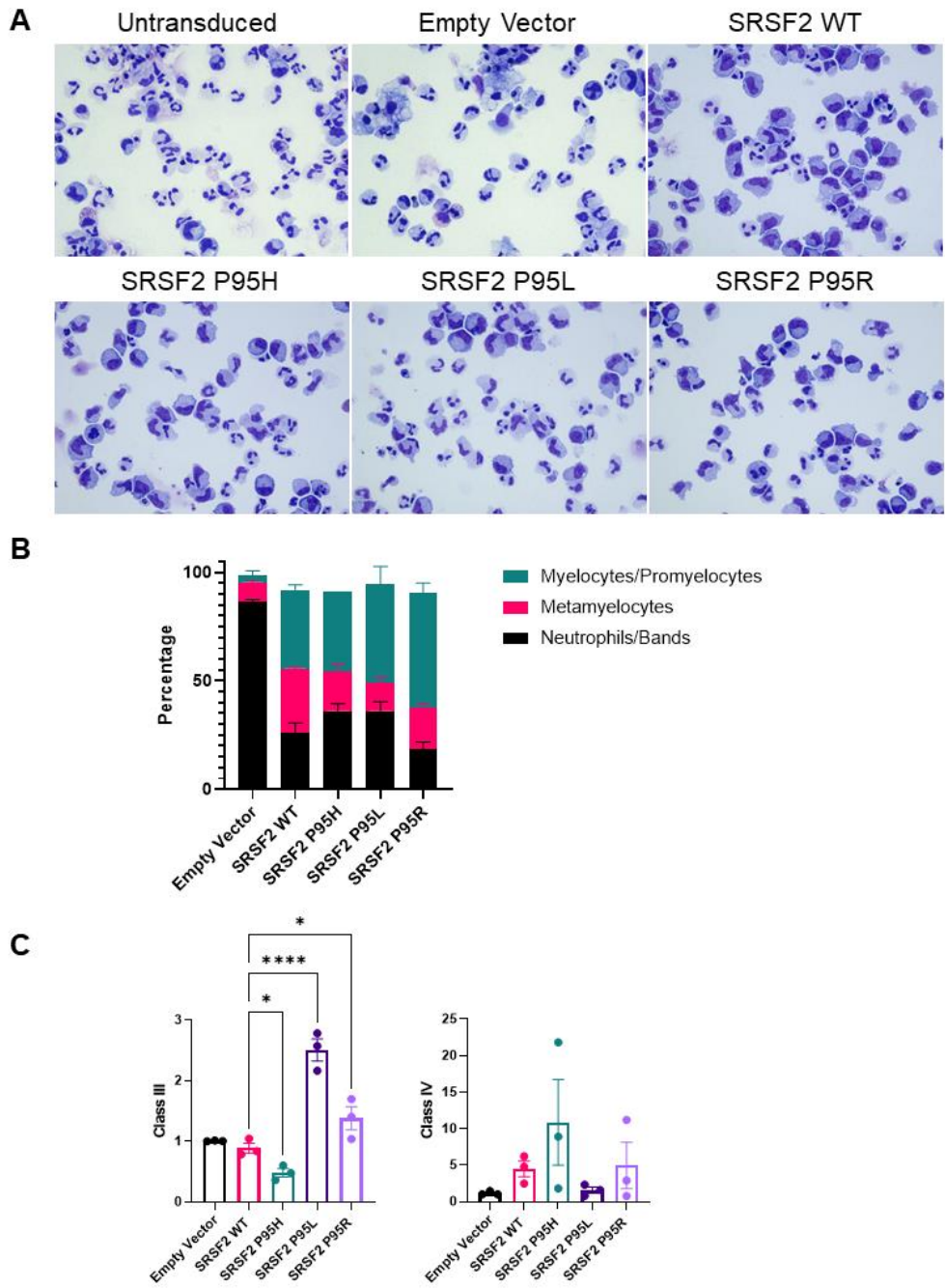
**Figure 3.19. Class IV:I Ratios in ESE-Deleted K562 Cells.** qPCR analysis of Class IV:I in (A) ESE1 and (B) ESE2 deletional mutants in the CSF3R minigene transfected into K562 cells.

### **Overexpression of SRSF2 Lead to Impaired Granulopoiesis in CD34+ Cells**

We next hypothesized that the promotion of Class IV transcripts by intron 17 excision through specific ESE recognition by SRSF2 would lead to impaired differentiation in neutrophilic precursors. To test our hypothesis, CD34+ cells from human umbilical cord blood were transduced with SRSF2 WT, SRSF2 P95H, SRSF2 P95L, or SRSF2 P95R and sorted for +mCherry based on equal MFI (**Appendix A, Figure 12**). Cells were successfully transduced as indicated by expression of FLAG antibody confirmed by western blot (**Figure 3.20**). Cells were terminally differentiated into neutrophils in the presence of G-CSF (10 ng/mL), and morphology was assessed on day 14 by cytopins (**Figure 3.21A**). Approximately 86% of cells transduced with the empty vector terminally differentiated into neutrophils (compared to 74% in untransduced cells). In contrast, only ~26% of cells expressing SRSF2 WT differentiated into neutrophils and 18-36% in the SRSF2 P95 mutants. The greatest phenotypic differences occurred in the promyelocyte and myelocyte differentials with ~3% in empty vector, 37% in SRSF2 WT, and 37-53% in SRSF2 P95 mutants (**Figure 3.21B**). Surprisingly, impaired neutrophil differentiation was also observed in cells with SRSF2 WT overexpression. We then hypothesize that impaired differentiation was a consequence of increased alternate *CSF3R* isoforms. Class III and Class IV relative expression was measured by qPCR where Class III was significantly elevated in cells expressing SRSF2 P95L, whereas Class IV was significantly elevated in cells expressing SRSF2 P95H (**Figure 3.21C**).



**Figure 3.20. Expression of Exogenous SRSF2 in CD34+ Cells.** Western blot showing expression of FLAG-tagged SRSF2 WT, SRSF2 P95H, SRSF2 P95L, and SRSF2 P95R in human CD34+ cells. Lysates were collected after sorting +mCherry cells two days post transduction.

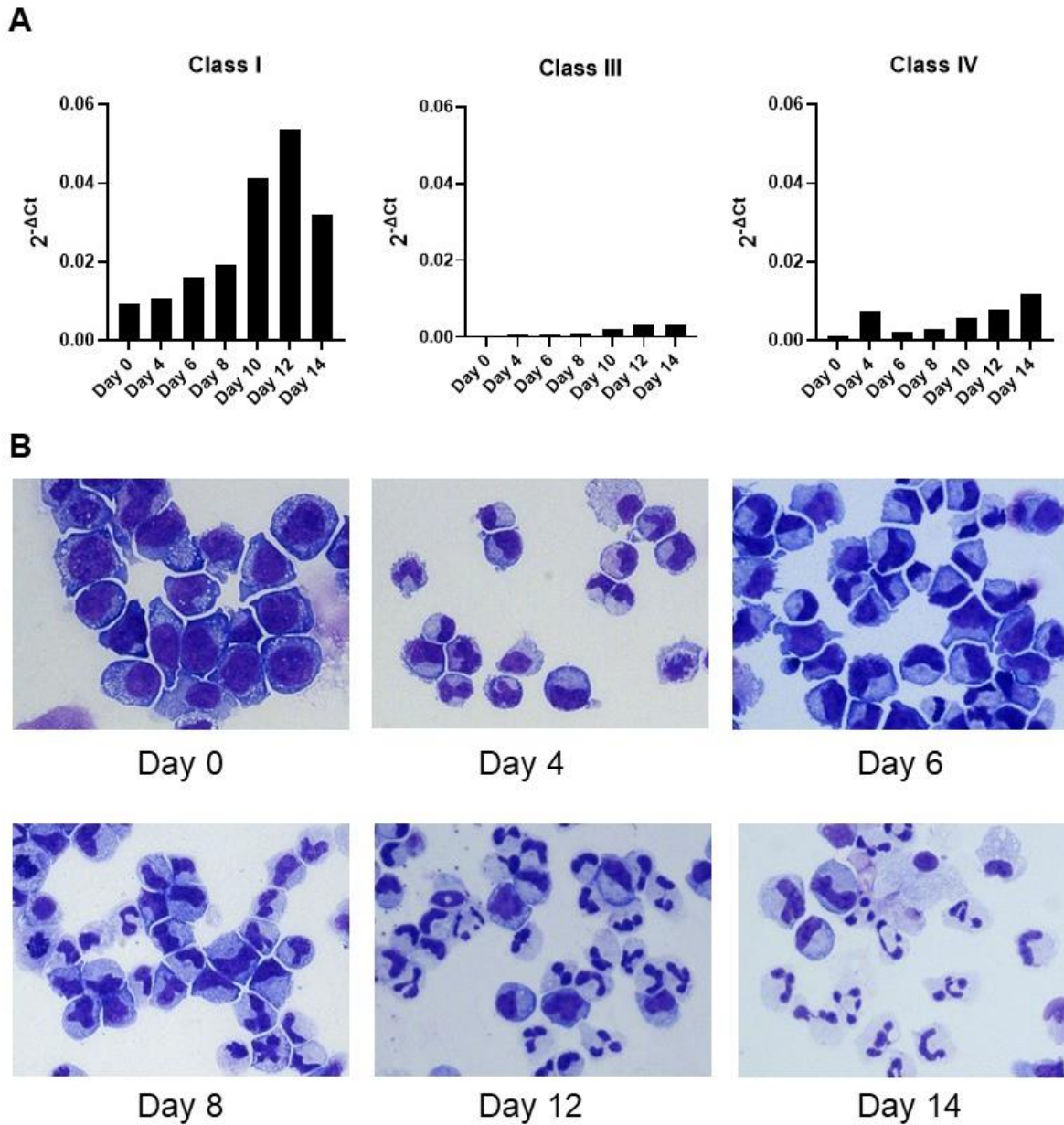


**Figure 3.21. Differentiation of CD34+ Cells with Expression of WT or Mutant SRSF2.** (A) Comparison of neutrophil morphology on day 14 between CD34+ cells transduced with SRSF2 WT, SRSF2 P95H, SRSF2 P95L, or SRSF2 P95R treated with 10 ng/mL G-CSF. (B) Neutrophil differentials from day 14. (C) Class III and Class IV expression on day 10. Error bars represent the mean with standard error of the mean from three independent experiments corrected for multiple comparisons by controlling the False Discovery Rate. \* $p \leq 0.0332$ , \*\*\*\* $p \leq 0.0001$ .

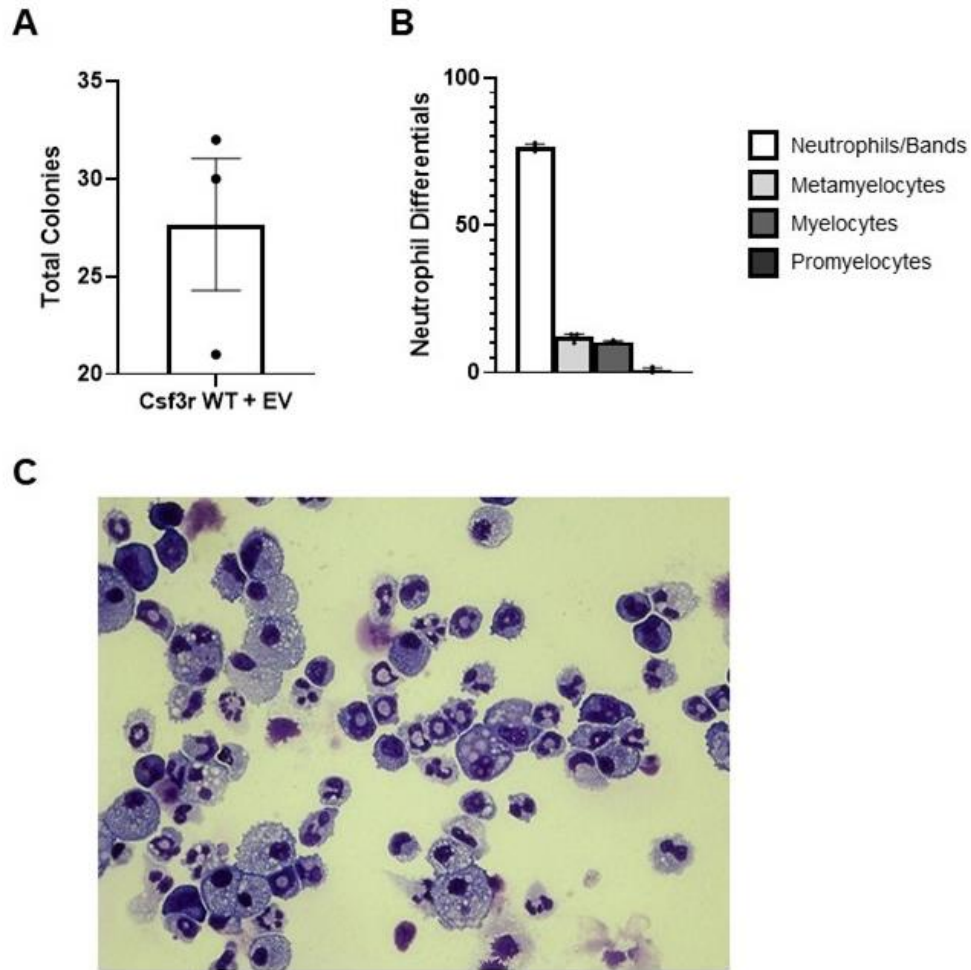
## **Alternative CSF3R Isoforms Do Not Rescue Granulopoiesis in Csf3r-Null Progenitor Cells in Response to Granulocytic Differentiating Factors**

Class I, Class III, and Class IV CSF3R are present in mixtures of heterodimers and homodimers in neutrophils. Neutrophils have a Class IV:I ratio of  $0.037 \pm 0.005$ , where Class IV homodimers represent 0.1% of all CSF3R dimers, 7.1% are Class IV:I heterodimers, and 92.8% are Class I homodimers. In normal CD34+ cells Class IV:I ratios are  $0.065 \pm 0.009$ , where Class IV homodimers represent 0.4% of all CSF3R dimers, 12.2% represent Class IV:I heterodimers, and 87.4% are Class I homodimers (34). Neutrophil precursors exhibit more Class IV and increase as cells differentiate. Similarly, Class III expression is also developmentally regulated in CD34+ cells and is the least expressed isoform (**Figure 3.22A and B**). In comparison to normal CD34+ cells, there is an overall 2.6-fold increase in Class IV expression in AML cells and a 2 to 3-fold increase in the percentage of heterodimers and a 5 to 20-fold increase in the percentage of Class IV homodimers. While Class I and IV CSF3R are the most studied isoforms, functional *ex vivo* studies have not been conducted to assess the function of individual isoforms in isolation. To investigate the impact of each isoform on granulopoiesis, LSK cells from *Csf3r*-null mice were harvested from the bone marrow and transduced with either empty vector control, human Class I, Class III, or Class IV. Cells were sorted for +Venus after 3 days of transduction and then cultured for 1 day in StemSpan SFEM II with murine SCF, human IL-11L, murine IL-3, murine Flt3L, and 1% penicillin/streptomycin. Five different media conditions were tested for LSK cell growth and viability. We found that the media containing the combination of cytokines listed above proliferated the fastest while maintaining >85% viability (**Appendix A, Figure 13**). Proliferation and differentiation ability of cells with a single CSF3R isoform were assessed by plating 100 cells onto MethoCult containing murine SCF, murine IL-3, murine IL-6, and human G-CSF. In a pilot experiment in which cells were grown in Methocult with G-CSF alone, add-back of human Class I showed some neutrophilic differentiation on day 8 but failed to form colonies (**Appendix A, Figure 14**). Cells were assessed on day 8 in which whole plates were morphologically analyzed

by cytopins and colonies were scored for CFU-G, CFU-GM, and CFU-M. Representative images of CFUs are shown in **Appendix A, Figure 15**. On average, cells from *Csf3r* WT mice had 57% CFU-GM, 28% CFU-M, and 14% CFU-G (**Figure 3.23A-C**). Colony scores show that Class I and Class IV-expressing cells produced on average a 3.5 to 4-fold increase in total number of colonies compared to empty vector control. In contrast, cells expressing Class III gave rise to a 2.2-fold decrease in total colonies compared to cells expressing Class I with no significant differences compared to empty vector control. This observation was expected given its hypoproliferative phenotype. We also expected higher colony numbers in Class IV expressing cells since it confers a hyperproliferative phenotype, which was observed when colony numbers were compared with Class I-expressing cells (**Figure 3.24A**). These cells showed a significant increase in CFU-Gs compared to Class III and empty vector control based on colony scores and flow cytometry (**Figure 3.25**). No differences were seen in CFU-M scores. CFU-GM were the most abundant colony type across Class I, Class III, and Class IV with Class I having the greatest number of colonies compared to empty vector (4.6-fold), Class III (2.5-fold) and Class IV (1.5-fold) (**Figure 3.24B**). Despite the absence of *Csf3r* in the empty vector control, terminal differentiation of neutrophils was still observed in the presence of these cytokines. As expected, Class I cells terminally differentiated into neutrophils with normal morphology, as depicted by the black arrows. Class III cells showed severely dysmorphic neutrophils with hyposegmented and pyknotic nuclei, as indicated by the blue arrows. Class IV cells showed a differentiation arrest at the metamyelocyte stage with normal morphology, as indicated by the red arrows (**Figure 3.24C**). Class I expressing cells had on average 88.7% cells that differentiated into bands/neutrophils, 6.3% metamyelocytes, and 5% myelocytes. Class III expressing cells showed 13.7% bands/neutrophils, 12.3% metamyelocytes, 32.7% myelocytes, and 41.3% promyelocytes. Class IV expressing cells showed 26.7% bands/neutrophils, 43.3% metamyelocytes, 23% myelocytes, and 7% promyelocytes. (**Figure 3.24D**). Interestingly, the Class IV cells had comparable cell differentials as the empty vector control.

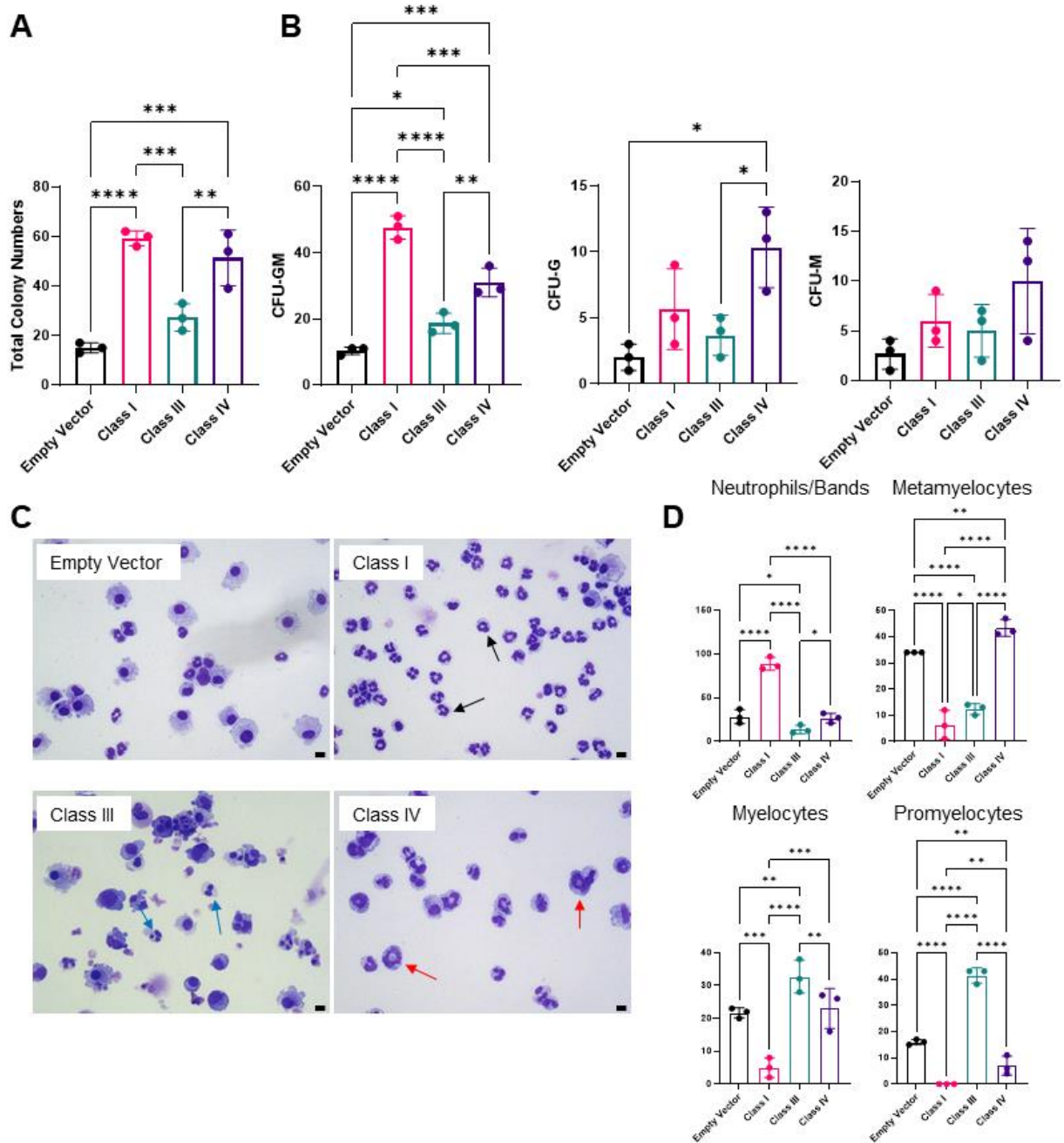


**Figure 3.22. CSF3R is Developmentally Regulated in Human CD34<sup>+</sup> Cells.** (A) *CSF3R* Class I, III, and IV mRNA expression in CD34<sup>+</sup> cells increase as cells differentiate into neutrophils. CD34<sup>+</sup> cells isolated from cord blood were cultured in StemSpan SFEM II supplemented with 10% FBS, 1% Pen/Strep, 100 ng/mL human SCF, 10 ng/mL human IL-3, 10 ng/mL human TPO, and 10 ng/mL human G-CSF on days 0-7. Cells were subsequently cultured in StemSpan SFEM II supplemented with 10% FBS, 1% Pen/Strep, and 10 ng/mL human G-CSF for days 7-14. cDNA was synthesized from RNA collected on days 0, 4, 6, 8, 10, 12, and 14 to measure Class I, III, and IV expression. (B) CD34<sup>+</sup> cells terminally differentiate into neutrophils by day 14 after G-CSF stimulation. Representative cytopsin images corresponding to the day in culture are shown.

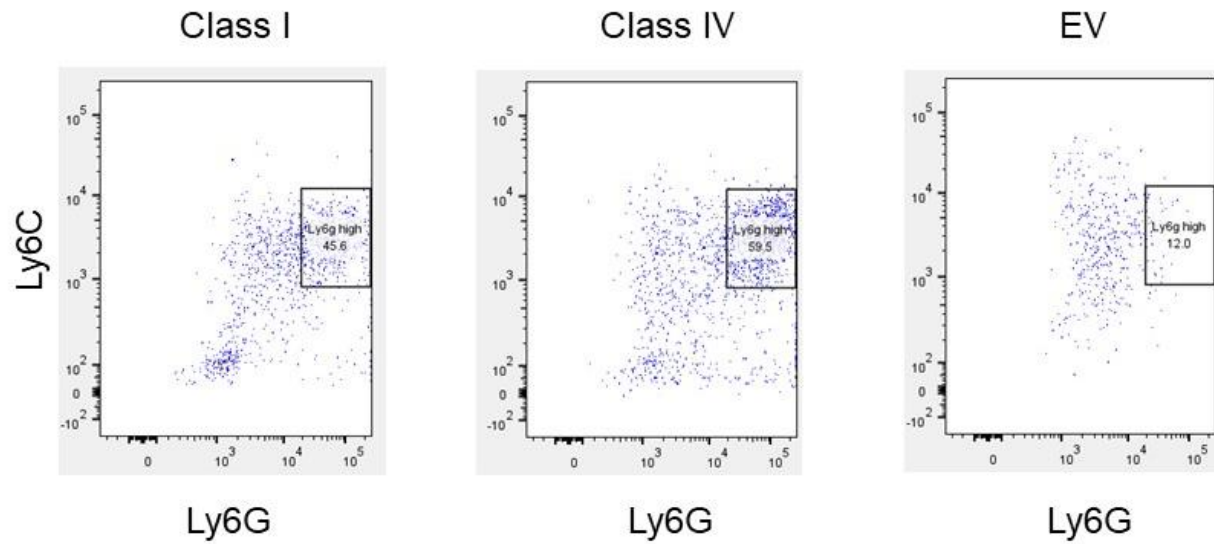


**Figure 3.23. Colony Scores from *Csf3r* WT LSK Cells.** (A) Total colony numbers of *Csf3r* WT LSK cells plated on Methocult M3534 supplemented with 100 ng/mL G-CSF were counted on day 8. (B) Neutrophil differential counts of *Csf3r* WT LSK cells on day 8. (C) Cytospin of whole colony assay plates on day 8.





**Figure 3.24. Comparison of *CSF3R* Isoforms and their Contributions to Colony Growth.** (A) Total colony counts and (B) colony scores from colony assays of LSK cells from *Csf3r*-null mice transduced with empty vector control, Class I, or Class IV and treated with G-CSF. Cells were plated in Methocult 3534 with 100 ng/mL G-CSF and colonies were scored on day 8. (C) Cytopins of total cells (empty vector, Class I, Class III, and Class IV) from CFU-G/M/GM assays on day 8 supplemented with 100 ng/mL G-CSF. Error bars represent the mean with standard deviation from three biological experiments. (D) Day 8 neutrophil differentials of *Csf3r*-null cells transduced with empty vector, Class I, Class III, or Class IV. Error bars represent the standard error of means from three independent experiments corrected for multiple comparisons by controlling the False Discovery Rate. \* $p \leq 0.0332$ , \*\* $p \leq 0.0021$ , \*\*\* $p \leq 0.0002$ , \*\*\*\* $p \leq 0.0001$ .



**Figure 3.25. Flow Cytometric Analysis for CD11b<sup>+</sup>CD11c<sup>-</sup> Ly6C<sup>low</sup>Ly6G<sup>+</sup> (Neutrophils) and CD11b<sup>+</sup>CD11c<sup>-</sup> Ly6C<sup>+</sup>Ly6G<sup>-</sup> (Monocytes).** Analysis of whole-plate CFU assays were done on day 8. Cells were stained with CD11b, CD11c, Ly-6G, and Ly-6C and the analysis was performed on FlowJo.

### 3.4 Discussion

As we reach the molecular age of MDS where we now have access to more sophisticated methods of sequencing and new discoveries in gene mutations, their cooperativity, timing of mutations, and single cell resolution of subclonal populations, incorporating mutational profiles in conjunction with the current standards for MDS stratification will be a critical component in providing risk and disease management for patients to better inform therapeutic strategies. These advancements will be more relevant in patients with low to intermediate risk MDS where there are more mutations and unfavorable somatic mutations. Currently, the only criteria for trial eligibility and clinical decision making are based on cytogenetic variables. Splicing factor SF3B1 and del(5q) were only recently added to the WHO classification. While there are challenges to incorporating new information into the established risk stratification models, the need for it is warranted and studies that have studied splicing factor mutations in depth have provided strong evidence for its association with patient prognosis. A future of hybrid genomics and clinical models will be paramount for accurately performing risk prediction models.

Since the discovery of splicing factor mutations in the past decade, hundreds of altered splicing events have been reported in MDS patients, murine, and human cell models. Despite these efforts, there have been little progress in elucidating a mechanism critical to disease pathogenesis. Splicing analyses all support findings that mutant splicing factors have altered functions that result in missplicing of target transcripts that can lead to improper function of resulting proteins. The change of function usually occurs as a result of a change in splice site recognition. However, many of the specific target transcripts that undergo these changes due to mutated splicing factors still need further investigation. For example, EZH2 is one example of a gene that has been well studied in the context of SRSF2 P95H where it induces a poison cassette exon that facilitates NMD. Given the global effects of splicing, it is surprising that only a few spliceosomal genes affect such a wide array of genes, which is a testament to how little we know

about how splicing regulation and mechanisms. However, the fact that a small list of splicing factor mutations can cause widespread abnormalities means that we can take advantage of the data provided by numerous transcriptome studies. Future studies will need to investigate specific gene targets and analyze how splice isoforms production is unbalanced. Not all consequences of missplicing results in NMD or a novel transcript. *CSF3R* is a good case in point, where splicing factor mutations alter the transcript abundance of certain isoforms that have implications in mildly driving diseases like MDS.

Although transcriptome analyses in patients have resulted in a wealth of information about the genomics of myeloid neoplasms, there is little overlap between the gene expression and missplicing events across different studies. One reason for this discrepancy may be due in part to the heterogeneous nature of co-occurring mutations and variable clonal composition. In addition, bulk-sequencing of RNA of whole bone marrow might dilute important splicing events like those seen in *CSF3R*, which are enriched in CD34+ cells and granulocytes. Nevertheless, combining genomic data from patients with specific transcript targets will be a useful tool in investigating a small, but impactful portion of the molecular mechanisms driving disease.

Given the importance of *CSF3R* in its role in hematopoiesis and regulation of neutrophil development, it is not surprising that its isoforms are altered under the conditions of mutated splicing factors. Class IV is the most studied alternatively spliced *CSF3R* isoform since its discovery and numerous studies have consistently shown that it is elevated in MDS and AML patients. For this reason, *CSF3R* provides a good model for not only understanding how it is differentially spliced by mutated splicing factors, but also how unbalanced isoform expression can contribute to disease progression. While these effects may be weak drivers for disease, accumulation of these isoform imbalances, which are usually tightly regulated for fine-tuning of neutrophil development, can be critical for disease pathogenesis. Furthermore, splicing effects are different depending on what splicing factors are mutated. Evidence of this observation was

shown in the minigene splicing assay where the splicing factors that affected Class IV the most were U2AF1, SRSF2, and LUC7L2. Not only do different splicing factors have different effects on *CSF3R* splicing, but different mutations in the same splicing factor can have opposing effects on splicing. We observed this phenomenon in U2AF1 S34F and Q157P mutations, lending more evidence to the fact that mutations need to be analyzed with more granularity to understand the scope of the contribution of splicing mechanisms to MDS. Our minigene data also suggests that splicing factor mutations have a stronger influence on Class IV splicing over Class III splicing, even though patient data suggests otherwise. To emphasize an earlier point, patients with myeloid neoplasms have numerous co-mutations and varying mutation loads that all contribute to splicing in a wide degree of ways that may not be fully recapitulated in our cell lines. The strength of our minigene assay comes from the fact that whatever observations were made about *CSF3R* splicing came from a direct relationship between the splicing factor mutation and the splicing events. We acknowledge that this relationship is simplified, and that alternate *CSF3R* isoform usage is not the only driver of disease, but rather a small part of an immensely complex network of gene regulation.

While the heterogeneity of patients exists, the fact that we were able to consistently see increased proportions of Class III and Class IV indicates that there is a strong correlation between splicing factor mutations and the production of alternate *CSF3R* isoforms, even in the background of numerous co-mutations and mutational burden. Throughout this dissertation, we measured Class III and Class IV expression using various methods. In the minigene splicing assay, we represented isoform expression as ratios for two reasons. The first reason is that K562 cells, being an erythroleukemia cell line, do not express *CSF3R* in enough quantities to be detected by qPCR. Therefore, ratios of either Class III or Class IV relative to Class I were used because the values were being measured from the same transcript. This method eliminates the need to compare expression to a housekeeping gene such as Actin because the expression values will

vary depending on the transfection efficiency. The second reason is that *CSF3R* isoforms exist in heterodimers and homodimers. Class IV:I heterodimers increase about 3-fold in AML patients compared to normal CD34+ cells. In the analysis of the patient data, Class I, Class III, and Class IV are represented as proportions. Since genes can produce a finite proportion of transcripts, we hypothesized that increased proportions of one isoform would mean decreased proportions of the canonical isoform (Class I). Indeed, this seemed to be the case in most patients with splicing factor mutations compared to patients without splicing factor mutations. Decrease of the canonical isoform in favor of producing more of the alternate isoforms might suggest that certain individuals with high expression of Class III or Class IV are more susceptible to leukemic transformation or acquiring additional mutations.

More in-depth investigations of the mechanisms driving *CSF3R* alternative splicing by U2AF1, SF3B1, and LUC7L2 are warranted, but outside the scope of this thesis. We chose to study how SRSF2 regulates *CSF3R* splicing because of convincing reports that show its involvement in promoting Class III and Class IV. In a separate study, mice with heterozygous P95H mutations showed decreased expression of *Csf3r* in conjunction with dysgranulopoiesis. SRSF2 is an essential protein that is involved in both constitutive and alternative splicing. SRSF2 functions to excise or retain exons by binding to ESEs through its RRM. Its RS rich domain also gets heavily phosphorylated to regulate splicing and has roles in spliceosome formation. Studies to date have only shown association of SRSF2 P95H mutations and increased relative Class III and Class IV expression but have not offered a mechanistic explanation for this observation. We used ESEfinder to locate all of the putative ESE motifs within *CSF3R* exon 17. One striking observation was that a large percentage of motifs were pyrimidine-rich that would recruit SRSF2 P95H to promote splicing of Class IV. The highest scoring motifs were also located in the region that gets spliced out to produce Class IV. The ITC studies did confirm that these motifs are important in splicing of *CSF3R* and that the RRM of SRSF2 does directly bind to these two motifs. Binding

affinities show that SRSF2 WT binds to ESE1 with greater affinity than SRSF2 P95H. In contrast, SRSF2 WT binds to ESE2 with less affinity than SRSF2 P95H. A simple interpretation of these dissociation values would suggest that ESE2, which has the higher motif score, recruits SRSF2 P95H with higher affinity than SRSF2 WT to splice out the intron promoting Class IV. However, the more likely explanation is a highly dynamic process where SRSF2 WT and SRSF2 P95H are in constant competition over binding to ESE1 and ESE2 (**Figure 3.26**). SRSF2 P95H outcompetes SRSF2 WT in binding to these motifs in most instances, leading to Class IV splicing. The possibility of other ESE motifs playing important roles in *CSF3R* splicing is still possible, given that SRSF2 binds to degenerate sequences. One of the caveats about the ESE scores is that there is only at best a rough correlation between scores and ESE activity. In other words, the maximum score does not necessarily mean the most effective ESE. Taking this into account, the higher scoring ESE2 motif does not necessarily mean that it is more effective than ESE1 or any of the other lower scoring motifs. It is possible that the pyrimidine-rich motifs scattered throughout the entire exon 17 collectively promote Class IV splicing. Future experiments to add back a purine-rich sequence in place of the pyrimidine-rich sequence will further validate the preferential binding of SRSF2 P95H to these motifs. Nonetheless, these studies still show that modulating alternative splicing by inhibiting SRSF2 binding to ESE1 or ESE2 could be a therapeutic approach.

Having established that SRSF2 P95H is increased in patients with myeloid neoplasms and that Class IV splicing is increased through the recognition of ESE1 and ESE2, we hypothesized that the prevalence of Class IV would lead to impaired neutrophil differentiation, thus contributing to disease progression. Expression of SRSF2 P95 mutations in CD34+ cells did impair neutrophil differentiation to a similar degree. Phenotypic assessment alone would suggest that all SRSF2 P95 mutations function in the same manner with similar downstream consequences. Relative expression of Class IV only indicates a significant increase in cells with SRSF2 P95H. SRSF2 P95L and SRSF2 P95R expressing cells both had similar levels of Class IV compared to the

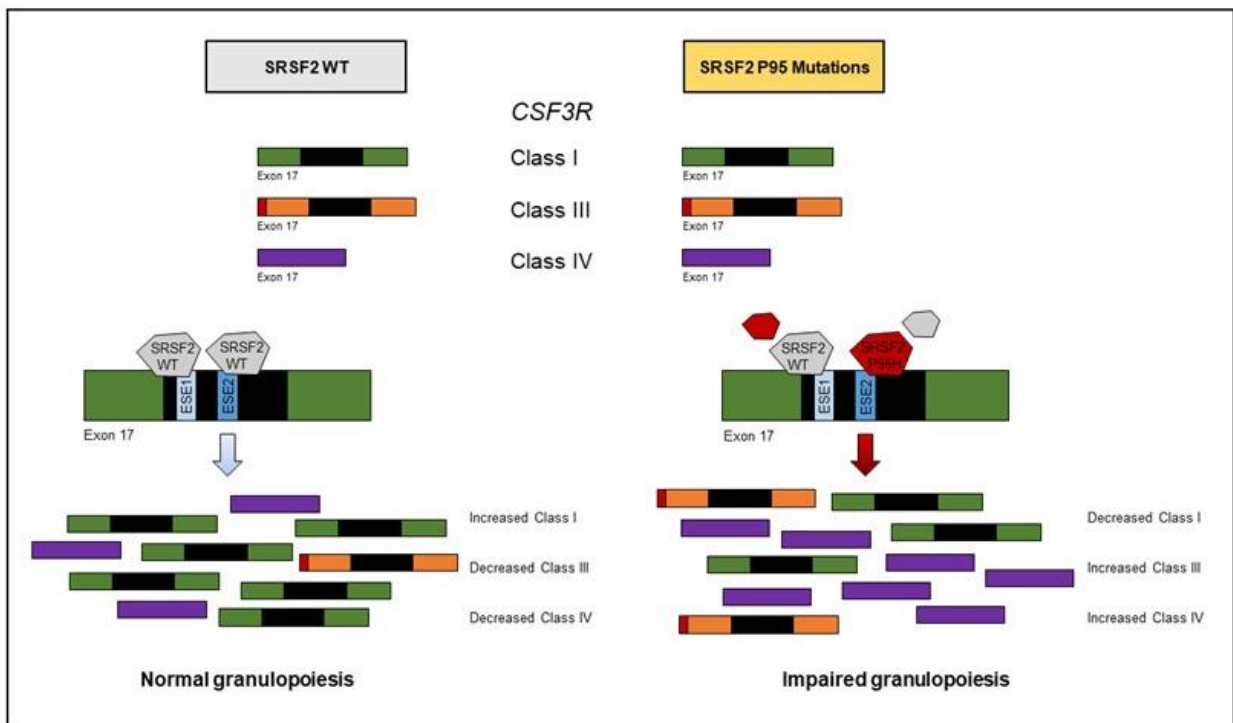
empty vector and SRSF2 WT. Class III showed the highest level in cells with SRSF2 P95L. These results call into question if different amino acid substitutions at the same residue also confer different functions. No studies to date have performed binding studies with the RRM of SRSF2 P95L and SRSF2 P95R, but they will be paramount in delineating possible minute differences between the P95 mutations. We might also expect a corresponding severity in the degree of differentiation impairment with the degree of Class IV splicing. Our differential results indicate that more Class IV does not correlate with the degree of maturation delay. Future studies to investigate the correlation of variant allele frequencies of SRSF2 mutations with Class III and Class IV might offer additional insight into the effect of mutational burden on *CSF3R* splicing. Since ESEs can be saturated with a finite number of SRSF2 protein, simply the presence of a mutated protein might be enough to outcompete the wildtype protein to significantly alter splicing.

Our add-back studies in LSK cells from mouse lacking *Csf3r* describe a Class III morphological phenotype that has never been reported. The tyrosine residues at the C-terminal end of CSF3R are crucial in providing the proper maturational signaling cascade for neutrophil development. It is expected that Class IV would confer a differentiation defective phenotype due to the ablation of the three of the tyrosines. Immature cells also proliferate faster, which is also seen in cells with increased Class IV. For Class III to present a more severe phenotype than Class IV implies that the additional 27 amino acid insertion between box 1 and box 2 renders the signaling cascade almost entirely ineffective at conferring the maturation signals needed for proper neutrophil development. While humans will always have a mixture of all three isoforms, future studies involving the importance of the frameshift caused by insertion of a partial intron 16, will be valuable from a mechanistic standpoint.

In conclusion, we have used various methods to describe for the first time the mechanism by which *CSF3R* is alternatively spliced by SRSF2. In addition, we also characterize the phenotype of Class III in a mouse model lacking *Csf3r*. The molecular mechanism driving MDS pathogenesis



is complex and dynamic, but it is known that splicing factor mutations cause aberrant splicing in many genes involved in hematopoiesis. Since they are founder lesions, targeting splicing factor mutations has become an attractive therapeutic target for modulating splicing. Modulating splicing can have important implications for disease management because splice isoforms of critical genes can contribute to disease progression by reducing the production of canonical transcripts. Unbalanced isoform usage from dysregulated splicing therefore plays a small but significant role in disease progression where its effects are compounded by co-occurring mutations. Advances in genomic information gained from sequencing will be instrumental in understanding MDS pathogenesis and better risk stratification for patients. Combining genomics with splicing of specific genes will provide better granularity of disease mechanisms that can inform new therapeutic strategies.



**Figure 3.26. Proposed Model of SRSF2 WT and SRSF2 P95H Binding Competition to ESEs.**

## Chapter 4: Perspectives and Future Directions

### Cell Line Models of SRSF2 P95 Mutations

SRSF2 regulation of *CSF3R* splicing was assessed in cell lines and through *ex vivo* differentiation of mouse LSK and human CD34+ cells. While minigenes are useful tools in examining splicing mechanisms in isolated conditions, artificial constructs do not mimic patient samples, even when these studies are conducted in other leukemia cell lines. For example, we measured endogenous Class IV:I ratios in U937 and NB4 cells but did not see any splicing changes in these cell lines even though SRSF2 P95 mutations were overexpressed. Splicing factor mutations all rely on a wildtype copy to function. To mimic this feature in cell lines, future studies could benefit from CRISPR/Cas9-mediated gene editing to create monoallelic SRSF2 mutations in human and mouse primary cells. Another approach to study the effect of SRSF2 P95 on *CSF3R* splicing is to knockdown SRSF2 WT and add-back SRSF2 P95H/L/R. The wildtype protein levels would function simply to maintain cell viability and the predominate expression of mutant SRSF2 may provide additional perspectives on Class III and Class IV splicing. It is still unclear if varying degrees of SRSF2 expression affect splicing in a dose-dependent manner. The add-back model would be one method of testing this hypothesis by varying the expression in cell lines. The same patient cohort used for the analysis of Class I, Class III, and Class IV isoforms can be further stratified by VAFs of SRSF2. Correlation of VAFs and Class III or Class IV expression in SRSF2-mutated MDS be measured through Pearson's correlation coefficient. One other important consideration is the splicing influence on Class IV by overexpression of SRSF2 WT. While puzzling, it is plausible to assume that overexpression of the wildtype protein affects Class IV splicing in a different mechanism than SRSF2 P95H. Knockdown studies may provide additional insights into how dosage effects influence Class IV splicing. In addition, patients without splicing factor mutations who show high Class III or Class IV can be assessed for SRSF2 expression. If overexpression of SRSF2 WT is seen in these patients, then elements that regulate

its expression, such as serine/arginine specific protein kinases (SRPKs) may be worth investigating.

### ***In Vivo* Models of Class III and Class IV on Granulopoiesis**

We were able to show in our add-back studies in murine LSK cells that expression of a single *CSF3R* isoform confers different granulocytic phenotypes. If *CSF3R* signaling is conserved in mice as it is in humans, future *in vivo* experiments would include transducing murine LSK cells with Class I, Class III, and Class IV and transplanting the cells into lethally irradiated *Csf3r*-null mice. Complete blood counts and bone marrow cytopspins would be assessed for differences in neutrophil differentiation upon G-CSF treatment.

### **The Role of LUC7L2 on *CSF3R* Splicing**

LUC7L2 has recently been characterized as a splicing factor that recognizes the 5'SS of mRNA that plays a role in cassette exon or intron splicing depending on the cellular context (86). Its role in MDS pathogenesis is still elusive, but another recent study reported that loss of LUC7L2 decreases genes related to glycolysis (122) and patients with low LUC7L2 expression showed missplicing in several mitochondrial genes (102). LUC7L2 and SRSF2, both SR proteins, showed increased splicing of Class IV in transfected cell lines and in patients with LUC7L2 haploinsufficiency and SRSF2 mutations. Patients harboring these abnormalities also have worse prognosis compared to other mutations. The top targets of missplicing from LUC7L2 deficiency include SMC5, ATG4B, TRA2A, IKZF1, SMEK1, and DYRK1A (unpublished data). Given its role in regulating many pathways, efforts to investigate how aberrant splicing of specific genes contribute to specific pathways leading to dysfunctional hematopoiesis are warranted.

### **Implications for Splicing Modulators as Therapeutics**

Advances in understanding MDS pathogenesis coupled with genomic sequencing landscapes have highlighted splicing factor mutations as a promising therapeutic target. Immunotherapies currently used to treat MDS are effective, but their use is limited to distinct subtypes with different

mutational characteristics. Current treatment goals for MDS are to improve peripheral blood counts and delay the progression of disease and onset of subsequent cancers. Treatment options vary depending on the level of risk. Disease management for patients with low-risk MDS include improving quality of life and preventing complications such as severe infections with supportive care. Patients in the high-risk category are put on regimens that delay disease progression and considered for stem cell transplant if suitable. Low intensity therapies consist of hypomethylating agents, immunosuppressive therapy, and immunomodulatory drugs. High intensity therapies include chemotherapy and hematopoietic stem cell transplantation. The highly heterogeneous nature of MDS makes treatment decisions challenging. At the same time, many potential targets can be developed for new therapeutics. The number of FDA approved drugs for MDS remains limited and not all agents have been shown to be effective in improving survival. Aside from bone marrow transplant, these therapies are not curative, and patients usually have to undergo additional treatment for secondary diseases.

#### *Splicing Modulators Targeting SF3B1-Mutated Cells*

One of the challenges in developing new therapies specific to MDS is the limitation of available preclinical *in vivo* mouse models. A greater understanding of the genetic landscape in MDS within the last decade has helped to overcome this barrier. Since splicing defects are one of the most common abnormalities in MDS, extensive efforts to target splicing factor mutations have been undertaken. In one study, murine models with SRSF2 or U2AF1 heterozygous mutations have been used to test the spliceosome inhibitor E7101. Mice with SRSF2<sup>P95H/+</sup> or U2AF1<sup>S34F/+</sup> heterozygous mice were found to have prolonged survival after exposure to the compound (123). Similarly, treatment with sudemycin compounds, which bind to SF3B1, in hematopoietic cells expressing U2AF1<sup>S34F/+</sup> reduced the expansion of U2AF1 S34F progenitor cells in a mouse transplant model (124). These studies have led to the hypothesis that there may be a degree of tolerability in splicing dysfunction within a cell, but further perturbances result in selective cell

death. E7107 was also the first spliceosome modulator to be tested in a phase I study. Even though the patients enrolled had solid tumors, investigators did observe a dose dependent change in mRNA splicing in cells from E7101 treated patients (125, 126).

Another compound that was tested in a phase I clinical trial is H3B-8800, which is an orally available modulator of SF3B1 that is molecularly distinct from E7101 but shares a pladienolide chemical backbone. Preclinical studies showed it was selective against SF3B1 mutated leukemia in xenograft models. Patients enrolled in this study were diagnosed with AML, MDS, and CMML (127). Results from the study showed that although no objective CR or PR were achieved, decreased RBC or platelet transfusion requirements were observed in 14% of enrolled patients (128). Other splicing modulators are under development but haven't been used in humans yet are sudemycin D6, Pladienolide-B, and FD-895.

#### *Splice-Switch Oligonucleotides*

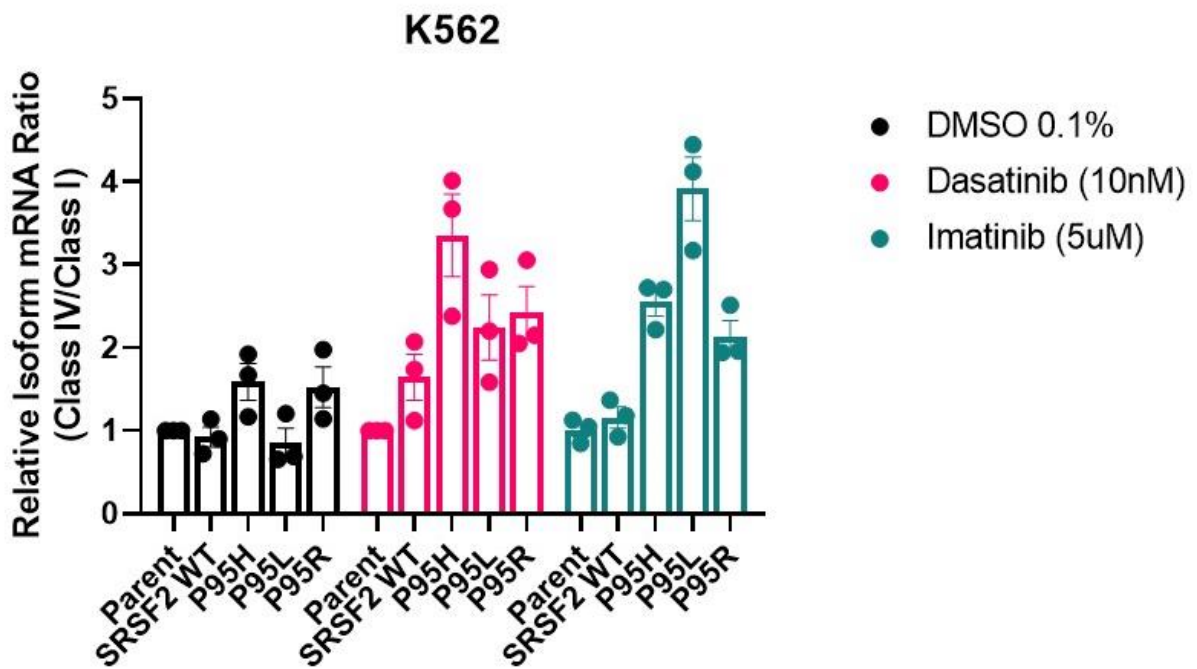
Splice-switch oligonucleotides (SSOs) are antisense oligonucleotides that are 15-30 nt in length that bind to splice sites or regulatory sequences like ESEs, ESSs, ISEs, and ISSs. Binding of SSOs to the RNA prevents splicing factors from physically accessing regulatory sites. For example, SSOs can be used to block cryptic splice sites to prevent the spliceosome from using the site induced by splicing factor mutations. This approach shifts the spliceosome to use the consensus splice sites, thereby shifting of gene splicing to a more productive isoform. Splicing inhibition can also be used to correct a reading frame by introducing exon skipping. Skipped exons effectively delete nucleotides containing the premature stop codon or duplications/deletions causing the frameshift and leave the remaining exons to re-ligate and produces a truncated protein that is partially functional. Forced skipped exons can also be used to delete toxic regions within mRNA, such as the poison cassette exon of EZH2. On the other hand, forced skipped exons can generate mRNA that undergoes NMD to eliminate a toxic mRNA entirely.

## *Risdiplam*

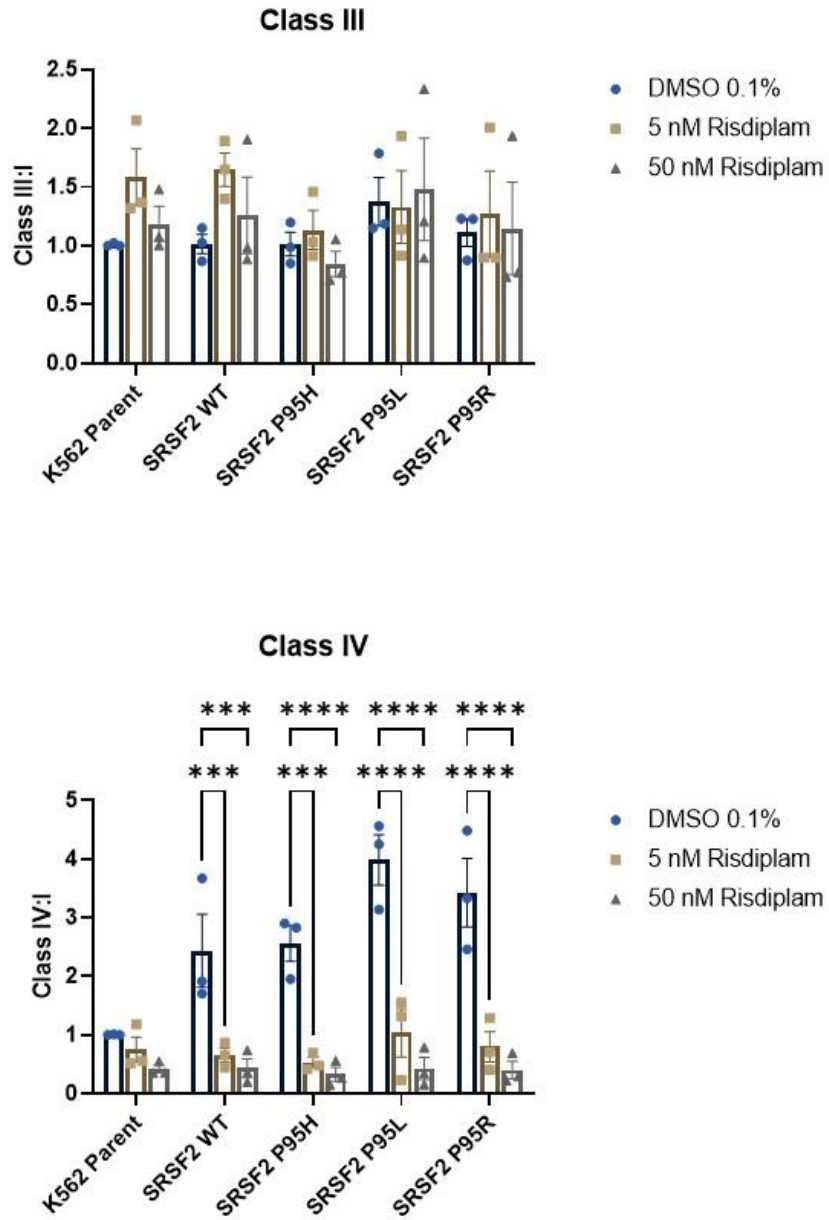
When we had first established a correlation between SRSF2 and *CSF3R* splicing, we began to search for methods to modulate splicing through post-translational modifications of SRSF2 and BCR-ABL+ K562 cells. The RS region of SRSF2 undergoes extensive phosphorylation to regulate splicing. Pilot experiments initially began with treating K562 cells with kinase inhibitors imatinib and dasatinib that had exogenous expression of SRSF2 P95H to test its effects on *CSF3R* splicing. Since dasatinib affects 6 phosphorylation sites on SRSF2 (S189-p, S191-p, S196-p, S206-p, S208-p, and S212-p), we hypothesized that inhibiting serine phosphorylation on these sites would decrease Class IV transcripts. However, results showed that inhibition of Tyr/Ser/Thr phosphorylation did not affect Class IV splicing (**Figure 4.1**).

Our next approach was to modulate splicing through a similar mechanism by which risdiplam is used to treat spinal muscular atrophy (SMA). Risdiplam is a recently FDA-approved drug for treating patients two months of age and older with SMA. In SMA, reduced levels in the genes *SMN1* and *SMN2* are responsible for disease phenotype. SMA represents a clinical case where a splicing modulator is highly effective at retaining a critical exon 7 in *SMN2* that rescues the phenotype. SMA patients have on average 2-4 copies of the *SMN2* gene, which influences *SMN* protein levels and disease severity. Risdiplam is a splice site modifying compound that promotes the retention of *SMN2* exon 7 by binding to a specific ESE sequence and the 5'SS of intron 7. Binding of the 5'SS improves its recognition by the U1 snRNP to enhance inclusion of exon 7 to produce a functional *SMN* protein. Studies on risdiplam have shown its specificity to *SMN2* splicing by deletional studies of its binding sequences. Since we were intrigued by the similarity of its mechanism to Class IV splicing, we treated K562 cells with exogenous SRSF2 P95 expression to investigate its effects on *CSF3R* splicing. Surprisingly, cells that expressed wildtype and mutated SRSF2 showed significantly decreased Class IV transcript levels compared to the controls when treated with either 5 nM or 50 nM of risdiplam (**Figure 4.2**). Class III levels were

unaffected, suggesting that risdiplam may be recognizing an ESE within exon 17 to retain the intron, promoting more Class I. Future experiments will include testing the compound on CD34+ cells with SRSF2 P95 expression in the presence of G-CSF to determine if the restoration of Class I will also be accompanied by restored neutrophil differentiation.



**Figure 4.1. Class IV:I *CSF3R* in K562 Cells Treated with Dasatinib and Imatinib.** K562 cells with stable exogenous expression of SRSF2 WT or SRSF2 P95 mutations were treated with 10 nM dasatinib or 5  $\mu$ M imatinib for 16 h. Class IV:I ratios were measured by qPCR.



**Figure 4.2. Class IV:I *CSF3R* in K562 Cells Treated with Risdiplam.** K562 cells with stable exogenous expression of SRSF2 WT or SRSF2 P95 mutations were treated with 5 nM or 50 nM risdiplam for 16 h. Class IV:I ratios were measured by qPCR.



## *Synthetic Introns*

A promising new therapeutic approach exploits synthetic introns to selectively target cells that possess splicing factor mutations while leaving healthy cells unaffected. A synthetic construct is inserted into an open reading frame that encodes a therapeutic protein to ensure that the protein would only be expressed in cells harboring the splicing factor mutation. These introns are roughly 100 nucleotides in length and have been studied in *SF3B1* mutant leukemia, breast cancer, and uveal melanoma cells where mutated cells were eradicated while wildtype cells were undisturbed (105). Current approaches to splicing modulation have not been successful at targeting cells bearing the spliceosomal mutations. Therapies inhibit both wildtype and mutant splicing factor cells to some degree. To date, synthetic introns represent the closest therapeutic intervention that selectively targets splicing factor-mutated cells and has prolonged survival in mouse models at the preclinical stage. These studies show promise because they can be extended to other types of splicing factor mutations and other cancers in which splicing abnormalities exist. In addition, synthetic introns may be beneficial in cancers that have mutated epigenetic factors which can directly or indirectly affect splicing. Granted, these studies were conducted in immunodeficient human xenograft mouse models so further testing is still warranted in animals with *Sf3b1* mutations with intact immune systems.

## Chapter 5: Concluding Remarks

### Development of MDS May Depend on Presence of Alternative Splice Isoforms

The ability of higher eukaryotes to generate protein diversity through alternative splicing allows them to regulate their development without expanding their genomes. Since alternative splicing relies on an intricate and precise modes of regulation, dysregulation of this system can lead to abnormal phenotypes and contributes to disease development. Human and mouse comparative transcriptome analyses have shown that alternative splicing is often associated with the recent creation or loss of exons. Exons can appear based on three mechanisms: exon shuffling, exonization of intronic sequences, or transition of a constitutive exon to an alternative exon. Exon shuffling occurs when a new exon is inserted into an existing gene or it is duplicated within the same gene. It creates a new chimeric protein that gives an evolutionary advantage to the organism. Evidence to support this observation is found in the mRNA, where there is a correlation between the borders of exons and protein domains. An example of this correlation is seen in the CSF3R domains where the exon boundaries align with the protein domains. The consequence of exon shuffling is increased genome size and the appearance of modular proteins. Modular proteins are associated with receptor proteins and extracellular matrix membrane associated proteins. In the case of CSF3R, the fibronectin domains in the extracellular domain are an example of modular proteins produced either by exon shuffling or duplication (129). Mutations that occur within the exon or in flanking introns compete with the authentic splice sites which lead to alternative splice site selection. They remain as minor splice sites since there is negative selection pressure. The ancestral site remains the major splice site so that the original protein is not disturbed. With so many transcripts being potentially made as a result of alternative splicing, a question remains as to whether or not these all make a functional protein.

Conservation of a specific alternative splicing pattern throughout evolution is strong evidence of biological function. Despite the importance of alternative splicing conservation, the number of

conserved constitutive exons still outnumber alternatively spliced exons. If new isoforms are created through mutations with no apparent function, but its low abundance is not toxic for the cell, then those isoforms have higher probabilities of being evolutionary conserved. Exons of low inclusive level are often associated with increased evolutionary changes. If the new transcript acquires a new function, random mutations will strengthen its regulatory sequences to increase its inclusion level or give it tissue specific attributes. The conservation of splice isoforms will depend on how beneficial they are to the cell (38). Class III and Class IV *CSF3R* may be examples in which random mutations gave rise to isoforms that were not completely deleterious to the cell, but have impaired functions on their own. Increased splicing of Class III and Class IV may drive production of Class I below a threshold that could render organisms to become susceptible to certain diseases like bone marrow failure disorders.

### ***CSF3R* Splicing is affected by MDS-Associated Splicing Factor Mutations**

The minigene splicing assays demonstrated that all MDS-associated splicing factor mutations that were tested, except for *SF3B1* K700E, affected Class IV *CSF3R* levels. *U2AF1* Q157P, *SRSF2* P95, and *LUC7L2* knockout increased Class IV while *U2AF1* S34F decreased Class IV. The same effects were also observed in patients with myeloid neoplasms, confirming the *in vitro* studies. These findings suggest that exon 17 of *CSF3R* contains alternative splice site sequences and regulatory elements that are particularly sensitive to splicing factor mutations. Given the function of Class IV in neutrophil development, the ability of *CSF3R* to recruit important splicing factors to facilitate its splicing is an important regulatory feature. Dysregulation of recruitment of splicing factors or their ability to properly recognize consensus splice site sequences on the pre-mRNA can enhance the production of alternate isoforms that have impaired signaling. The production of Class III and Class IV in the presence of splicing factor mutations could also be a mechanism by which the cells attempt to maintain homeostasis. Rather than have complete loss of function of Class I *CSF3R*, the cells produce alternate isoforms that have reduced functional

capacity. Since the majority of CSF3R exist as heterodimers, the presence of Class I may be transducing intact signaling that is above a threshold to preserve normal hematopoiesis, albeit in a reduced state that results in impaired neutrophil differentiation. Problems arise when the ratios of Class III:I or Class IV:I reach a point above a threshold to where the reduced expression of Class I is unable to preserve normal maturational signaling. For example, Class III and Class IV alone were unable to rescue the phenotype of *Csf3r*-null mouse LSK cells in response to G-CSF. Add-back of Class I alone produced terminally differentiated neutrophils with normal phenotype, which suggests that Class III and Class IV are not essential in normal neutrophil development. Therefore, modulating the levels of Class III and Class IV in hematological malignancies may delay disease progression.

### **SRSF2 Regulates CSF3R Class IV Splicing Through ESE Recognition**

One striking observation was that the number of CCNG ESE motifs outnumbered GGNG ESE motifs by more than 50% in CSF3R exon 17. The two top-scoring motifs according to ESEfinder are located within the intronic region of exon 17 that is excluded to produce Class IV CSF3R. These motifs are usually situated within exons in close proximity to introns with weak splice site signals and/or require ESEs for their inclusion. SRSF2 P95H recognizes CCNG motifs in greater affinity than GGNG motifs which influences exon inclusion or exclusion. In contrast, SRSF2 WT can recognize both motifs with equal affinity. I hypothesized that SRSF2 P95H would bind to ESE1 and ESE2 with greater affinity compared to SRSF2 WT. Surprisingly, SRSF2 P95H bound with greater affinity to ESE2 but not ESE1 compared to SRSF2 WT. This observation suggests that ESE1 may have additional motifs that facilitate binding of SRSF2 WT despite the presence of CCNG motifs. While only the two top-scoring motifs were tested using ITC, the probability of SRSF2 P95H recognizing other lower-scoring motifs is likely. Since SRSF2 WT can recognize both motifs with equal affinity, the number of CCNG motifs may need to outnumber GGNG motifs in the context of a mutated SRSF2 in order to preserve partial function of CSF3R by promoting

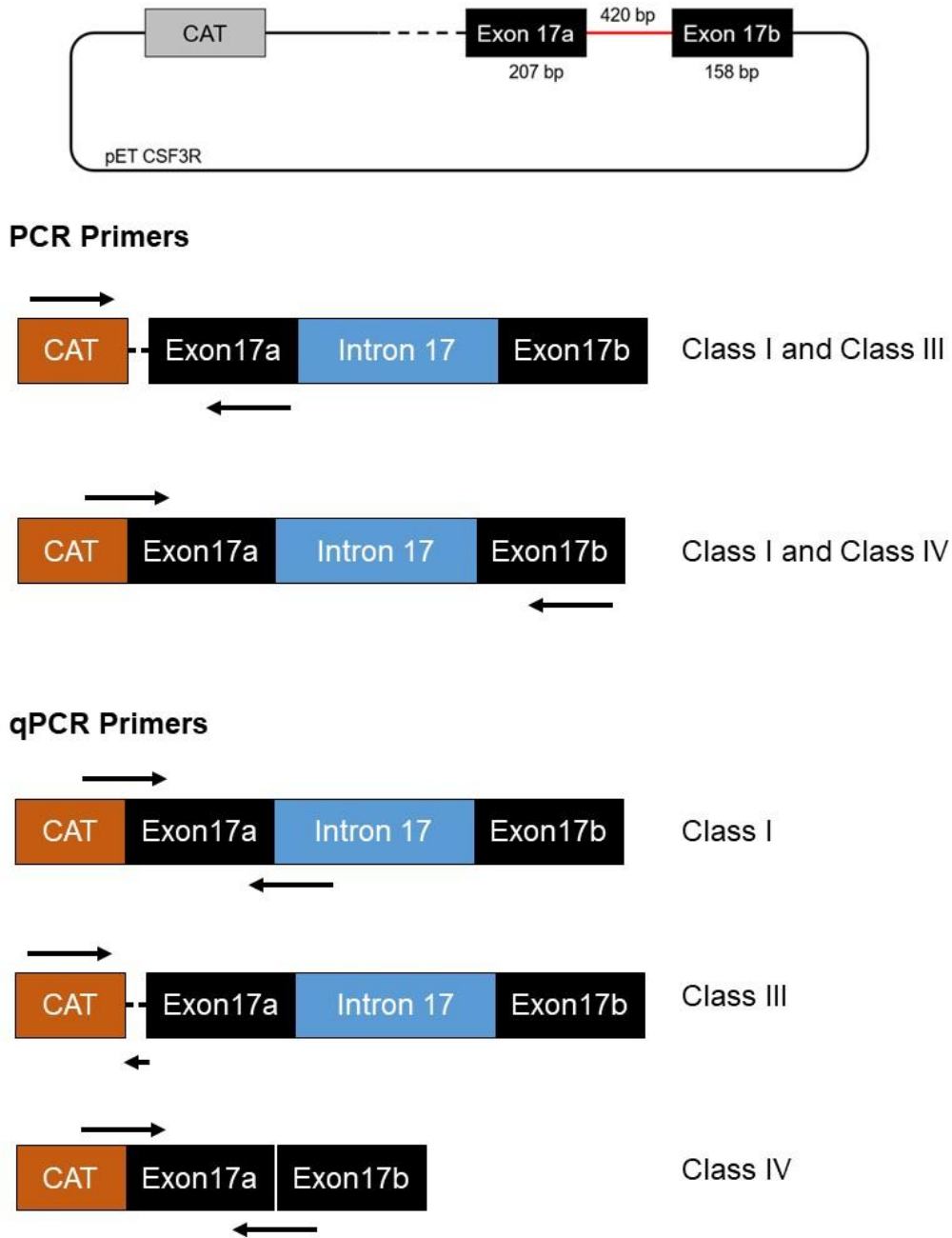
Class IV splicing. While mutated SRSF2 relies on a wildtype allele, the splicing efficiency of the remaining copy may have reduced splicing efficiency to produce Class I. In addition to demonstrating the direct binding of SRSF2 to ESE1 and ESE2, I also showed that deletion of either motif leads to reduced Class I and Class IV splicing, indicating that these sequences are important in regulating the inclusion of intron 17. Future studies will include mutating the CCNG motifs to GGNG in ESE1 and ESE2 to determine if the binding affinities can be reversed.

## **Conclusion**

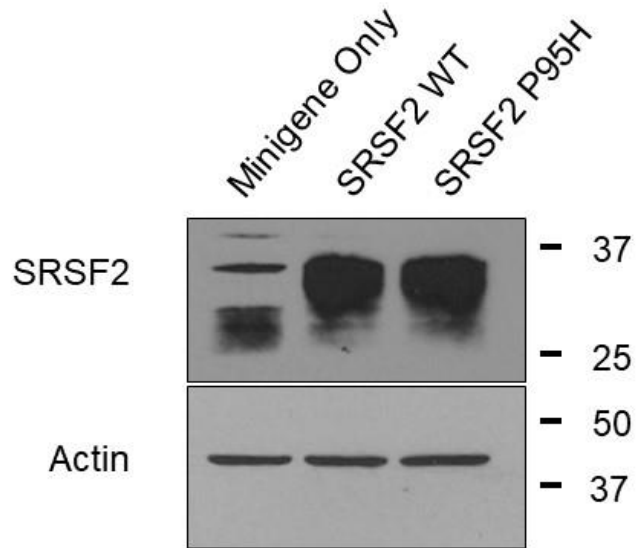
The discovery of splicing factor mutations has opened a wide range of opportunities in interrogating the missplicing of specific genes as well as developing therapeutics to modulate splicing. We have shown, using *CSF3R* as a model, that splicing aberrations resulting in imbalanced isoform production can contribute to dysgranulopoiesis. Alternative splice isoforms of *CSF3R* have specific roles in neutrophil development that may have been a consequence of the evolution of splice site sequences and the emergence of new exons. My studies have shown that the predominance of alternative isoforms can lead to disruption of neutrophil differentiation and possibly co-drive MDS progression. Indeed, *CSF3R* represents one gene out of hundreds that are critical in regulating hematopoiesis. The investigation of other splice variants will provide additional insights into discovering therapeutics to correct the splicing defects. Splicing modulators hold promise as an MDS-specific therapy that show positive effects in eliminating cells harboring heterozygous mutations. While the results of these preclinical studies have shown promise in mouse models, the results of phase I clinical trials are often disappointing. Incongruencies in studies across disease models emphasize the complexities of not only gene regulation, but also differences in hematopoiesis. While the molecular mechanisms driving MDS remain elusive, the identification of splicing factor mutations as drivers of disease have spearheaded a concerted effort across multiple studies that have accelerated the use of genomics to identify new gene targets. Future advancements in drug development will need to rely on a

combination of genomics and existing WHO criteria to better understand how to treat MDS and its subtypes.

## Appendix A

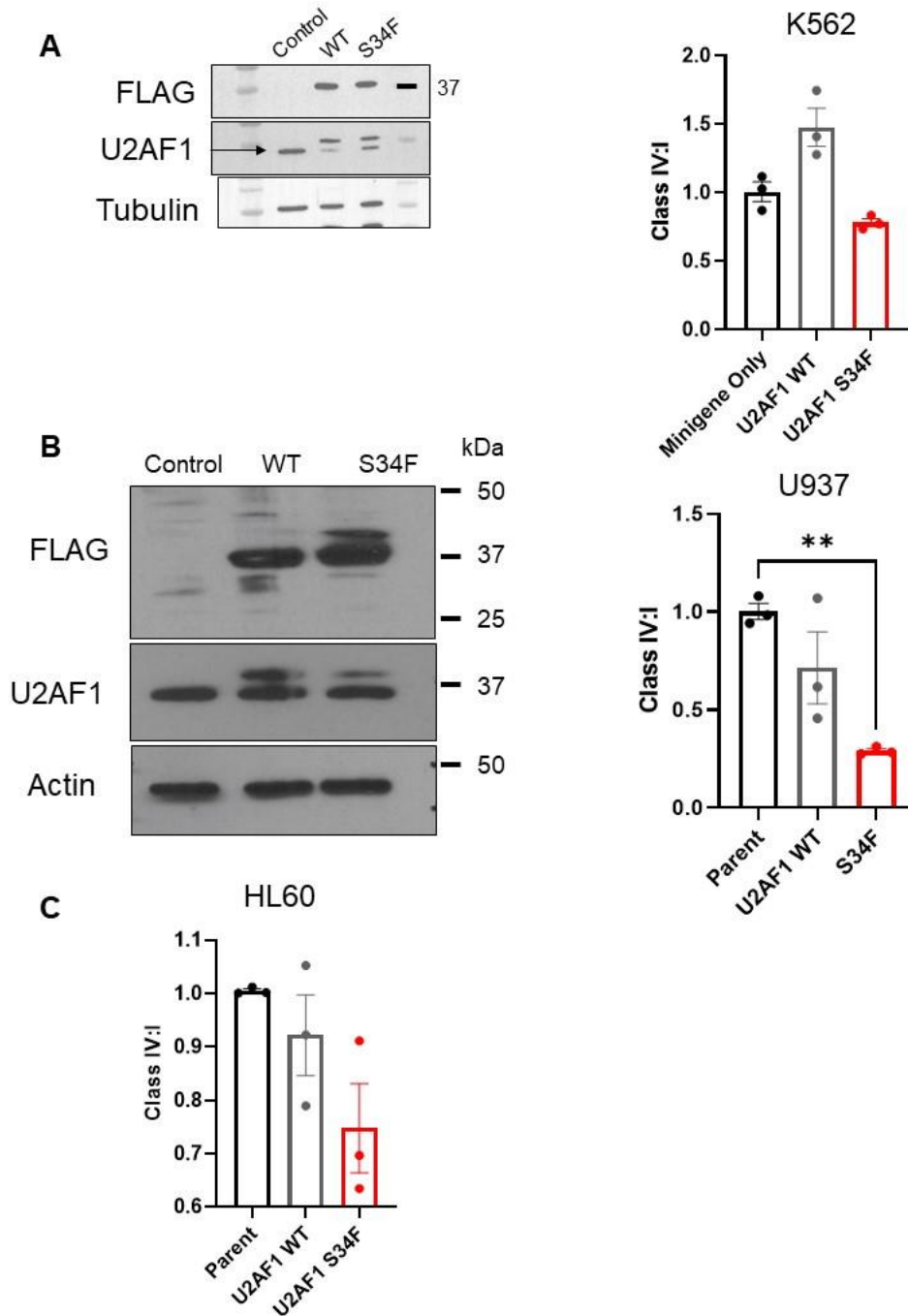


**Figure 1. Schematic of the *CSF3R* Minigene.** The *CSF3R* minigene is comprised of a CAT primer anchoring gene, 253 bp of intron 16 (represented by the dotted line), and exon 17 in a pET backbone. Class I contains a retained intron (red line) that is 420 bp long between Exon 17a and Exon 17b. Forward primers for PCR or qPCR are located within the CAT gene to distinguish amplification from endogenous *CSF3R*. Reverse primers are situated in sequences that are specific to Class I, III, or IV.

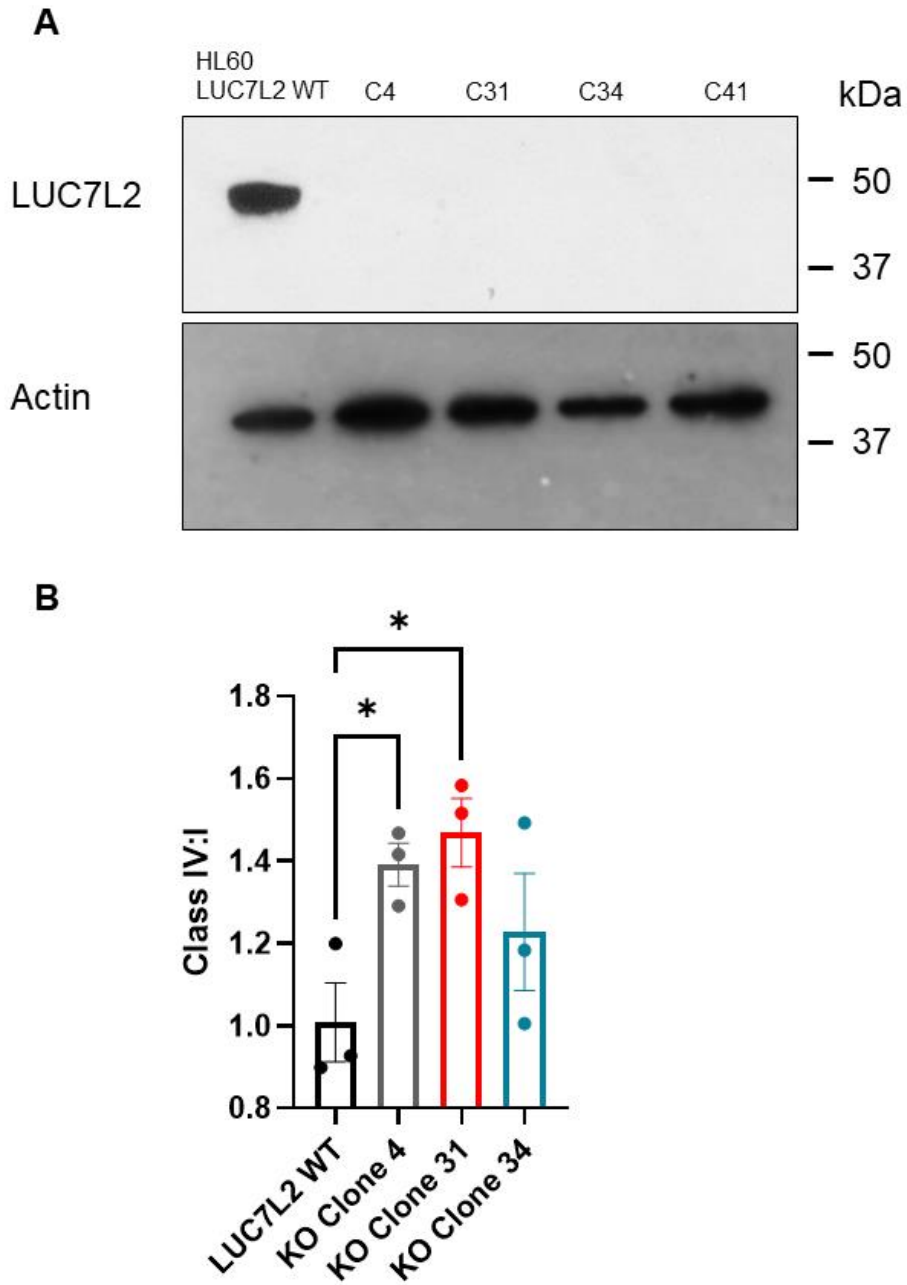


**Figure 2. Transient Overexpression of SRSF2 in K562 Cells.** K562 cells were electroporated with 1  $\mu$ g of the *CSF3R* minigene and either SRSF2 WT or SRSF2 P95H. Cell lysates were made following a 18 h transfection and a western blot was performed to assess the protein expression of SRSF2 in K562 cells compared to a minigene only transfected control.

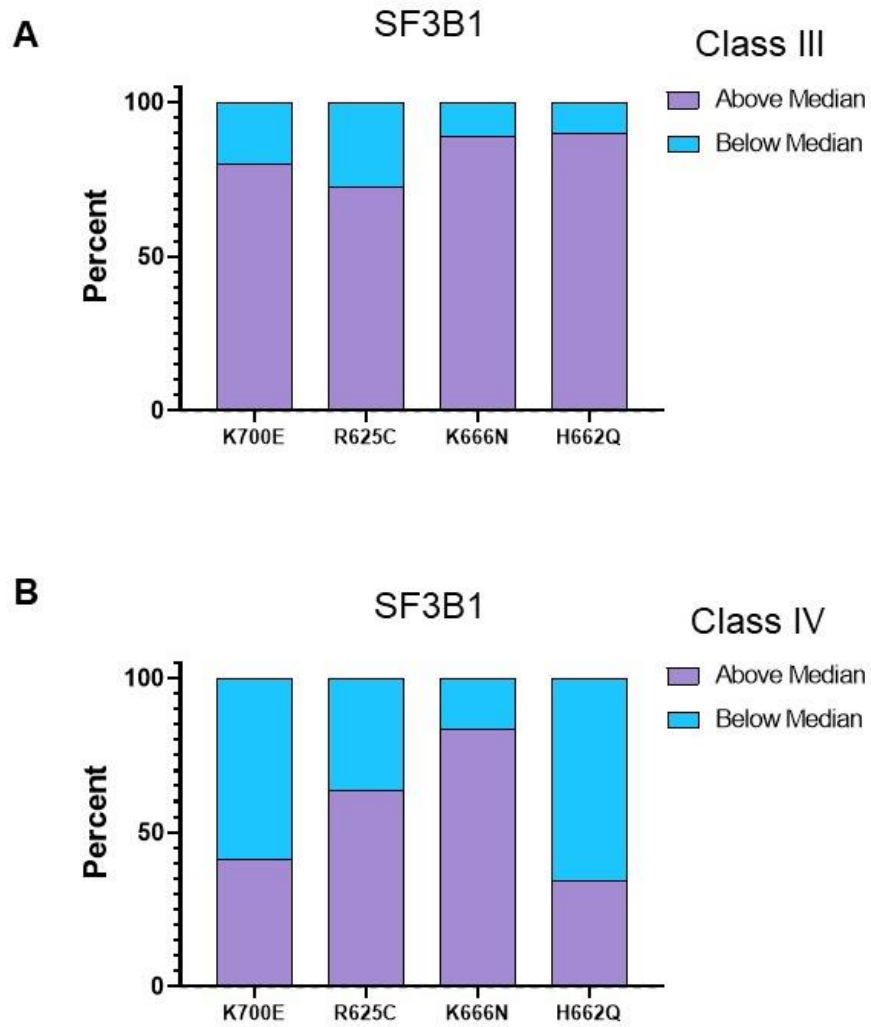




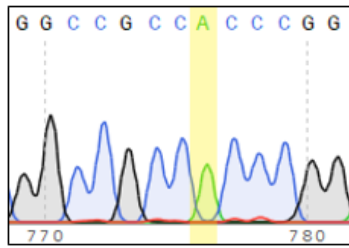
**Figure 3. *CSF3R* Class IV:I Validation in Other Leukemia Cell Lines with Stable Expression of U2AF1.** Western blots showing stable expression of FLAG tagged U2AF1 WT and U2AF1 S34F in (A) K562 cells and (B) U937 cells all show a decrease in Class IV:I in U2AF1 S34F expressing cells. (C) HL60 cells with stable expression of U2AF1 WT and U2AF1 S34F show a decrease in Class IV:I by qPCR. Minigene *CSF3R* was measured in K562 cells and endogenous *CSF3R* was measured in U937 and HL60 cells. Error bars represent the mean with standard error of the mean from three independent experiments. \*\* $p \leq 0.01$ .



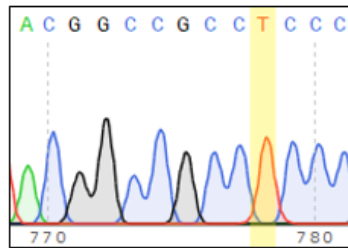
**Figure 4. *CSF3R* Class IV:I ratios in LUC7L2 knockout HL60 cells.** (A) Western blot showing absence of LUC7L2 protein in four clones (C4, C31, C34, and C41) in HL60 cells. (B) qPCR of cDNA of three clones collected 18 h post transfection. Class IV:I ratios were measured using Class I and Class IV specific qPCR primers by SYBR green fluorescence. Error bars represent the mean with standard error of the mean from three independent experiments. \* $p \leq 0.05$ .



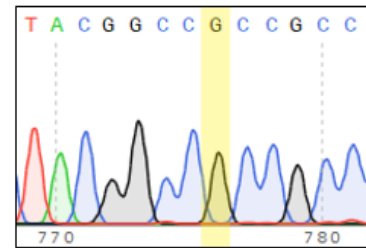
**Figure 5. Correlation of SF3B1 Mutations with Class III and Class IV Expression.** The percentage of patients with high or low Class III (A) or Class IV *CSF3R* (B) were characterized according to the subset of SF3B1 mutations.



P95H  
CCC>CAC



P95L  
CCC>CTC



P95R  
CCC>CGC

**Figure 6. Sequence Confirmation of SRSF2 P95 Mutations.** Nucleotide substitutions c.284C>A, c.284C>T, or c.284C>G result in missense mutations in SRSF2 that change the amino acid from a proline to either a histidine, leucine, or arginine at position 95.

### **SRSF2 WT RRM domain sequence from Uniprot**

>sp|Q01130|SRSF2\_HUMAN Serine/arginine-rich splicing factor 2 OS=Homo sapiens OX=9606  
GN=SRSF2 PE=1 SV=4

MSYGRPPPDVEGMTSLKVDNLTYRTSPDTRLRRVFEKYGRVGDVYIPRDRYTKESRGFAFVRFHDKRDAE  
DAMDAMDGAVLDGRELRVQMARYGRPPDSHHS

### **Non-optimized Codon from IDT for RRM domain of SRSF2 WT**

ATG AGC TAC GGC AGA CCC CCC CCC GAC GTG GAG GGC ATG ACC AGC CTG AAG GTG GAC A  
AC CTG ACC TAC AGA ACC AGC CCC GAC ACC CTG AGA AGA GTG TTC GAG AAG TAC GGC AG  
A GTG GGC GAC GTG TAC ATC CCC AGA GAC AGA TAC ACC AAG GAG AGC AGA GGC TTC GCC  
TTC GTG AGA TTC CAC GAC AAG AGA GAC GCC GAG GAC GCC ATG GAC GCC ATG GAC GGC G  
CC GTG CTG GAC GGC AGA GAG CTG AGA GTG CAG ATG GCC AGA TAC GGC AGA CCC CCC G  
AC AGC CAC CAC AGC

### **Optimized codon from IDT for RRM domain of SRSF2 WT**

ATG AGT TAT GGG AGG CCC CCA CCA GAC GTC GAG GGC ATG ACC TCT CTT AAA GTA GAC AA  
C CTC ACT TAT CGC ACC TCA CCA GAT ACA TTG AGG AGA GTA TTC GAG AAA TAT GGG CGG G  
TT GGC GAC GTA TAC ATT CCG AGG GAT AGG TAC ACC AAA GAG AGT CGC GGA TTC GCT TTC  
GTT CGC TTT CAC GAT AAG AGA GAC GCC GAA GAT GCG ATG GAC GCA ATG GAT GGT GCG G  
TG CTG GAC GGT AGG GAG TTG CGC GTA CAA ATG GCG CGC TAT GGC CGA CCC CCC GAC TC  
C CAT CAC TCT

Restriction site: NdeI & EcoRI (Blue)

Cutting site between two domain: (Yellow)

Extra His tag: (Green)

TEV cutting site: (Purple)

### **gBlock from IDT**

ATATATCATATG CAC CAC CAT CAT CAT CAC CAT TCC TCT GGT GTG GAC CTT GGA ACG GAA  
AAT CTG TAT TTT CAA

ATG AGT TAT GGG AGG CCC CCA CCA GAC GTC GAG GGC ATG ACC TCT CTT AAA GTA GAC AA  
C CTC ACT TAT CGC ACC TCA CCA GAT ACA TTG AGG AGA GTA TTC GAG AAA TAT GGG CGG G  
TT GGC GAC GTA TAC ATT CCG AGG GAT AGG TAC ACC AAA GAG AGT CGC GGA TTC GCT TTC  
GTT CGC TTT CAC GAT AAG AGA GAC GCC GAA GAT GCG ATG GAC GCA ATG GAT GGT GCG G  
TG CTG GAC GGT AGG GAG TTG CGC GTA CAA ATG GCG CGC TAT GGC CGA CCC CCC GAC TC  
C CAT CAC TCT

TAA CATTGGAAGTGGATAACGGATCCGAATTCATATAT

**Figure 7. gBlock Sequence of SRSF2 WT RRM.** SRSF2 WT RRM sequence was confirmed on Uniprot and codons were optimized using on the IDT website.

### SRSF2 P95H RRM domain sequence from Uniprot

>sp|Q01130|SRSF2\_HUMAN Serine/arginine-rich splicing factor 2 OS=Homo sapiens OX=9606  
GN=SRSF2 PE=1 SV=4

MSYGRPPPDVEGMTSLKVDNLTyrTSPDtlRRVFEKYGRVGDVYIPRDRYTKESRGFAFVRFHDKRDAE  
DAMDAMDGAVLDGRELRVQMARYGRHPDSHHS

### Non-optimized Codon from IDT for RRM domain of SRSF2 P95H

ATG AGC TAC GGC AGA CCC CCC CCC GAC GTG GAG GGC ATG ACC AGC CTG AAG GTG GAC A  
AC CTG ACC TAC AGA ACC AGC CCC GAC ACC CTG AGA AGA GTG TTC GAG AAG TAC GGC AG  
A GTG GGC GAC GTG TAC ATC CCC AGA GAC AGA TAC ACC AAG GAG AGC AGA GGC TTC GCC  
TTC GTG AGA TTC CAC GAC AAG AGA GAC GCC GAG GAC GCC ATG GAC GCC ATG GAC GGC G  
CC GTG CTG GAC GGC AGA GAG CTG AGA GTG CAG ATG GCC AGA TAC GGC AGA CAC CCC G  
AC AGC CAC CAC AGC

### Optimized codon from IDT for RRM domain of SRSF2 P95H

ATG AGT TAT GGT CGG CCC CCT CCA GAC GTT GAA GGT ATG ACT AGT CTG AAA GTT GAC AA  
C CTC ACA TAC AGA ACC TCC CCT GAT ACG CTG CGG AGA GTA TTC GAA AAG TAC GGA AGA G  
TT GGC GAC GTA TAC ATC CCG AGA GAC AGG TAT ACG AAG GAA AGT AGG GGT TTT GCG TTT  
GTG CGC TTC CAT GAT AAG CGC GAC GCA GAG GAT GCC ATG GAT GCT ATG GAC GGA GCG G  
TT CTG GAC GGA CGG GAG CTC AGA GTT CAA ATG GCG CGG TAC GGA CGC CAC CCG GAC AG  
C CAC CAT TCA

Restriction site: NdeI & EcoRI (Blue)

Cutting site between two domain: (Yellow)

Extra His tag: (Green)

TEV cutting site: (Purple)

### gBlock from IDT

ATATATCATATG CAC CAC CAT CAT CAT CAC CAT TCC TCT GGT GTG GAC CTT GGA ACG GAA  
AAT CTG TAT TTT CAA

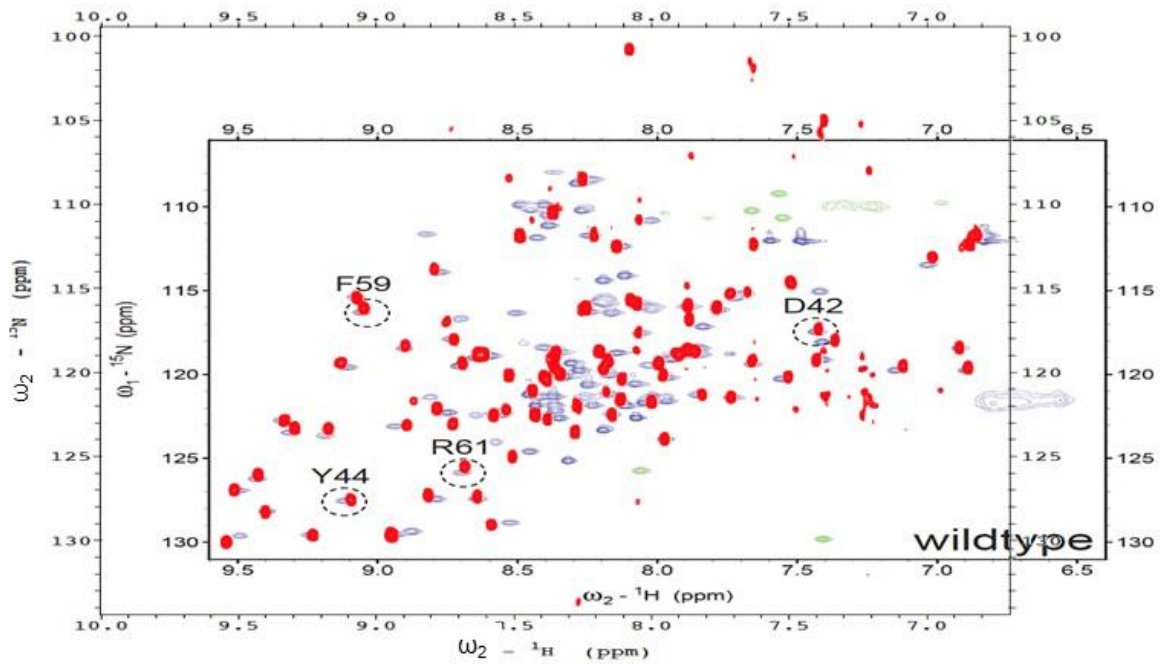
ATG AGT TAT GGT CGG CCC CCT CCA GAC GTT GAA GGT ATG ACT AGT CTG AAA GTT GAC AA  
C CTC ACA TAC AGA ACC TCC CCT GAT ACG CTG CGG AGA GTA TTC GAA AAG TAC GGA AGA G  
TT GGC GAC GTA TAC ATC CCG AGA GAC AGG TAT ACG AAG GAA AGT AGG GGT TTT GCG TTT  
GTG CGC TTC CAT GAT AAG CGC GAC GCA GAG GAT GCC ATG GAT GCT ATG GAC GGA GCG G  
TT CTG GAC GGA CGG GAG CTC AGA GTT CAA ATG GCG CGG TAC GGA CGC CAC CCG GAC AG  
C CAC CAT TCA

TAA CATTGGAAGTGGATAACGGATCCGAATTCATATAT

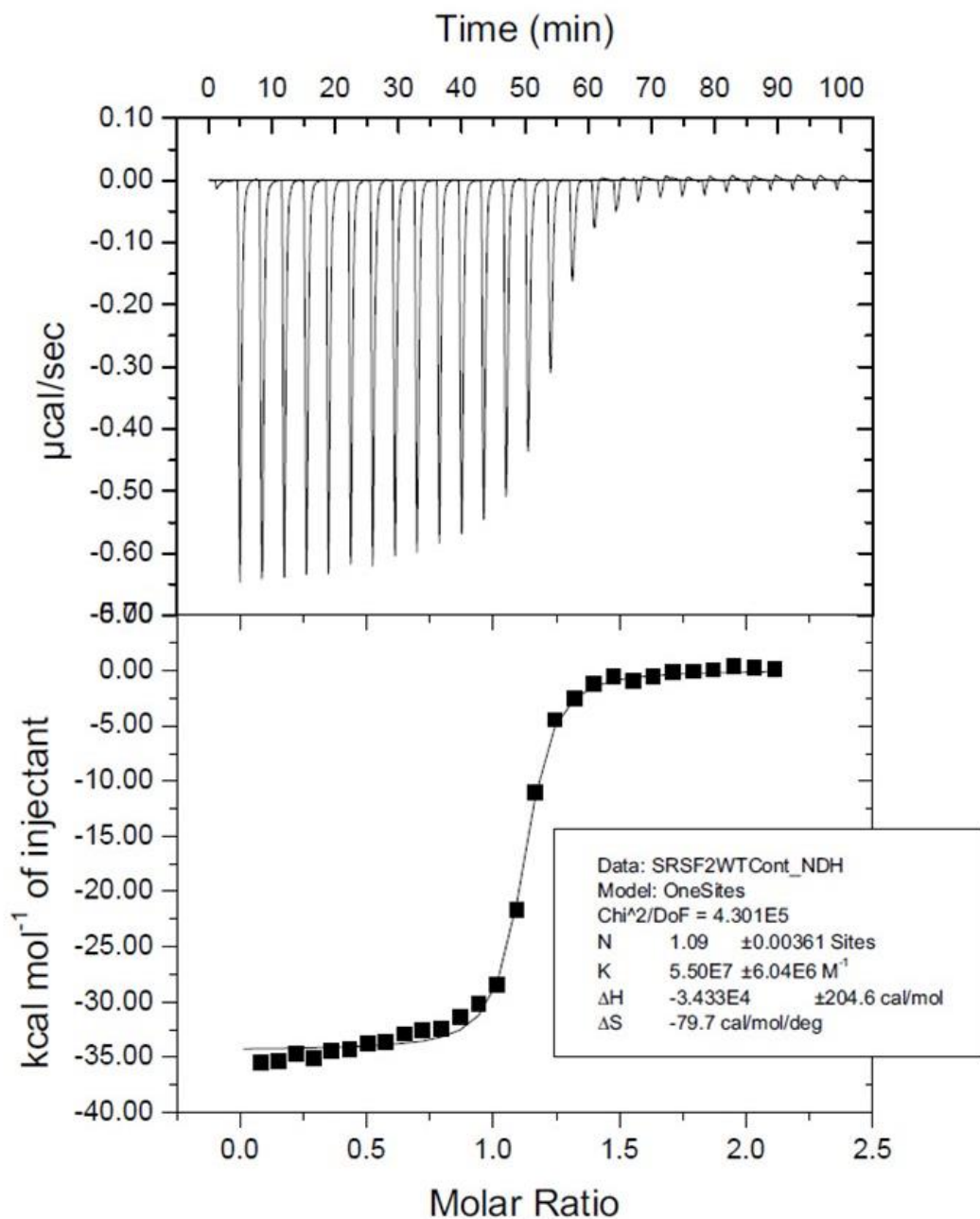
**Figure 8. gBlock Sequence of SRSF2 P95H RRM.** SRSF2 P95H RRM sequence was confirmed on Uniprot and codons were optimized using on the IDT website. The red letters depict the location of the amino acid and nucleotide substitutions.



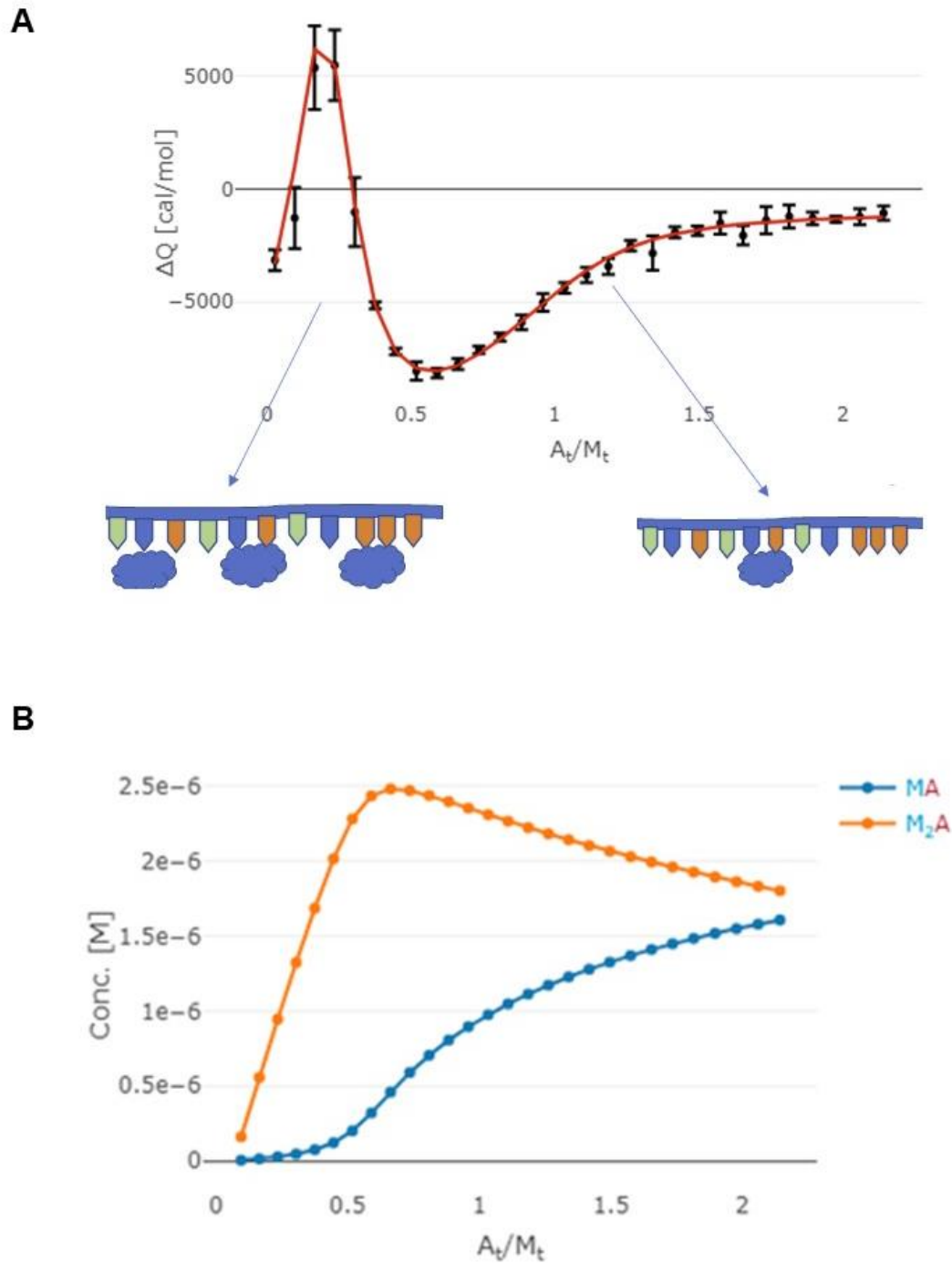
## SRSF2 WT



**Figure 9. Protein Folding of SRSF2 WT.** Our results of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra with an NMR titration of  $^{15}\text{N}$ -labelled SRSF2 WT RRM with unlabelled 5'-UCCAGU-3' were overlaid with the spectra reported by Daubner, *et al.*

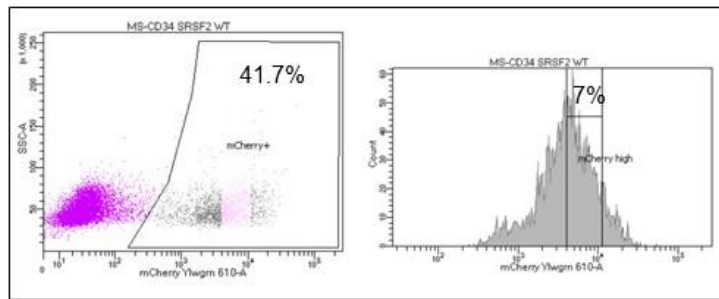


**Figure 10. ITC of SRSF2 WT Titrated with the Control RNA Sequence.** Binding affinity of SRSF2 WT RRM for the 5'-UCCAGU-3' RNA measured by ITC. Raw heat values are depicted above, and the corresponding binding curve is depicted below.

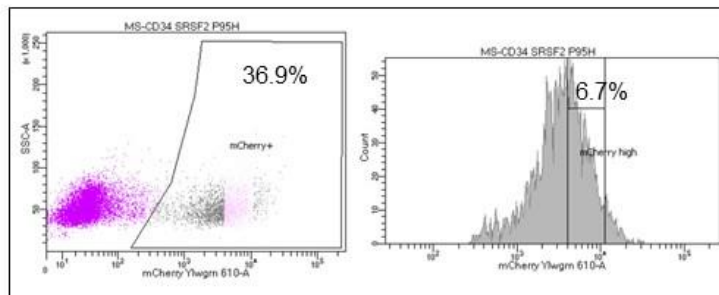


**Figure 11. Validation of the Binding Model for SRSF2 RRMs with ESE1.** (A) The initial region of the isotherm could represent a weaker binding site that only occurs when the protein is in excess. When there is only one strong binding site, the final 1:1 stoichiometry could prevail when the RNA is in excess. (B) Titration of SRSF2 with a shortened ESE1 demonstrated a prevailing 2:1 isotherm with a growing 1:1 isotherm, validating our binding model.

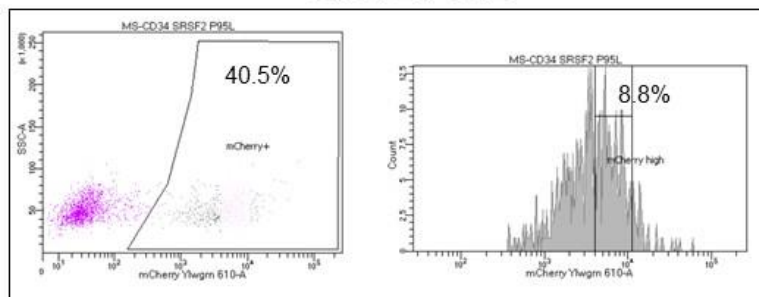
### SRSF2 WT



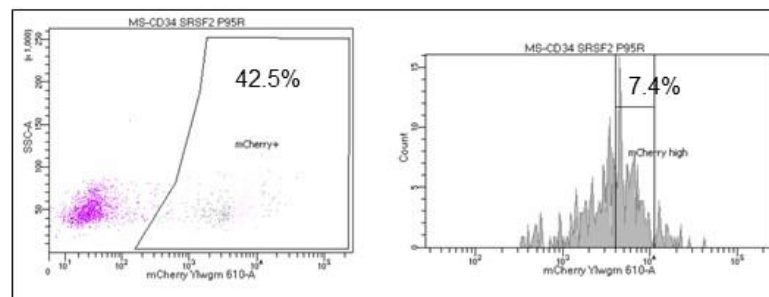
### SRSF2 P95H



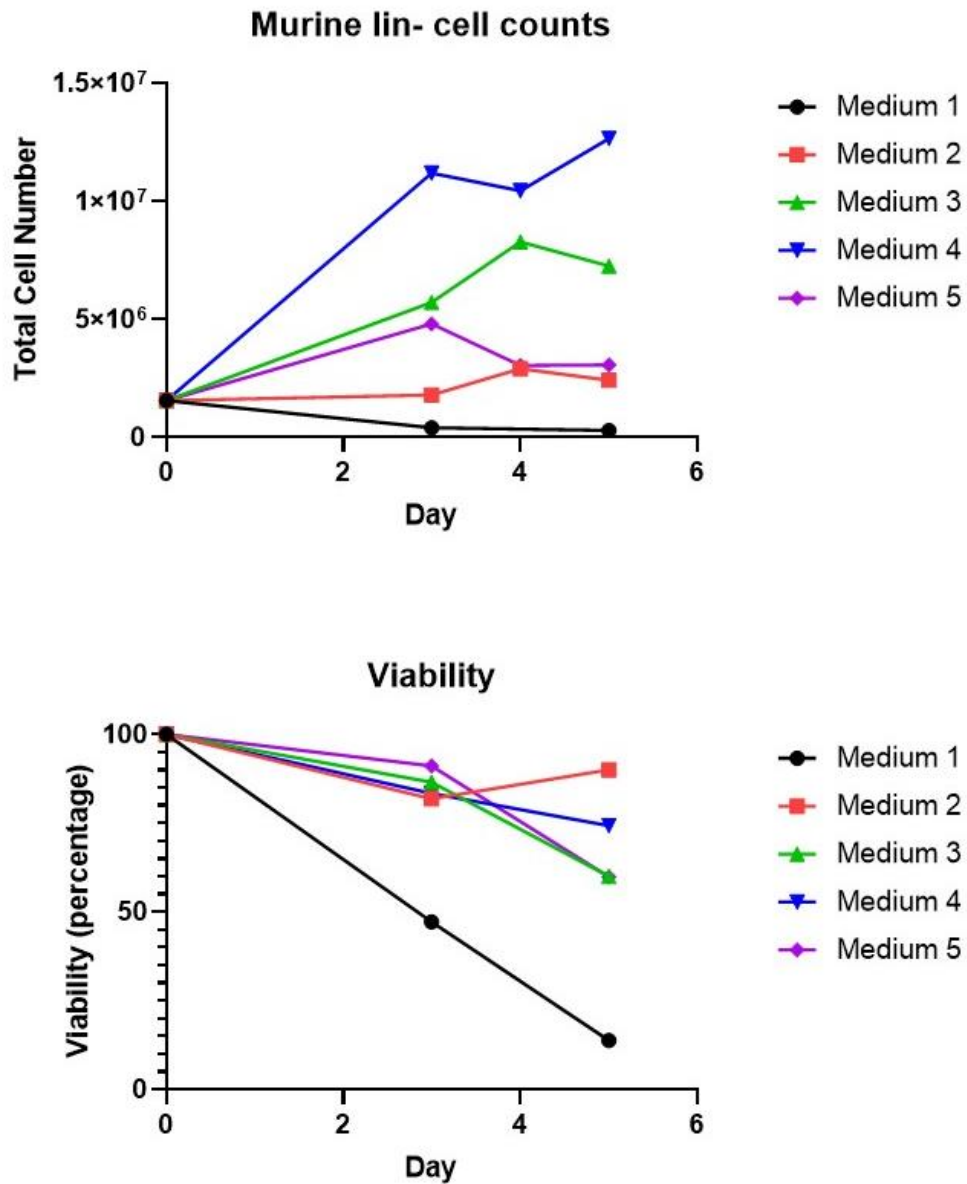
### SRSF2 P95L



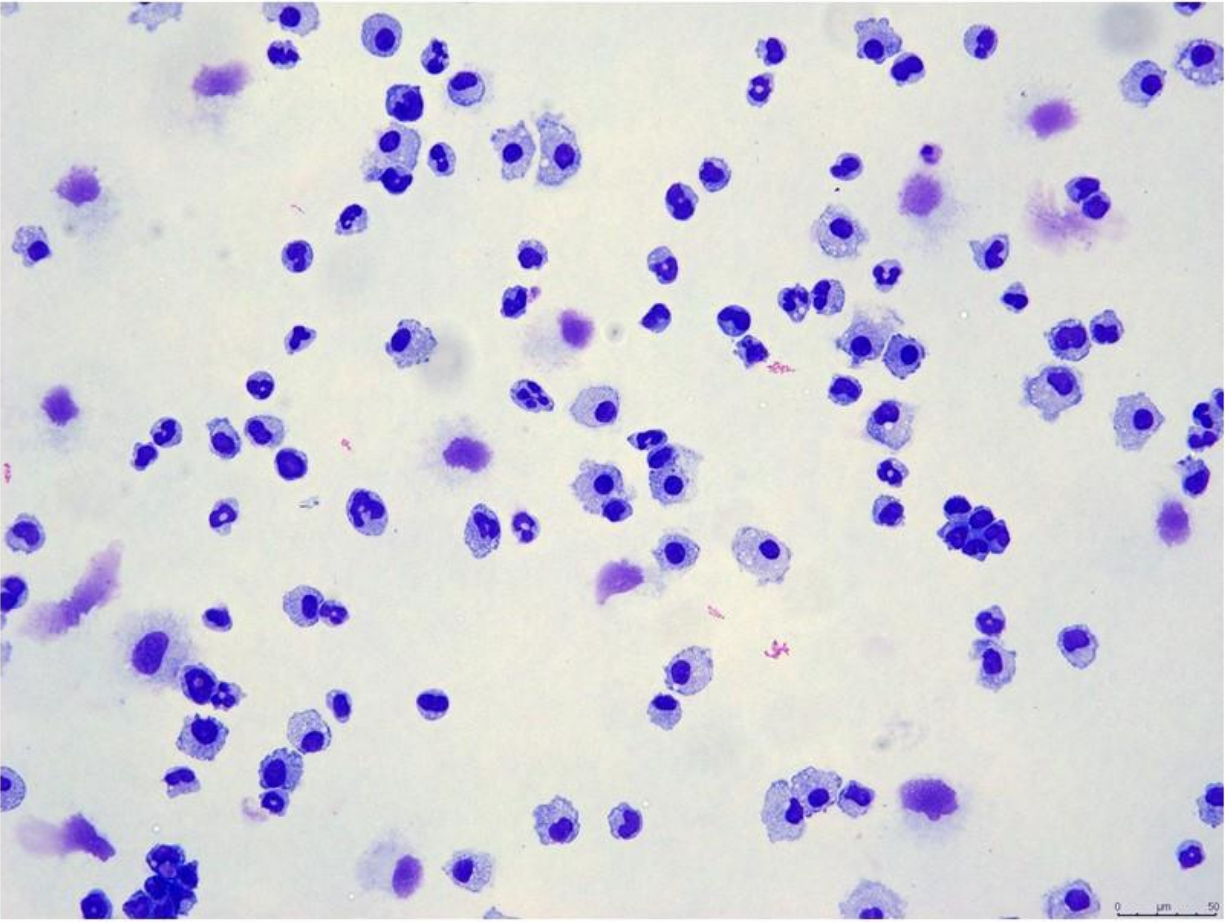
### SRSF2 P95R



**Figure 12. Gating Strategy for Sorting CD34+ Cells Transduced with SRSF2.** Cells were sorted for equal MFI on +mCherry expression by flow cytometry.

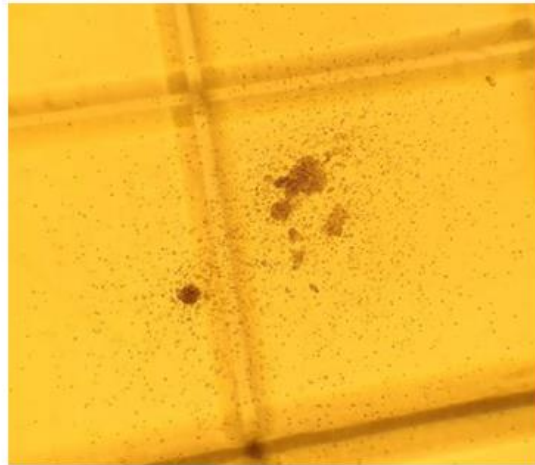


**Figure 13. Media Conditions for Optimal Growth and Viability of Murine LSK Cells.** Cells were grown in the following conditions: Medium 1: StemPro-34 with StemPro Nutrient Supplement, 10 ng/mL mSCF, 100 ng/mL mFlt3L, and 10 ng/mL mTPO; Medium 2: IMDM with heat inactivated 10% FBS, 20 ng/mL mSCF, 100 ng/mL mFlt3L, and 10 ng/mL mTPO; Medium 3: StemSpan SFEM II with 10 ng/mL mL-3, 25 ng/mL mSCF, 20 ng/mL mL-6, and 25 ng/mL mTPO; Medium 4: StemSpan SFEM II with 100 ng/mL mSCF, 100 ng/mL mFlt3L, 100 ng/mL hIL-11, and 20 ng/mL mL-3; Medium 5: OptimMEM with 10% FBS, 50 ng/mL mSCF, 10 ng/mL mL-3, and 10 ng/mL mL-6.



**Figure 14. Cytopspin of Mouse LSK Cells with Class I CSF3R in the Presence of G-CSF Alone.** Cells were grown in Methocult 3531 with the addition of human G-CSF (100 ng/mL) alone. Cytopspins of whole plates on day 8 of culture show some neutrophil differentiation without colony formation.





CFU-G



CFU-GM



CFU-M

**Figure 15. Representative CFU-G, CFU-GM, and CFU-M Images.** Individual colonies were imaged on a stereo microscope at day 8. Images of CFU-G were taken at 4X and CFU-GM and CFU-M were taken at 8X.

## Appendix B

### PCR Primers

Gene	Forward	Reverse
CSF3R Class I-III (PCR minigene)	TGTGTGGGGAGCGTGGATTC	CGGCTTCTTTTCATCCTCCTCC
CSF3R Class I-IV (PCR minigene)	GAGGACCCACAAGGATGCCTT	CCCAAGAGTGTCTATAACAACAACAA AAACTGC
CSF3R Class I (qPCR, minigene)	GAGGACCCACAAGGATGCCTT	CAGCAGCTGCCCATAAAGGAC
CSF3R Class III (qPCR, minigene)	CTTCTACACACCCATGTCCCGC	AGCACGGCTCATTTTCAGATGTC
CSF3R Class IV (qPCR, minigene)	GAGGACCCACAAGGATGCCTT	GCCCAGCCTGATCGCT
CSF3R Class I (qPCR, endogenous)	CCACAATCATGGAGGAGGATGC	CAGCAGCTGCCCATAAAGGAC
CSF3R Class III (qPCR, endogenous)	CAGGAAGAATCCCCTCTGGC	GTCTGCCCCAGCCACTG
CSF3R Class IV (qPCR, endogenous)	ATGCCTTCCAGCTGCCC	GCCCAGCCTGATCGCT
Actin	GAATCAATGCAAGTTCGGTTCC	TCATCTCCGCTATTAGCTCCG



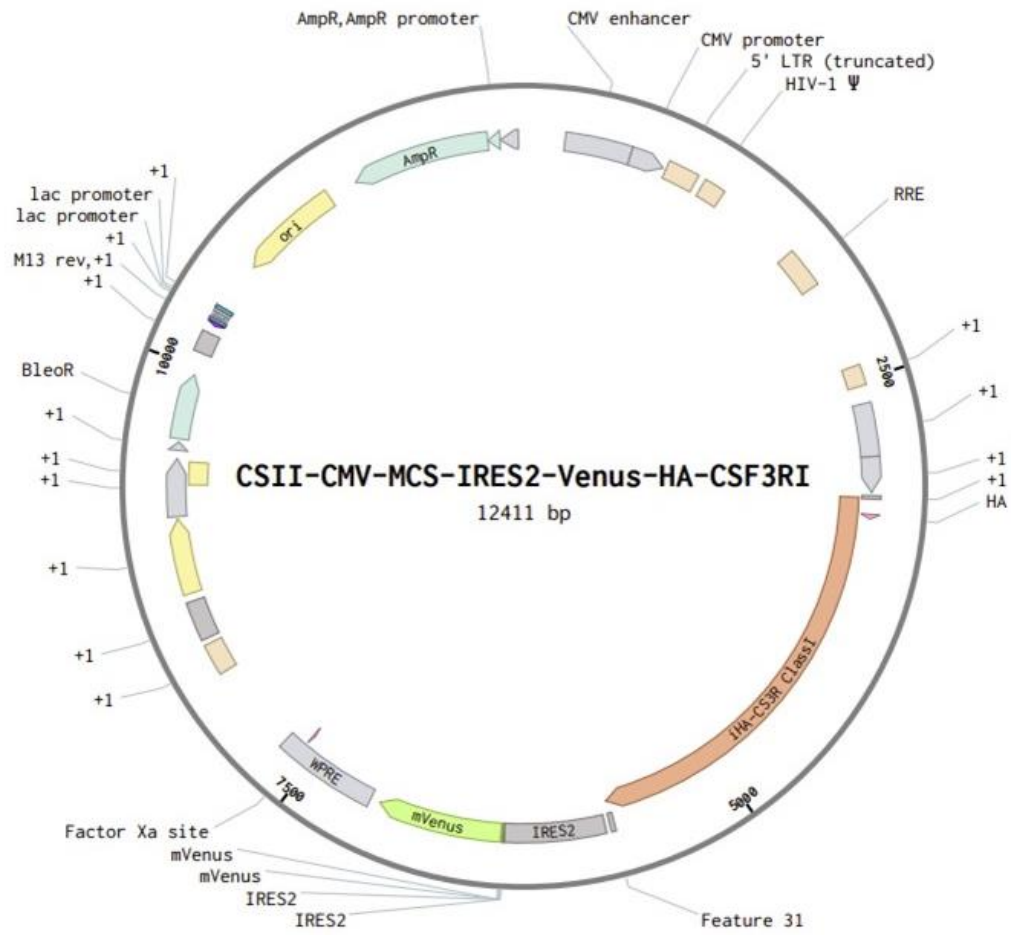
## Appendix C

### Select Gene and Vector Maps

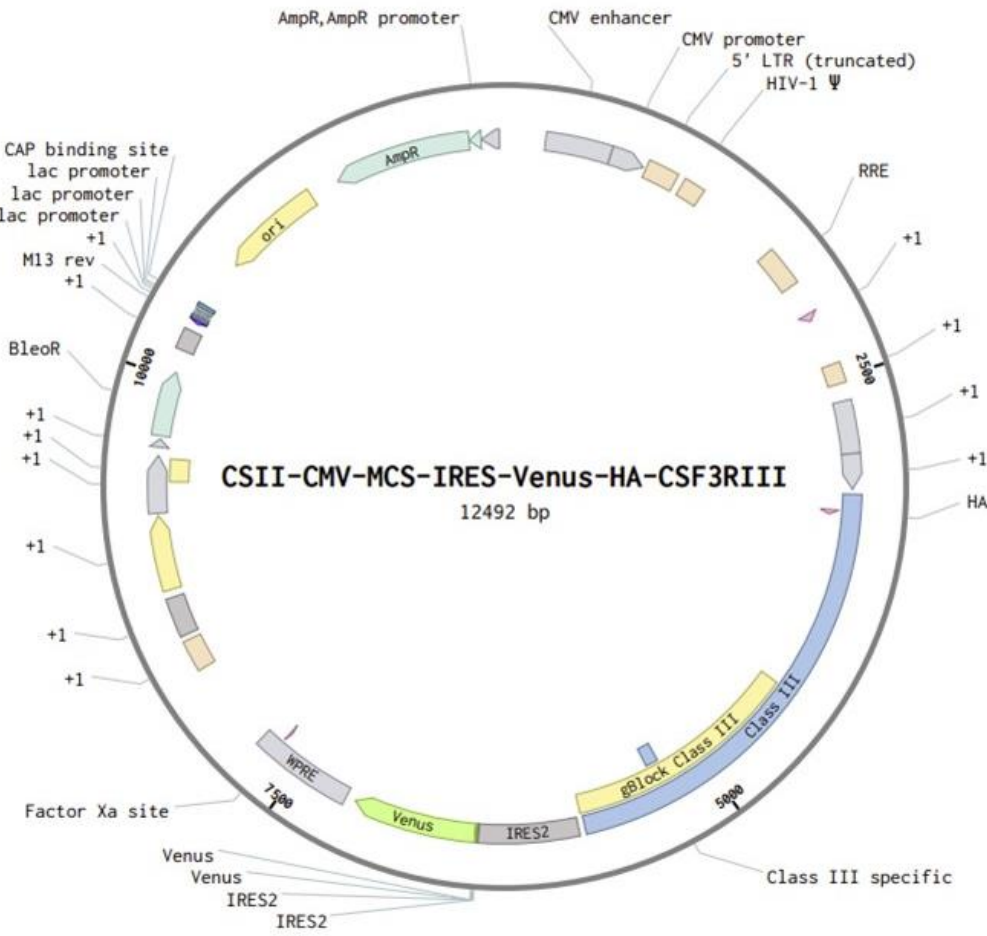
#### hCSF3R RNA all exons full length (3419 bp)



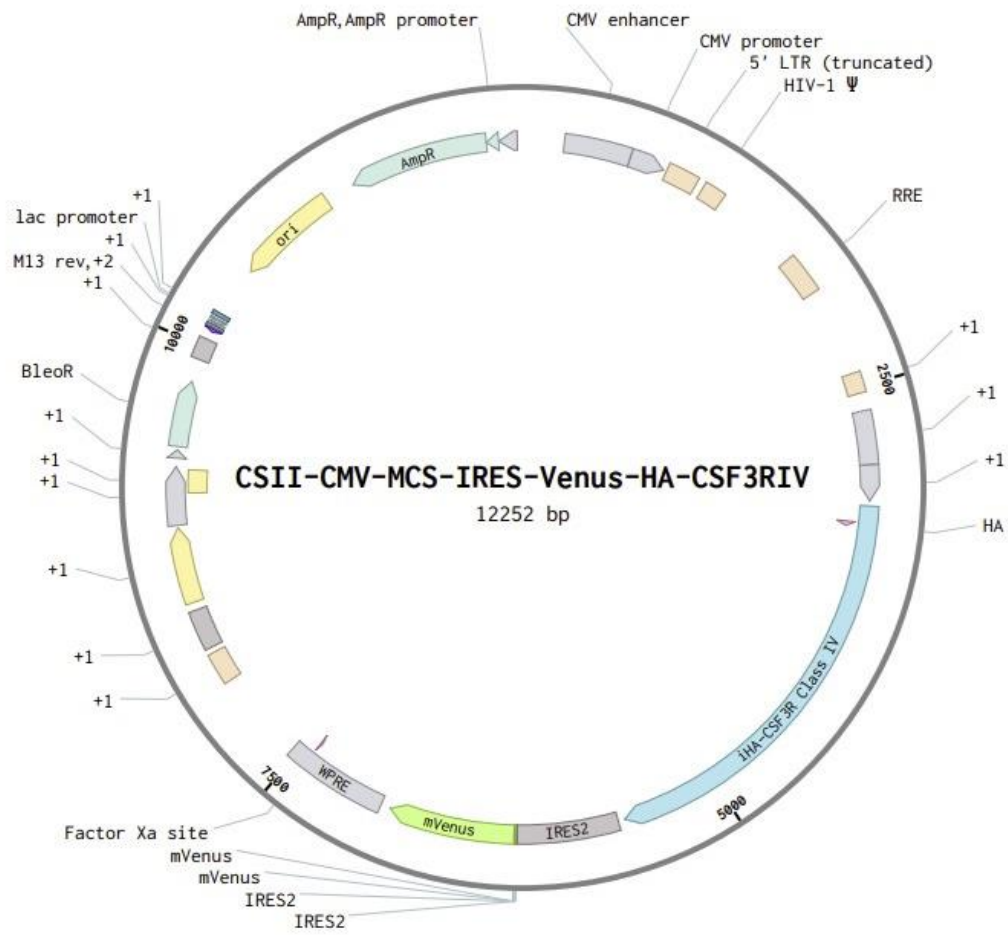
# CSII-CMV-MCS-IRES2-Venus-HA-CSF3RI (12411 bp)



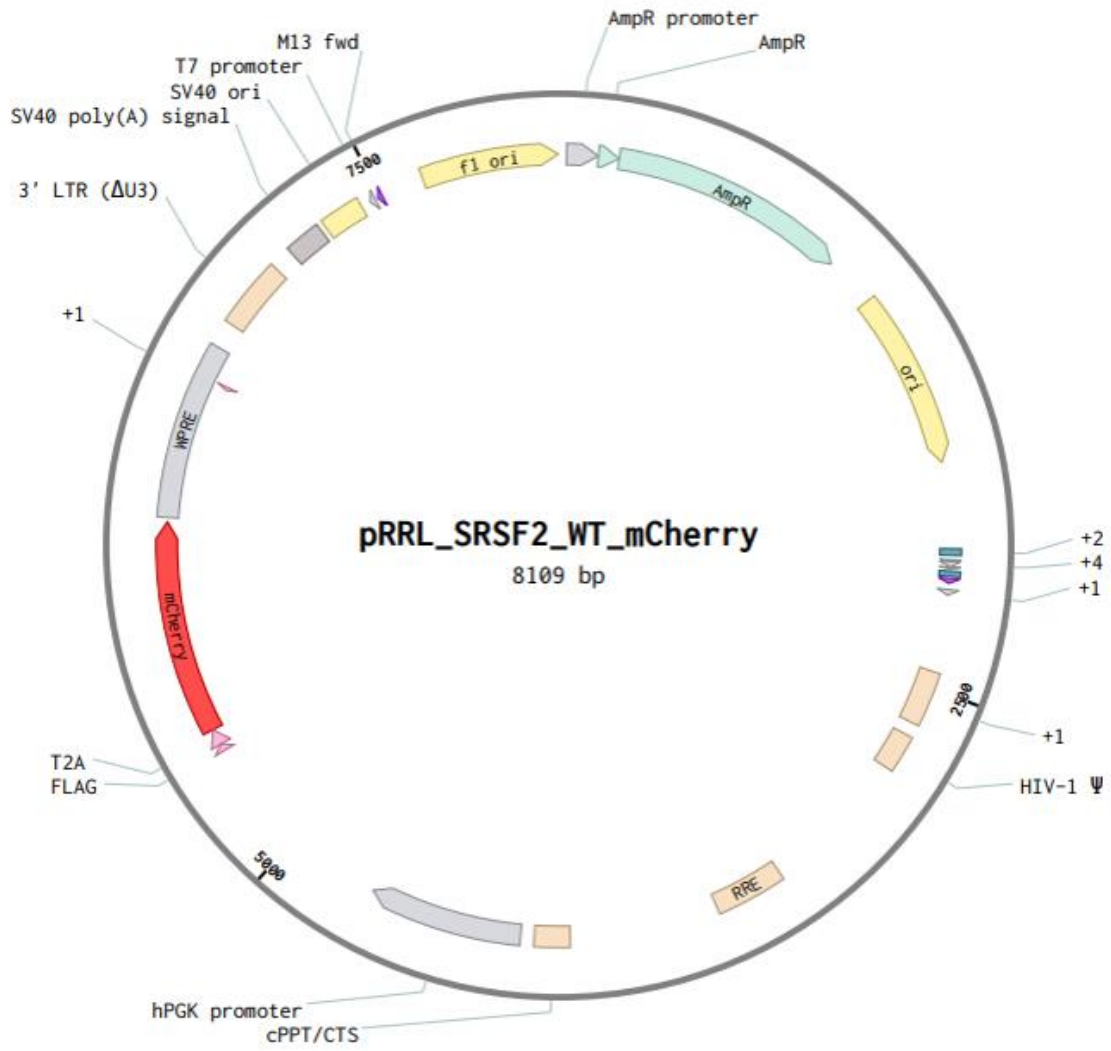
CSII-CMV-MCS-IRES-Venus-HA-CSF3RIII (12492 bp)



# CSII-CMV-MCS-IRES-Venus-HA-CSF3RIV (12252 bp)



# pRRL\_SRSF2\_WT\_mCherry (8109 bp)



## References

1. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64-9.
2. Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. 2012;119(14):3211-8.
3. Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Abu Kar S, Jerez A, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood*. 2012;119(14):3203-10.
4. Cogle CR. Incidence and Burden of the Myelodysplastic Syndromes. *Current Hematologic Malignancy Reports*. 2015;10(3):272-81.
5. Ma X, Does M, Raza A, Mayne ST. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer*. 2007;109(8):1536-42.
6. Ogawa S. Genetics of MDS. *Blood*. 2019;133(10):1049-59.
7. Ogawa S. Genetic basis of myelodysplastic syndromes. *Proc Jpn Acad Ser B Phys Biol Sci*. 2020;96(3):107-21.
8. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28(2):241-7.
9. Voso MT, Gurnari C. Have we reached a molecular era in myelodysplastic syndromes? *Hematology Am Soc Hematol Educ Program*. 2021;2021(1):418-27.
10. Platzbecker U, Kubasch AS, Homer-Bouthiette C, Prebet T. Current challenges and unmet medical needs in myelodysplastic syndromes. *Leukemia*. 2021;35(8):2182-98.
11. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-27; quiz 99.
12. Malcovati L, Papaemmanuil E, Ambaglio I, Elena C, Gallì A, Della Porta MG, et al. Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia. *Blood*. 2014;124(9):1513-21.
13. Nagata S, Fukunaga R. Granulocyte colony-stimulating factor and its receptor. *Progress in Growth Factor Research*. 1991;3(2):131-41.
14. Shinjo K, Takeshita A, Ohnishi K, Ohno R. Granulocyte Colony-Stimulating Factor Receptor at Various Differentiation Stages of Normal and Leukemic Hematopoietic Cells. *Leukemia & Lymphoma*. 1997;25(1-2):37-46.
15. Lieschke GJ, Grail D, Hodgson G, Metcalf D, Stanley E, Cheers C, et al. Mice Lacking Granulocyte Colony-Stimulating Factor Have Chronic Neutropenia, Granulocyte and Macrophage Progenitor Cell Deficiency, and Impaired Neutrophil Mobilization. *Blood*. 1994;84(6):1737-46.
16. Seto Y, Fukunaga R, Nagata S. Chromosomal gene organization of the human granulocyte colony-stimulating factor receptor. *Journal of immunology (Baltimore, Md : 1950)*. 1992;148(1):259-66.
17. Fukunaga R, Ishizaka-Ikeda E, Pan CX, Seto Y, Nagata S. Functional domains of the granulocyte colony-stimulating factor receptor. *EMBO J*. 1991;10(10):2855-65.
18. Ziegler SF, Bird TA, Morella KK, Mosley B, Gearing DP, Baumann H. Distinct regions of the human granulocyte-colony-stimulating factor receptor cytoplasmic domain are required for proliferation and gene induction. *Mol Cell Biol*. 1993;13(4):2384-90.
19. Akbarzadeh S, Ward AC, McPhee DO, Alexander WS, Lieschke GJ, Layton JE. Tyrosine residues of the granulocyte colony-stimulating factor receptor transmit proliferation and differentiation signals in murine bone marrow cells. *Blood*. 2002;99(3):879-87.
20. Nicholson SE, Starr R, Novak U, Hilton DJ, Layton JE. Tyrosine residues in the granulocyte colony-stimulating factor (G-CSF) receptor mediate G-CSF-induced differentiation of murine myeloid leukemic (M1) cells. *J Biol Chem*. 1996;271(43):26947-53.

21. Hermans MHA. Signaling mechanisms coupled to tyrosines in the granulocyte colony-stimulating factor receptor orchestrate G-CSF-induced expansion of myeloid progenitor cells. *Blood*. 2002;101(7):2584-90.
22. Kendrick TS, Lipscombe RJ, Rausch O, Nicholson SE, Layton JE, Goldie-Cregan LC, et al. Contribution of the Membrane-distal Tyrosine in Intracellular Signaling by the Granulocyte Colony-stimulating Factor Receptor. *J Biol Chem*. 2004;279(1):326-40.
23. Yoshikawa A, Murakami H, Nagata S. Distinct signal transduction through the tyrosine-containing domains of the granulocyte colony-stimulating factor receptor. *EMBO J*. 1995;14(21):5288-96.
24. Fukunaga R, Ishizaka-Ikeda E, Nagata S. Growth and differentiation signals mediated by different regions in the cytoplasmic domain of granulocyte colony-stimulating factor receptor. *Cell*. 1993;74(6):1079-87.
25. Nicholson SE, Oates AC, Harpur AG, Ziemiecki A, Wilks AF, Layton JE. Tyrosine kinase JAK1 is associated with the granulocyte-colony-stimulating factor receptor and both become tyrosine-phosphorylated after receptor activation. *Proceedings of the National Academy of Sciences*. 1994;91(8):2985.
26. Tian SS, Lamb P, Seidel HM, Stein RB, Rosen J. Rapid activation of the STAT3 transcription factor by granulocyte colony-stimulating factor. *Blood*. 1994;84(6):1760-4.
27. Tian SS, Tapley P, Sincich C, Stein RB, Rosen J, Lamb P. Multiple signaling pathways induced by granulocyte colony-stimulating factor involving activation of JAKs, STAT5, and/or STAT3 are required for regulation of three distinct classes of immediate early genes. *Blood*. 1996;88(12):4435-44.
28. de Koning JP, Soede-Bobok AA, Schelen AM, Smith L, van Leeuwen D, Santini V, et al. Proliferation Signaling and Activation of Shc, p21Ras, and Myc Via Tyrosine 764 of Human Granulocyte Colony-Stimulating Factor Receptor. *Blood*. 1998;91(6):1924-33.
29. Fukunaga R, Seto Y, Mizushima S, Nagata S. Three different mRNAs encoding human granulocyte colony-stimulating factor receptor. *Proceedings of the National Academy of Sciences of the United States of America*. 1990;87(22):8702-6.
30. Tweardy DJ, Anderson K, Cannizzaro LA, Steinman RA, Croce CM, Huebner K. Molecular cloning of cDNAs for the human granulocyte colony-stimulating factor receptor from HL-60 and mapping of the gene to chromosome region 1p32-34. *Blood*. 1992;79(5):1148-54.
31. Awaya N, Uchida H, Miyakawa Y, Kinjo K, Matsushita H, Nakajima H, et al. Novel variant isoform of G-CSF receptor involved in induction of proliferation of FDCP-2 cells: relevance to the pathogenesis of myelodysplastic syndrome. *J Cell Physiol*. 2002;191(3):327-35.
32. White SM, Alarcon MH, Tweardy DJ. Inhibition of granulocyte colony-stimulating factor-mediated myeloid maturation by low level expression of the differentiation-defective class IV granulocyte colony-stimulating factor receptor isoform. *Blood*. 2000;95(11):3335-40.
33. Larsen A, Davis T, Curtis BM, Gimpel S, Sims JE, Cosman D, et al. Expression cloning of a human granulocyte colony-stimulating factor receptor: a structural mosaic of hematopoietin receptor, immunoglobulin, and fibronectin domains. *Journal of Experimental Medicine*. 1990;172(6):1559-70.
34. White SM, Ball ED, Ehmann WC, Rao AS, Tweardy DJ. Increased expression of the differentiation-defective granulocyte colony-stimulating factor receptor mRNA isoform in acute myelogenous leukemia. *Leukemia*. 1998;12(6):899-906.
35. Mehta HM, Futami M, Glaubach T, Lee DW, Andolina JR, Yang Q, et al. Alternatively spliced, truncated GCSF receptor promotes leukemogenic properties and sensitivity to JAK inhibition. *Leukemia*. 2014;28(5):1041-51.
36. Bernard T, Gale RE, Linch DC. Analysis of granulocyte colony stimulating factor receptor isoforms, polymorphisms and mutations in normal haemopoietic cells and acute myeloid leukaemia blasts. *British journal of haematology*. 1996;93(3):527-33.

37. Wong Justin JL, Ritchie W, Ebner Olivia A, Selbach M, Wong Jason WH, Huang Y, et al. Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. *Cell*. 2013;154(3):583-95.
38. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics*. 2010;11(5):345-55.
39. Keightley M-C, Lieschke GJ. Splicing dysfunction and disease: The case of granulopoiesis. *Seminars in Cell & Developmental Biology*. 2018;75:23-39.
40. Wang B, Mehta H. Cytokine receptor splice variants in hematologic diseases. *Cytokine*. 2020;127:154919.
41. Lance A, Druhan LJ, Vestal CG, Steuerwald NM, Hamilton A, Smith M, et al. Altered expression of CSF3R splice variants impacts signal response and is associated with SRSF2 mutations. *Leukemia*. 2020;34(2):369-79.
42. Sloand EM, Yong ASM, Ramkissoon S, Solomou E, Bruno TC, Kim S, et al. Granulocyte colony-stimulating factor preferentially stimulates proliferation of monosomy 7 cells bearing the isoform IV receptor. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(39):14483-8.
43. Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. 1977;74(8):3171-5.
44. Sharp PA. Split genes and RNA splicing. *Cell*. 1994;77(6):805-15.
45. Sharp PA. The discovery of split genes and RNA splicing. *Trends in Biochemical Sciences*. 2005;30(6):279-81.
46. Larsen A, Davis T, Curtis BM, Gimpel S, Sims JE, Cosman D, et al. Expression cloning of a human granulocyte colony-stimulating factor receptor: a structural mosaic of hematopoietin receptor, immunoglobulin, and fibronectin domains. *The Journal of experimental medicine*. 1990;172(6):1559-70.
47. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010;463(7280):457-63.
48. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry*. 2003;72:291-336.
49. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*. 2009;10(11):741-54.
50. Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*. 2009;136(4):701-18.
51. Dvinge H, Bradley RK. Widespread intron retention diversifies most cancer transcriptomes. *Genome Medicine*. 2015;7(1):45.
52. Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual review of biochemistry*. 2015;84:291-323.
53. Wang Y, Liu J, Huang BO, Xu Y-M, Li J, Huang L-F, et al. Mechanism of alternative splicing and its regulation. *Biomed Rep*. 2015;3(2):152-8.
54. Cáceres EF, Hurst LD. The evolution, impact and properties of exonic splice enhancers. *Genome Biology*. 2013;14(12):R143.
55. El Marabti E, Younis I. The Cancer Spliceome: Reprogramming of Alternative Splicing in Cancer. *Frontiers in molecular biosciences*. 2018;5:80.
56. Sahebi M, Hanafi MM, van Wijnen AJ, Azizi P, Abiri R, Ashkani S, et al. Towards understanding pre-mRNA splicing mechanisms and the role of SR proteins. *Gene*. 2016;587(2):107-19.
57. Phelan MM, Goult BT, Clayton JC, Hautbergue GM, Wilson SA, Lian L-Y. The structure and selectivity of the SR protein SRSF2 RRM domain with RNA. *Nucleic Acids Res*. 2012;40(7):3232-44.
58. Zhou Z, Fu X-D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*. 2013;122(3):191-207.



59. Graveley BR. Sorting out the complexity of SR protein functions. *RNA* (New York, NY). 2000;6(9):1197-211.
60. Ghosh G, Adams JA. Phosphorylation mechanism and structure of serine-arginine protein kinases. *FEBS J*. 2011;278(4):587-97.
61. Giannakouros T, Nikolakaki E, Mylonis I, Georgatsou E. Serine-arginine protein kinases: a small protein kinase family with a large cellular presence. 2011;278(4):570-86.
62. Busch A, Hertel KJ. Evolution of SR protein and hnRNP splicing regulatory factors. 2012;3(1):1-12.
63. Eswaran J, Horvath A, Godbole S, Reddy SD, Mudvari P, Ohshiro K, et al. RNA sequencing of cancer reveals novel splicing alterations. *Sci Rep*. 2013;3:1689.
64. Inoue D, Bradley RK, Abdel-Wahab O. Spliceosomal gene mutations in myelodysplasia: molecular links to clonal abnormalities of hematopoiesis. *Genes & development*. 2016;30(9):989-1001.
65. Darman RB, Seiler M, Agrawal AA, Lim KH, Peng S, Aird D, et al. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell reports*. 2015;13(5):1033-45.
66. Alsafadi S, Houy A, Battistella A, Popova T, Wassef M, Henry E, et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nature Communications*. 2016;7:10615.
67. Cazzola M, Rossi M, Malcovati L. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. *Blood*. 2013;121(2):260.
68. García-Moreno JF, Romão L. Perspective in Alternative Splicing Coupled to Nonsense-Mediated mRNA Decay. *Int J Mol Sci*. 2020;21(24):9424.
69. Shirai Cara L, Ley James N, White Brian S, Kim S, Tibbitts J, Shao J, et al. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer cell*. 2015;27(5):631-43.
70. Ilagan JO, Ramakrishnan A, Hayes B, Murphy ME, Zebari AS, Bradley P, et al. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome research*. 2015;25(1):14-26.
71. Yoshida H, Park S-Y, Sakashita G, Nariai Y, Kuwasako K, Muto Y, et al. Elucidation of the aberrant 3' splice site selection by cancer-associated mutations on the U2AF1. *Nature communications*. 2020;11(1):4744-.
72. Wang H, Guo Y, Dong Z, Li T, Xie X, Wan D, et al. Differential U2AF1 mutation sites, burden and co-mutation genes can predict prognosis in patients with myelodysplastic syndrome. *Scientific Reports*. 2020;10(1):18622.
73. Akef A, McGraw K, Cappell SD, Larson DR. Ribosome biogenesis is a downstream effector of the oncogenic U2AF1-S34F mutation. *PLoS Biol*. 2020;18(11):e3000920-e.
74. Zhang J, Zhao H, Wu K, Peng Y, Han X, Zhang H, et al. Knockdown of spliceosome U2AF1 significantly inhibits the development of human erythroid cells. *J Cell Mol Med*. 2019;23(8):5076-86.
75. Moon H, Cho S, Loh TJ, Jang HN, Liu Y, Choi N, et al. SRSF2 directly inhibits intron splicing to suppresses cassette exon inclusion. *BMB reports*. 2017;50(8):423-8.
76. Meggendorfer M, Roller A, Haferlach T, Eder C, Dicker F, Grossmann V, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood*. 2012;120(15):3080-8.
77. Thol F, Kade S, Schlarmann C, Löffeld P, Morgan M, Krauter J, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 2012;119(15):3578-84.
78. Zhang J, Lieu YK, Ali AM, Penson A, Reggio KS, Rabadan R, et al. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proceedings of the National Academy of Sciences of the United States of America*. 2015;112(34):E4726-34.

79. Kon A, Yamazaki S, Nannya Y, Kataoka K, Ota Y, Nakagawa MM, et al. Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood*. 2018;131(6):621-35.
80. Daubner GM, Cléry A, Jayne S, Stevenin J, Allain FH-T. A syn–anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. 2012;31(1):162-74.
81. Xiao R, Sun Y, Ding J-H, Lin S, Rose DW, Rosenfeld MG, et al. Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol Cell Biol*. 2007;27(15):5393-402.
82. Komeno Y, Huang YJ, Qiu J, Lin L, Xu Y, Zhou Y, et al. SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing. *Mol Cell Biol*. 2015;35(17):3071-82.
83. Singh H, Lane AA, Correll M, Przychodzen B, Sykes DB, Stone RM, et al. Putative RNA-splicing gene LUC7L2 on 7q34 represents a candidate gene in pathogenesis of myeloid malignancies. *Blood cancer journal*. 2013;3:e117.
84. Hosono N, Makishima H, Jerez A, Yoshida K, Przychodzen B, McMahon S, et al. Recurrent genetic defects on chromosome 7q in myeloid neoplasms. *Leukemia*. 2014;28(6):1348-51.
85. Jerez A, Sugimoto Y, Makishima H, Verma A, Jankowska AM, Przychodzen B, et al. Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis. *Blood*. 2012;119(25):6109-17.
86. Daniels NJ, Hershberger CE, Gu X, Schueger C, DiPasquale WM, Brick J, et al. Functional analyses of human LUC7-like proteins involved in splicing regulation and myeloid neoplasms. *Cell reports*. 2021;35(2):108989.
87. Kotini AG, Chang C-J, Boussaad I, Delrow JJ, Dolezal EK, Nagulapally AB, et al. Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. *Nat Biotechnol*. 2015;33(6):646-55.
88. Inaba T, Honda H, Matsui H. The enigma of monosomy 7. *Blood*. 2018;blood-2017-12-822262.
89. Hisasue M, Okayama H, Okayama T, Suzuki T, Mizuno T, Fujino Y, et al. Hematologic abnormalities and outcome of 16 cats with myelodysplastic syndromes. *Journal of veterinary internal medicine*. 2001;15(5):471-7.
90. Weiss DJ, Smith SA. Primary myelodysplastic syndromes of dogs: a report of 12 cases. *Journal of veterinary internal medicine*. 2000;14(5):491-4.
91. Pinheiro D, Mawhin MA, Predecki M, Woollard KJ. In silico analysis of myeloid cells across animal kingdom reveal neutrophil evolution by colony stimulating factors. *eLife*. 2020;9.
92. Yuan Q, Zhou Z, Lindell SG, Higley JD, Ferguson B, Thompson RC, et al. The rhesus macaque is three times as diverse but more closely equivalent in damaging coding variation as compared to the human. *BMC Genetics*. 2012;13(1):52.
93. Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, et al. Toward a Phylogenetic Classification of Primates Based on DNA Evidence Complemented by Fossil Evidence. *Molecular Phylogenetics and Evolution*. 1998;9(3):585-98.
94. Pellagatti A, Armstrong RN, Steeples V, Sharma E, Repapi E, Singh S, et al. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood*. 2018;132(12):1225-40.
95. Pollyea DA, Harris C, Rabe JL, Hedin BR, De Arras L, Katz S, et al. Myelodysplastic syndrome-associated spliceosome gene mutations enhance innate immune signaling. *Haematologica*. 2019;104(9):e388-e92.
96. Chen S, Benbarche S, Abdel-Wahab O. Splicing factor mutations in hematologic malignancies. *Blood*. 2021;138(8):599-612.
97. Gurnari C, Pagliuca S, Visconte V. Alternative Splicing in Myeloid Malignancies. *Biomedicines*. 2021;9(12).

98. Liu M, Wang F, Zhang Y, Chen X, Cao P, Nie D, et al. Gene mutation spectrum of patients with myelodysplastic syndrome and progression to acute myeloid leukemia. *Int J Hematol Oncol*. 2021;10(2):IJH34.
99. Qiu J, Zhou B, Thol F, Zhou Y, Chen L, Shao C, et al. Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators. *RNA (New York, NY)*. 2016;22(10):1535-49.
100. Jiang L, Wang L, Shen C, Zhu S, Lang W, Luo Y, et al. Impact of mutational variant allele frequency on prognosis in myelodysplastic syndromes. *Am J Cancer Res*. 2020;10(12):4476-87.
101. Jiang W, Chen L. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Comput Struct Biotechnol J*. 2020;19:183-95.
102. Hershberger CE, Moyer DC, Adema V, Kerr CM, Walter W, Hutter S, et al. Complex landscape of alternative splicing in myeloid neoplasms. *Leukemia*. 2021;35(4):1108-20.
103. Saez B, Walter MJ, Graubert TA. Splicing factor gene mutations in hematologic malignancies. *Blood*. 2017;129(10):1260-9.
104. Wheeler EC, Vora S, Mayer D, Kotini AG, Olszewska M, Park SS, et al. Integrative RNA-omics Discovers GNAS Alternative Splicing as a Phenotypic Driver of Splicing Factor–Mutant Neoplasms. *Cancer Discovery*. 2022;12(3):836-55.
105. North K, Benbarche S, Liu B, Pangallo J, Chen S, Stahl M, et al. Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. *Nat Biotechnol*. 2022.
106. Schneider-Poetsch T, Chhipi-Shrestha JK, Yoshida M. Splicing modulators: on the way from nature to clinic. *The Journal of Antibiotics*. 2021;74(10):603-16.
107. Marima R, Francies FZ, Hull R, Molefi T, Oyomno M, Khanyile R, et al. MicroRNA and Alternative mRNA Splicing Events in Cancer Drug Response/Resistance: Potent Therapeutic Targets. *Biomedicines*. 2021;9(12):1818.
108. Rahman MA, Nasrin F, Bhattacharjee S, Nandi S. Hallmarks of Splicing Defects in Cancer: Clinical Applications in the Era of Personalized Medicine. *Cancers*. 2020;12(6):1381.
109. Daguinet E, Dujardin G, Valcarcel J. The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO reports*. 2015;16(12):1640-55.
110. Crews LA, Balaian L, Delos Santos NP, Leu HS, Court AC, Lazzari E, et al. RNA Splicing Modulation Selectively Impairs Leukemia Stem Cell Maintenance in Secondary Human AML. *Cell stem cell*. 2016;19(5):599-612.
111. Brunner AM, Steensma DP. Targeting Aberrant Splicing in Myelodysplastic Syndromes: Biologic Rationale and Clinical Opportunity. *Hematol Oncol Clin North Am*. 2020;34(2):379-91.
112. Fujita K-I, Ishizuka T, Mitsukawa M, Kurata M, Masuda S. Regulating Divergent Transcriptomes through mRNA Splicing and Its Modulation Using Various Small Compounds. *Int J Mol Sci*. 2020;21(6):2026.
113. Hautin M, Mornet C, Chauveau A, Bernard D, Corcos L, Lippert E. Splicing Anomalies in Myeloproliferative Neoplasms: Paving the Way for New Therapeutic Venues. *Cancers*. 2020;12(8).
114. Nurtdinov RN, Artamonova II, Mironov AA, Gelfand MS. Low conservation of alternative splicing patterns in the human and mouse genomes. *Human molecular genetics*. 2003;12(11):1313-20.
115. Xiang G, Keller CA, Heuston E, Giardine BM, An L, Wixom AQ, et al. An integrative view of the regulatory and transcriptional landscapes in mouse hematopoiesis. 2020;30(3):472-84.
116. Schmitt CE, Lizama CO, Zovein AC. From transplantation to transgenics: Mouse models of developmental hematopoiesis. *Exp Hematol*. 2014;42(8):707-16.
117. Parekh C, Crooks GM. Critical differences in hematopoiesis and lymphoid development between humans and mice. *J Clin Immunol*. 2013;33(4):711-5.
118. Gao Y, Vasic R, Halene S. Role of alternative splicing in hematopoietic stem cells during development. *Stem Cell Investig*. 2018;5:26-.

119. Hodson DJ, Screen M, Turner M. RNA-binding proteins in hematopoiesis and hematological malignancy. *Blood*. 2019;133(22):2365-73.
120. Bapat A, Keita N, Martelly W, Kang P, Seet C, Jacobsen JR, et al. Myeloid Disease Mutations of Splicing Factor SRSF2 Cause G2-M Arrest and Skewed Differentiation of Human Hematopoietic Stem and Progenitor Cells. *Stem Cells* 2018;36(11):1663-75.
121. Federmann B, Abele M, Rosero Cuesta DS, Vogel W, Boiocchi L, Kanz L, et al. The detection of SRSF2 mutations in routinely processed bone marrow biopsies is useful in the diagnosis of chronic myelomonocytic leukemia. *Human pathology*. 2014;45(12):2471-9.
122. Jourdain AA, Begg BE, Mick E, Shah H, Calvo SE, Skinner OS, et al. Loss of LUC7L2 and U1 snRNP subunits shifts energy metabolism from glycolysis to OXPHOS. *Molecular cell*. 2021;81(9):1905-19.e12.
123. Lee SC, Dvinge H, Kim E, Cho H, Micol JB, Chung YR, et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nature medicine*. 2016;22(6):672-8.
124. Shirai CL, White BS, Tripathi M, Tapia R, Ley JN, Ndonwi M, et al. Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. *Nature Communications*. 2017;8:14060.
125. Hong DS, Kurzrock R, Naing A, Wheler JJ, Falchook GS, Schiffman JS, et al. A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) spliceosome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors. *Investigational New Drugs*. 2014;32(3):436-44.
126. Eskens FALM, Ramos FJ, Burger H, O'Brien JP, Piera A, de Jonge MJA, et al. Phase I Pharmacokinetic and Pharmacodynamic Study of the First-in-Class Spliceosome Inhibitor E7107 in Patients with Advanced Solid Tumors. *Clinical Cancer Research*. 2013;19(22):6296-304.
127. Seiler M, Yoshimi A, Darman R, Chan B, Keaney G, Thomas M, et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nature medicine*. 2018;24(4):497-504.
128. Steensma DP, Wermke M, Klimek VM, Greenberg PL, Font P, Komrokji RS, et al. Results of a Clinical Trial of H3B-8800, a Splicing Modulator, in Patients with Myelodysplastic Syndromes (MDS), Acute Myeloid Leukemia (AML) or Chronic Myelomonocytic Leukemia (CMML). *Blood*. 2019;134(Supplement\_1):673-.
129. Liongue C, Ward AC. Evolution of Class I cytokine receptors. *BMC Evolutionary Biology*. 2007;7(1):120.