

**NORTHWESTERN UNIVERSITY**

CDK-Dependent Phosphorylation and Disease-Associated Mutations in Distinct  
Regions of Splicing Factor 3B1 (SF3B1) Influence its Interaction with Chromatin and  
RNA.

A DISSERTATION

Submitted to the Graduate School in Partial Fulfilment of the requirements

for the degree of

Doctor of Philosophy

Field of Driskill Graduate Program in Life Sciences

By

Tushar Murthy

Evanston, Illinois

December 2018

© Copyright by Tushar Murthy 2018

## ABSTRACT

Splicing factor 3B1 (SF3B1) is a core splicing protein that stabilizes the interaction between the U2 snRNA and the branch point (BP) in the RNA target during splicing. SF3B1 is heavily phosphorylated at its N terminus and a substrate of cyclin-dependent kinases (CDKs). Although SF3B1 phosphorylation coincides with splicing catalysis, the functional significance of SF3B1 phosphorylation is largely undefined. Here, we show that SF3B1 phosphorylation follows a dynamic pattern during cell cycle progression that depends on CDK activity. SF3B1 is known to interact with chromatin, and we found that SF3B1 maximally interacts with nucleosomes during G1/S and that this interaction requires CDK2 activity. In contrast, SF3B1 disassociated from nucleosomes at G2/M, coinciding with a peak in CDK1-mediated SF3B1 phosphorylation. Thus, CDK1 and CDK2 appear to have opposing roles in regulating SF3B1–nucleosome interactions. Importantly, these interactions were modified by the presence and phosphorylation status of linker histone H1, particularly the H1.4 isoform. Performing genome-wide analysis of SF3B1–chromatin binding in synchronized cells, we observed that SF3B1 preferentially bound exons. Differences in SF3B1 chromatin binding to specific sites, however, did not correlate with changes in RNA splicing, suggesting that the SF3B1–nucleosome interaction does not determine cell cycle–dependent changes to mRNA

splicing. Our results define a cell cycle stage–specific interaction between SF3B1 and nucleosomes that is mediated by histone H1 and depends on SF3B1 phosphorylation. Importantly, this interaction does not seem to be related to SF3B1’s splicing function and, rather, points toward its potential role as a chromatin modifier.

The HEAT domain-containing C terminus of SF3B1 contains frequent mutations that have been discovered in several neoplastic processes including myelodysplastic syndrome (MDS). These mutations are almost always hemizygous and mutually exclusive with other splicing factor mutations, suggesting their role as driver mutations during disease pathogenesis. SF3B1 mutations are associated with the selection of a novel BP and cryptic 3’ splice site (ss) upstream of the canonical sequences. However, the precise molecular mechanism by which mutant SF3B1 expression results in altered BP and 3’ss selection remains unclear. While current hypotheses imply that mutant SF3B1 expression results in mis-splicing due to obligate changes in BP and 3’ss selection, such hypotheses are not supported by transcriptomic analyses of clinical samples, cell lines and isogenic murine models. In order to better define factors that influence BP selection by mutant SF3B1-containing spliceosomes, we generated an isogenic murine embryonic stem cell (mESC) line with the hemizygous *Sf3b1* K700E

mutation using CRISPR-Cas9-based gene editing. A control homozygous *Sf3b1* K700K line was also generated for comparison. To enable a statistically robust and unbiased evaluation of BP choice by wild-type (WT) and mutant SF3B1, mESC lines were combined with a massively parallel reporter assay (MPRA) containing millions of splicing reporter minigenes with degenerate sequences in the 3' intronic region. Our results demonstrate that while both WT and mutant SF3B1 utilize canonical BP sequences (YTAAY) at similar frequencies, mutant SF3B1 is capable of preferentially utilizing non-canonical BP sequences that vary at -1 position relative to the BP adenosine. Our findings provide crucial insights into the biochemical mechanism by which mutant SF3B1 expression results in altered splicing.

## ACKNOWLEDGEMENTS

This thesis would not have been possible without my mentor, Dr. Alex Minella. I am grateful to Alex for his mentorship, guidance and his commitment to my success. Alex was extremely patient, always had an open mind, encouraged new ideas and did not let temporary setbacks keep us from achieving our goals. I also want to thank Alex for giving me the freedom to pursue my interest in careers outside the academia.

I am grateful to Dr. Manoj Pillai for providing intellectual and material support and for his guidance that was crucial for the completion of this work. Manoj always made time to answer my questions and patiently guided me through complex bioinformatics analyses.

I am thankful to my thesis committee members Dr. John Crispino, Dr. Chonghui Cheng and Dr. Seth Corey for their insights and suggestions on how to make my research more impactful.

Special thanks to Minella lab members: Dr. Tanushri Sengupta, Dr. Ka Tat Siu, Dr. Alison Meyer, Theresa Bluemn, Dr. Kelsey Swartz and Scott Wood. Kelsey and Scott were my support system in and outside the lab. They were always there for me whenever I need their help; whether it was brainstorming about an experiment, taking care of my experiments during the weekends when I was away in Chicago to see my wife or when I really needed company to go get hot chicken wings.

I owe a special thanks to Dr. Sridhar (Sid) Rao. Sid treated me like I was his own graduate student and made me feel like a part of his lab, especially after Alex left for

Amgen. Sid's optimistic attitude is an inspiration and I would consider it a success if I could be half as optimistic as him. I also want to thank every member of his lab for their support and especially Mike Reimer, who taught me everything there is to know about Next-Gen sequencing. I cannot forget the support and generosity shown to me by Dr. Gilbert White and everyone at the BloodCenter of Wisconsin without whom it would have been impossible for me to complete my thesis research.

I don't think I could have made it through graduate school without the love and support of special friends like Soham Thorat, Dr. Kshitij Khatri, Sushant Dhiman, Omkar Bhapkar and Dr. Jesus Izaguirre who were there with me through thick and thin.

I want to thank my family, and especially my parents and sister, for their unconditional love and support and for always believing in me. My mother taught me the value of patience which helped me get through testing times. I would not have pursued a PhD if it was not for my father who encouraged me to apply to graduate school when I was unsure of myself. My father has taught me, through example, to never give up. I will never forget his encouraging words over the years. I also want to thank my mother and father-in-law, Cindy and Robert Warner, for their love and generosity.

I owe a huge debt of gratitude to my wife Mallory for going through this experience with me. Thank you for supporting me and my ambitions, being there for me through all the ups and downs and making me believe in myself. Thank you for all the sacrifices you've made while we lived apart for 4 years.

Lastly, I want to dedicate this work to my grandmother who passed away in 2017. I wish she was here to share this important moment with me.

## PREFACE

Chapter 1 briefly introduces basic concepts in RNA splicing, the function of the spliceosome machinery during splicing, the current state of our knowledge regarding the role of recurrent mutations in splicing factor 3B 1 (SF3B1) in promoting disease pathogenesis and the contribution of cell cycle kinases-mediated phosphorylation to SF3B1 function and regulation.

Chapter 2 is derived from an original publication titled 'Cyclin-dependent kinase 1 (CDK1) and CDK2 have opposing roles in regulating interactions of splicing factor 3B1 with chromatin', with minor modifications (1). Chapter 3 includes data derived from the original manuscript titled 'Degenerate minigene library analysis enables identification of altered branch point choice by mutant splicing factor 3B1 (SF3B1)', currently under review for publication. Experimental procedures used to generate data for Chapter 2 and 3 are provided at the end of the respective chapters.

Chapter 2 and Chapter 3 end with their own 'Conclusion' section that provides context for the data with respect to existing knowledge in the field and provides future directions to build upon these findings.

## TABLE OF CONTENTS

Abstract		3
Acknowledgements		6
Preface		8
<b>Chapter 1</b>		
1. Introduction		
1.1	Overview of splicing	10
1.2	Spliceosome mutations in disease	17
1.3	The role of SF3B1 mutations in MDS pathogenesis	20
1.4	Cyclins, Cyclin Dependent Kinases and Cell Cycle Regulation	23
1.5	SF3B1 phosphorylation by cell cycle kinases	26
<b>Chapter 2</b>		
<b>CDK-regulated SF3B1-chromatin interactions</b>		
2. Results:		
2.1	SF3B1 phosphorylation is dynamic during cell cycle progression	31
2.2	SF3B1-nucleosome interactions are dynamic during cell cycle progression	36
2.3	SF3B1-chromatin interactions are dependent on CDK activity	43
2.4	SF3B1 K700E mutation alters chromatin binding	53
2.5	SF3B1-chromatin interactions in vitro depend on linker histone H1 and phosphorylation of both SF3B1 and H1	55
2.6	Integrated analysis of SF3B1 genome occupancy and cell cycle-dependent splicing	62
	<b>Conclusions</b>	74
	<b>Experimental Procedures</b>	79
<b>Chapter 3</b>		
<b>Study of mutant SF3B1-induced altered splicing</b>		
3. Results		
3.1	Generation of isogenic Sf3b1-mutant mouse embryonic stem cells (mESCs)	89
3.2	Aberrant splicing of endogenous transcripts in <i>Sf3b1</i> K700E mESCs	97
3.3	Massively parallel reporter assay (MPRA) for the unbiased analysis of BP usage in mutant Sf3b1 expressing mESCs	101
	<b>Conclusions</b>	108
	<b>Experimental Procedures</b>	115
	<b>References</b>	121

## Chapter 1

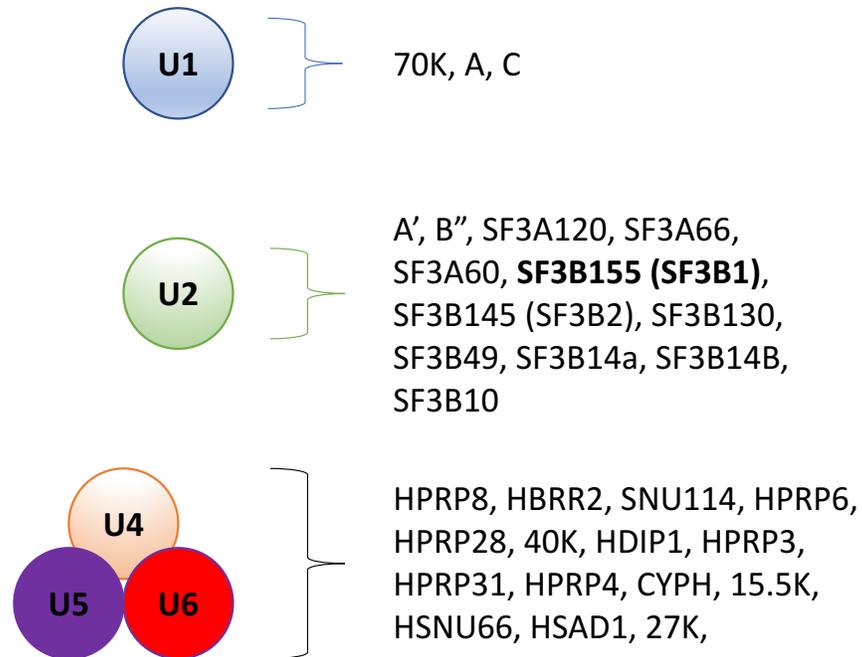
### Introduction

#### 1.1 Overview of RNA splicing:

A majority of eukaryotic genes contain coding sequences (exons) that separated by non-coding sequences (introns). These genes are expressed as precursor mRNAs (pre-mRNAs) that contain both introns and exons. In order to translate information contained within the expressed mRNA molecule into proteins, pre-mRNAs are processed into mRNA by the removal of introns and ligation of exons by a process known as RNA splicing. While most introns are removed and most exons are incorporated into the spliced mRNA, many pre-mRNAs are alternatively spliced to produce mRNA variants in which different combinations of exons are ligated together. In some instances, introns are retained within the mature mRNA. More than 75% of genes are estimated to undergo alternative splicing (AS) and 50% of these AS events result in altered reading frames (ARFs) (2). ARFs allow a single stretch of DNA to be transcribed in more than one reading frame. Thus, by enabling a single gene to encode for multiple protein isoforms, splicing can increase the complexity of the proteome. Given the important

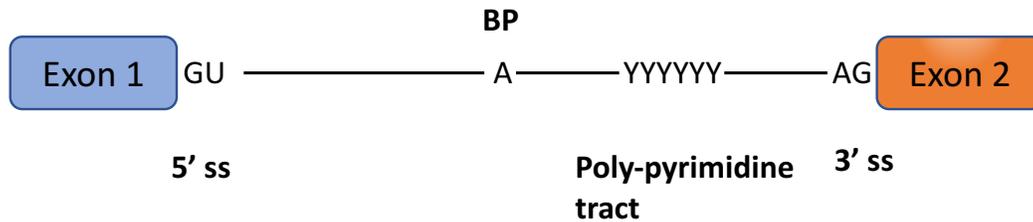
downstream implications of this process, RNA splicing has evolved to be a tightly regulated process.

In eukaryotes, RNA splicing is achieved with the help of a large number of proteins and multi-protein complexes called the spliceosome machinery. Among the many proteins and complexes within the spliceosome machinery, large multi-protein-RNA complexes known as small nuclear ribonucleoproteins (snRNPs) play a major role. In most eukaryotes, two unique spliceosome machineries consisting of unique proteins, complexes and snRNPs exist: the less abundant U12-type spliceosome that splices a small and rare class of introns and the U2-type spliceosome that splices a majority of introns (3). For the purposes of this thesis, we will focus on the U2-type spliceosome. The U2-type spliceosome consists of five snRNP molecules viz. U1, U2, U4, U5 and U6 (Figure 1). Each snRNP consists of a distinct set of proteins complexed with a unique RNA molecule. Together, through the systematic binding and dissociation from pre-mRNA, these snRNPs help orchestrate RNA splicing.



**Figure.1.** Schematic showing the five snRNP complexes and their constituent proteins that, together, play a critical role in the splicing of U2 type introns. SF3B1, our protein of interest, is represented in bolded text. U4, U5, U6 are shown in a trimeric complex that associates with the pre-mRNA during splicing.

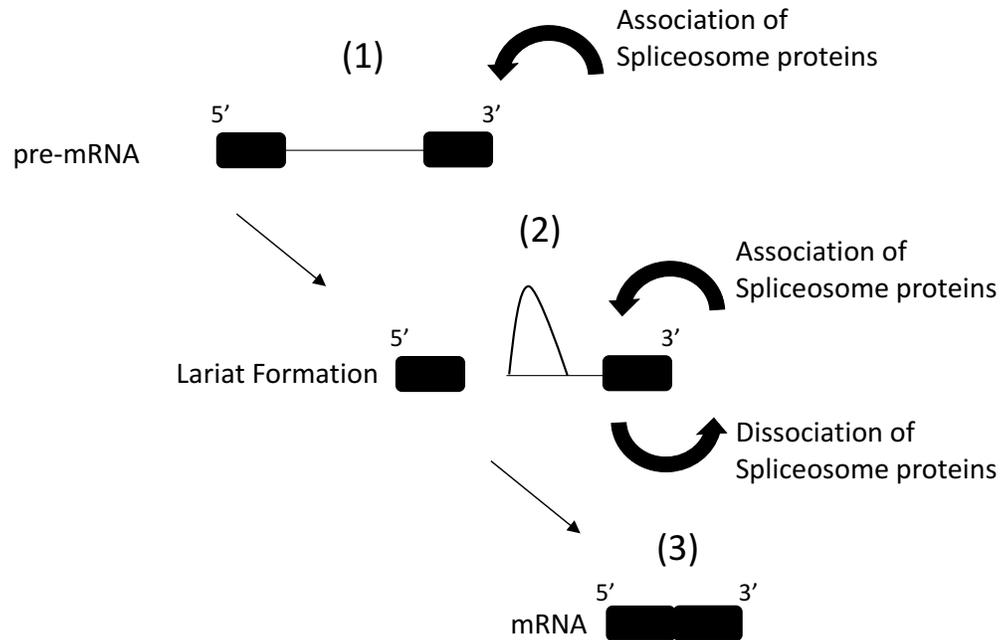
The assembly of the spliceosome machinery on the pre-mRNA molecule relies on short conserved sequences contained within the pre-mRNA (Figure 2). In the case of eukaryotes, a GU dinucleotide sequence at the 5' end of an intron represents the 5' splice-site (5'ss) and an AG dinucleotide sequence at the 3' end of an intron represents the 3' splice-site (3'ss). Located 18-40 nucleotides upstream of the 3'ss is the branch-point (BP) which, in most cases, is an adenosine nucleotide. Between the 3'ss and the BP lies a stretch of pyrimidines. This stretch is called the poly-pyrimidine tract. Other *cis* acting elements within the pre-mRNA include motifs within introns and exons known as exonic splice enhancers (ESEs) and intronic splice enhancers (ISEs) (3).



**Figure.2.** Schematic showing conserved sequences within the pre-mRNA that facilitate the assembly of spliceosome proteins and influence splicing outcomes. Shown are two exons (Blue and Orange boxes) are separated by an intron represented by the black line. Conserved 5' splice site (5'ss) GU dinucleotide, 3' splice site (3'ss) AG dinucleotide, the branch-point (BP) adenosine (A) and the poly-pyrimidine tract (YYYYYY) are shown within the intron.

The process of splicing begins with intron and exon definition that dictates which introns are removed and which exons are ligated. The first step in this process is the formation of the E complex, where U1 snRNP binds the 5'ss. Formation of the E complex is followed by the binding of U2 auxiliary factor 1 (U2AF1) and U2 auxiliary factor 2

(U2AF2) to the 3'ss and the poly-pyrimidine tract, respectively. Next, the recruitment and stable binding of U2 snRNP to the BP results in the formation of the A complex. The assembly of spliceosome proteins on pre-mRNA early during the process of splicing greatly influences splicing outcomes. During these early steps the spliceosome machinery determines which introns are to be removed and which exons are to be spliced. After the formation of the A complex, the pre-assembled U4/5/6 tri-snRNP complex binds the pre-mRNA and gives rise to the catalytic B complex. Through a series of conformational changes and RNA-protein binding changes, U1 and U4 snRNPs dissociate from the pre-mRNA and thus give rise to the activated B complex ( $B^{act}$ ). The B complex is further catalytically activated by the binding of the DEAD-box RNA helicase PRP2 which generates the  $B^*$  complex. In the first catalytic step, the 2' hydroxyl group (OH) of the BP adenosine carries out a nucleophilic attack on the 5'ss. This results in cleavage of the 5'ss and ligation of the 5' end of the intron to the BP adenosine forming a lariat which yields the C complex. In the second catalytic step, the 3'ss undergoes a nucleophilic attack from the 3' OH of the 5' exon that leads to their ligation and the removal of the intron. The spliceosome then dissociates and individual components are recycled to take part in subsequent rounds of splicing.



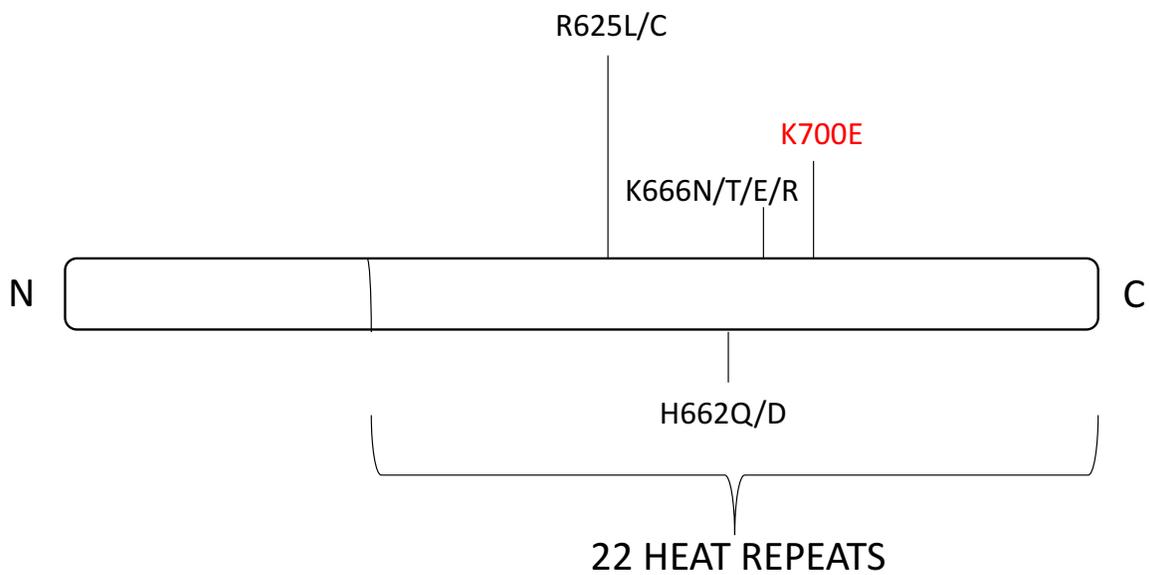
**Figure 3.** Schematic showing the 3 major steps during splicing catalysis. (1) Spliceosome proteins assemble on the pre-mRNA where exons are represented by black boxes and introns by the black line separating them. (2) A series of steps involving the association and dissociation of spliceosome proteins from the mRNA culminate in the formation of the lariat resulting from the ligation of the 5'OH of the intron to the BP adenosine. (3) In the final steps of splicing, introns are removed and exons are ligated together. Spliceosome proteins are recycled for use in the next round of splicing.

## 1.2 Spliceosome mutations in disease:

Recently, it has been discovered that a number of genes encoding spliceosome proteins, active during the early stages of splicing, are mutated in cancers (4). These spliceosome proteins include: U2AF1 and U2AF2 that make contact with the 3'ss and the poly-pyrimidine tract, respectively; Serine/Arginine rich splicing factor 2 (SRSF2) that binds ESEs; Zinc Finger CCCH-Type, RNA Binding Motif and Serine/Arginine Rich 2 (ZRSR2) that interacts with U2AF proteins; and Splicing Factor 3B1 (SF3B1) that is found within the multi-protein U2 snRNP complex.

SF3B1 is a core splicing factor and stabilizes the interaction between the BP adenosine and the U2 snRNP (5,6). Whole-genome sequencing has identified recurrent SF3B1 mutations in multiple neoplastic processes. SF3B1 hotspot mutations are found in uveal melanoma, chronic lymphocytic leukemia (CLL) and breast cancer, albeit at lower frequencies. SF3B1 mutations are found at high frequencies in myelodysplastic syndrome (MDS) (4,7-10). MDS is a group of clonal stem cell disorders that are characterized by ineffective hematopoiesis and cytopenia. Of the different MDS subtypes, SF3B1 mutations are most frequently associated with refractory anemia with ring-sideroblasts (RARS) MDS subtype that is characterized by anemia and characteristic

morphologic atypia, with immature erythroid progenitors containing peri-nuclear, iron laden mitochondria (4,7). Of note, 75% of all MDS patients with ring sideroblasts (RS) were found to harbor SF3B1 mutations, while only 6% of MDS patients without RS were found to harbor SF3B1 mutations. SF3B1 mutations are almost always hemizygous and mutually exclusive i.e. they rarely occur in combination with other driver mutations found in MDS patients. Majority of SF3B1 mutations are found to be contained within the C-terminus HEAT-domains of the SF3B1 protein and are missense mutations. These include K666N, K666Q, K666R and K700E (Figure 4). The K700E mutation, resulting from an A→G substitution, is the most frequently occurring SF3B1 mutation (7).



**Figure 4.** Diagram showing the most frequently found SF3B1 mutations and their locations within the SF3B1 protein.

### **1.3 The role of SF3B1 mutations in MDS pathogenesis:**

The prevailing hypothesis in the field of MDS biology is that SF3B1 mutations result in altered splicing of genes that are critical during the differentiation and maturation of hematopoietic stem cells and give rise to a clonal stem cell population with altered differentiation potential. This hypothesis has been supported by work that has demonstrated that CD34+ hematopoietic stem cells from MDS patients with SF3B1 mutations show aberrant splicing of genes involved in differentiation, iron homeostasis, mitochondrial metabolism and RNA splicing (11). Subsequent studies that analyzed splicing at the transcriptomic level by performing RNA sequencing (RNA-seq) of CLL, uveal melanoma and breast cancer patient samples demonstrated that expression of mutant SF3B1 induces alternate 3'ss selection through the use of an alternate BP (12). This alternate 'cryptic' 3'ss was shown to be 19-20nt upstream of the canonical 3'ss and a large number of aberrantly spliced mRNAs were shown to be degraded by non-sense-mediated decay (NMD) (12,13). These findings have been validated in MDS patient samples by us and other groups (14-16).

Since HEAT domains play a role in mediating protein interactions, it was hypothesized that mutations within the N-terminus HEAT domains of SF3B1 alter the protein structure

in a way that disrupts its interaction with other proteins in the spliceosome machinery leading to altered splicing and disease pathogenesis. A structure-based analysis of SF3B1 protein has provided critical insights into the role of HEAT domains in facilitating SF3B1 interaction with other proteins in the SF3B complex and other spliceosome proteins like U2AF65. This study revealed that frequently occurring mutations within SF3B1 HEAT domains alter the three-dimensional conformation of the HEAT domains and SF3B1 interactions with U2AF65, potentially influencing splicing outcomes. However, the authors also demonstrated that the SF3B1 K700E mutation, specifically, does not alter SF3B1-U2AF65 interactions. This suggests that the mechanism behind aberrant splicing due to SF3B1 K700E expression is more complex than previously imagined (17).

More recently, two groups have reported findings from knock-in mouse models of *Sf3b1* K700E (15,16). It was reported that mutant *Sf3b1* mice develop macrocytic anemia due to ineffective erythropoiesis and defects within the hematopoietic stem cell (HSC) compartment. However, in both studies, mice developed anemia without the presence of RS which is the hallmark of human disease with SF3B1 mutations. While aberrant splicing with cryptic 3'ss usage and increased NMD was reported in both *Sf3b1*

K700E mouse models, there was no significant overlap between aberrantly spliced transcripts in mice versus human MDS samples. This suggests that there are key differences in splicing between humans and mice. Indeed, we have reported that while there is significant conservation of exonic sequences between humans and mice, intronic regions are less conserved (14). We also demonstrated that in a number of mRNA transcripts, aberrantly spliced due to cryptic 3'ss usage, these cryptic 3'ss were found within RNA secondary structures. Cryptic 3'ss within RNA secondary structures were accessible by mutant SF3B1 but not wild-type (WT) SF3B1. Together, these findings suggest that a more systematic and unbiased study of splicing is required if we are to develop a proper understanding of MDS pathogenesis due to mutant SF3B1.

While significant advances have been made in understanding the role of SF3B1 mutations in MDS pathogenesis, key questions remain unanswered. Isoforms generated due to cryptic 3'ss represent only a small proportion of the total transcript. This finding is contrary to current hypotheses that imply that hemizygous SF3B1 mutations are absolute and cause mis-splicing due to obligate changes in BP and 3'ss selection. Also, it is not clear what factors lead to the cryptic 3'ss and alternate BP selection. Are there sequence and structural determinants within the pre-mRNA that

predispose certain transcripts to aberrant splicing by mutant SF3B1? Can the identification of the sequence and structure determinants be used to predict and identify genes and pathways most likely to be affected by the expression of mutant SF3B1? Can the lack of conservation of these sequence and structural determinants between humans and mice explain why mouse models do not replicate human disease? Answers to these questions will provide critical insights that can help pinpoint the exact mechanism by which SF3B1 mutations contribute to MDS pathogenesis.

#### **1.4 Cyclins, Cyclin Dependent Kinases and Cell Cycle Regulation:**

The progress of dividing cells through different stages of the cell cycle is a tightly regulated process. This process relies on the presence and activity of two protein families: cyclins and cyclin-dependent kinases (CDKs). CDKs are serine/threonine kinases whose activity depends on their binding to their respective cyclin binding partners. Each member of the CDK family of proteins contains a conserved catalytic domain consisting of an ATP-binding pocket, a cyclin-binding domain and an activation loop motif. CDKs are activated as well as inhibited by phosphorylation at distinct serine and threonine residues. Cell cycle stage-specific association of different cyclins with

CDKs allows the timely transition and progression of cells from one stage of the cell cycle to another. While it was originally believed that each cyclin and CDK protein has a unique function and is essential for normal progression through the cell cycle, studies that utilized knock-out mouse models revealed an extensive compensatory mechanism where the loss of one cyclin or CDK is compensated by the expression and activity of another cyclin or CDK.

Early in G1 phase of the cell cycle, cyclin D binds and activates CDK4/6 and initiates the phosphorylation of the retinoblastoma protein (RB). Phosphorylation of RB results in the dissociation of RB-bound E2F transcription factors which initiate the expression of genes required for cell cycle progression. Among the E2F responsive genes is cyclin E. Cyclin E binds to CDK2 in late G1 and hyperphosphorylates RB. This leads to further activation of E2F-mediated gene expression and entry into S-phase. Cyclin A-CDK2 promote DNA replication during S phase and cyclin A-CDK1 activity then allows entry into G2/M, followed by cyclin B-CDK1-mediated completion of mitosis (18,19).

Originally thought to be master regulators of cell cycle, cyclin-CDK complexes have been shown to play additional roles beyond the cell cycle. These include the regulation of transcription, DNA damage response, chromatin organization and RNA splicing.

Phosphorylation of RNA polymerase II C-terminal domain has been shown to regulate its activity during transcription initiation and elongation (20). Cyclin-CDK activity has been shown to be essential for the resolution of DNA breaks by the DNA damage response machinery (21). Several cyclins are expressed in the nucleus and have been shown to phosphorylate chromatin-associated proteins in complex with CDKs. For example, CDK-mediated phosphorylation of linker histone H1 contributes to the regulation of chromatin organization and structure (22,23). EZH2, an epigenetic regulator that catalyzes the methylation of histone proteins, is a cyclin-CDK1/2 substrate and undergoes phosphorylation at distinct threonine residues in a cell cycle dependent manner (24). Furthermore, cyclin E associates with spliceosome machinery proteins and their phosphorylation by CDK1/2 has been shown to be associated with splicing catalysis (25-27). These findings have led to the recognition of cyclins and CDKs as multifaceted proteins with important roles in the regulation of cell cycle and beyond.

### 1.5 SF3B1 phosphorylation by cell cycle kinases:

Nearly 20 years ago it was reported that cyclin E associates with U2 snRNP and its component proteins including SF3B1, SF3B2, and SF3A1. SF3B1 has been shown to be phosphorylated *in vitro* by CDK1 and CDK2 that are active during G1/S and G2/M phases of the cell cycle, respectively (25-27). Subsequent studies demonstrated that another cell cycle kinase, DYRK1A, also phosphorylates SF3B1. These findings suggest a strong link between cell cycle and the spliceosome machinery (28). Mass-spectrometry (MS) studies have shown that the N-terminus domain of SF3B1 contains nearly 31 Serine/Threonine-Proline (S/T-P) CDK consensus sites. These findings have complicated efforts aimed at understanding the precise outcomes of cell cycle-dependent phosphorylation of SF3B1 (Figure 5) (29). While a large majority of these sites have not been experimentally validated, a number of them have been shown to be phosphorylated *in vivo* and to be associated with specific functions. SF3B1 phosphorylated at threonine 313 (T313) is found in nuclear speckles and associated with active spliceosomes (30). This finding is particularly interesting as splicing is known to follow a specific cell cycle-dependent program that results in alternative splicing of key transcripts that play important roles during cell cycle progression. Moreover, it

suggests that phosphorylation of SF3B1 by cell cycle kinases may be important in facilitating the cell cycle stage-specific program of alternative splicing (31). Threonines 244, 248 and 313 are known to be phosphorylated by CDK2 and phosphorylation at these sites has been shown to be responsible for mediating the interaction between SF3B1 and a chromatin associated protein called nuclear inhibitor of protein phosphatase 1 (NIPP1) (27). The phosphorylation-dependent interaction of SF3B1 with a chromatin associated protein is not entirely surprising. A large body of evidence suggests that splicing occurs co-transcriptionally and members of the RNA splicing machinery interact with chromatin during transcription (32). Indeed, Kfir et al. recently demonstrated that SF3B1 interacts with nucleosomes near exons in an RNA-independent manner. Using genome-wide occupancy data for SF3B1 in combination with splicing analyses and knock-down approaches, the authors contend that SF3B1 occupancy of chromatin determines splicing outcomes (33). Given that CDKs are also known to interact with chromatin and phosphorylate various chromatin-associated proteins, it led us to consider the possibility that phosphorylation of SF3B1 by cell cycle kinases during cell cycle progression regulates its association with chromatin and in turn

influences the cell cycle stage specific program of splicing near SF3B1-occupied regions of the genome.

Figure 5.

#	SITE	MOTIF	#	SITE	MOTIF
1	S129	Il <u>s</u> PE	18	T326	GE <u>t</u> Pt
2	T142	GK <u>t</u> PD	19	T328	tP <u>t</u> PG
3	T207	DQ <u>t</u> PG	20	T341	DE <u>t</u> PA
4	T211	GA <u>t</u> PK	21	T350	Gst <u>t</u> PV
5	T223	AE <u>t</u> PG	22	T354	VL <u>t</u> PG
6	T227	GH <u>t</u> Ps	23	T362	IG <u>t</u> PA
7	T235	DE <u>t</u> PG	24	T369	MA <u>t</u> Pt
8	T244*	sE <u>t</u> PG	25	T371	tP <u>t</u> PG
9	T248*	GA <u>t</u> PG	26	T379	sM <u>t</u> PE
10	T257	DP <u>t</u> Ps	27	T426	IR <u>t</u> PA
11	T261	sH <u>t</u> PA	28	T434*	tA <u>t</u> Pt
12	T267	AA <u>t</u> PG	29	T436	tP <u>t</u> PL
13	T273	GD <u>t</u> PG	30	S488	TL <u>s</u> PE
14	T278	HA <u>t</u> PG	31	T508	NG <u>t</u> PP
15	T296	DE <u>t</u> PK	32	S541	LM <u>s</u> PT
16	T303	RD <u>t</u> PG	33	T1021	RL <u>t</u> PI
17	T313*	AE <u>t</u> PR	34	T1170	AV <u>t</u> PL

**Figure 5.** Proline-directed serine/threonine phosphorylation sites in SF3B1 listed by PhosphositePlus tool (34) and previously identified by mass spectrometry (MS). All serine and threonine residues are shown in lowercase letters. Predicted CDK phosphosites have been underlined. (\*) indicates that MS data was confirmed in studies utilizing site-specific methods (27-29). Six threonine residues were mutated to alanine in order to understand their impact on SF3B1-nucleosome interactions and are denoted by bolded and italicized text.

The goal of this thesis is to understand how hotspot mutations in the C-terminus domain of SF3B1 affect its function during splicing and contribute to disease pathogenesis. Another goal is to determine whether the phosphorylation of residues within the N-terminus domain of SF3B1 by cell cycle kinases plays a role in orchestrating the cell cycle stage-specific program of alternative splicing. The following chapters will discuss our systematic approach to answering questions surrounding SF3B1 mutations in disease pathogenesis and the phosphorylation-dependent regulation of SF3B1 function during splicing.

## Chapter 2: Results

### CDK-regulated SF3B1-chromatin interaction

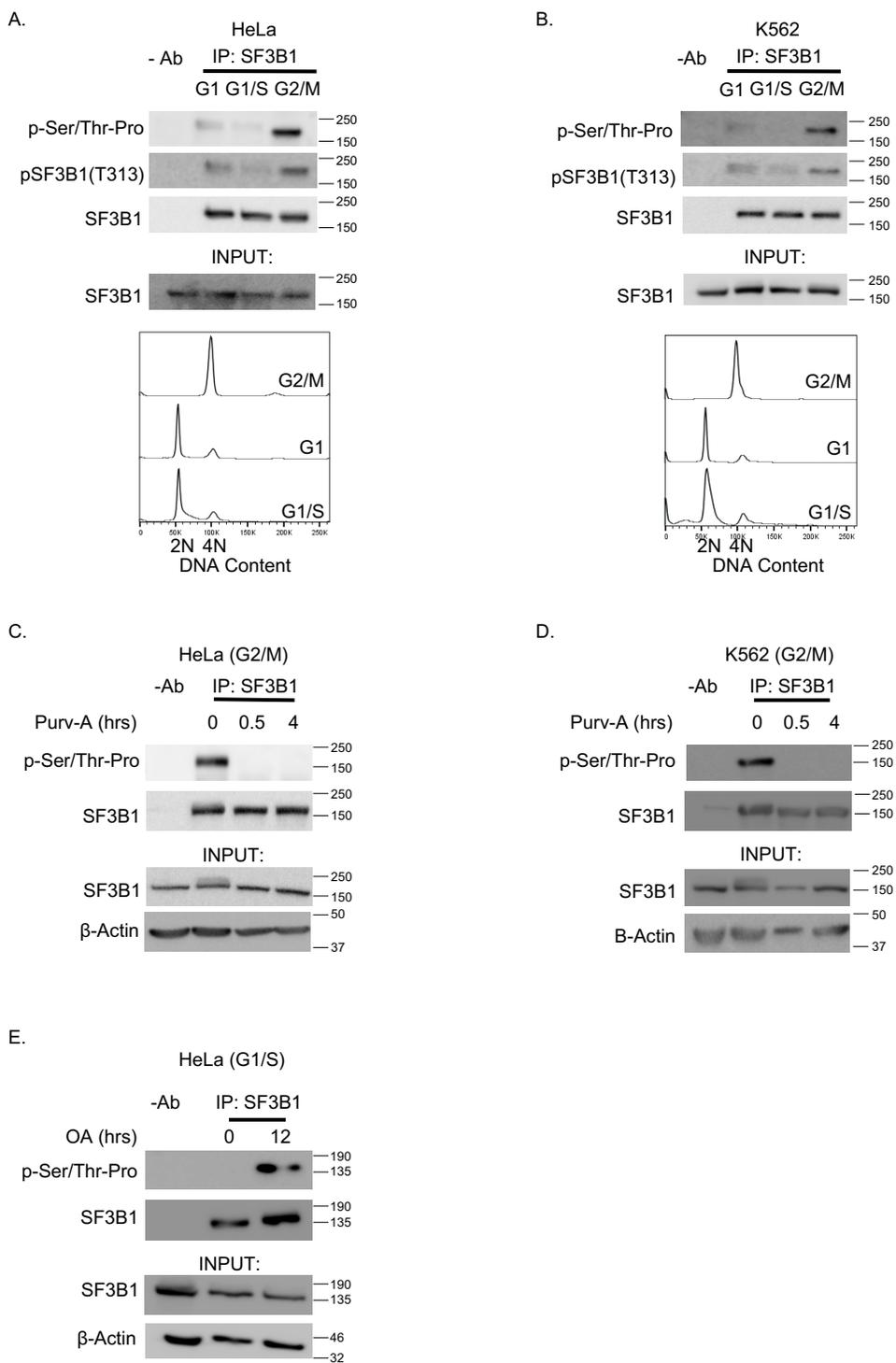
#### 2.1 SF3B1 phosphorylation is dynamic during cell cycle progression.

SF3B1 contains numerous serine or threonine residues juxtaposed to a proline at the +1 position, representing potential sites for phosphorylation by proline-directed kinases, including cyclin-CDKs, glycogen synthase kinase 3 (GSK3), and mitogen-activated protein kinases (MAPK). Although several mass spectrometry (MS) studies have demonstrated SF3B1 phosphorylation sites clustering within the N-terminal domain of the protein (Figure 5), the functional significance of SF3B1 phosphorylation is largely undefined. SF3B1 phosphorylation is coupled with splicing catalysis (26); specifically, phosphorylation of threonine 313, a cyclin E-CDK2 substrate, is associated with active splicing (30). Cyclin E-CDK2-mediated phosphorylation of SF3B1 at threonines 244, 248 and 313 has been shown to mediate interaction with NIPP1, a protein phosphatase that localizes within the nucleus (27). In order to understand first whether SF3B1 phosphorylation changes during cell cycle progression, we used mitotic arrest to synchronize two human cell lines, HeLa and K562, and measured SF3B1 expression

and serine/threonine phosphorylation (using anti-phospho-serine/threonine-proline (pSer/Thr-Pro) and anti-phospho-threonine 313 (pSF3B1) (30) antibodies by immunoblot. We found a highly dynamic pattern of SF3B1 phosphorylation through cell cycle progression, wherein SF3B1 is phosphorylated at low levels in G1, declining further at G1/S, but is highly phosphorylated at G2/M (Figures 6A and 6B). Based on previous MS data that demonstrated CDK-mediated phosphorylation of SF3B1, we hypothesized that the increased phosphorylation of SF3B1 in G2/M was dependent on CDK1 activity. We further hypothesized that the subsequent decrease in SF3B1 phosphorylation following mitotic exit was mediated by phosphatases active during re-entry into G1. To test these hypotheses, we first treated HeLa and K562 cells arrested in G2/M with Purvalanol-A, a specific CDK1 (CDC2) inhibitor (35,36). Brief pharmacologic inhibition of CDK1 (at 0.5 and 4 hours) completely blocks detectable SF3B1 phosphorylation (Figures 6C and 6D). Inhibition of phosphatase activity in G1/S cells using okadaic acid (OA) treatment that inhibits both PP2A and PP1 phosphatases (37) results in increased SF3B1 phosphorylation (Figure 6E). Together, these results demonstrate that SF3B1 phosphorylation is dynamic during the course of cell cycle progression, peaking at G2/M in a CDK-dependent manner and then decreasing due to

phosphatase activity as cells progress through G1/S

Figure 6.



**Figure 6. SF3B1 phosphorylation is dynamic during cell cycle progression.**

SF3B1 was immunoprecipitated from whole-cell lysates of synchronized (A) HeLa and (B) K562 cells. Immunoblot analysis to assay SF3B1 phosphorylation was performed using an antibody that detects phosphorylation at all serine/threonine residues with a proline at the +1 position (pSer/Thr-Pro) and a site-specific antibody that detects phosphorylation at threonine 313 of SF3B1 (Thr313). Lower panels in (A) and (B) show flow cytometric analysis of cells synchronized and harvested for analysis in G2/M, G1 and G1/S phases. Cells were labeled with propidium iodide (PI) to measure DNA content and model cell cycle. Histogram peaks representing diploid and tetraploid DNA content are labeled as 2N and 4N respectively. G2/M arrested (C) HeLa and (D) K562 cells were treated with Purvalanol-A (Purv-A) for indicated times and SF3B1 was immunoprecipitated from whole-cell lysates followed by immunoblot analysis to assay SF3B1 phosphorylation. (E) G1/S HeLa cells were treated with 20 nM okadaic acid (OA) for 12 hours and SF3B1 was immunoprecipitated from whole-cell lysates. SF3B1 phosphorylation was determined as shown.

## **2.2 SF3B1-nucleosome interactions are dynamic during cell cycle progression.**

Work by several groups has demonstrated that SF3B1 interacts with chromatin or chromatin-associated proteins (38,39). Of note, Kfir et al. (33) showed that SF3B1 associates with mono-nucleosomes near exons and positively influences splicing of these occupied exons. Moreover, splicing is known to be coordinated with cell cycle progression, such that the expression and splicing of specific transcripts are regulated in a stage-specific manner (31). In light of these data and our finding of dynamic phosphorylation of SF3B1, we hypothesized that changes to phosphorylation of SF3B1 during cell cycle progression influences its association with nucleosomes. To test this, we first prepared whole cell lysates and mono-nucleosome-enriched fractions from mitotically arrested and synchronized HeLa and K562 cells. Mono-nucleosome fractions containing nuclear proteins and chromatin were prepared by isolating nuclei and digesting them with micrococcal nuclease (MNase), such that the resulting DNA fragments are approximately 150 base pair (bp) size (length of DNA wrapped around a single nucleosome is 147 bp) (Figure 7A inset). We compared SF3B1 protein abundance in nucleosome-enriched versus whole cell lysates of synchronized cells. In nucleosome-enriched lysates, SF3B1 levels are significantly reduced in G2/M, when

SF3B1 phosphorylation peaks (Figure 6), and increased in G1 and G1/S. However, SF3B1 protein abundance remained unchanged between G1, G1/S and G2/M in whole cell lysates (Figures 7A-7B), supporting our hypothesis that cell cycle dependent phosphorylation influences SF3B1 association with nucleosomes.

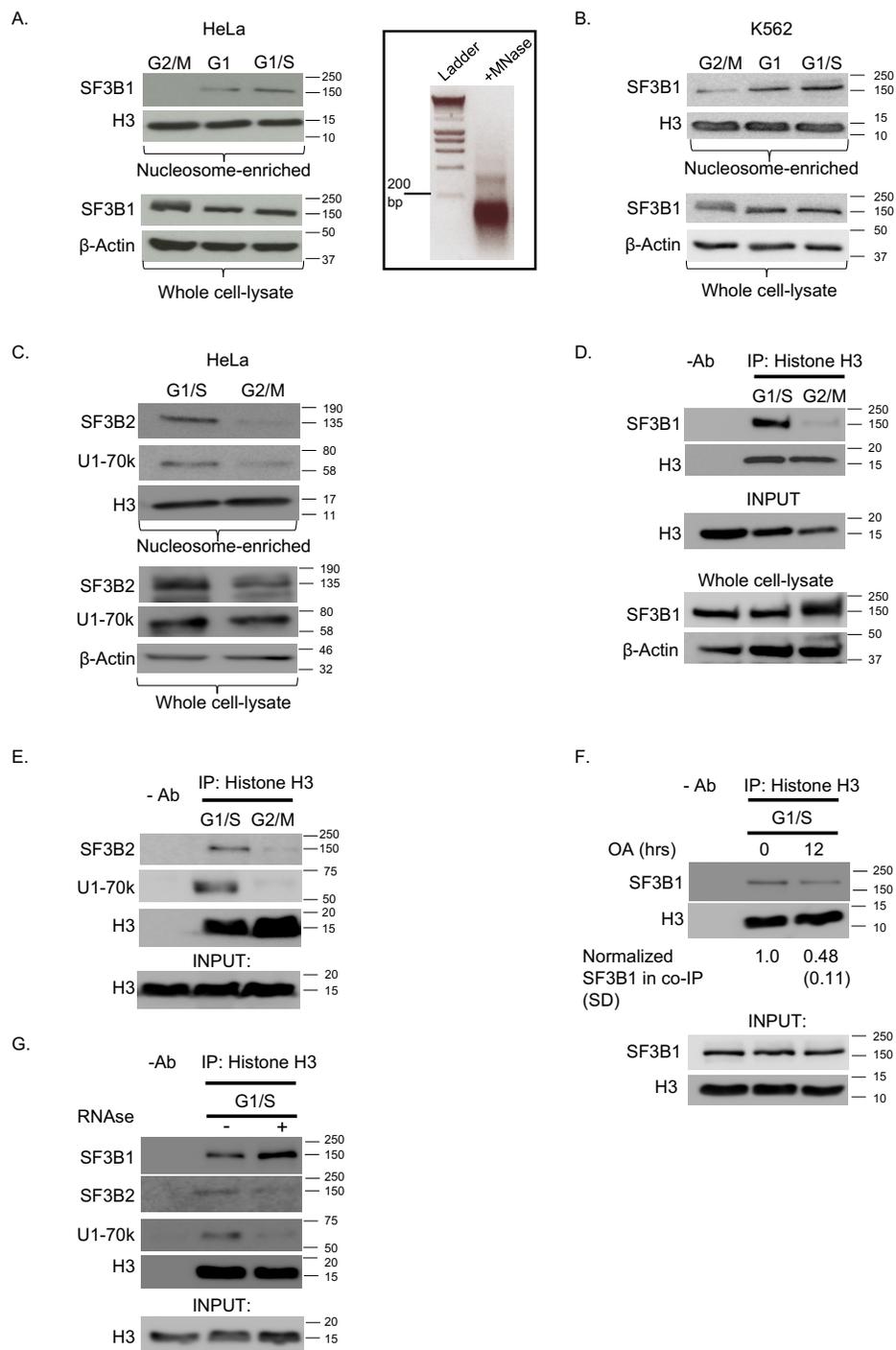
Next, we asked whether the change in chromatin association was specific to SF3B1 or if other spliceosome proteins also exhibit cell cycle-dependent chromatin association. We hence determined the total abundance and nucleosome association of SF3B2, a component of the U2 snRNP (similar to SF3B1), and also for U1-70k (or snRNP70), a subunit of a distinct snRNP (U1). Both SF3B2 and U1-70k followed the same cell cycle-dependent chromatin association as SF3B1 (Figure 7C). Next, to confirm that this cell cycle-related change in splicing factor abundance within chromatin-enriched lysates was directly linked to a change in interaction with nucleosomes, we examined the association of SF3B1, SF3B2 and U1-70K with the core nucleosome protein, histone H3, in synchronized cells. As expected, SF3B1, SF3B2 and U1-70K did not associate with histone H3 in G2/M cells but showed increased histone association in G1/S (Figure 7D-7E). Treatment of G1/S HeLa cells with OA resulted in decreased SF3B1-nucleosome interaction (Figure 7F). Since OA treatment did not alter cell cycle distribution (not shown), this result suggests that the dynamics of SF3B1 phosphorylation and not merely the progression of cells through G1/S regulate SF3B1-nucleosome interactions.

Splicing is a co-transcriptional process, and hence it is possible that the observed

cell cycle-dependent interaction between spliceosome proteins and chromatin are dependent on nascent, transcribed RNA or snRNA within spliceosome. To address this, we tested the effect of RNase on the interaction of the spliceosome proteins with histone H3 and chromatin. While RNase-A treatment significantly diminished the interaction between histone H3 and SF3B2 and U1-70k in G1/S-phase, the SF3B1 interaction with nucleosomes remained intact, consistent with an RNA-independent mode of association during G1/S (Figure 7G), consistent with previous reports (33).

Taken together, our data demonstrate that SF3B1 interactions with nucleosomes are dynamic during cell cycle progression. SF3B1 interacts with nucleosomes during G1 and G1/S stages where it demonstrates lower level phosphorylation, whereas its interaction with nucleosomes is greatly diminished during G2/M when CDK1 activity peaks, suggesting that SF3B1 phosphorylation during G2/M directly contributes to its dissociation from nucleosomes. Importantly, unlike other spliceosomal proteins we tested (both U2 snRNP components and non-U2 components), this cell cycle-dependent interaction of SF3B1 with histone is RNA-independent and hence likely not solely due to transcription.

Figure 7.



**Figure 7. SF3B1-nucleosome interactions are dynamic during cell cycle progression.** SF3B1 protein abundance was assayed in nucleosome-enriched and whole-cell lysates from (A) HeLa and (B) K562 cells synchronized in G1, G1/S and G2/M. Histone H3 (H3) and beta-Actin were used as nuclear and whole-cell lysate loading controls respectively. Inset: DNA was isolated from micrococcal nuclease (Mnase)-digested nuclear lysates using phenol: chloroform extraction and run on an agarose gel to assess the size of the DNA fragments. (C) Total protein abundance of SF3B2 and U1 subunit-70K was assayed in nucleosome-enriched and whole-cell lysates of HeLa cells synchronized in G1/S and G2/M. U1 subunit-70K Santa Cruz Biotechnology sc-390988 and SF3B2 Abcam ab56800 antibodies were used in this experiment. H3 and beta-Actin were used as nuclear and whole-cell lysate loading controls respectively. (D) H3 was immunoprecipitated from nucleosome-enriched lysates of HeLa cells synchronized in G1/S and G2/M. SF3B1 co-immunoprecipitation was assayed by immunoblot assay. (E) H3 was immunoprecipitated from nucleosome-enriched lysates from HeLa cells synchronized in G1/S and G2/M. SF3B2 and U1-70K-subunit co-immunoprecipitation was assayed by immunoblot using the U1 subunit-70K antibody from Dr. Doug Black's Lab and SF3B2 (Novus 79848) antibody. (F) G1/S HeLa

cells were treated with 20 nM okadaic acid (OA) for 12 hours and H3 was immunoprecipitated from nucleosome-enriched lysates. SF3B1 co-immunoprecipitation (co-IP) was assayed by immunoblot assay. SF3B1 in co-IP was quantified and normalized to immunoprecipitated histone H3. The experiment was performed in triplicate and relative SF3B1 abundance in co-IP quantified using NIH ImageJ and standard deviation (SD) calculated. (G) H3 was immunoprecipitated from nucleosome-enriched lysates of HeLa cells synchronized in G1/S and G2/M. SF3B1, SF3B2 and U1-70K subunit co-immunoprecipitation was assayed after RNase-A treatment of the immunoprecipitation reaction as shown. The immunoblot displayed is representative of three independent assays.

### **2.3 SF3B1-chromatin interactions are dependent on CDK activity.**

We next hypothesized that dynamic SF3B1-nucleosome interactions are dependent on CDK-dependent phosphorylation. We tested this hypothesis by pharmacologic inhibition of CDK1 activity in G2/M and of CDK2 activity in G1/S. Treatment of G2/M cells with Purvalanol-A restored the interaction between SF3B1 and nucleosomes (Figure 8A). In contrast, inhibition of CDK2 activity in G1/S by treatment with Roscovitine (40) results in diminished SF3B1-nucleosome interaction (Figures 8B and 8D), suggesting that CDK1 and CDK2 play opposing roles in regulating SF3B1-nucleosome interactions during cell cycle progression. Another kinase, DYRK1A, is also active during G1 and is known to phosphorylate SF3B1 at threonine 434 (28). We hence evaluated the role of DYRK1A in regulating SF3B1-nucleosome interactions. Inhibition of DYRK1A by treatment with a selective inhibitor, Harmine (41), did not affect SF3B1-nucleosome interaction, while the inhibitor did elicit an increase in cyclin D2 protein, which is normally destabilized in response to DYRK1A-mediated phosphorylation (42) (Figures 8C-8D). These data suggest that CDK1 and CDK2 play a selective role in regulating SF3B1-nucleosome interactions during cell cycle progression.

Previous studies have implicated specific threonine residues within the N-

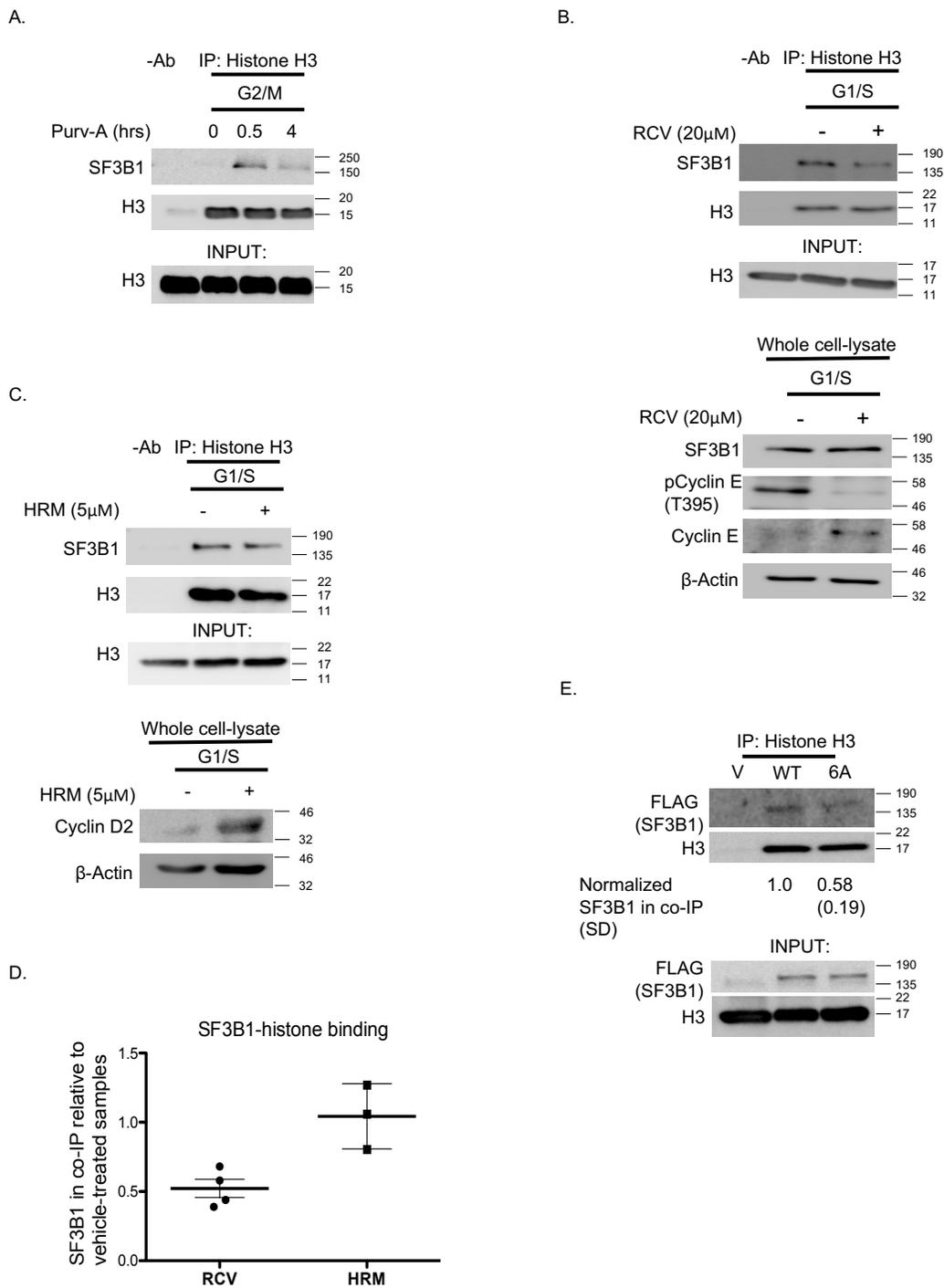
terminal region of SF3B1 in mediating its interaction with other proteins, including NIPP1 (27). While our data from pharmacologic inhibition of kinase activity addressed the role of overall phosphorylation, the role of phosphorylation of SF3B1 at specific amino acid residues in regulating its interaction with nucleosomes is unclear. Mass spectrometry (MS) has identified CDK-dependent phosphorylation of SF3B1 at multiple sites in addition to those CDK sites previously described as regulating the NIPP1 interaction, including threonines 142, 211, 257, 261, 426 and 434, all of which have +1 prolines (29). We hence mutated all 6 of these MS-identified threonine residues (Figure 1) (28,29). On probing the SF3B1-nucleosome interactions in K562 cells expressing wild-type (WT) or the compound phospho-mutant SF3B1 (6A), we found partially decreased nucleosome interaction of 6A suggesting that aggregate phosphorylation of SF3B1's N-terminus is required for regulation of SF3B1-nucleosome interactions (Figure 8E). Consistent with this result we found only a modest decrease in cyclin E-CDK2 mediated phosphorylation of 6A when compared to WT (Figure 9), likely reflective of the large number of N-terminal SF3B1 phosphosites contributing to overall phosphorylation of the protein. Indeed, single point mutations in SF3B1 phosphosites showed no alteration in nucleosome interactions, whereas a more extensive mutant

containing 25 alanine substitutions within predicted S-P/T-P phosphosites was not expressed sufficiently in cells to study (data not shown).

To determine whether 6A expression had any functional effects on splicing, we used a previously described splicing reporter construct (43). The reporter construct consists of a beta-galactosidase ( $\beta$ -Gal) gene and a luciferase gene separated by an intron that contains three in-frame stop codons. When spliced normally, the stop codons are spliced out of the mRNA expressed from the reporter construct generating a fusion protein with both  $\beta$ -Gal and luciferase enzyme activity. In case of inefficient splicing leading to a retained intron with the three in-frame stop codons in the mRNA the resultant protein has only the  $\beta$ -Gal enzyme activity. The ratio of luciferase to  $\beta$ -Gal enzyme activity measured using a fluorometric assay provides a measure of splicing efficiency. Using this reporter system, we found that the luciferase to  $\beta$ -Gal ratio was lower in cells overexpressing WT SF3B1 compared to the empty vector control. This suggested that mere overexpression of WT protein results in altered splicing. However, when comparing luciferase to  $\beta$ -Gal enzyme activity ratios in WT and 6A overexpressing cells, we did not observe any significant differences (Figure 10B). 6A expression in K562 cells also did not alter the cell cycle status of these cells and did not result in any

significant changes in apoptotic cell death compared to cells overexpressing WT SF3B1 (Figure 10C). Taken together, these data demonstrate that while the SF3B1 6A mutant showed a modest decrease in binding to chromatin, it did not result in any observable effects on splicing and overt cell cycle defects.

Figure 8.

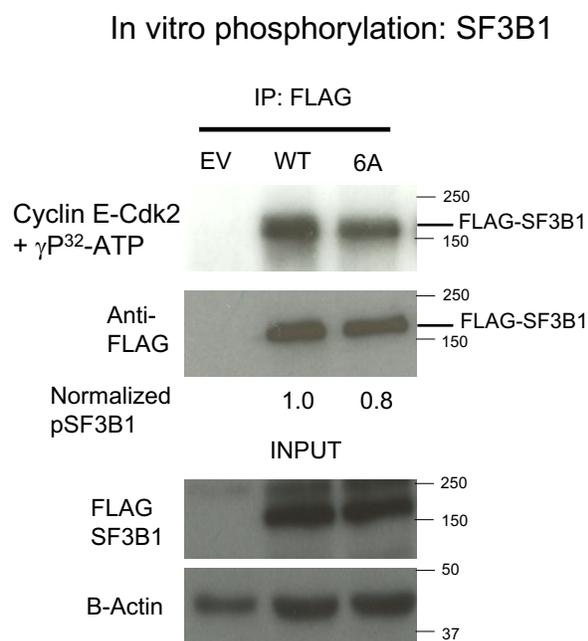


**Figure 8. SF3B1-chromatin interactions are dependent on CDK activity.**

(A) Histone H3 was immunoprecipitated from nucleosome-enriched lysates of HeLa cells synchronized in G2/M and treated with CDK1 inhibitor Purvalanol-A (Purv-A) for the indicated times. SF3B1 co-immunoprecipitation was assessed by immunoblot analysis. No significant changes in cell cycle status of Purv-A treated cells were observed by flow cytometric analysis (not shown). (B) Upper panel- H3 was immunoprecipitated from nucleosome-enriched lysates of HeLa cells synchronized in G1/S and treated with CDK2 inhibitor Roscovitine (RCV) for 12 hours. SF3B1 co-IP was assessed by immunoblot analysis. No significant changes in cell cycle status of RCV treated cells were observed by flow cytometric analysis. Lower panel- RCV-mediated inhibition of CDK2 activity was confirmed by analysis of cyclin E auto-phosphorylation (44) using similar conditions as in panel (B). (C) Upper panel- H3 was immunoprecipitated from G1/S HeLa cells treated with DYRK1a inhibitor Harmine (HRM) for 12 hours. SF3B1 co-immunoprecipitation was assessed by immunoblot assay. Lower panel - DYRK1a inhibition was confirmed by analysis of cyclin D2 protein levels in whole cell lysates of HRM treated G1/S HeLa cells by immunoblot assay. (D) SF3B1 abundance in H3 co-IPs from drug-treated relative to vehicle-treated cells was

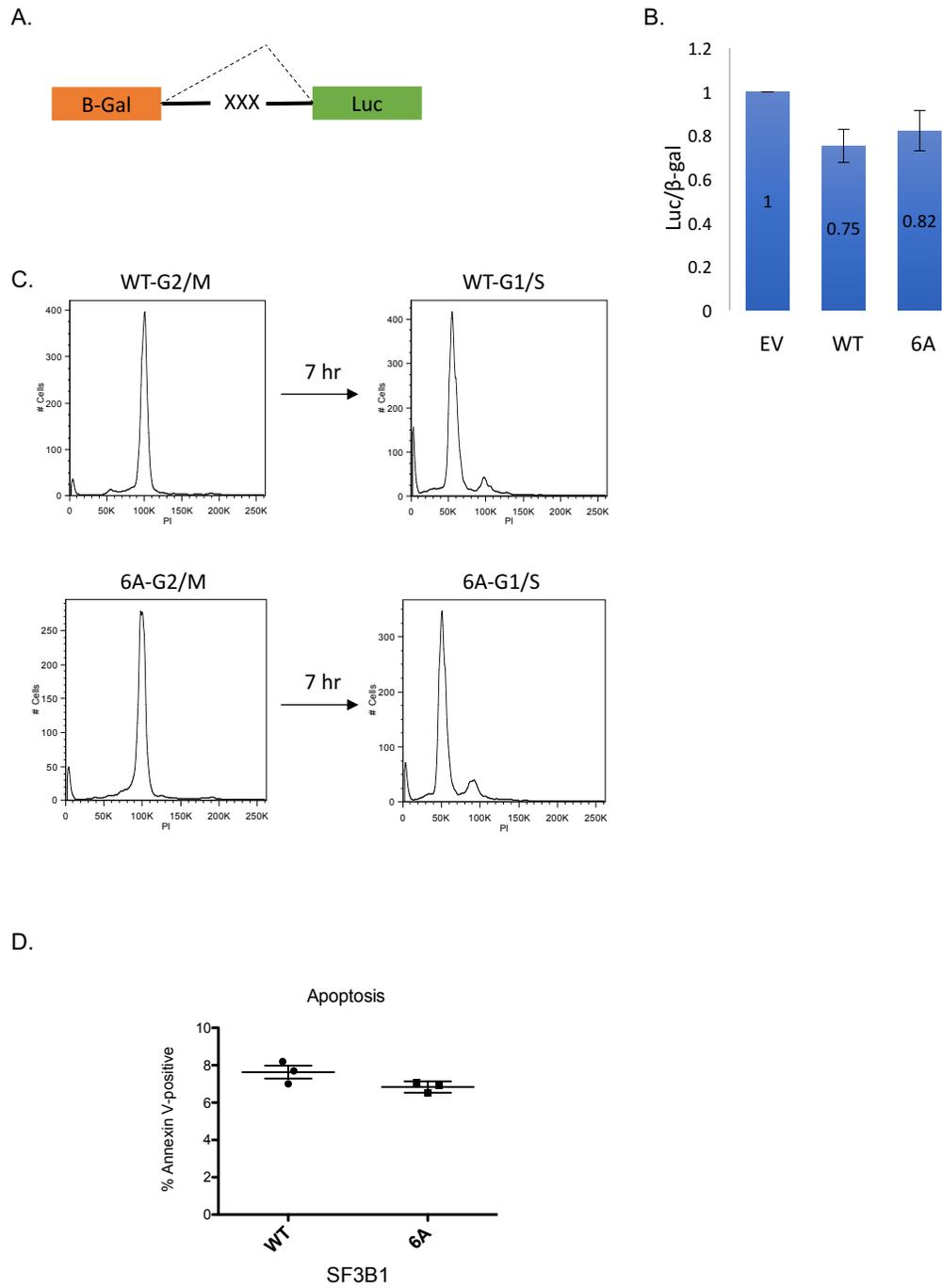
quantified from 3-4 experiments and is displayed with means (thick bars) and SDs (thin bars). (E) FLAG-tagged wild-type (WT) and compound phosphosite mutant SF3B1 (6A) were overexpressed in K562 cells by retroviral transduction. Cells were synchronized in G1/S, and H3 was immunoprecipitated from nucleosome-enriched lysates. SF3B1 co-immunoprecipitation was assayed by immunoblot analysis using an antibody against the FLAG tag in triplicate experiments.

Figure 9.



**Figure 9.** Flag-tagged Wild-type (WT), compound phosphomutant (6A) SF3B1 or empty vector (EV) were overexpressed in K562 cells and immunoprecipitated using an anti-Flag antibody. Immunoprecipitated proteins were phosphorylated in vitro with purified cyclin E-CDK2 in the presence of gamma-P<sup>32</sup>-ATP. SF3B1 phosphorylation was assayed by western blot followed by autoradiography. Equal expression and pull-down of WT and 6A SF3B1 were confirmed by immunoblot assays using indicated antibodies.

Figure 10.

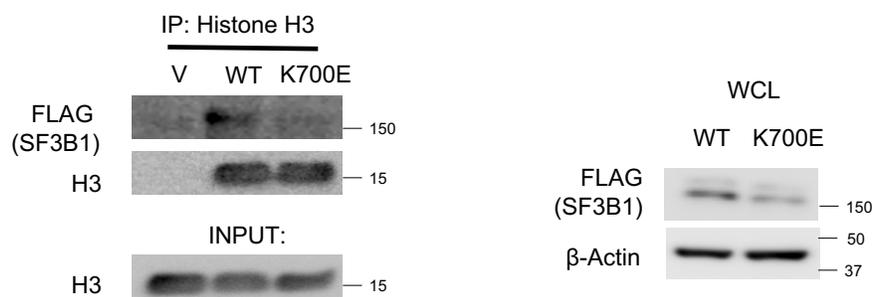


**Figure 10. Functional effects of mutant SF3B1 expression on splicing efficiency, cell cycle progression and apoptosis.** (A) Diagram showing the splicing reporter that was used to measure splicing efficiency. The reporter construct consists of a beta-galactosidase gene (orange box) and a luciferase gene (green box) separated by an intron that contains three in-frame stop codons (XXX). (B) HEK293 cells were co-transfected with the splicing reporter and WT SF3B1 or the 6A mutant. Cells were lysed and beta-galactosidase and luciferase activities were measured fluorometrically. (C) K562 cells were transduced with WT SF3B1 and the 6A mutant. Comparable cell cycle entry of WT and 6A expressing at 7 hours following G2/M arrest is shown. (Data shown are representative of triplicate assays). (D) Baseline levels of apoptosis in WT and expressing cells measured by Annexin-V staining are shown. (Shorter bars represent standard deviations and longer represent medians from triplicate experiments).

## 2.4 SF3B1 K700E mutation alters chromatin binding:

Mutations predominantly within the N-terminus HEAT domain of SF3B1 are associated with several neoplastic processes, with SF3B1 K700E mutations being the most frequent (4,8-10). Transcriptomic analyses of patient samples, cell lines and murine models have shown that expression of mutant SF3B1 results in aberrant splicing of a large number of genes (12,14,16). Additionally, recent reports have also shown that SF3B1 binding to chromatin, enriched near exons, influences splicing outcomes of SF3B1-occupied exons (33). In light of these data, we decided to determine whether the SF3B1 K700E HEAT domain mutation results in altered SF3B1 binding to chromatin and contributes to the mechanism by which mutant SF3B1 causes aberrant splicing. As before, we overexpressed WT and the SF3B1 K700E mutant (SF3B1<sup>MUT</sup>) in K562 cells. On probing the interaction between WT and SF3B1<sup>MUT</sup> with chromatin in nucleosome enriched lysates, we observed impaired interaction between SF3B1<sup>MUT</sup> and histone H3. Given that SF3B1 interaction with chromatin has been shown to influence splicing outcomes, our finding suggests that altered binding of the disease-associated SF3B1 mutant with chromatin may potentially contribute to aberrant splicing observed due to mutant SF3B1 expression. (12-16,33) (Figure 11)

Figure 11.



**Figure 11. SF3B1 K700E mutation alters binding with chromatin:** FLAG-tagged wild-type (WT) and disease-associated SF3B1 K700E mutant (SF3B1<sup>MUT</sup>) were overexpressed in K562 cells by retroviral transduction. Histone H3 was immunoprecipitated from nucleosome-enriched lysates. SF3B1 co-immunoprecipitation was assayed by immunoblot analysis using an antibody against the FLAG-tag (Left Panel). SF3B1 expression was confirmed in whole cell lysates (Right Panel).

## **2.5 SF3B1-chromatin interactions in vitro depend on linker histone H1 and phosphorylation of both SF3B1 and H1.**

A large number of nucleosome binding proteins influence chromatin organization. SF3B1 is part of the SF3B complex, which in turn forms part of a large multi-protein complex, the U2 snRNP that is also known to contain chromatin-associated proteins (45). It is thus unclear whether the interaction between SF3B1 and nucleosomes is direct or depends on other proteins known to complex with nucleosomes. To help determine which of these scenarios is more likely in vivo, we reconstituted SF3B1-nucleosome binding in vitro. Given that the majority of serine/threonine residues identified as potential substrates for CDKs in SF3B1 are located within an N-terminal domain that excludes the HEAT-containing domain (Figure 5), we sub-cloned this fragment (1-500 aa or SF3B1<sup>1-500</sup>) for use in these studies. Examining previously published mass spectrometry-based SF3B1 interactome datasets, we noted that histone H1 is part of the complex that contains SF3B1 and other nucleosome proteins (33,46). H1 is involved in chromatin organization during cell cycle progression and is a known CDK substrate during cell cycle progression (22). Using co-immunoprecipitation, we first confirmed that SF3B1, histone H1 and histone H3 are in complex in vivo (Figure

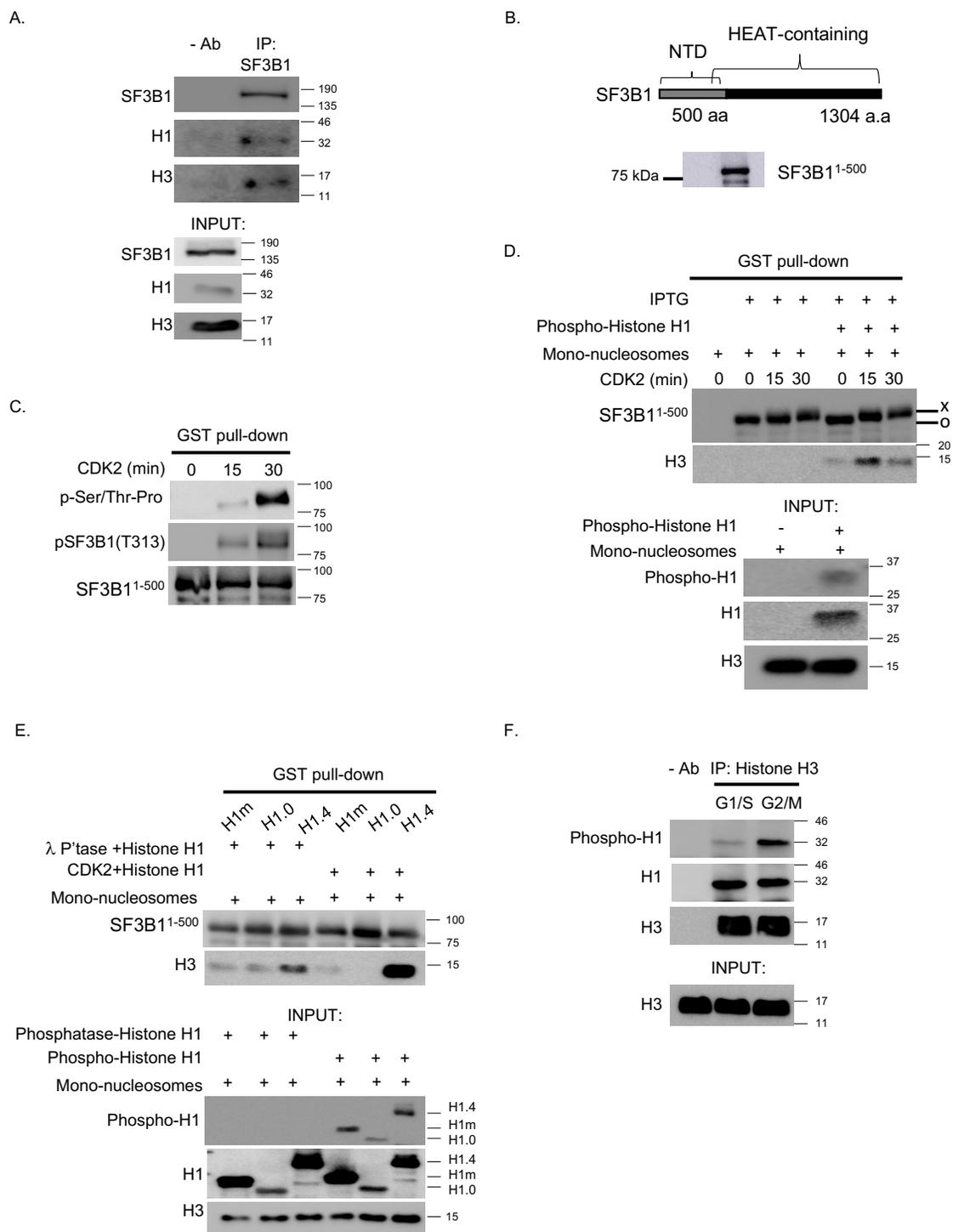
12A). The SF3B1<sup>1-500</sup> with N-terminal GST tag was then expressed in E.coli and isolated for in vitro binding assays (Figure 12B). We confirmed that SF3B1<sup>1-500</sup> could be phosphorylated by purified cyclin E-CDK2 in vitro (Figure 12C). Using an in vitro binding assay in which SF3B1<sup>1-500</sup> is first phosphorylated in vitro using purified cyclin E-CDK2 and then incubated with purified HeLa mono-nucleosomes in the presence or absence of purified calf thymus histone H1, also phosphorylated in vitro by cyclin E-CDK2, we found phosphorylation of both histone H1 and SF3B1<sup>1-500</sup> is required for robust binding of the latter to purified mono-nucleosomes. However, prolonged SF3B1<sup>1-500</sup> incubation with cyclin E-CDK2 in vitro led to reduced binding (Figure 12D, upper panel), suggesting that lower amounts of SF3B1 phosphorylation are permissive to chromatin interactions, whereas hyper-phosphorylated SF3B1 impedes the interaction.

Histone H1 has multiple isoforms: Histones H1.1, H1.2, H1.3, H1.4 and H1.5 are ubiquitously expressed in a cell cycle-dependent manner, while histone H1.0 and the H1.X isoforms are expressed mainly in differentiated cells independent of cell cycle (22). Among the ubiquitous isoforms, H1.4 undergoes phosphorylation during S-phase and mitosis. H1.4 phosphorylation has been detected at serine residues 172 and 187 in interphase cells, and additional phosphorylations at serine 27 and threonines 18, 146

and 154 have been detected in mitotic cells (47). Importantly, H1.4 is known to maintain S-phase progression, as selective depletion of it leads to a decrease in cell cycling and S-phase (48). Also, with affinity purification and MS, H1.4 was recently identified as a specific interacting partner of SF3B1 (46). We thus hypothesized that H1.4 promotes SF3B1-nucleosome interactions during G1/S. To test this, we performed separate in vitro binding assays of SF3B1-nucleosome interactions using either native calf thymus histone H1 isoform mix, purified human histone H1.4 or histone H1.0. As shown in Figure 12E (upper panel), we found that the presence of purified histone H1.4 most enhances the interaction between SF3B1<sup>1-500</sup> and mono-nucleosomes compared to the other linker histone preparations. This result suggests that the dynamics of linker histone isoform interactions with mono-nucleosomes during cell cycle progression contribute to the regulation of SF3B1-chromatin interactions. Moreover, we found that phosphorylation of H1.4 by CDK2 results in the most robust SF3B1-mono-nucleosome interaction in vitro (Figure 12E). Notably, H1 in complex with H3 undergoes an increase in phosphorylation from G1/S to G2/M in vivo (Figure 12F), suggesting that while H1.4 phosphorylation promotes SF3B1<sup>1-500</sup> interaction with mono-nucleosomes in vitro, high levels of phosphorylation on both SF3B1 and H1, evident during G2/M, diminish SF3B1-

nucleosome interactions. In this way, phosphorylation of both SF3B1 and linker histone H1 may enable switch-like regulation of U2-chromatin interaction during cell cycle progression, by permitting histone binding when both are phosphorylated at lower levels (e.g. during G1/S) and disfavoring binding when they are hyper-phosphorylated during progression to G2/M.

Figure 12.



**Figure 12. SF3B1-chromatin interactions are dependent on the phosphorylation status of both SF3B1 and linker histone H1.**

(A) SF3B1 was immunoprecipitated from nucleosome-enriched HeLa lysates. Co-IP of histone H3 and histone H1 were assessed by western blot analysis. Histone H1 (Active Motif 39707) antibody was used in this experiment. Data shown are representative of three independent experiments. (B) Upper panel - Schematic highlighting N-terminal domain (NTD) of SF3B1 protein (SF3B1<sup>1-500</sup>) that is encoded by an inducible expression construct used in panels B-F is shown in relationship to the HEAT motif-containing region spanning most of the protein (amino acids 463-1304) (17). Lower panel - N-term GST-tagged SF3B1<sup>1-500</sup> expression in E.coli after IPTG induction was assessed by immunoblot analysis using an antibody that recognizes an N-terminus region of SF3B1 (Abcam ab172634). (C) SF3B1<sup>1-500</sup> expressed in E.coli was isolated from whole cell lysates after IPTG induction and phosphorylated in vitro using purified cyclin E-CDK2 for the indicated incubation times. SF3B1 phosphorylation was assessed by immunoblot analysis. (D) Upper panel - N-term GST-SF3B1 expressed in E.coli was phosphorylated in vitro with purified cyclin E-CDK2 for the indicated times and incubated in vitro with purified mono-nucleosomes in the presence or absence of phosphorylated native calf thymus histone H1. \* - slower migrating species of SF3B1<sup>1-500</sup> with 30 min cyclin E-

CDK2 incubation, compared to  $^{\circ}$  – faster migrating species without phosphorylation in vitro. N-term GST-SF3B1-mono-nucleosome interactions were determined by immunoblot analysis. Lower panel - equal levels of total H3 and H1 and phosphorylated histone H1 added to the in vitro binding reaction were confirmed by immunoblot analysis. Histone H1 (Active Motif 39707) antibody was used, and H1 phosphorylation was assessed using the phospho-serine/threonine-Proline (Abcam ab9344) antibody. Representative result of three independent replicates is shown. (E) Upper panel - N-term GST-SF3B1 expressed in E.coli was phosphorylated in vitro with purified cyclin E-CDK2 for 15 minutes and incubated in vitro with purified mono-nucleosomes in the presence of lambda phosphatase treated or cyclin E-CDK2-phosphorylated native calf-thymus H1 containing a mix of isoforms (H1m), purified human histone H1.0 or purified human histone H1.4. Lower panel – immunoblot showing levels of H3, total H1 and phosphorylated H1. Mononucleosomes and linker histone H1 (H1m, H1.0, or H1.4) were added to the binding assays as shown in equal quantities by mass. Data displayed are representative of two independent experiments. (F) H3 was immunoprecipitated from nucleosome-enriched lysates of G1/S- and G2/M-synchronized HeLa cells. Abundance and phosphorylation of co-immunoprecipitated

histone H1.4 was assessed by immunoblot analysis. Antibodies recognizing H1.4 specifically (histone H1.4 (Abcam ab105522)) and phosphorylated H1 (phospho-histone H1 (Abcam ab4270)) were used. Representative result of three independent replicates is shown.

## **2.6 Integrated analysis of SF3B1 genome occupancy and cell cycle-dependent splicing.**

Our biochemical data demonstrate that SF3B1-nucleosome interactions are regulated by CDK-dependent phosphorylation of SF3B1 and linker histone H1. To understand the functional significance of this SF3B1-nucleosome interaction (and specifically how this interaction relates to stage-specific splicing programs), we performed SF3B1 chromatin immunoprecipitation followed by sequencing (ChIP-Seq) and paired-end RNA sequencing (RNA-Seq) using G1/S and G2/M HeLa cells. Mono-nucleosome-enriched fractions from synchronized HeLa cells (G1/S and G2/M, in biological duplicates) were immunoprecipitated with antibodies against SF3B1. Illumina compatible libraries were prepared from DNA fragments that immunoprecipitated with SF3B1, sequenced and analyzed per informatics pipeline detailed in Materials and Methods. Input libraries were

prepared prior to IP. Control libraries (nonspecific antibody control and input) were also prepared. Resulting reads were pre-processed and mapped to the hg19 genome assembly using STAR Aligner. Correlation coefficients of biological replicate samples were confirmed prior to further downstream analysis (Figure 13). Peak calls and comparative analysis were performed using HOMER (49).

G1/S cells had higher specific SF3B1-binding peaks relative to G2/M (7386 vs. 4096) in comparison to input controls. We also found that in G1/S cells, SF3B1 peaks are over-represented within exons, compared to G2/M cells, in which SF3B1 was found to occupy primarily intergenic sites (Figure 14A). Similar to a previous observation by Kfir et al. (33), the density of peaks in exons when normalized to proportional length in human genome was significantly higher in exons than those in introns or intergenic areas in G1/S, but not G2/M cells. (Figure 14B). A modest enrichment of exonic reads was noted relative to surrounding intronic regions in both G1/S and G2M (Figure 15). G1/S and G2/M peaks were largely distinct, with only 167 (1.45%) peaks overlapping between the two datasets. G1/S-specific peaks also showed a distribution skewed towards intragenic regions, within both exons and introns (Figure 14C), consistent with our observations for total peaks in G1/S. When analyzed for DNA motifs, both G1/S and

G2/M datasets showed highly significant enrichment of non-overlapping DNA motifs (Figure 16). In summary, our ChIP-Seq results confirm our biochemical data showing increased affinity of SF3B1 for chromatin during G1/S compared to G2/M. Importantly, the increased binding in G1/S is primarily in transcribed intragenic areas (including transcription start and end sites).

Next, paired-end RNA-Seq was performed using matched samples from synchronized HeLa cells that were used for the ChIP-seq analysis. cDNA libraries were prepared from poly-A enriched total RNA and paired end sequencing was performed (75 bp, average depth of approx.122 million). Reads were aligned to hg19 using STAR aligner and first analyzed for changes in gene expression by Cufflinks (50). In line with previous reports, significant differences were found in gene expression profiles of G1/S and G2/M (1248 differentially expressed genes enriched in cell cycle and cell death/apoptosis pathways). We then used the rMATS (51) algorithm to identify alternative splicing events in G1/S and G2/M samples. A total of 48,105 alternative splicing events (distributed across 5 different types - A3SS: alternative 3' splice site; A5SS: alternative 5' splice site; RI: retained intron; MXE: mutually exclusive exon; SE: skipped exon) were detected between G1/S and G2/M, of which 2340 met statistical cut-off (Figure 14D,

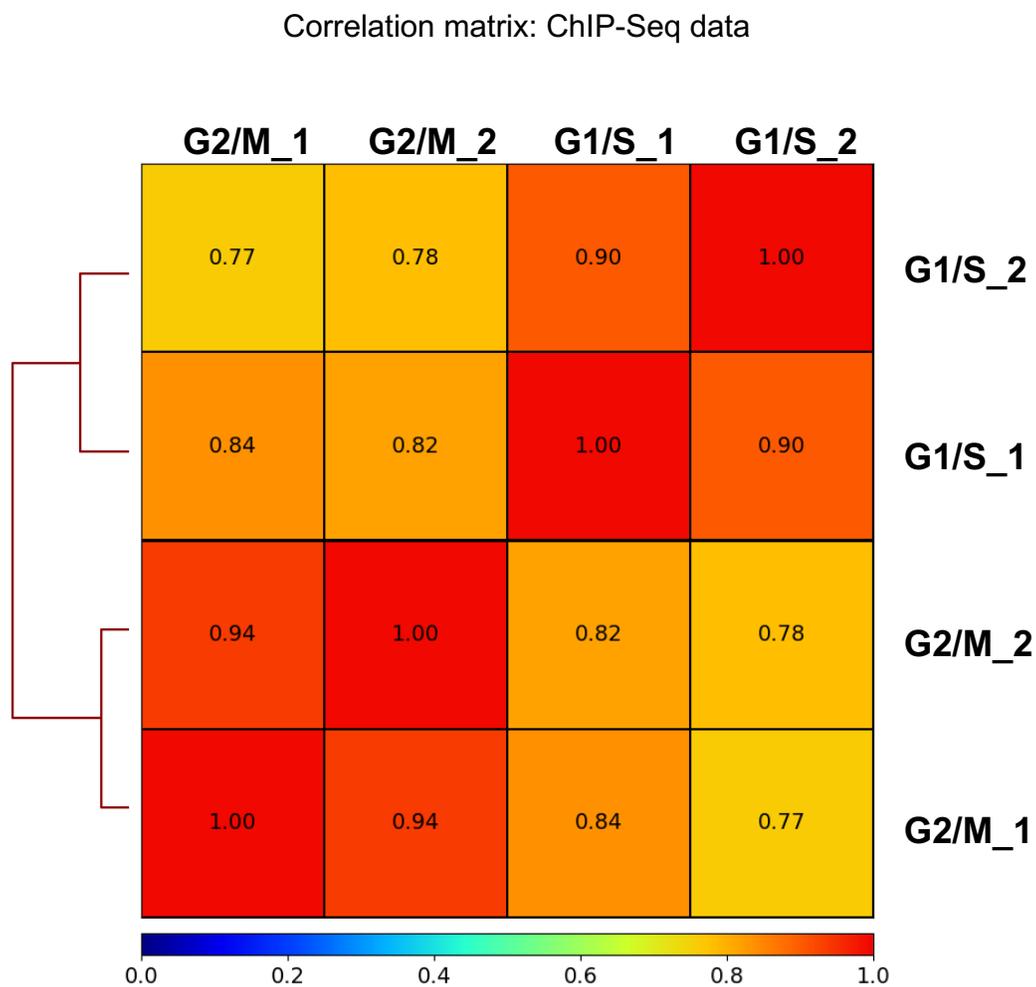
FDR < 0.05 and delta PSI > 5%). Our results are in agreement with previous reports that suggested specific changes to RNA transcriptome linked to change in cell cycle (31).

To test our hypothesis that SF3B1-chromatin occupancy affects cell cycle specific transcriptomes by influencing splicing and/or gene expression in proximity to the site of gene occupancy, we determined overlap between genes enriched in three different datasets: G1/S-specific chromatin binding (ChIP-Seq), gene expression (Cufflinks) and alternative splicing (rMATS analysis). To determine if SF3B1 occupancy of chromatin positively influences transcription of genes in those regions or change alternative splicing, we determined overlap of genes co-localized with G1/S-specific peaks with transcripts or alternative splicing events increased in G1/S compared to G2/M. In contrast to our expectations, as shown in Figure 14E, there appeared to be little overlap among these gene-sets, not reaching statistical significance by Fisher's exact test. To examine the possibility that SF3B1 binding of nucleosomes decreased transcription or alternative splicing, a similar analysis was performed for transcripts and splicing events that decreased in G1/S compared to G2/M (Figure 17). Little overlap was evident in this analysis as well. Importantly, no G2/M peaks were found to be

specific when compared to G1/S when analyzed with similar HOMER parameters.

Taken together, our genome-wide integrative analyses suggest that change in SF3B1-chromatin interactions during cell cycle progression does not reflect accompanying changes to transcription or splicing of genes at the same loci.

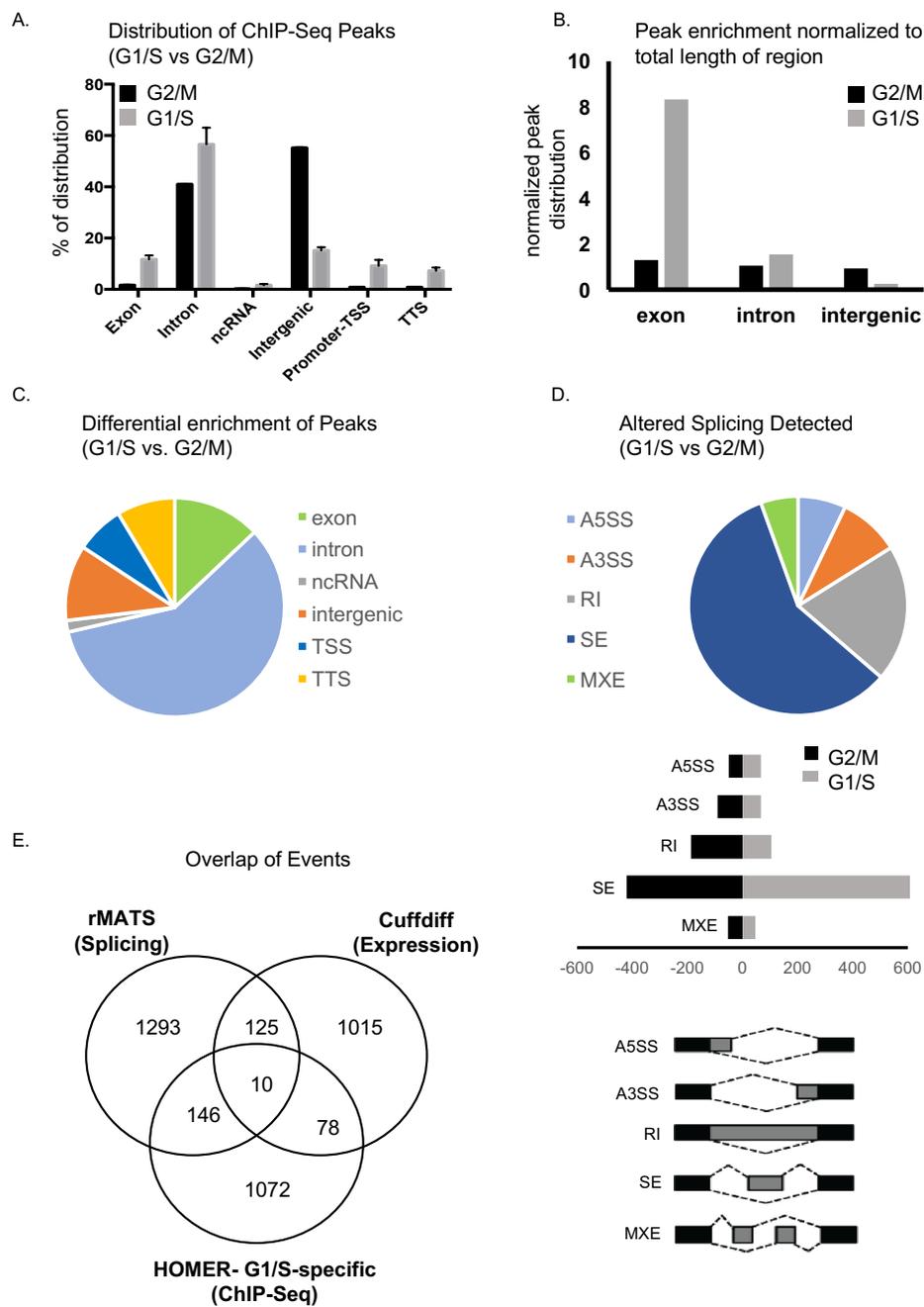
Figure 13.



**Figure 13.** Matrix demonstrating correlation coefficients of ChIP-Seq datasets for the samples shown, as derived using deepTools2 (52).

\*Generated by Dr. Manoj Pillai, Yale University.

Figure 14.



**Figure 14. Integrative analysis of SF3B1 genome occupancy and cell cycle-dependent splicing.**

(A) Distribution of ChIP-Seq peaks as a proportion of genomic regions in G1/S and G2/M. Peaks determined by HOMER algorithm located within promoter regions (promoter-TSS), transcriptional termination site-containing regions (TTS), exons, introns, ncRNA (non-coding RNA) and intergenic regions are shown as a percentage.

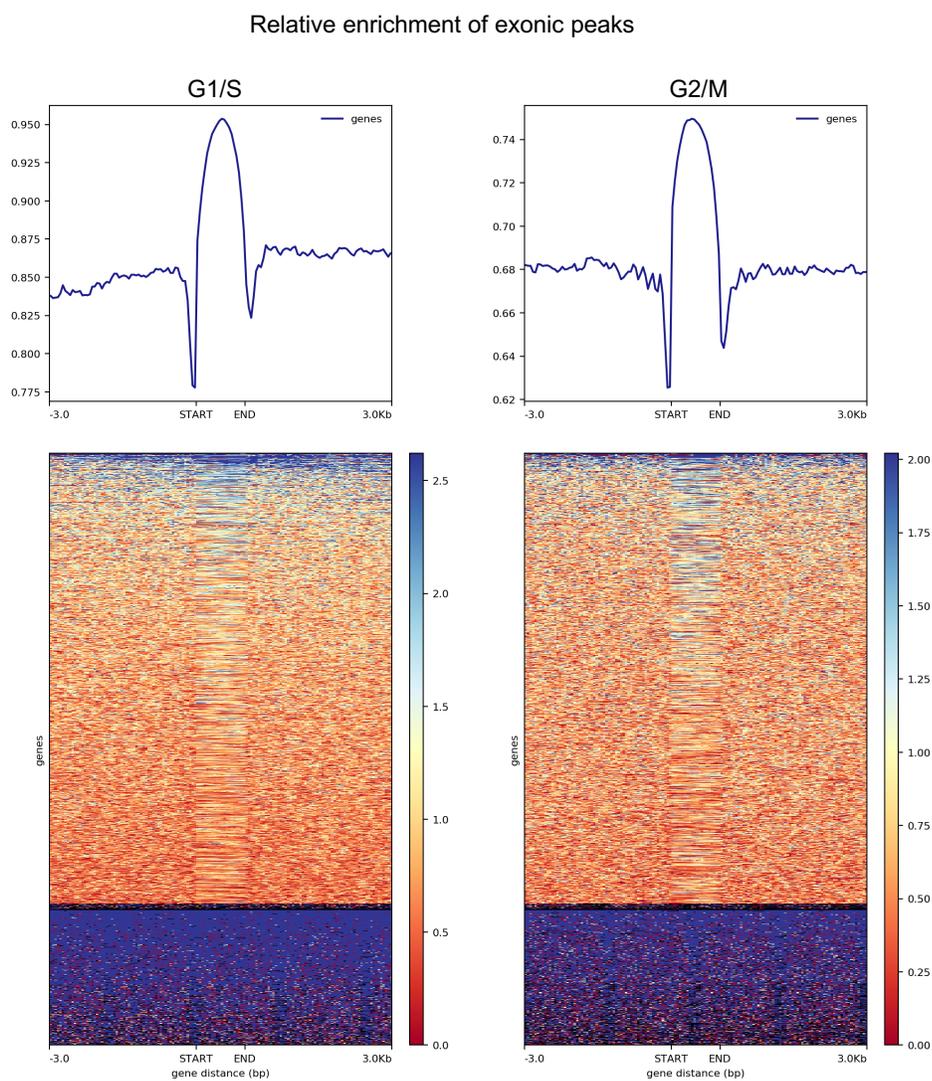
(B) Enrichment of peaks normalized to relative length of genomic regions. Length of Peaks determined to be in exon, intron or intergenic regions were normalized to the total length of these regions in the hg19 genome build. Values for G1/S and G2/M are plotted side by side. (C) Differential enrichment of peaks (G1/S vs G2/M). A total of 456 peaks were found to be differentially enriched in G1/S compared to G2/M peaks using the HOMER algorithm. Their relative distribution across genomic regions is plotted. (D)

Analysis of altered splicing in G1/S vs. G2/M. Using the rMATS algorithm, aligned RNA-Seq files were analyzed for altered splicing. Distribution of 2340 events that met statistical cut-off (FDR <0.05 and delta PSI > 5%) across 5 different types of events (A3SS: alternative 3' splice site; A5SS: alternative 5' splice site; RI: retained intron; MXE: mutually exclusive exon; SE: skipped exon) are shown on the left. Distribution of

these events (over represented in G1/S or G2/M) is shown in the right panel. (E) Overlap of genes between rMATS, Cuffdiff and HOMER analysis of G1/S and G2/M datasets. Genes with significant changes in splicing (FDR <0.05) in any of the 5 splice event types (A3SS, A5SS, SE, MXE or RI) with a higher isoform ratio in G1/S compared to G2/M were included in the rMATS set. Cuffdiff set includes genes significantly over-expressed in G1/S compared to G2/M ( $p < 0.05$ ). The HOMER set contained genes co-localized with intragenic peaks (290 of the total 456) enriched in G1/S compared to G2/M. Overlaps were not found to be significant by Fisher's exact test.

\*Generated by Dr. Manoj Pillai, Yale University.

Figure 15



**Figure 15.** Relative enrichment of exonic peaks (normalized to 1.0 Kb length, between START and END in upper plot) compared to surrounding intronic regions (3.0 Kb upstream and 3.0 Kb downstream. Results for G1/S shown on left and G2/M on right.

\*Generated by Dr. Manoj Pillai, Yale University.

Figure 16.

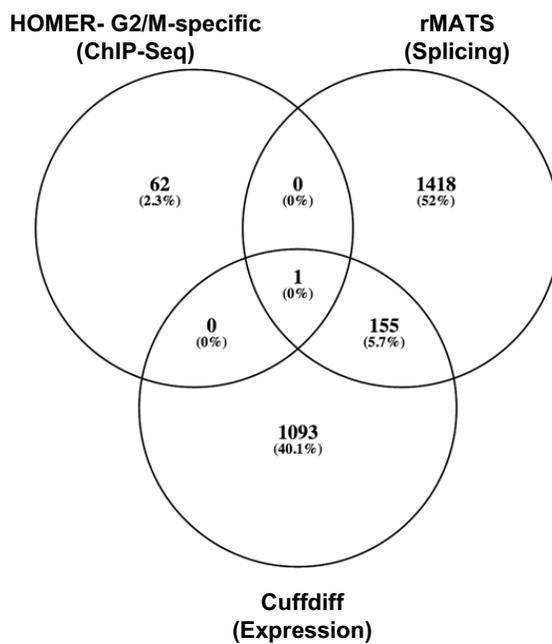
SF3B1-occupied DNA motifs in G1/S and G2/M (8/10/12 bp Motif Length)

	G1/S	p-value	G2/M	p-value
1.		1E-22		1E-84
2.		1E-21		1E-63
3.		1E-19		1E-39
4.		1E-19		1E-33
5.		1E-15		1E-30
6.		1E-15		1E-25
7.		1E-14		1E-23
8.		1E-14		1E-23
9.		1E-14		1E-22
10.		1E-12		1E-19

**Figure 16.** Top 10 SF3B1-occupied DNA binding motifs enriched in G1/S and G2/M synchronized cells, shown with corresponding p-values.

Figure 17.

Association G2/M-specific SF3B1 occupancy and cell cycle-dependent splicing and gene expression changes



**Figure 17.** Overlap of genes between rMATS, Cuffdiff and HOMER analysis of G1/S and G2/M datasets, focusing on G2/M-specific intragenic peaks.

\*Generated by Dr. Manoj Pillai, Yale University.

**Conclusions:**

SF3B1 is phosphorylated by CDKs (26,27) and SF3B1 phosphorylation is known to be both associated with active spliceosomes and important for mediating protein-protein interactions (26,30). Our data now provide a new function for SF3B1 phosphorylation by CDKs in regulating SF3B1-nucleosome interactions. SF3B1 phosphorylation at proline-directed serine/threonine sites peaks at G2/M, which causes it to dissociate from nucleosomes. Consistent with the varied functions of cyclin B/CDK1 in mitosis initiation, we speculate that CDK1-dependent SF3B1 disassociation from chromatin may be necessary to ensure efficient chromatin condensation, given the size of the spliceosome, which could serve as an impediment to this process if remaining associated.

We have also learned that SF3B1-nucleosome interactions in vitro are dependent on the presence and phosphorylation of linker histone H1, especially the H1.4 isoform. Our in vivo data and previously reported MS-based studies (36) demonstrate that H1.4, which plays an important role in the regulation of chromatin organization and transcription, interacts with SF3B1. H1.4 methylation at lysine 26 promotes its binding to HP1 and L3MBTL1 leading to the recruitment of these factors,

with roles in heterochromatin formation and transcriptional repression, to chromatin (53,54). H1.4 acetylation at lysine 34, on the other hand, is known to co-localize on active promoters with a transcriptional activator, TAF1, and have a positive effect on transcription (55). Similar to SF3B1, H1.4 also undergoes phosphorylation during cell cycle and contains a number of CDK substrate motifs. Phosphorylation at threonine 187, a CDK substrate motif, is associated with RNA polymerase mediated transcription (47) and phosphorylation of H1.4 at serine 27 inhibits HP1 binding and potentially heterochromatin formation (53). Given that SF3B1 and histone H1 exist in a complex together, are both phosphorylated during cell cycle progression by CDKs and the recognized role of H1 in regulating transcription and chromatin organization, the potential for SF3B1 and histone H1 acting collaboratively on chromatin in regulating chromatin structure and transcription should be further explored. Additionally, phosphorylated H1.4 has been found in the nucleolus, suggesting a sequestration function to regulate the activity of binding partners (56). Studies utilizing high-resolution imaging to examine whether H1 isoforms interact with SF3B1 differentially during cell cycle progression and alter SF3B1 subcellular localization may provide additional mechanistic detail on the regulation of SF3B1-chromatin interactions.

To understand the functional outcomes of cell cycle dependent SF3B1-nucleosome interactions, we performed ChIP-seq and RNA-seq in synchronized cells. ChIP-seq revealed a cell cycle stage-specific pattern of SF3B1 chromatin occupancy with increased binding within transcribed regions (exons, introns as well as transcription start and termination sites) in G1/S compared G2/M with genome-wide analysis (Figure 5A). While RNA-Seq of matched samples revealed cell cycle stage-specific splicing and expression changes, in agreement with a previous report (31), we found no evidence that SF3B1 occupancy on chromatin within these intragenic regions influences their expression level or alternative splicing in corresponding transcripts in a cell cycle stage-specific manner. Our findings were surprising since they differ from a previous report by Kfir et al. (33) that suggested a direct, positive effect on the splicing of exons located near SF3B1-bound regions of nucleosomes. It is important to note some differences between the design of these two studies: Kfir et al's conclusions were not based on correlation of genome occupancy and changes in splicing, but rather by comparing exon utilization after SF3B1 knock-down or trichostatin A treatment. Our results are based on direct correlation of SF3B1 occupancy to changes to the transcriptome in two cell cycle states with the maximal identified difference in SF3B1-nucleosome interaction. We did

not observe significant alterations in cell cycle kinetics or apoptosis with overexpressing the compound phosphosite-mutant (6A) SF3B1 versus wild-type, though a relatively limited impact of these amino acid substitutions on overall phosphorylation and histone binding may account for the absence of an obvious phenotype with overexpression of this mutant protein. Further investigations into the functional importance of SF3B1-nucleosome interactions and their possible role in regulation of transcription (independent of splicing) and downstream biology are warranted, especially in light of our finding that the disease-associated SF3B1 K700E mutant demonstrates altered interaction with chromatin.

SF3B1 associates with a number of proteins involved in control of chromatin modifications and transcription. For example, SF3B1 interacts with NIPP1 in a phosphorylation-dependent manner, and NIPP1 is implicated in the regulation of EZH2 occupancy at promoter regions (57,58). SF3B1 and PHF5a are found together in the SF3B complex and PHF5a has been shown to regulate RNA polymerase-dependent transcription of pluripotency genes (45). Furthermore, it has been shown that splicing promotes the recruitment of methyltransferase HYPB/Setd2 to chromatin resulting in histone H3 lysine 36 methylation at actively transcribed intron-containing genes (59). A

potential non-splicing related role for cell cycle-dependent SF3B1-nucleosome interactions might be to enable key chromatin modifications by facilitating the recruitment of chromatin modifying proteins and complexes to transcriptionally active chromatin. Thus, our data may point to a previously unappreciated role for CDKs: coordinators of chromatin modifications via the modulation of SF3B1-nucleosome interactions during cell cycle progression.

**Experimental procedures:****Cell Culture:**

HeLa cells were cultured in DMEM (Gibco Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Gemini Bioproducts) and 1X penicillin/streptomycin (Pen/Strep) (Gibco Life Technologies). K562 cells were cultured in RPMI-1640 (Gibco Life Technologies) supplemented with 10% FBS and 1X Pen/Strep. HeLa and K562 cells were synchronized in G2/M by culturing in their respective growth media with 40 ng/mL Nocodazole (Sigma-Aldrich, M1404) for 20 hours. For G1 and G1/S synchronizations, HeLa cells were cultured in growth media with 40 ng/mL Nocodazole for 16 hours. For all synchronization experiments, cycle status was verified using flow cytometry as described below. Cells were then washed with growth media once and cultured in growth media for 20 hours (G1) or 24 hours (G1/S). K562 cells were cultured in growth media with 40 ng/mL Nocodazole for 20 hours, washed once with growth media and cultured in growth media for 4 hours (G1) or 7.5 hours (G1/S). Pharmacologic inhibitors were added to G1/S synchronized cells for the indicated times following release from Nocodazole treatment at the following concentrations: Roscovitine (Sigma-Aldrich, R7772) 20  $\mu$ M; Purvalanol-A (Sigma-Aldrich, P4484) 15

$\mu\text{M}$ ; Harmine (Sigma-Aldrich, 286044) 5  $\mu\text{M}$ ; Okadaic Acid (OA) (Sigma-Aldrich, O7760) 20 nM. Retroviral supernatants were prepared by transfecting Phoenix cells (G. Nolan, Stanford) by calcium phosphate precipitation method. K562 cells were spinoculated with retroviral supernatants for 90 minutes at room temperature (RT).

**Flow cytometry:**

For cell cycle analysis, cells were fixed with 75% ethanol and incubated at 4°C overnight. Cells were then washed once with 1X phosphate-buffered saline (PBS) (Gibco, Life Technologies) and stained with 10  $\mu\text{g}$  propidium iodide (PI) (Sigma-Aldrich 81845), 0.1% bovine serum albumin (BSA) (Sigma-Aldrich A2153) and 10  $\mu\text{g}$  RNase-A (Sigma-Aldrich, R6513) in 1X PBS. Flow cytometry was performed on LSRII flow cytometer (BD Biosciences) and analysis used FlowJo software. Apoptosis detection was performed using Annexin V Apoptosis Detection Kit I (BD Biosciences).

**Plasmids:**

Codon optimized FLAG-tagged SF3B1 (11) was cloned into BamHI and EcoRI sites of pBABE-Puro and pUC57 plasmids. Threonine-to-alanine point mutations in SF3B1 were introduced by site-directed mutagenesis using the following primers:

T142A: 5' GCC GAC GGA GGC AAA GCA CCT GAT CCA AAG ATG 3'; T211A: 5' CAG ACA CCT GGA GCA GCT CCA AAG AAA CTG AGT 3'; T257 261A: 5' AGC AAG ATT TGG GAC CCC GCA CCT TCT CAT GCT CCA GCT GGC GCA GCA 3'; T426A: 5' TAC GTC CCA ATC AGA GCT CCC GCC AGG AAA CTG 3'; T434A 5' AGG AAA CTG ACC GCT GCA CCC ACT CCT CTG 3'. N-term SF3B1 (amino acids 1-500) were PCR amplified from pBABE-Puro-SF3B1 and sub-cloned into pGEX-6P2 (GE Healthcare Lifescience) using BamHI and EcoRI sites.

**Immunoprecipitation (IP) and Immunoblots:**

Whole cell lysates were prepared by lysing cells in RIPA buffer (150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% Sodium Deoxycholate, 50mM Tris pH7.5 and 2mM EDTA) supplemented with protease and phosphatase inhibitors (PPI) (10 mg/ml each of aprotinin, leupeptin and pepstatin, 50 mM sodium fluoride and 1 mM sodium

orthovanadate). Immunoprecipitations were performed by rotating lysates at 4 °C overnight with the indicated antibodies (2 µg/sample) conjugated to Protein Sepharose A (Life Technologies) or Protein Sepharose G (GE Healthcare) beads for antibodies produced in rabbit or mouse respectively. Beads were washed 3x with RIPA buffer and equal amounts of samples were electrophoresed and transferred to nitrocellulose membranes. Immunoblots were performed using the indicated antibodies. Nucleosome-enriched lysates were prepared as described in (33) with some modifications. Cells were lysed using nucleosome lysis buffer (60 mM KCl, 15 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.1 mM EGTA, 15 mM Tris-HCl pH 7.5, 0.2% NP-40) supplemented with 0.1 mM DTT and PPI. Lysates were then passed through a sucrose cushion (50 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 25 mM KCl, 1.2M Sucrose) by centrifugation at 3.5g for 15 minutes at 4 °C. The pellet (containing the nuclei) was resuspended in the digestion buffer (50 mM Tris-HCl pH 7.5, 4mM MgCl<sub>2</sub>, 1mM CaCl<sub>2</sub>, 0.32 M Sucrose) with 200U micrococcal nuclease (Mnase) (Worthington Biochemical) and incubated for 10 minutes at 37 °C. Digestion was stopped by the addition of 1 mM each of EDTA and EGTA. Nuclei were pelleted by centrifugation at 3.5g for 10 minutes at 4 °C. Pellet was resuspended in 500 µl solubilization buffer (50mM HEPES pH 7.6, 500mM LiCl, 1mM

EDTA, 0.7% sodium deoxycholate, 0.1% SDS and 1% NP-40) and rotated at 4 °C for 1 hour followed by centrifugation at 12,000g for 10 minutes at 4 °C. The pellet was discarded and the supernatant, the nucleosome enriched fraction, was used for immunoprecipitations and immunoblot assays. DNA was extracted from a small aliquot of the nucleosome-enriched fraction using phenol: chloroform: isoamyl alcohol (Thermo Fisher Scientific 15593031) and run on a 2% agarose gel to confirm the expected size of the digested DNA (147 bp and multiples thereof). For immunoprecipitations, nucleosome enriched fractions were diluted to 1 ml in dilution buffer (0.005% SDS, 0.1% Triton X-100, 1.2 mM EDTA, 1.67 mM Tris HCl pH8.0, 167 mM NaCl) and incubated with antibodies (2 µg/sample) conjugated to Sepharose A or G beads with rotation at 4 °C overnight. Beads were washed with dilution buffer 3x and immunoblots were performed as described above. RNAse-A treatment was performed as described in (18).

The following antibodies were used: SF3B1 (Abcam ab172634 (immunoblot), MBL D221-3 (IP and ChIP)), phospho-serine/threonine-Proline (Abcam ab9344), phospho-threonine 313 SF3B1 (Cell Signaling Technologies 25009s), histone H3 (Abcam ab1791 (immunoblot), Active Motif 61475 (IP)), histone H1 (Active Motif 39707),

histone H1.4 (Abcam ab105522), phospho-histone H1 (Abcam ab4270), U1-70k (Abcam ab51266, Santa Cruz Biotechnology sc-390988 and an antibody produced and provided as a gift by Dr. Douglas Black's Lab), beta-Actin (Sigma-Aldrich A5441), SF3B2 (Abcam ab56800 and Novus 79848), Anti-Flag tag (Sigma-Aldrich F1804).

### **GST-tagged protein expression and immunoaffinity purification:**

pGEX-6P2-N-term-SF3B1<sup>1-500</sup> was transformed into BL21 competent E.coli cells (Life Technologies C-6000-03) by heat-shock method. Cells were grown in LB-media at 37°C until they reached an O.D of 0.4-0.6. SF3B1<sup>1-500</sup> expression was induced by treatment with 1 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) (Fisher Scientific BP16201) for 4 hours at 37°C. Cells were pelleted by centrifugation and frozen at -80°C. Pellets were thawed and lysed in 1% Triton X-100 in 1X PBS, sonicated using a Diagenode Bioruptor (12 cycles of 10s pulse, 30s rest). Glutathione agarose beads (25  $\mu$ L/ml lysate) (Thermo 16100) were washed 2X with 1X PBS (10X bead volume). Lysates were incubated with glutathione beads at 4°C overnight with rotation. Beads were washed 3X with 1X PBS (10X bead volume).

**In vitro Binding Assay:**

N-Term, GST-tagged SF3B1 (SF3B1<sup>1-500</sup>) was expressed in bacteria and bound to glutathione beads as described above. Beads were then washed 3X with lambda-phosphatase wash buffer (25 mM Tris-HCl pH 7.5, 50 mM NaCl, 2.5 mM MnCl<sub>2</sub>, 2 mM DTT) and incubated with lambda-phosphatase (NEB P0753) for 20 minutes according to the manufacturer's protocol. Phosphorylation by cyclin E-CDK2 in vitro was performed as described below. Next, beads were washed 1X with binding buffer (0.1% Triton X-100, 1.2 mM EDTA, 1.67 mM Tris HCl pH8.0, 100 mM NaCl), resuspended in 1ml binding buffer followed by incubation with 1 µg purified histone H1 (Sigma-Aldrich, Calf Thymus), recombinant human histone H1.0 (NEB M2501) or recombinant human histone H1.4 (MyBioSource MBS963223) and 1 µg purified HeLa mono-nucleosomes (EpiCypher 16-0002) at 4 °C with rotation for 12 hours. Histone H1 was treated with lambda-phosphatase for 20 minutes according to the manufacturer's protocol or phosphorylated for 30 minutes in kinase assay buffer as described below. Beads were washed 3X with binding buffer followed by immunoblot analysis as described before.

**In vitro kinase reactions:**

Substrates for cyclin E-CDK2 included FLAG-tagged full-length SF3B1 expressed in mammalian cells as above or GST-SF3B1<sup>1-500</sup>. Immunoprecipitation used either protein sepharose G or glutathione beads, and these were then washed 3X with kinase wash buffer (25 mM Tris-HCl pH 7.5, 70 mM NaCl, 10 mM MgCl<sub>2</sub>, 1mM DTT) and resuspended in kinase reaction buffer (kinase wash buffer, 10 μM ATP (NEB P0756), [ $\gamma$ -<sup>32</sup>P]ATP for phosphorylation of full-length SF3B1 only (5 μCi per reaction; 1Ci=37GBq), 0.5 μg active cyclin E-CDK2 (Millipore 14-475) followed by incubation at 37 °C.

**RNA sequencing:**

Synchronized HeLa cells were resuspended in TRIzol (Thermo Fisher Scientific 15596026). Total RNA was extracted from TRIzol using the manufacturer's protocol and quantified using a Qubit fluorometer (Thermo Fisher Scientific). Poly-A mRNA was enriched using NEBNext Poly-A mRNA isolation kit (NEB E7490). Libraries for paired-end sequencing were prepared using NEBNext Ultra RNA Library Preparation Kit (NEB E7530) and sequenced on Illumina NextSeq 500. Approximately 511 million reads

passing filter were obtained with an average of approximately 122 million reads per sample. Reads were aligned to human genome hg19.

### **ChIP sequencing:**

Nucleosome-enriched lysates were prepared from synchronized HeLa cells as described above. Lysates were diluted to 1 ml with dilution buffer and incubated with 7.5 µg/sample SF3B1 antibody (MBL D221-3) at 4 °C with rotation overnight. Sheep anti-mouse M-280 Dynabeads (Life Technologies 11201D) were washed and pre-cleared by incubation with 0.5% BSA in PBS at 4 °C with rotation overnight. Beads were then added to the immunoprecipitated lysates and incubated at 4 °C with rotation for 4 hours. Beads were washed four times with dilution buffer, magnetically isolated and resuspended in 150 µl SDS elution-buffer (1% SDS, 10mM EDTA, 50 mM Tris pH 8.0), then eluted by incubation at 65 °C overnight. Samples were treated with RNase-A and proteinase-K followed by DNA isolation using phenol: chloroform: isoamyl alcohol. Libraries for sequencing were prepared using NEBNext Ultra DNA Library Preparation Kit (NEB E7370) and sequenced on Illumina NextSeq 500. Approximately 393 million

total reads passing filter were generated with an average of 75 million reads passing filter per sample. Reads were aligned to human genome hg19.

### **Bioinformatic Analysis:**

All Illumina reads were preprocessed with Trimmomatic (60) and aligned with Bowtie2 (61) to hg19 genome (ChIP-Seq) or STAR (62) to hg19 transcriptome (RNA-Seq). Peak calling for individual ChIP-Seq files was performed with HOMER (49) using the respective input files as controls. For differential peak calls and overlap of peaks, biological replicates were merged and analyzed using HOMER (getDifferentialPeaks or mergePeaks functions of HOMER). Differential enrichment of reads in the exonic region and correlation between ChIP-Seq files were performed using deepTools2 (14). Differential gene expression was determined from aligned RNA-Seq files using Cufflinks (50). Analysis of differential splicing was performed using rMATS (51). Custom perl scripts were used to integrate results of HOMER, Cufflinks and rMATS analyses. Detailed parameters for analysis and scripts will be made available on request.

## Chapter 3

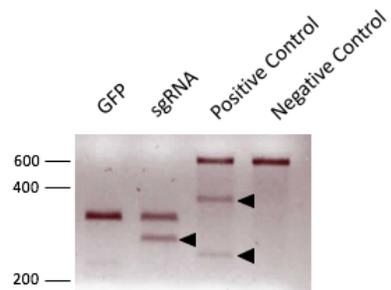
### Study of mutant SF3B1-induced altered splicing

#### 3.1 Generation of isogenic *Sf3b1*-mutant mouse embryonic stem cells (mESCs):

To study the effect of mutant SF3B1 expression on splice-site usage and splicing outcomes in a scalable system, we generated isogenic mESCs expressing the *Sf3b1* K700E mutation from its endogenous locus using CRISPR-Cas9 genome editing and homology directed repair (HDR). We selected a guide-RNA (gRNA) that directed Cas9 to exon 15 of *Sf3b1*, close to lysine 700, where it introduced bi-allelic double-stranded breaks (DSBs). The ability of this gRNA-directed Cas9 to introduce double-stranded breaks at this locus was confirmed by SURVEYOR mutation detection assay (Figure 18A). We then introduced the Cas9-gRNA complex into the CCE mESC line (63,64) in conjunction with a HDR vector. The HDR vector contained a neomycin (Neo) resistance gene flanked by sequences homologous to the genomic region surrounding and containing the K700E (AAA→GAA) mutation. Although Cas9 introduced bi-allelic DSBs, we consistently found that only one allele underwent HDR to incorporate the repair vector. The other allele was found to contain insertions and deletions (INDELS), most likely the result of repair via non-homologous end joining (NHEJ) (Figure 18B).

Figure 18.

(A)



(B)

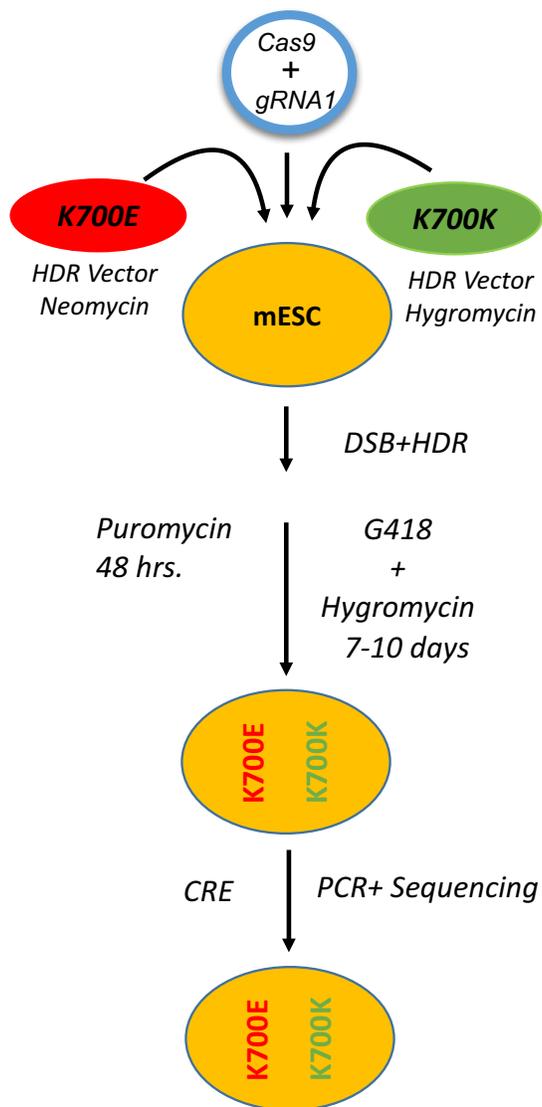


**Figure 18. Generation of isogenic Sf3b1-mutant mouse embryonic stem cells (mESCs).**

(A) SURVEYOR mutation detection assay to assess double-stranded breaks (DSBs) introduced by the Cas9-gRNA complex. Arrowheads, indicative of mismatches resulting from INDELS, point to cleavage products of PCR fragments amplified from genomic DNA of mESCs transfected with Cas9-gRNA. (B) Results from Sanger sequencing of genomic DNA isolated from mESC edited using the single HDR vector strategy were aligned with Sf3B1 genomic DNA sequence using SerialCloner® software.

In order to faithfully replicate the human disease, where SF3B1 mutations are almost always hemizygous, we developed a dual HDR vector strategy to ensure that while one *Sf3b1* allele was edited by the introduction of the K700E (AAA—>GAA) mutation, the other allele did not contain INDELS (Figure 19). Using this strategy, we generated hemizygous *Sf3b1* K700E/K700K mutant mESCs (K700E) and control *Sf3b1* K700K/K700K mESCs (K700K). Recombination of the repair vectors was observed at similar frequencies for both K700K and K700E editing at 37% and 40.3%, respectively (Figure 20A). Editing was confirmed by PCR analysis of genomic DNA (Figure 20B). Sequencing of genomic DNA after Cre-mediated excision of the resistance genes confirmed successful editing at the *Sf3b1* locus in both K700E and K700K cells (Figure 20C). We confirmed the expression of full-length *Sf3b1* at the mRNA and protein levels in both K700E and K700K cells (Figure 20D and 20E). *Sf3b1* expression was confirmed by Sanger sequencing of cDNA prepared from K700E and K700K cells (Figure 20F).

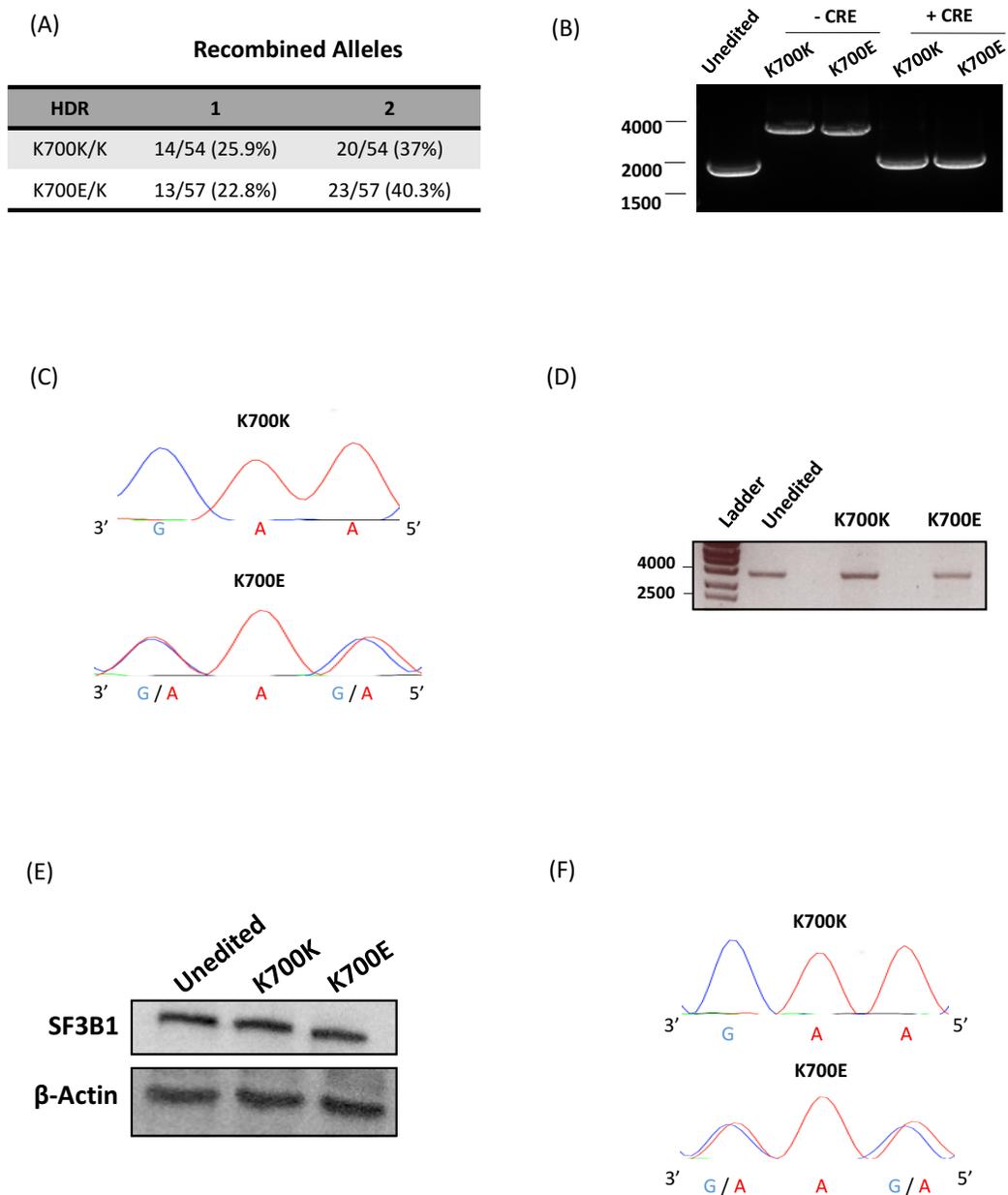
Figure 19.



**Figure 19. Generation of isogenic *Sf3b1*-mutant mouse embryonic stem cells (mESCs) using a dual HDR vector strategy:**

Targeting of mESCs to introduce hemizygous mutations at the *Sf3b1* locus using CRISPR-Cas9 and a dual vector HDR strategy. The Cas9-gRNA complex was introduced into mESCs in conjunction with two HDR vectors. One vector contained the neomycin resistance gene flanked by homology arms containing the K700E mutation, while the other vector contained a hygromycin resistance gene containing a synonymous K700K mutation in the homology arm. Similarly, a control *Sf3b1* K700K/K700K mESC line was generated.

Figure 20.



**Figure 20. Characterization of isogenic *Sf3b1*-mutant mouse embryonic stem cells (mESCs):**

(A) Frequency of recombination at one or both alleles resulting from the editing of mESCs using the dual HDR vector strategy. (B) PCR analysis of genomic DNA isolated from unedited, K700K and K700E mESCs pre and post Cre excision of the antibiotic resistance genes. (C) Confirmation of successful editing by Sanger sequencing of genomic DNA isolated from K700K and K700E mESCs. (D) Analysis of the expression of full-length *Sf3b1* mRNA in unedited, K700K and K700E mESCs by PCR amplification from cDNA. (E) SF3B1 protein expression by western blot analysis from cell lysates extracted from unedited, K700K and K700E mESCs. (F) Analysis of *Sf3b1* mRNA expression from edited alleles by Sanger sequencing of cDNA isolated from K700K and K700E mESCs.

### 3.2 Aberrant splicing of endogenous transcripts in *Sf3b1* K700E mESCs:

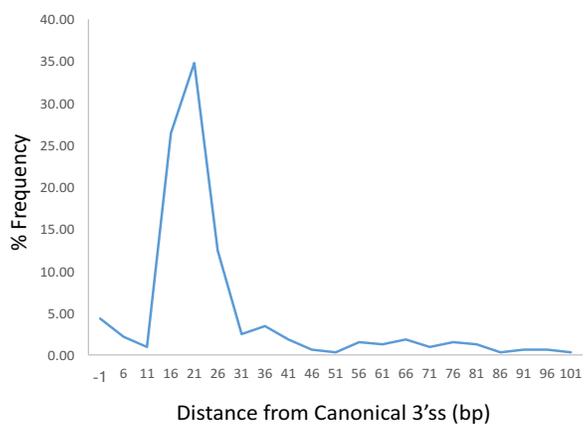
Recent studies investigating the impact of SF3B1 mutations on splicing outcomes have used MDS and cancer patient samples, mouse models and engineered cell lines to conclude that mutant SF3B1 expression results in aberrant splicing due to the selection of a cryptic 3' splice site (3'ss) upstream of the canonical 3'ss and altered BP usage (12-16). To understand whether the *Sf3b1* K700E mutation in our mESC model similarly affects splicing outcomes globally, we performed paired-end RNA-sequencing (RNAseq). Illumina-compatible libraries were prepared and total RNA from K700K and K700E mESCs were sequenced. Paired-end reads were mapped to the mouse genome (mm9) to identify splice junctions (SJ). Further analysis of novel SJ led to the identification of 2370 cryptic 3'ss in K700E cells. These cryptic 3'ss were enriched 10-30 bp immediately upstream of the canonical 3'ss (Figure 21A) consistent with previous reports (12-16). Computationally identified cryptic 3'ss usage in 3 candidate genes was validated by RT-PCR analysis (Figure 21B).

We then used rMATS, a computational tool that identifies differential alternative splicing events between samples by using reads across SJ and exon coverage, to identify differential alternative splicing events (A5'ss: altered 5' splice site, A3'ss:

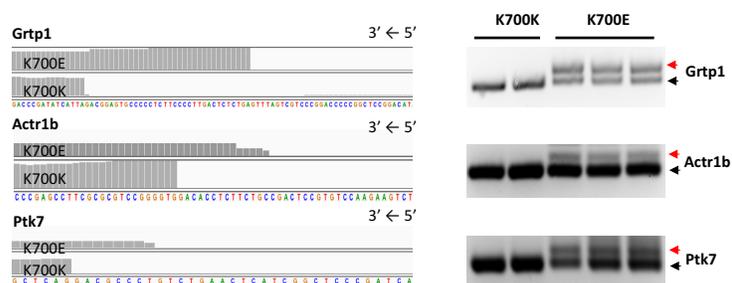
alternate 3' splice site, RI: retained intron, SE: skipped exon, MXE: mutually exclusive exons) between the K700K and K700E mESCs. Of the 39,635 differential alternative splicing events detected, 4943 events were found to be significantly altered ( $p < 0.05$ ). SE were found to be the most frequent differential alternative splicing event at 48.69% while A3'ss accounted for 7.12%, consistent with MDS patient and cell line data previously analyzed (14) (Figure 21C). Taken together, these data demonstrate that changes in splicing due to Sf3b1 K700E expression in mESCs are similar to those observed in previously reported models and that our mESC model is suitable for the systematic study of altered splicing outcomes caused due to the expression of Sf3b1 K700E mutation.

Figure 21.

(A)

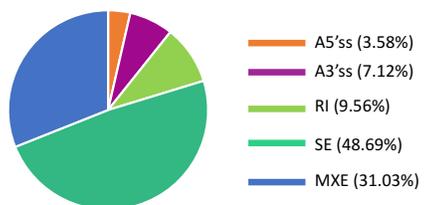


(B)



(C)

EVENT	DETECTED	ALTERED (p<0.05)
A5'SS	1413	177
A3'SS	2370	352
RI	2065	473
SE	29137	2407
MXE	4650	1534
Total	39635	4943



**Figure 21. Aberrant splicing of endogenous transcripts in Sf3b1 K700E mESCs:**

(A) Distribution of the distances of cryptic 3'ss in relation to their canonical 3'ss are shown for K700E mESCs, as determined by RNA-seq.

(B) Computationally identified cryptic 3'ss selection in four candidate genes validated by RT-PCR using RNA from two K700K and three K700E mESC clones. Red and black arrows represent isoforms associated with cryptic 3'ss and canonical 3'ss selection, respectively.

(C) Frequency of differential alternative splicing events detected and altered ( $p < 0.05$ ) between K700K and K700E mESCs. Of A5'ss (Alternative 5'ss), A3'ss (Alternative 3'ss), RI (Retained intron), SE (Skipped exon), MXE (Mutually exclusive exons), SE were the most frequent differential alternative splicing events.

### **3.3 Massively parallel reporter assay (MPRA) for the unbiased analysis of BP usage in mutant Sf3b1 expressing mESCs:**

While our mESC model and other isogenic murine models have consistently shown that mutant SF3B1 results in aberrant splicing due to the use of a cryptic 3'ss upstream of the canonical 3'ss, the precise molecular mechanism behind the cryptic 3'ss selection remains unclear. One proposed model suggests that aberrant splicing due to cryptic 3'ss usage results from an altered BP selection (13). The model is plausible given the proposed function of SF3B1 in stabilizing the interaction between U2 snRNP and the BP. However, it is based on the analysis of a limited number of putative BP sequences. Moreover, the genome-wide computational analysis of BP choice is fraught with challenges. Firstly, genome-wide analysis of BP choice by the sequencing of BP-containing intermediary intronic sequences (lariats) is challenging due to the low detectable frequencies of such lariats and due to technical challenges associated with current lariat sequencing methodologies. Secondly, such analyses have limited utility given the relatively small proportion of aberrantly spliced transcripts that result from cryptic 3'ss usage. In order to overcome these challenges and better understand altered BP selection due to mutant SF3B1, we decided to combine the power of our isogenic

mESC model with a massively parallel splicing reporter assay (MPRA) (65). The MPRA utilizes a library containing millions of synthetic mini-gene reporters, each containing a unique 20-nucleotide barcode. Each minigene consists of exons separated by introns that contain two 3' splice acceptor (SA) sites: SA1 and SA2. Each minigene also consists of a cryptic SA (SA<sub>CRYPT</sub>) upstream of SA2. Additionally, SA1 is flanked on each side by two stretches of degenerate sequences that are 25 bp long (25A and 25B) (Figure 22A). It has been previously shown that SA2 is the predominant splice acceptor in a majority of minigenes in the library and SA1 selection is strongly influenced by motifs present within the degenerate 25A and 25B sequences (65). Thus, we hypothesized that using the minigene library in combination with our mESC model would enable the unbiased identification of sequences and motifs that influence altered splice site and BP usage due to mutant SF3B1.

K700E and K700K mESCs were transiently transfected with the minigene library and total RNA was isolated from the cells 48 hours after transfection. Illumina-compatible libraries were prepared using minigene-specific primers and paired-end sequencing was performed. Reads were matched to individual minigenes using their unique 20-nucleotide barcodes. An analysis of SA usage in K700E and K700K cells

revealed that while SA2 was the predominant SA utilized in both cell types, K700E cells used SA1 and SA<sub>CRYPT</sub> at higher frequencies compared to K700K cells (Figure 22B). Interestingly, in K700E cells, this increase came at the expense of the canonical SA2.

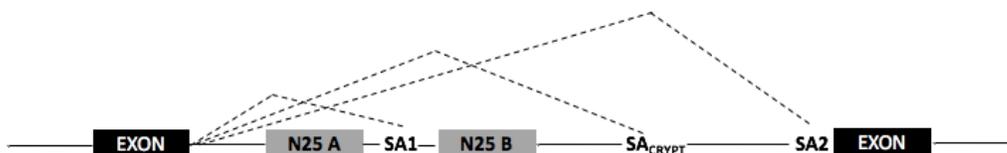
In order to determine whether specific sequence motifs upstream or downstream of a 3' splice acceptor site influence 3' ss selection differently in K700E and K700K cells, we decided to assess whether motifs within the degenerate 25A and 25B regions promoted enhanced SA1 usage. This was done by calculating the odds-ratio of unique 6-mers present within the degenerate regions. An odds ratio is a measure of the association between a variable (a unique 6-mer) and the occurrence of a fixed outcome (splicing using SA1). A higher OR would reflect a high probability that the presence of a 6-mer results in an increase in SA1 usage and would thus be predicted to act as a splicing enhancer. Conversely, a lower OR would reflect a lower probability that the presence of a 6-mer results in SA1 usage and would thus be predicted to act as a suppressor of splicing. Using the MPRA in an overexpression system, our colleagues reported that there were no significant differences in the pattern of OR for 6-mers in HEK293T cells overexpressing WT or SF3B1 K700E. 6-mers with higher OR, that were predicted

enhancers of splicing, conformed to the putative consensus BP motif YTAAY. 6-mers with lower OR that were predicted splicing repressors were G-rich sequence as has been shown previously (65) (Personal communication from Dr. Manoj Pillai, Yale University). Hence, we decided to focus only on non-canonical BP sequences are utilized differentially by WT and mutant SF3B1. We identified 6-mers that were associated with a higher OR in K700E cells compared to K700K cells. Analysis of 6-mers with the highest differential OR ( $\Delta\log_2$  K700E vs. K700K) revealed that these motifs contained non-canonical BP sequences. Interestingly, these BP sequence were enriched in motifs that differed from the canonical YTAAY sequence at the -1 position relative to the BP adenosine (Figure 22C and 22D). These findings suggest that K700E and K700K cells differ in BP usage only in the presence of a non-canonical BP. This effect, however, is modest but consistent with previous observations of modest changes in splicing resulting from cryptic 3'ss selection. More importantly, our findings point to key differences in BP motifs that influence splice site selection by WT and mutant SF3B1 and prove the utility of our mESC model in combination with the MPRA. Insights gained from our findings can be translated to humans and can potentially help narrow down the list of potential genes and molecular pathways affected by mutant SF3B1

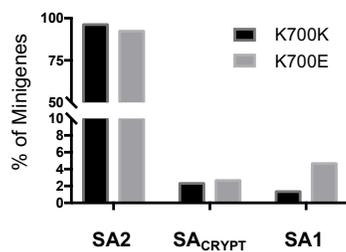
expression during disease pathogenesis.

Figure 22.

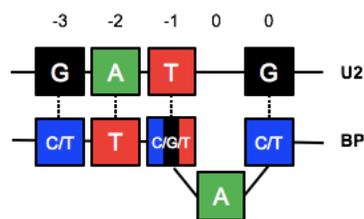
(A)



(B)



(C)



(D)

6mer	Log <sub>2</sub> _OR (K700K)	log <sub>2</sub> _OR (K700E)	Δ_log <sub>2</sub>
CCTIAT	0.08	0.91	0.83
TTTIAC	0.28	1.02	0.74
ATCATA	0.11	0.83	0.71
TAAAGAT	0.03	0.72	0.69
TTTTIA	0.51	1.15	0.64
ACTCAC	1.16	1.78	0.62
TTCTIA	0.43	1.04	0.61
TTIACC	0.25	0.85	0.60
CACTGA	0.44	1.01	0.57
CGCTTA	0.17	0.72	0.55

**Figure 22. Massively parallel reporter assay (MPRA) for the unbiased analysis of****BP usage in mutant Sf3b1 expressing mESCs:**

(A) Schematic showing the minigene splicing reporter library. Shown are the three splice acceptor sites SA1, SA2 and SA<sub>CRYPT</sub> along with the 25 bp degenerate regions 25A and 25B that lie upstream and downstream of SA1 respectively. (B) Differential splice acceptor site usage in K700K and K700E cells is shown. Bars represent the proportion of minigene transcripts that were found to utilize each splice acceptor site.

(C) Schematic showing the base-pairing between the U2 snRNA and the non-canonical BP sequence YTAAY varying at the -1 and -3 positions and favored by K700E cells.

(D) 6-mer sequences that enhance SA1 splice site usage and with the highest odds ratio (OR) difference between K700E and K700K cells are shown. +3 to -1 positions relative to the BP adenosine are shown in red. Non-canonical BP sequences varying from the canonical YTAAY sequence are underlined.

\*Panels (B), (C), (D) generated by Dr. Panoj Pillai.

**Conclusions:**

Mutations within the HEAT domains of U2 snRNP component SF3B1 have been frequently found in multiple neoplastic processes such as MDS, CLL, uveal melanoma and breast cancers (4,7-10). SF3B1 mutations are found in a significant majority (75%) of MDS cases that are characterized by the presence of ring sideroblasts (4,7). SF3B1 mutations are almost always hemizygous missense mutations and mutually exclusive with other splicing factor mutations (in U2AF1, SRSF2 and ZRSR2). This suggests that these mutations are driver mutations and can initiate disease pathogenesis by themselves (66).

SF3B1 is a core component of the U2 snRNP, makes contact with the BP adenosine during splicing catalysis, and stabilizes the interaction between U2 snRNA and the BP adenosine (5,6). Given this role, it has been postulated that SF3B1 mutations alter splicing outcomes by altering BP and 3'ss usage. It is believed that SF3B1 mutations give rise to a clonal hematopoietic stem and progenitor cell population, with altered differentiation potential, by altering the splicing of key genes important during hematopoietic differentiation. Transcriptomic analyses in cell lines, patient samples and mouse models of mutant SF3B1 have revealed that expression of

mutant SF3B1 results in altered splicing due to aberrant BP selection and consequently the selection of a cryptic 3'ss that is upstream of the canonical 3'ss (12-16). Through the study of a small set of BPs and 3'ss, these studies suggested that the presence of a weak poly-pyrimidine tract, interrupted with adenosines directly upstream of the cryptic 3'ss, leads to altered BP and 3'ss selection by mutant SF3B1 (12,13). Current hypotheses imply that mutant SF3B1 expression results in obligate usage of an altered BP and cryptic 3'ss selection. However, these hypotheses are not supported by experimental observations where transcripts generated due to cryptic 3'ss usage are observed at low frequencies and not at a 50% frequency that would be expected due to hemizygous mutant SF3B1 expression.

Technical limitations of current computational methods, low detectable frequencies of BP-containing splicing intermediaries (lariats) and limited numbers of cryptic 3'ss and associated BPs available for analysis in patient-derived datasets have hindered the identification of the precise molecular mechanism behind altered BP and 3'ss selection due to mutant SF3B1 expression. While significant progress has been made in understanding how mutant SF3B1 expression contributes to MDS pathogenesis, key questions remain unanswered. To address some of these questions,

we developed an experimental system that faithfully reproduces the genetics of disease-associated mutant SF3B1. We generated isogenic mESCs expressing the hemizygous *Sf3b1* K700E mutation from its endogenous locus using CRISPR-Cas9. Upon systematic characterization of these cells, we found that mutant SF3B1 expression results in altered splicing due to the selection of a cryptic 3'ss directly upstream of the canonical 3'ss, consistent with previous reports. In order to define factors influencing BP and cryptic 3'ss selection by WT and mutant SF3B1 containing spliceosomes, we decided to combine the power of our mESC line with a MPRA that allows unbiased analysis of splicing. The MPRA consists of a library of millions of minigenes that vary only in fixed 25 bp regions upstream and downstream of a 3'ss. Using this MPRA in an overexpression system, our colleagues discovered that WT and mutant SF3B1-containing spliceosomes do not differ substantially in their utilization of transcripts containing the canonical BP motif YTAAY. This led us to examine the utilization of non-canonical BP motifs. Careful analysis revealed preferential usage of non-canonical BPs only by mutant and not WT SF3B1-containing spliceosomes. These non-canonical BPs were highly enriched in motifs that varied at the -1 position relative to the BP adenosine. Our findings point to a crucial difference between WT and mutant

SF3B1 whereby weaker interactions between the U2snRNP and non-canonical BPs are likely stabilized by mutant SF3B1 but not WT SF3B1. Based on these findings, we propose the following model of BP usage by mutant SF3B1 expressing cells: (1) canonical BP motifs are utilized at similar frequencies by both WT and mutant SF3B1-containing spliceosomes and (2) non-canonical BP motifs, when present, can be preferentially utilized by mutant SF3B1 in the presence of cryptic 3'ss upstream of the canonical 3'ss. This model, unlike previously postulated models, allows occasional use of new BPs by mutant SF3B1 and potentially explains why isoforms generated due to cryptic 3'ss account for only a small proportion of the total transcript.

While transcriptomic analysis of patient-derived samples can provide useful information, such analyses have demonstrated limited utility in understanding molecular mechanisms that drive altered splicing due to mutant SF3B1 expression. Here, we have demonstrated the utility of an isogenic mESC model of a SF3B1 disease-associated mutation combined with the MPRA. Our results demonstrate that this system can, in an unbiased manner, help define the biochemical mechanism by which splicing factor mutations alter splicing outcomes. Additionally, this system overcomes challenges associated with the study of endogenous transcripts that has hindered the

comprehensive analysis of BP and splice site usage in the context of splicing factor mutations. Use of our isogenic mESCs in combination with MPRA has several advantages: (1) our isogenic mESCs recapitulate human disease genetics; (2) the pluripotent capacity of mESCs allows their directed differentiation into mature erythroid cells and the systematic study of splicing during hematopoietic differentiation in a stage-specific manner; (3) mESCs are scalable and amenable to large-scale biochemical and genome-wide analyses; (4) the large number of minigenes reporters used in the MPRA allow the unbiased and statistically robust study of a large number of 3'ss and BPs.

Therapeutic approaches that target the spliceosome broadly have not yet achieved proof-of-concept in the treatment of SF3B1-mutant disease. Therefore, a significant next step would be to understand SF3B1 mutations convey gain-of-function pathologies. Identification of perturbed genes and pathways that underlie disease pathology will allow the development of specific and targeted therapies. While our results provide an important next step in this direction, significant efforts will be needed before these insights can be translated into the clinic. As a first step, it is necessary to determine whether our findings in a mouse model-system translate directly to humans,

especially since knock-in Sf3b1 K700E mouse models do not recapitulate the classic features (ring sideroblasts) of human MDS (15,16).

A number of strategies can be employed to build further upon the insights gained from our mESC model system. Improved methods for the culture and manipulation of human induced-pluripotent stem cells (iPSCs) present a unique opportunity to study disease pathology in pluripotent human cells. For example, MPRA performed in gene-edited iPSCs expressing mutant SF3B1 from their endogenous locus would allow the direct comparison of splicing in mouse and human model systems and help explain why mouse models do not faithfully recapitulate human SF3B1-mutant disease. Recently, Chang et al described the use of patient-derived iPSCs for studying the role of splicing factor mutations in disease pathology and reported that these cells represent a more robust model of disease compared to mouse models. Using a patient-derived iPSC model, the authors identified a greater percentage of genes mis-spliced due to P95H mutations in SRSF2, compared to previously reported mouse models (67). Rules of splicing established by using gene-edited iPSCs or patient-derived iPSC model systems in combination with MPRA can be used to train bioinformatics tools and machine-learning algorithms. These trained algorithms can then be applied to the transcriptomic

analyses of patient samples or patient-derived iPSCs to parse the endogenous transcriptomes and generate a tractable list of the most likely genes whose mis-splicing drives MDS pathology. Importantly, given their pluripotent potential, iPSCs can be differentiated into an erythroid phenotype which allows MPRA and downstream analyses in a hematopoietic stage-specific manner. Such analyses are important if we are to develop an understanding of the early, intermediate and late changes during disease pathogenesis that are a direct result of mutant-SF3B1 expression.

In summary, we have demonstrated the advantages of using a system of gene-edited cells in combination with a splicing reporter library for the unbiased and systematic study of altered splicing biochemistry conferred by mutations in the HEAT domain of SF3B1. Using this system, we have extended our current understanding of altered splice-site usage in SF3B1-mutated diseases and laid the foundation for future studies that will enable the identification of factors driving disease pathology.

**Experimental Procedures:**

**Cell Culture:** CCE mouse embryonic stem cells (Kindly gifted by Dr. Mitchell Weiss, St. Jude's Children's Research Hospital) were cultured on tissue-culture dishes coated with 0.1% gelatin (Sigma G9136) (w/v in sterile water) in DMEM (Gibco® 11965) media supplemented with 15% fetal bovine serum (Gemini Bio-products 100-125), 1X penicillin/streptomycin (Gibco® 15140-122), GlutaMAX® (Gibco ® 35050061), 0.4mM 1-Thioglycerol (Sigma M1735) and leukemia inhibitory factor (LIF).

**Gene Editing:** CCE mESCs cultured on inactivated MEFs were adapted to gelatin by passaging them two times on cell culture dishes coated with 0.1% gelatin.  $5 \times 10^6$  mESCs were transfected with 4ug of pX459-V2-sgRNA, 3ug pL452-Hygro-K700K, 3ug pL452-Neo-K700E or pL452-Neo-K700K using Lipofectamine®200 Reagent (Thermo Fisher Scientific) and plated into 10cm cell-culture dishes coated with 0.1% gelatin with mESC growth media. 24 hours after transfection, media was replaced with fresh media containing 2ug/mL Puromycin (Sigma Aldrich P8833). 48 hours after culture in growth media containing Puromycin, cells were rinsed once with PBS (GIBCO) and cultured in fresh mESC growth media containing 350ug/mL G418 (Thermo Fisher Scientific).

10131027) and 250ug/mL HygromycinB (Thermo Fisher Scientific 10687010) for 10 days. Growth media containing G418 and HygromycinB was replaced 2-3 days. After 10 days, individual mESC colonies were isolated and cultured on 0.1% gelatin with mESC growth medium until ready to screen.

**Surveyor Assay:** 5 million CCE mESCs were transfected with 10ug pX-459-V2 sgRNA or a control GFP plasmid using Lipofectamine® (Thermo Fisher Scientific) reagent according to manufacturer's protocol and cultured for 24 hours on 0.1% gelatin in complete mESC growth medium. 24 hours post-transfection, growth medium was replaced with fresh medium containing 2 ug/mL Puromycin and cells were cultured for 48 hours. 48 hours after Puromycin selection, fresh growth medium was added to cells and the cells were grown until 50-60% confluent. Cells were harvested by trypsinization and washed with PBS once. Genomic DNA (gDNA) was harvested from pX-459-V2 sgRNA transfected cells and untransfected control cells using Qiagen Puregene Core Kit B according to manufacturer's protocol. The genomic locus expected to be targeted by pX-459-V2 sgRNA was amplified from gDNA isolated from transfected and control cells by PCR using primers indicated below. The PCR-amplified fragment was

separated on a 1.5% agarose gel and separated using the IBD gel-purification kit according to manufacturer's protocol. Gel purified DNA fragments were treated with DNA endonuclease provided in the SURVEYOR Assay Kit according to the manufacturer's protocol. Cleavage of the DNA fragments was assessed by gel electrophoreses using a 1.5% agarose gel.

**Plasmids and Primers:** Guide RNA (sgRNA) (5'GCTCAAGCCCCTATGGA 3') used for targeting the Sf3b1 K700E locus in CCE cells was designed using the Zhang Lab MIT CRISPR design tool (crispr.mit.edu). sgRNA was cloned into the BbsI restriction site of pX459-V2 Cas9 from *S.pyogens* (Kindly gifted by Feng Zhang, Addgene plasmid 62988). Repair vectors were generated by cloning homology arms into the Sall-EcoRI and BamHI-NotI restriction sites of pL452-Neo and pL452-Hygro vectors (68). pL452-Hygro vector was generated by replacing the Neomycin resistance gene within pL452-Neo with a Hygromycin resistance gene using BclI and BsmI restriction sites. The following primers were used for the SURVEYOR assay: FWD 5' TGAAAGACTTCTCACGCAAATTT 3' REV 5' TAGTATAGTAGTTGGCATATTCT 3'. The following primers were used for screening CCE clones for the integration of the

repair vectors and confirming CRE-mediated excision: FWD 5' GCCACTGTTGATTGATGAAGA 3' REV 5'GGCTGAGCACACCTTTAATT3'. The following primers were used for amplifying full length Sf3b1 cDNA: FWD 5' CCTCTGTTGATGTCCCCTACAC 3' REV 5'CTGCTGCGCCTAGATTACCC3'. Cryptic 3' splice-site usage identified by RNA-seq was validated using the following primers: Grtp1: FWD 5' GGACGCTCTTGTTGGAAGGA 3', REV 5' CAGGGTCCACAACACACCCAT 3'; Actr1b: FWD 5' CCACAAAAGGACGAGGCTCT 3', REV 5' AGACTTGTGGATGGCGAAGG 3'; Ptk7 FWD 5' TTAGGAAGCCCCAAGACAGC 3', REV 5' GTCCCGTCATACACCTCCAC 3'; Gga3 FWD 5' AAAGTGAAGTTGCAGCCACC 3', REV 5' CCACAGGAGGAAACTGGTCC 3'.

**Western Blots and Antibodies:** Whole cell lysates were prepared by lysing cells in NP-40 lysis buffer (0.5% Nonidet P-40, 10 mM Tris pH 7.4, 0.15M NaCl) supplemented with protease and phosphatase inhibitors (PPI) (10 mg/ml each of aprotinin, leupeptin and pepstatin, 50 mM sodium fluoride and 1 mM sodium orthovanadate). Equal amounts of cell lysates were run on a denaturing polyacrylamide gel and transferred to

nitrocellulose membranes. Immunoblots were performed using the indicated antibodies. SF3B1 (Abcam ab172634), beta-Actin (Sigma-Aldrich A5441).

**Bioinformatics Analysis (Provided by Dr. Manoj Pillai, Yale University):** To align the transcript reads, a reference minigene library was first constructed from the 3'SS minigene plasmid library using Starcode algorithm (<https://github.com/gui11aume/starcode>). The 20 bp unique sequence in the second exon was determined from the read 2 (R2) of paired Illumina reads and used a barcode to identify minigenes and to match them to transcripts. Transcripts were aligned to minigene sequences using the BLASTn algorithm ([blast.ncbi.nlm.nih.gov](http://blast.ncbi.nlm.nih.gov)) using the parameters [ `-task blastn -outfmt 6` ]. Those without alignments or >5 mismatches were discarded. Splice junctions identified by the alignment were adjusted if needed for variations in length of 25A or 25B length for consistency (a small fraction of minigenes had 25A region with less than 25 bp, and similarly some minigenes were noted to have less than 25 bp in 25B). After such adjustment, frequencies (Splice\_fraction) for individual splice positions (SA1, SA2, SAcryptic as well as other minor splice positions and those in 25A/25B) were computed.

The splice fraction of each splice sites was determined using the formula

$$\text{Splice\_fraction} = \frac{\text{No. of transcripts for a splice position in a minigene}}{\text{Total transcripts in that minigene}}$$

To determine odds ratio (OR) of individual 6-mers, methodology as described previously was followed(1). Briefly, all possible combinations of a 6mers from A,T,G and C were generated (n=4096). Degenerate sequences (25A and 25B) were scanned for the presence of each of the kmers. The minigenes with the presence of the kmer (atleast once) were then checked for presence or absence of SA1 splice site and divided into a SA1+ and SA1- set. Probability of SA1 use for each kmer was then determined.

## Bibliography:

1. Murthy, T., Bluemn, T., Gupta, A. K., Reimer, M., Jr., Rao, S., Pillai, M. M., and Minella, A. C. (2018) Cyclin-dependent kinase 1 (CDK1) and CDK2 have opposing roles in regulating interactions of splicing factor 3B1 with chromatin. *The Journal of biological chemistry* **293**, 10220-10234
2. Kovacs, E., Tompa, P., Liliom, K., and Kalmar, L. (2010) Dual coding in alternative reading frames correlates with intrinsic protein disorder. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5429-5434
3. Will, C. L., and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**
4. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W. K., Miyawaki, S., Sugano, S., Haferlach, C., Koeffler, H. P., Shih, L. Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S., and Ogawa, S. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-69
5. Gozani, O., Potashkin, J., and Reed, R. (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Molecular and cellular biology* **18**, 4752-4760
6. Gozani, O., Feld, R., and Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes & development* **10**, 233-243
7. Papaemmanuil, E., Cazzola, M., Boulton, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J. S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., Godfrey, A. L., Rapado, I., Cvejic, A., Rance, R., McGee, C., Ellis, P., Mudie, L. J., Stephens, P. J., McLaren, S., Massie, C. E., Tarpey, P. S., Varela, I., Nik-Zainal, S., Davies, H. R., Shlien, A., Jones, D., Raine, K., Hinton, J., Butler, A. P., Teague, J. W., Baxter, E. J., Score, J., Galli, A., Della Porta, M. G., Travaglino, E., Groves, M., Tauro, S., Munshi, N. C., Anderson, K. C., El-Naggar, A., Fischer, A., Mustonen, V., Warren, A. J., Cross, N. C., Green, A. R., Futreal, P. A., Stratton, M. R., and Campbell, P. J. (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *The New England journal of medicine* **365**, 1384-1395
8. Wang, L., Lawrence, M. S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D. S., Zhang, L., Zhang, W., Vartanov, A. R., Fernandes, S. M., Goldstein, N. R., Folco, E. G., Cibulskis, K., Tesar, B., Sievers, Q. L., Shefler, E., Gabriel, S., Hacohen, N., Reed, R., Meyerson, M., Golub, T. R., Lander, E. S., Neuberg, D., Brown, J. R., Getz, G., and Wu, C. J. (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *The New England journal of medicine* **365**, 2497-2506

9. Harbour, J. W., Roberson, E. D., Anbunathan, H., Onken, M. D., Worley, L. A., and Bowcock, A. M. (2013) Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nature genetics* **45**, 133-135
10. Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G. R., Yates, L. R., Papaemmanuil, E., Beare, D., Butler, A., Cheverton, A., Gamble, J., Hinton, J., Jia, M., Jayakumar, A., Jones, D., Latimer, C., Lau, K. W., McLaren, S., McBride, D. J., Menzies, A., Mudie, L., Raine, K., Rad, R., Chapman, M. S., Teague, J., Easton, D., Langerod, A., Lee, M. T., Shen, C. Y., Tee, B. T., Huimin, B. W., Broeks, A., Vargas, A. C., Turashvili, G., Martens, J., Fatima, A., Miron, P., Chin, S. F., Thomas, G., Boyault, S., Mariani, O., Lakhani, S. R., van de Vijver, M., van 't Veer, L., Foekens, J., Desmedt, C., Sotiriou, C., Tutt, A., Caldas, C., Reis-Filho, J. S., Aparicio, S. A., Salomon, A. V., Borresen-Dale, A. L., Richardson, A. L., Campbell, P. J., Futreal, P. A., and Stratton, M. R. (2012) The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-404
11. Dolatshad, H., Pellagatti, A., Fernandez-Mercado, M., Yip, B. H., Malcovati, L., Attwood, M., Przychodzen, B., Sahgal, N., Kanapin, A. A., Lockstone, H., Scifo, L., Vandenberghe, P., Papaemmanuil, E., Smith, C. W., Campbell, P. J., Ogawa, S., Maciejewski, J. P., Cazzola, M., Savage, K. I., and Boultonwood, J. (2015) Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* **29**, 1092-1103
12. Darman, R. B., Seiler, M., Agrawal, A. A., Lim, K. H., Peng, S., Aird, D., Bailey, S. L., Bhavsar, E. B., Chan, B., Colla, S., Corson, L., Feala, J., Fekkes, P., Ichikawa, K., Keaney, G. F., Lee, L., Kumar, P., Kunii, K., MacKenzie, C., Matijevic, M., Mizui, Y., Myint, K., Park, E. S., Puyang, X., Selvaraj, A., Thomas, M. P., Tsai, J., Wang, J. Y., Warmuth, M., Yang, H., Zhu, P., Garcia-Manero, G., Furman, R. R., Yu, L., Smith, P. G., and Buonamici, S. (2015) Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell reports* **13**, 1033-1045
13. Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., Dutertre, M., and Stern, M. H. (2016) Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nature communications* **7**, 10615
14. Kesarwani, A. K., Ramirez, O., Gupta, A. K., Yang, X., Murthy, T., Minella, A. C., and Pillai, M. M. (2017) Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures. *Oncogene* **36**, 1123-1133
15. Obeng, E. A., Chappell, R. J., Seiler, M., Chen, M. C., Campagna, D. R., Schmidt, P. J., Schneider, R. K., Lord, A. M., Wang, L., Gambe, R. G., McConkey, M. E., Ali, A. M., Raza, A., Yu, L., Buonamici, S., Smith, P. G., Mullally, A., Wu, C. J., Fleming, M. D., and Ebert, B. L. (2016) Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer cell* **30**, 404-417

16. Mupo, A., Seiler, M., Sathiaselvan, V., Pance, A., Yang, Y., Agrawal, A. A., Iorio, F., Bautista, R., Pacharne, S., Tzelepis, K., Manes, N., Wright, P., Papaemmanuil, E., Kent, D. G., Campbell, P. C., Buonamici, S., Bolli, N., and Vassiliou, G. S. (2017) Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia* **31**, 720-727
17. Cretu, C., Schmitzova, J., Ponce-Salvatierra, A., Dybkov, O., De Laurentiis, E. I., Sharma, K., Will, C. L., Urlaub, H., Luhrmann, R., and Pena, V. (2016) Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. *Molecular cell* **64**, 307-319
18. Lim, S., and Kaldis, P. (2013) Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development (Cambridge, England)* **140**, 3079-3093
19. Satyanarayana, A., and Kaldis, P. (2009) Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. *Oncogene* **28**, 2925-2939
20. Larochelle, S., Amat, R., Glover-Cutter, K., Sanso, M., Zhang, C., Allen, J. J., Shokat, K. M., Bentley, D. L., and Fisher, R. P. (2012) Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature structural & molecular biology* **19**, 1108-1115
21. Trovesi, C., Manfrini, N., Falcettoni, M., and Longhese, M. P. (2013) Regulation of the DNA damage response by cyclin-dependent kinases. *Journal of molecular biology* **425**, 4756-4766
22. Hergeth, S. P., and Schneider, R. (2015) The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO reports* **16**, 1439-1453
23. Contreras, A., Hale, T. K., Stenoien, D. L., Rosen, J. M., Mancini, M. A., and Herrera, R. E. (2003) The dynamic mobility of histone H1 is regulated by cyclin/CDK phosphorylation. *Molecular and cellular biology* **23**, 8626-8636
24. Zeng, X., Chen, S., and Huang, H. (2011) Phosphorylation of EZH2 by CDK1 and CDK2: a possible regulatory mechanism of transmission of the H3K27me3 epigenetic mark through cell divisions. *Cell cycle (Georgetown, Tex.)* **10**, 579-583
25. Seghezzi, W., Chua, K., Shanahan, F., Gozani, O., Reed, R., and Lees, E. (1998) Cyclin E associates with components of the pre-mRNA splicing machinery in mammalian cells. *Molecular and cellular biology* **18**, 4526-4536
26. Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998) Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes & development* **12**, 1409-1414
27. Boudrez, A., Beullens, M., Waelkens, E., Stalmans, W., and Bollen, M. (2002) Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1. *The Journal of biological chemistry* **277**, 31834-31841
28. de Graaf, K., Czajkowska, H., Rottmann, S., Packman, L. C., Lilischkis, R., Luscher, B., and Becker, W. (2006) The protein kinase DYRK1A phosphorylates the splicing factor SF3b1/SAP155 at Thr434, a novel in vivo phosphorylation site. *BMC biochemistry* **7**, 7

29. Chi, Y., Welcker, M., Hizli, A. A., Posakony, J. J., Aebersold, R., and Clurman, B. E. (2008) Identification of CDK2 substrates in human cell lysates. *Genome biology* **9**, R149
30. Girard, C., Will, C. L., Peng, J., Makarov, E. M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Luhrmann, R. (2012) Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nature communications* **3**, 994
31. Dominguez, D., Tsai, Y. H., Weatheritt, R., Wang, Y., Blencowe, B. J., and Wang, Z. (2016) An extensive program of periodic alternative splicing linked to cell cycle progression. *eLife* **5**
32. Herzel, L., Ottoz, D. S. M., Alpert, T., and Neugebauer, K. M. (2017) Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature reviews. Molecular cell biology* **18**, 637-650
33. Kfir, N., Lev-Maor, G., Gleich, O., Alajem, A., Datta, A., Sze, S. K., Meshorer, E., and Ast, G. (2015) SF3B1 association with chromatin determines splicing outcomes. *Cell reports* **11**, 618-629
34. Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research* **43**, D512-520
35. Gray, N. S., Wodicka, L., Thunnissen, A. M., Norman, T. C., Kwon, S., Espinoza, F. H., Morgan, D. O., Barnes, G., LeClerc, S., Meijer, L., Kim, S. H., Lockhart, D. J., and Schultz, P. G. (1998) Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science (New York, N.Y.)* **281**, 533-538
36. Bain, J., McLauchlan, H., Elliott, M., and Cohen, P. (2003) The specificities of protein kinase inhibitors: an update. *The Biochemical journal* **371**, 199-204
37. Bialojan, C., and Takai, A. (1988) Inhibitory effect of a marine-sponge toxin, okadaic acid, on protein phosphatases. Specificity and kinetics. *The Biochemical journal* **256**, 283-290
38. Isono, K., Mizutani-Koseki, Y., Komori, T., Schmidt-Zachmann, M. S., and Koseki, H. (2005) Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. *Genes & development* **19**, 536-541
39. Cavellan, E., Asp, P., Percipalle, P., and Farrants, A. K. (2006) The WSTF-SNF2h chromatin remodeling complex interacts with several nuclear proteins in transcription. *The Journal of biological chemistry* **281**, 16264-16271
40. Meijer, L., Borgne, A., Mulner, O., Chong, J. P., Blow, J. J., Inagaki, N., Inagaki, M., Delcros, J. G., and Moulinoux, J. P. (1997) Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. *European journal of biochemistry* **243**, 527-536
41. Gockler, N., Jofre, G., Papadopoulos, C., Soppa, U., Tejedor, F. J., and Becker, W. (2009) Harmine specifically inhibits protein kinase DYRK1A and interferes with neurite formation. *The FEBS journal* **276**, 6324-6337
42. Thompson, B. J., Bhansali, R., Diebold, L., Cook, D. E., Stolzenburg, L., Casagrande, A. S., Besson, T., Leblond, B., Desire, L., Malinge, S., and Crispino, J. D. (2015) DYRK1A controls

- the transition from proliferation to quiescence during lymphoid development by destabilizing Cyclin D3. *The Journal of experimental medicine* **212**, 953-970
43. Nasim, M. T., and Eperon, I. C. (2006) A double-reporter splicing assay for determining splicing efficiency in mammalian cells. *Nature protocols* **1**, 1022-1028
  44. Clurman, B. E., Sheaff, R. J., Thress, K., Groudine, M., and Roberts, J. M. (1996) Turnover of cyclin E by the ubiquitin-proteasome pathway is regulated by cdk2 binding and cyclin phosphorylation. *Genes & development* **10**, 1979-1990
  45. Strikoudis, A., Lazaris, C., Trimarchi, T., Galvao Neto, A. L., Yang, Y., Ntziachristos, P., Rothbart, S., Buckley, S., Dolgalev, I., Stadtfeld, M., Strahl, B. D., Dynlacht, B. D., Tsirigos, A., and Aifantis, I. (2016) Regulation of transcriptional elongation in pluripotency and cell differentiation by the PHD-finger protein Phf5a. *Nature cell biology* **18**, 1127-1138
  46. Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M. P., Parzen, H., Szpyt, J., Tam, S., Zarraga, G., Pontano-Vaites, L., Swarup, S., White, A. E., Schweppe, D. K., Rad, R., Erickson, B. K., Obar, R. A., Guruharsha, K. G., Li, K., Artavanis-Tsakonas, S., Gygi, S. P., and Harper, J. W. (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505-509
  47. Zheng, Y., John, S., Pesavento, J. J., Schultz-Norton, J. R., Schiltz, R. L., Baek, S., Nardulli, A. M., Hager, G. L., Kelleher, N. L., and Mizzen, C. A. (2010) Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II. *J Cell Biol* **189**, 407-415
  48. Sancho, M., Diani, E., Beato, M., and Jordan, A. (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS genetics* **4**, e1000227
  49. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589
  50. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-515
  51. Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., and Xing, Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E5593-5601
  52. Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dundar, F., and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-165
  53. Daujat, S., Zeissler, U., Waldmann, T., Happel, N., and Schneider, R. (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27

- phosphorylation blocks HP1 binding. *The Journal of biological chemistry* **280**, 38090-38095
54. Trojer, P., Li, G., Sims, R. J., 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D., Erdjument-Bromage, H., Tempst, P., Nimer, S. D., Wang, Y. H., and Reinberg, D. (2007) L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* **129**, 915-928
  55. Kamieniarz, K., Izzo, A., Dundr, M., Tropberger, P., Ozretic, L., Kirfel, J., Scheer, E., Tropel, P., Wisniewski, J. R., Tora, L., Viville, S., Buettner, R., and Schneider, R. (2012) A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation. *Genes & development* **26**, 797-802
  56. Kalashnikova, A. A., Winkler, D. D., McBryant, S. J., Henderson, R. K., Herman, J. A., DeLuca, J. G., Luger, K., Prenti, J. E., and Hansen, J. C. (2013) Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. *Nucleic acids research* **41**, 4026-4035
  57. Nuytten, M., Beke, L., Van Eynde, A., Ceulemans, H., Beullens, M., Van Hummelen, P., Fuks, F., and Bollen, M. (2008) The transcriptional repressor NIPP1 is an essential player in EZH2-mediated gene silencing. *Oncogene* **27**, 1449-1460
  58. Van Dessel, N., Beke, L., Gornemann, J., Minnebo, N., Beullens, M., Tanuma, N., Shima, H., Van Eynde, A., and Bollen, M. (2010) The phosphatase interactor NIPP1 regulates the occupancy of the histone methyltransferase EZH2 at Polycomb targets. *Nucleic acids research* **38**, 7500-7512
  59. de Almeida, S. F., Grosso, A. R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., Andrau, J. C., Ferrier, P., and Carmo-Fonseca, M. (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nature structural & molecular biology* **18**, 977-983
  60. Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114-2120
  61. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359
  62. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21
  63. Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986) Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323**, 445-448
  64. Keller, G., Kennedy, M., Papayannopoulou, T., and Wiles, M. V. (1993) Hematopoietic commitment during embryonic stem cell differentiation in culture. *Molecular and cellular biology* **13**, 473-486
  65. Rosenberg, A. B., Patwardhan, R. P., Shendure, J., and Seelig, G. (2015) Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163**, 698-711

66. Inoue, D., and Abdel-Wahab, O. (2016) Modeling SF3B1 Mutations in Cancer: Advances, Challenges, and Opportunities. *Cancer cell* **30**, 371-373
67. Chang, C. J., Kotini, A. G., Olszewska, M., Georgomanoli, M., Teruya-Feldstein, J., Sperber, H., Sanchez, R., DeVita, R., Martins, T. J., Abdel-Wahab, O., Bradley, R. K., and Papapetrou, E. P. (2018) Dissecting the Contributions of Cooperating Gene Mutations to Cancer Phenotypes and Drug Responses with Patient-Derived iPSCs. *Stem cell reports* **10**, 1610-1624
68. Liu, P., Jenkins, N. A., and Copeland, N. G. (2003) A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome research* **13**, 476-484