An investigation of the role of mutations in the spliceosome machinery in the myelodysplastic syndromes

A thesis submitted for the degree of *Doctor of Philosophy*

Michaelmas, 2022

Juseong Lee

Nuffield Division of Clinical Laboratory Sciences

Radcliffe Department of Medicine

Jesus College, University of Oxford

Abstract

The myelodysplastic syndromes (MDS) are common myeloid malignancies. Mutations in splicing factor genes (including *SF3B1*, *SRSF2* and *U2AF1*) occur in over half of MDS patients and result in aberrant pre-mRNA splicing of many target genes, indicating that aberrant spliceosome function plays a key role in the pathogenesis of MDS. However, the molecular mechanisms through which the splicing factor mutations drive the MDS phenotype are not fully understood. A previous study from our group has shown that the *U2AF1*^{S34F} mutations induces aberrant splicing of *STRAP* in cells of the erythroid lineage. We have identified a splicing event of *STRAP*, identical to the one caused by the *U2AF1*^{S34F} mutation, in the erythroid precursors of *SRSF2* mutant MDS cases. Functional studies demonstrated that knockdown of *STRAP* leads to inactivation of p38 MAPK and downregulation of CSDE1-bound transcripts, suggesting that underlying mechanism of ineffective erythropoiesis in MDS with low expression of *STRAP* is caused by *SRSF2* mutations. We have also performed an analysis on global splicing alteration using transcriptomic data of splicing factor mutant MDS patients, which revealed that splicing factor mutations (*SF3B1* and *SRSF2*) reshape the mRNA splicing landscape in MDS. This analysis identified that alternative splicing is cell-type dependent and splicing factor mutations alter the pattern of whole mRNA splicing within each bone marrow subpopulation. *SRSF2* is the most frequently mutated splicing factor gene with adverse prognosis in MDS. *SRSF2* mutations commonly co-occur with mutations of other specific genes, most frequently *TET2* and *ASXL1*. We performed single-cell transcriptomic analysis using the 10X Genomics platform on haematopoietic stem and progenitor cells of MDS patients harbouring *SRSF2* mutations and co-mutations of *TET2* and/or *ASXL1*. Cell population composition analysis revealed differences in population abundance across the genotype groups. Differential gene expression analysis unveiled genotype-specific gene expression signatures and dysregulated pathways.

Acknowledgements

I would like to acknowledge several people for their support and guidance throughout my time at University of Oxford. First and foremost, I would like to thank Prof. Jacqueline Boultwood for supporting me to complete my DPhil. I also would like to thank my supervisors, Dr. Andrea Pellagatti and Dr. Hamid Dolatshad. I deeply thank Dr. Pellagatti for being extremely patient and supportive while we continued to pursue our research. My sincere thankfulness to Dr. Dolatshad that he always has been a wonderful mentor. I could not have made my DPhil this far without their guidance, firstly through helping me immeasurably with my daily research, but also by inspiring me with the quality of their own research.

I owe Dr. Shalini Singh and Dr. Richard Armstrong for they taught me with valuable knowledge and advice on many experiments used in this thesis. My thankfulness also goes to Dr. Sean Wen, Dr. Sally-Ann Clark and Dr. Alba Rodriguez as well, for their wonderful collaboration. Particularly, the computational work for single-cell analysis was carried out with collaboration with Dr. Sean Wen (Chapter 5). I am grateful to my dear colleagues, Dr. Dharamveer Tatwavedi, Aik Seng Ng, Dr. Tianyu Sun, Dr. Vickram S. Tittrea and Dr. Nora E. Rahmani who supported me always as great friends. I would like to wish them all the best in their future endeavours. I also gratefully acknowledge Jesus College and Radcliffe of Department of Medicine for supporting this research with the scholarships and awards.

Last but not the least, my deepest thankfulness goes to my family. Their love and prayers have been instrumental in the success of completing this thesis. I specially would like to thank my wife, Hojin, for encouraging and being with me always.

Table of Contents

List of figures

List of Table

Abbreviations

5hmC	5-hydroxymethylcytosine
A3SS	Alternative 3’ splice site
A5SS	Alternative 5’ splice site
AML	Acute Myeloid Leukaemia
ANOVA	Analysis of variance
ASO	Antisense oligonucleotides
Asx	Additional sex combs
ASXL1	Additional sex combs-like 1
B2M	β2-microglobulin
BFU-E	Burst-Forming Unit- erythroid
BM	Bone Marrow
BSA	Bovine serum albumin
CCUS	Clonal cytopenia of undetermined significance
CFU	Colony-forming cell assay
CFU-E	Colony Forming Unit-erythroid
CH	Clonal hematopoiesis
CHIP	Clonal haematopoiesis of indeterminate potential
CLL	Chronic lymphocytic leukemia
CT	Cycle threshold
DE	Differential gene expression
DNMT3A	DNA methyltransferase 3A
EPO	Erythropoietin

ERY	Erythroid
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FAB	French–American–British
FACS	Fluorescent activated cell sorting
FDR	False discovery rate
FMOs	Fluorescence Minus One controls
GEO	Gene Expression Omnibus
GO	Gene Ontology
GRA	Granulocytic
GSEA	Gene Set Enrichment Analysis
HD	Healthy donors
HEAT	Huntington, elongation factor 3, PR65/A, TOR
HEL cell	Human Erythroid Leukaemia cell
HMA	DNA hypomethylating agents
hnRNPs	Heterogeneous nuclear ribonucleoproteins
HSC	Hematopoietic Stem Cell
HSCT	Haematopoietic stem cell transplant
HSPC	Hematopoietic Stem and Progenitor Cell
ICC	International Consensus Classification
IL-3	Interleukin 3
IPSS	International Prognostic Scoring System
IPSS-R	Revised-IPSS
ISE	Intronic splicing enhancer

ISS	Intronic splicing silencer
JMML	Juvenile myelomonocytic leukaemia
MDS	Myelodysplastic Syndrome
MDS del(5q)	MDS with isolated del(5q)
MDS-EB	MDS with excess blasts
MDS-MLD	MDS with multilineage dysplasia
MDS-RS	MDS with ring sideroblasts
MDS-RS-MLD	MDS with ring sideroblasts with multilineage dysplasia
MDS-RS-SLD	MDS with ring sideroblasts with single-lineage dysplasia
MDS-SLD	MDS with single-lineage dysplasia
MDS-U	MDS unclassifiable
MON	Monocytic
MPN	Myeloproliferative neoplasm
MXE	mutually exclusive exons
NCT	Non-template controls
NGS	Next-generation sequencing
NMD	Nonsense-mediated decay
OGT	O-linked N-acetylglucosamine transferase
PCA	Principal component analysis
PMSF	Phenylmethylsulfonyl fluoride
pre-mRNA	Precursor mRNA
PRMT	Protein arginine methyltransferases
PSI	Percent Spliced In
PTM	Post-translational modifications

qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RBD	RNA binding domain
RBP	RNA binding protein
RI	Retained introns
RNA	Ribonucleic Acid
RNA-Seq	RNA Sequencing
RRM	RNA recognition motif
RS	Ring Sideroblasts
sAML	Secondary Acute Myeloid Leukaemia
SCF	Stem Cell Factor
scRNA-seq	Single-cell RNA sequencing
SE	Exon skipping
SF3B1	Splicing factor 3B subunit 1
SMA	Spinal muscular atrophy
SMN	Survival motor neuron
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
SR	Serine/Arginine-rich
SRSF2	Serine and arginine Rich Splicing Factor 2
SSO	Splice-switching antisense oligonucleotide
TBST	Tris-buffered saline with Tween-20 detergent
U2AF	U2 snRNP auxiliary factor
U2AF1	U2 small nuclear RNA auxiliary factor 1
UMAP	Uniform manifold approximation and projection

VAF	Variant allele frequency
WHO	World Health Organization
WIMM	Weatherall Institute of Molecular Medicine
WT	Wild Type
ZRSR2	Zinc Finger CCCH-Type, RNA Binding Motif And Serine/Arginine Rich 2

1 Chapter One - Introduction

1.1 Myelodysplastic syndrome (MDS)

The myelodysplastic syndromes (MDS) are a type of myeloid neoplasms characterized by ineffective haematopoiesis, myelodysplasia and increased risk of transformation to acute myeloid leukaemia (AML) (Arber et al., 2016). The etymology of “dysplasia" has an origin from the ancient Greek for abnormal (dys) formation (plasis), which refers to the abnormal development of cells within tissues or organs and their abnormal structure caused as a result. In 1982, the French–American–British (FAB) group used the term of “myelodysplasia” to denote morphological features and disorders of the myeloid system from hematopoietic stem cell, which was popularised as the term MDS (Bennett et al., 1982).

Approximately 4 to 5 per 100,000 people are diagnosed with this condition annually, and a median age of onset is about 70 years. Symptoms of MDS are generally present as blood cytopenias, such as anaemia, neutropenia, thrombocytopenia and bone marrow dysplasia. Notably, around 30% cases of MDS may develop into AML, depending on subtypes (Cazzola, 2020).

Diagnostic criteria for MDS are persistent cytopenia in one or more peripheral-blood lineages, morphologic dysplasia in one or more bone marrow cell lineages and genetic evidence of clonality. MDS can be further divided into subtypes based on the number of dysplastic lineages, ring sideroblasts (RS), blast percentage, and cytogenetic abnormality (Fenaux et al., 2021, Arber et al., 2016, Malcovati et al., 2013). Historically, MDS was stratified into 5 subtypes by FAB (Bennett et al., 1982) but World Health Organization

(WHO) revised it in 2001 and subsequently in 2008 to better encapsulate the morphologic information of MDS to support better clinical decisions. In addition, mutations in the spliceosome gene *SF3B1* were discovered, which led to the expansion of the category of MDS-RS. In the 2016 WHO classification, MDS is categorized in 6 subtypes namely; MDS with single-lineage dysplasia (MDS-SLD), MDS with multilineage dysplasia (MDS-MLD), MDS with ring sideroblasts (MDS-RS) with single-lineage dysplasia (MDS-RS-SLD) or multilineage dysplasia (MDS-RS-MLD), MDS with isolated del(5q) (MDS del(5q), MDS with excess blasts (MDS-EB) which is divided into type 1 or type 2 (MDS-EB-1 and MDS-EB-2) and unclassifiable (MDS-U) (Arber et al., 2016). Recently, a revised WHO classification and International Consensus Classification (ICC) for myeloid neoplasms were published in 2022 (Khoury et al., 2022, Arber et al., 2022). However, there are concerns regarding the implications of these classifications for routine clinical care and in clinical research of MDS (Zeidan et al., 2022).

MDS and AML exist along a pathological spectrum, and International Prognostic Scoring System (IPSS) is a significant standard for assessing prognosis of adult patients with MDS, to predict long-term outcome (Greenberg et al., 1997). IPSS was first developed in 1997 and later revised (IPSS-R) by the International Working Group for the Prognosis (Greenberg et al., 2012). Patients are assigned into 1 of 5 cytogenetic groups, based on the risk of mortality and AML transformation, and scoring scales accordingly to disease progression. IPSS-R is also used to estimate the risk of transformation to AML and expected survival (**Table 1-1**).

Table 1-1. Revised International Prognostic Scoring System.

* This table is adapted from Cazzola et al. (Cazzola, 2020).

Prognostic variable	*Scores to be assigned to variable values*						
	0	0.5	1	1.5	2	3	4
Cytogenetics	Very good	-	Good	-	Intermediate	Poor	Very poor
Bone marrow blasts, %	≤2	-	>2% - <5%	-	5% - 10%	>10%	-
Hb, g/dL	≥10	-	8- <10	<8	-	-	-
Platelet count, 10^9/L	≥100	50 - <100	<50	-	-	-	-
Absolute neutrophil count, 10^9/L	≥0.8	<0.8	-	-	-	-	-

IPSS-R prognostic risk category	*Total score*	*Median survival (years)*	*Median time for 25% of patients to undergo evolution to AML (years)*
Very low *(19% of all patients)*	≤1.5	8.8	Not reached
Low *(38% of all patients)*	>1.5 - 3	5.3	10.8
Intermediate *(20% of all pts)*	>3 - 4.5	3.0	3.2
High *(13% of all patients)*	>4.5 - 6	1.6	1.4
Very high *(10% of all patients)*	>6	0.8	0.7

Dichotomization into lower-risk vs higher-risk MDS	*Total score*	*Median survival (years)*	*Median time for 25% of patients to undergo evolution to AML (years)*
Lower-risk MDS *(63% of all pts)*	≤3.5	5.9	Not reached
Higher-risk MDS *(37% of all pts)*	>3.5	1.5	1.4

1.2 Genetic and molecular pathogenesis of MDS

Next-generation sequencing (NGS) uncovered that MDS patients carry multiple somatic mutations within the whole coding region, including driver mutations that enhance clonal selection and passenger mutations which are not related to disease progression (Nagata and Maciejewski, 2019, Makishima et al., 2017). In MDS patients, the recurrent mutated genes are involved in vital cellular process including mRNA splicing (*SF3B1, SRSF2, U2AF1* and *ZRSR2*), DNA methylation (*DNMT3A, IDH1/2* and *TET2*), chromatin modification (*ASXL1*, *BCOR* and *EZH2*) and transcription (*RUNX1*, *ETV6* and *GATA2*), checkpoint/cell cycle and signaling (**Table 1-2**) (Yoshida et al., 2011, Ogawa, 2019, Cazzola, 2020, Haferlach et al., 2014). Among the mutated genes aforementioned, 6 driver mutations (*SF3B1, TET2*, *SRSF2, ASXL1*, *DNMT3A*, and *RUNX1*) are mutated in more than 10% of MDS patients with other additional mutations occurring at lower frequency (Haferlach et al., 2014, Makishima et al., 2017). Particularly, Makishima et al. identified the mean number of driver mutations is two or three per patient.

Table 1-2. Most commonly mutated genes in MDS patients.

* This data is adapted from Cazzola et al. (Cazzola, 2020).

RNA splicing	SF3B1 (mutated in >10% of patients) SRSF2 (mutated in >10% of patients) U2AF1 (mutated in 5-10% of patients) ZRSR2 (mutated in 5-10% of patients) PRPF8 LUC7L2 U2AF2
Epigenetic regulators (DNA methylation and histone modification)	TET2 (mutated in >10% of patients) ASXL1 (mutated in >10% of patients) DNMT3A (mutated in >10% of patients) EZH2 (mutated in 5-10% of patients) BCOR (mutated in 5-10% of patients) IDH2 (mutated in 5-10% of patients) IDH1 (mutated in 2-4% of patients) PHF6 (mutated in 2-4% of patients) BCORL1 ATRX EP300 ZBTB33
Transcription regulation	RUNX1 (mutated in >10% of patients) CUX1 (mutated in 2-4% of patients) ETV6 (mutated in 2-4% of patients) CEBPA (mutated in 2-4% of patients) GATA2 WT1 KDM3A
DNA repair control	TP53 (mutated in 5-10% of patients) PPM1D BRCC3
Signaling	CBL (mutated in 5-10% of patients) NRAS (mutated in 5-10% of patients) KRAS (mutated in 2-4% of patients) NF1 (mutated in 2-4% of patients) PTPN11 (mutated in 2-4% of patients) JAK2 (mutated in 2-4% of patients) MPL (mutated in 2-4% of patients) SH2B3 KIT GNB1
Cohesin complex	STAG2 (mutated in 5-10% of patients) CTCF RAD21 SMC3 SMC1A
Miscellanea	DDX41 (mutated in 2-4% of patients) SETBP1 (mutated in 2-4% of patients) ETNK1 (mutated in 2-4% of patients) NPM1 KMT2C CSNK1A1

1.2.1 Growth and propagation of myelodysplastic clones

Through diverse mechanisms, early or founder mutations in haematopoietic stem cells (HSC) lead to dysregulated clonal evolution during haematopoiesis. This process is characterised by key transitory stages with different clinical features in the haematopoietic system (Chen et al., 2019, Ogawa, 2019, Cazzola, 2020).

In the first stage, a somatic mutant clone grows. Early driver mutations occur in HSCs creating mutant clones of haematopoietic stem and progenitor cells (HSPCs) with growth advantage over wild-type cells in the bone marrow niche.

Secondly, the mutant clones transit to pre-MDS phase of "clonal haematopoiesis of indeterminate potential" (CHIP). In brief, clonal outgrowth of hematopoietic cells carrying mutations is referred as "clonal haematopoiesis", and "indeterminate potential" implies uncertainty of the medical condition. Patients in CHIP state have cancer-associated mutant cells in the blood without persistent cytopenia and are excluded from other groups of haematopoietic cancer patients. As mutant stem cells migrate to other bone marrow district with acquired growth advantage, if the hematopoietic cells harbour at least 4% of somatic mutation in bone marrow cells, equating to 2% of variant allele frequency (VAF) for the mutation at least, it can be considered as CHIP (DeZern et al., 2019, Cazzola, 2020, Valent, 2019). Somatic mutations in epigenetic regulator genes (*DNMT3A*, *TET2*, or *ASXL1*) are found commonly in CHIP state, while mutations in spliceosome genes (*SF3B1*, *SRSF2*, or *U2AF1*) are only observed in minor cases. This implies that patients with mutation in epigenetic regulator genes are in a more stable state, while those with spliceosome mutation tend to be developed into MDS or clonal cytopenia of undetermined significance (CCUS).

In the next stage, the mutant cells with growth advantages expand steadily to be dominant in the bone marrow. This stage is related to the occurrence of additional somatic mutations and is considered as CCUS, predicative of higher risk of development to MDS (Kwok et al., 2015). This phase can be the last precursor state of MDS, unless the criteria is sufficient to diagnose MDS, although there is genetic overlap between MDS and MDS precursor conditions. In this phase, as additional somatic mutations occur, clinical presentations of MDS include increased blasts, morphologic dysplasia, and cytogenetic aberrations. To be considered as MDS, one or more MDS-related somatic mutations must be detected in bone marrow or peripheral blood cells, and in one or more peripheral blood cell lineages, cytopenia should be observed at least for 4 months. Lastly, more than 10% of cells in one or more bone marrow cell lineages must have morphologic dysplasia in MDS (Arber et al., 2016).

1.2.2 Splicing factor mutations in MDS

Accurate mRNA splicing is tightly governed by spliceosome for normal haematopoietic cell development (Baralle and Giudice, 2017, Li et al., 2021, de Jong, 2019). In the last decade, genomic advancements have uncovered multiple mutations in splicing factor genes as early driver-mutations in haematological neoplasms (Yoshida et al., 2011, Haferlach et al., 2014). The recurrent somatic mutations in splicing machinery most commonly occur during the initial steps of spliceosome assembly, called as E/A splicing complexes (**Figure 1-1**), including SF3B1, SRSF2, U2AF1 and ZRSR2. It is known that all these splicing factor mutations in the components of E/A complexes are implicated in preventing the precise recognition of 3' splice sites thus giving rise to aberrant splicing of mRNA transcripts (Sperling et al., 2017).

1.2.2.1 A brief overview of mRNA splicing

In molecular biology, the flow of genetic information is often depicted as 'the central dogma of biochemistry', where information contained in DNA flows into RNA via transcription and is ultimately translated into proteins. The whole process from DNA to protein translation which governs function and behaviour of cell is extremely complicated. In eukaryotic organisms, most genes are transcribed from the template strand of DNA to pre-mRNA, the first transcript from a protein coding gene. Pre-mRNAs then undergo a modification process called mRNA splicing during which introns are excised and remaining (either adjacent or joining) exons are joined to produce mature mRNA (Sharp, 1994).

Most of human genes contain multiple introns and exons, and more than 90% of human genes go through splicing process (Pan et al., 2008). This maturation step of mRNA is undoubtedly an essential process of post-transcriptional modification, allowing for significant diversity and complexity of functional proteins in spite of the relatively limited genome size of approximately 20,000 genes in human (Graveley, 2001).

It is also noted that mRNA splicing is implicated in various diseases (Kim et al., 2018), including neurodegenerative disorders (Lee et al., 2016a), cancer (Bonnal et al., 2020), immune and infectious diseases (Alieva et al., 2014), cardiovascular diseases (Lara-Pezzi et al., 2013), and metabolic conditions (Malakar et al., 2016). In particular, mutations in splicing factor genes are relevant to the haematological malignancies (Saez et al., 2017, Cerasuolo et al., 2020).

1.2.2.2 Splicing machinery

The mRNA splicing is initiated at both ends of introns (**Figure 1-1**) called 5' splice site and 3' splice site. The consensus sequence of GURRGU (R; purine), at the 5' splice site, is present in majority of mammals introns. At the 3' end, the consensus sequence, CAG, is frequently observed. To promote recognition of 3' splice site and a branch point, polypyrimidine track ((Y)n) is located at the upstream of 3' splice site and at the downstream of the branch point in mammals ((Y)nNCAG). Also, the branch point adenosine resides at 20 - 30 nucleotides upstream of the 3' splice site to support the formation of intron lariat. Gao et al. have determined that the consensus sequence around branch point is YUNAY (A; branch point, Y; pyrimidine and N; any nucleotide) in human (Gao et al., 2008). Initially, the cleavage at 5' splice site and the formation of lariat

structure occur by 2’-5’ phosphodiester bond formation with guanine residue. The cleavage at 3’ splice and ligation, then, occur to produce mature mRNA.

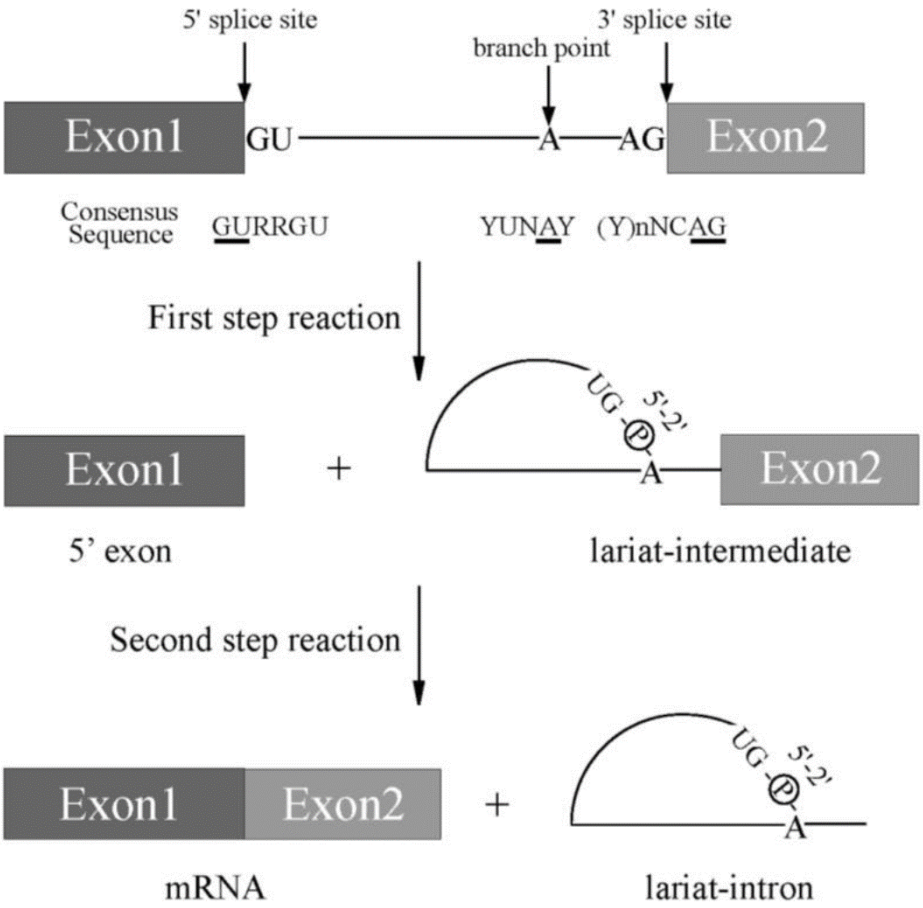


Figure 1-1. A scheme for splicing reaction with two steps.

Boxes show exons, and lines between the boxes represent introns. R and Y stand for purine and pyrimidine residues, respectively. N indicates any nucleotides. Conserved 5’ and 3’ splice sites, and Adenosine residue used for branch nucleotide are underlined.

* This figure is adapted from Kataoka et al. (Kataoka et al., 2021)

Splicing of pre-mRNA is regulated by a large ribonucleoprotein complex, known as the spliceosome. The assembly of the spliceosome on pre-mRNA occurs with major factors of the 5 uridine rich small nuclear ribonucleoproteins (snRNP) which are named as U1, U2,

U4, U5 and U6 (**Figure 1-2**). Each protein complex is composed of multiple proteins and small nuclear RNAs (snRNA) which share their names with snRNPs. This spliceosome set is called ‘major spliceosome’ which is associated with intron removal at around 99.5% of introns (known as U2-type introns) (Turunen et al., 2013). Another spliceosome regulating the remaining 0.5% of introns is called ‘minor spliceosome’ which consists of U11, U12, U4atac, U5 (the common to the major spliceosome) and U6atac (**Figure 1-2**). The introns regulated by minor spliceosome are called as U12-type intron.

The differences of sequences in both U2-type and U12-type introns are substantial. The sequences between 5’ and 3’ splice sites within U2-type introns are typically GU and AG, respectively (**Figure 1-2**), while U12-type introns have AU and AC at the 5’ and 3’ splice sites, respectively. The sequences around branch points are also considerably different between U2-type and U12-type introns (Turunen et al., 2013).

For both major and minor spliceosomes, the assembly on mRNA is a multiple-steps process (**Figure 1-2**). At the 5’ splice site, the snRNP U1 recognises and binds to mRNA by RNA-RNA pairing (E phase). Subsequently, U2 snRNP is recruited to the branch point interacting with U2 snRNP auxiliary factors (U2AF), such as U2AF1 and U2AF2, to bind to the 3’ splice site (A phase). Next, the tri-snRNP, U4/U5/U6, join the spliceosome complex (pre-B phase). The spliceosome is activated, and then U1 and U4 snRNPs exit (B and B^{act} phase). The assembled spliceosome is remodelled to its catalytically active conformation with the help of PRP2 (B* phase). At the 5’ splice site, the cleavage takes place and the lariat structure of intron is formed (C1 phase). Lastly, at the 3’ splice site, introns are excised, and exons are joined together via transesterification reactions (C2 phase) (Wahl et al., 2009).

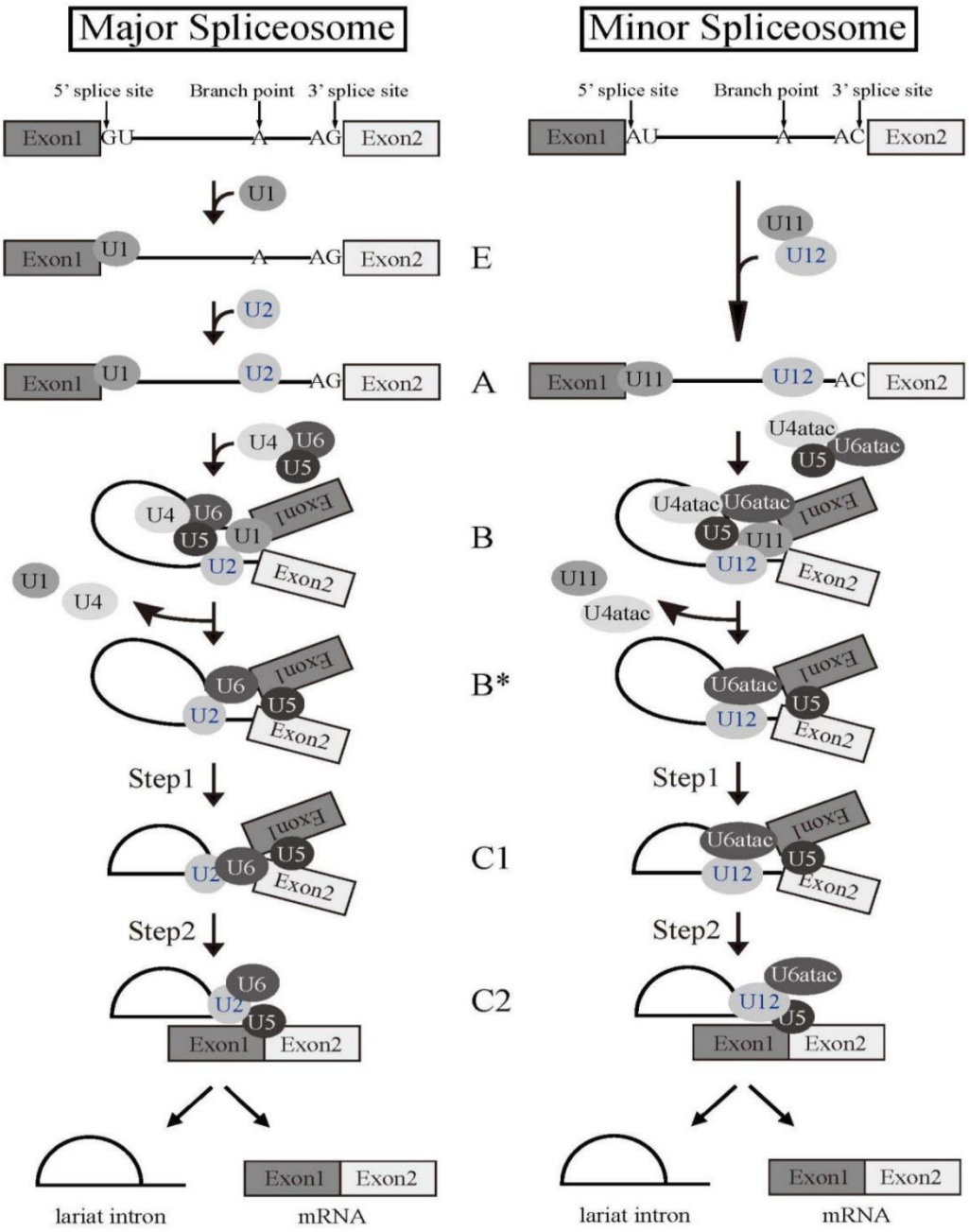


Figure 1-2. Splicing reaction and formation of spliceosome - schematic representations of major and minor spliceosome formations.

Both splicing reactions take place stepwise in a spliceosome. Spliceosomal Uridine-rich small nuclear ribonucleoproteins (U snRNPs) are indicated with their names. The name of each spliceosome intermediate complex is shown in the middle.

* This figure is adapted from Kataoka et al. (Kataoka et al., 2021)

1.2.2.3 *SF3B1* mutations

SF3B1 is a protein encoded by the splicing factor 3B subunit 1 (*SF3B1*) gene and it is a core component of the U2 snRNP in the spliceosome. SF3B1 is involved in the recognition of the branch point by interacting with pre-mRNAs in a sequence-independent manner - approximation (21-34 base) of the branch point adenosine to the 5' splice site, and facilitates binding of U2 snRNP to pre-mRNA (Stanley and Abdel-Wahab, 2022).

SF3B1 is the most commonly mutated gene among splicing factor genes in haematological malignancies such as MDS, chronic lymphocytic leukaemia (CLL) (Chen et al., 2021). In MDS, the presence of *SF3B1* mutation could be a hallmark for the diagnosis of a distinct subtype of MDS called, MDS with ring sideroblasts (MDS-RS), according to the 2016 WHO classification (Arber et al., 2016) and it confers prolonged overall and leukaemia-free survival (Yoshida et al., 2011). The most common mutation hotspots in *SF3B1* gene are at K700, E622, R625, H662, and K666 which are located within the C-terminal Huntington, elongation factor 3, PR65/A, TOR (HEAT) domains (residues 622–781) which is known to be involved in a range of protein–protein and protein–RNA interactions (Teng et al., 2017, Cretu et al., 2016). Interestingly, the different hotspot mutations in *SF3B1* are related to different subtypes of cancer (Malcovati et al., 2020) and each of them causes a different sensitivity to splicing-based drugs (Bonnal et al., 2020). For instance, mutations at K700, the most frequently mutated amino-acid residue, are observed in over 90% of MDS-RS patients while mutations at R625 take place in melanomas (Stanley and Abdel-Wahab, 2022). Mutations at K666 are associated with high-risk MDS and AML.

Recent studies have shown that *SF3B1* mutations are associated with defects in branch point recognition and activation of cryptic 3' splice sites (Dolatshad et al., 2016). For example, we have demonstrated that *SF3B1* mutations lead to aberrant splicing of *ABCB7*

gene, an iron transporter, promoting cryptic 3' splice sites usage. This study suggested that the abnormally spliced *ABCB7*, targeted by the nonsense-mediated decay (NMD), causes marked downregulations of *ABCB7* in MDS. This is linked to the formation of RS that results from abnormal accumulation of iron around mitochondria. Recently, Bondu et al. have identified abnormally spliced *ERFE*, the erythroid hormone erythroferrone, in MDS patients harbouring *SF3B1*. These abnormal splicing events in *ERFE* are caused by a cryptic 3' splice site adding 12 additional nucleotides, which contributes to increase in the expression of the variant ERFE protein. Consequently, it leads to suppression of hepcidin, an iron homeostasis regulator, and abnormal deposition of iron, which is responsible for the ineffective erythropoiesis (Bondu et al., 2019).

1.2.2.4 *SRSF2* mutations

Serine and arginine rich splicing factor 2 (SRSF2), which was originally called SC35 and SRp30b, is a member of the serine/arginine-rich (SR) protein family. SRSF2 is responsible for recognising and binding to exonic splicing enhancer (ESE) sequences with consensus motif of SSNG (S; Cytosine or Guanine and N; any nucleotide) within pre-mRNA via its RNA binding domain (RBD, AA 1-101) to allow appropriate exonic splicing regulation. Also, SRSF2 protein plays a crucial role in interactions of RNA binding proteins (RBP) to promote spliceosome assembly such as the U2AF complex and the U1 snRNP (Graveley and Maniatis, 1998, Liu et al., 2000, Daubner et al., 2012).

SRSF2 mutations occur frequently in MDS patients (20% ~30%), AML patients (10% ~14%) and CMML patients (40 ~ 50%) (Chen et al., 2021). *SRSF2* mutations are associated with poorer prognosis in MDS and increased risk of progression to sAML. The

presences of *TET2* mutations and *ASXL1* mutations are the most commonly observed in MDS patients harbouring *SRSF2* mutations. The mutation hotspot is located at the P95 residue in most of *SRSF2* mutant MDS. Wild-type SRSF2 proteins tend to bind evenly to CCNG and GGNG motifs within the ESE sequences. Kim et al., however, have revealed that *SRSF2* mutations in myeloid malignancies lead to the alteration of the binding preferences, reducing its ability to recognise G-rich ESEs while enhancing the binding preference for C-rich ESEs (Kim et al., 2015).

In many studies, the abnormally spliced targets genes caused by *SRSF2* mutations have been investigated. The downstream effect of the target genes has been identified to be associated with impaired haematopoiesis in myeloid malignancies (Kim et al., 2015, Liang et al., 2018, Yoshimi et al., 2019, Lee et al., 2018). For example, mutations in *SRSF2* gene cause aberrant splicing of *EZH2* in MDS, the inclusion of a poison exon, which is targeted by NMD, resulting in downregulation of *EZH2* expression level. This dysregulation leads to ineffective haematopoietic cell differentiation in MDS with *SRSF2* mutations (Kim et al., 2015). A recent study has shown that the cooperation of the pathogenic mutations, *SRSF2* and *IDH2* (an epigenetic regulator), skewed epigenetic regulations through aberrant splicing of *INTS3*, a member of the integrator complex (Yoshimi et al., 2019) in AML. This type of abnormal splicing events in *INTS3* were observed only in *IDH2* and *SRSF2* double-mutant cells, suggesting that coordinated consequences of alteration in splicing and epigenetic process promote leukemogenesis. Interestingly, Liang et al. have shown that mutations in *SRSF2* result in “splicing-cascade” via altering other RNA processing and splicing genes in *SRSF2* mutant cells, implying functional association between splicing factor genes to regulate global mRNA splicing process (Liang et al., 2018).

1.2.2.5 *U2AF1* mutations

U2 auxiliary factor (U2AF) is a non-snRNP protein consisting of two subunits. U2AF1, also known as U2AF35, is a 35 kDa protein which is a small subunit of U2AF heterodimeric complex and U2AF2, also called as U2AF65, is a 65 kDa protein, a large subunit of U2AF. U2AF plays a role of recognizing and binding polypyrimidine tract and AG dinucleotide, the consensus 3' splice sites, along with SF3b complex in the initial phase of spliceosome assembly (Gozani et al., 1998, Wu et al., 1999).

In haematological malignancies, *U2AF1* mutation is deemed as one of the main drivers of leukemogenesis, especially in MDS (5% ~ 21%), CMML (8% ~ 17%) and AML (3% ~ 12%) (Zhao et al., 2022). *U2AF1* mutation is also frequently identified in solid tumours as well, especially in lung adenocarcinoma (Imielinski et al., 2012). MDS patients harbouring *U2AF1* mutations also are expected to have adverse clinical outcomes and increased rate of progression to AML. The most common *U2AF1* mutations occur at the residues S34 and Q157 within the conserved zinc-finger domains. Ilagan et al. have reported that *U2AF1* mutations lead to specific alteration in 3' splice site preference depending on the residues of mutations (C/A >> T at the -3 position in $U2AF1^{S34}$ mutant cells and G >> A at the +1 position in $U2AF1^{Q157}$ mutant cells) (Ilagan et al., 2015).

In one of our recent studies, Yip et al. have determined that $U2AF1^{S34F}$ mutation in MDS HSPCs cause aberrant splicing of *STRAP* and *H2AFY,* impairing erythroid differentiation and granulomonocytic differentiation in a lineage-specific manner, respectively (Yip et al., 2017). Longer IRAK4 isoform (50 kDa), which is an isoform of IRAK4 (37 kDa) with additional inclusion of exon 4, also has been identified in *U2AF1* mutant MDS and AML (Smith et al., 2019). These variants are generated by aberrantly splicing in *U2AF1* mutant

cells interacting with MyD88 to mediate activation of NF-κB and MAPK signaling, which is essential for leukemic cell growth.

1.2.2.6 *ZRSR2* mutations

The removal on introns and the ligation of exons are governed by the spliceosome. Most species have two types of parallel but analogous spliceosomes, the major (U1, U2, U4, U5 and U6 snRNAs) and the minor (U11, U12, U4atac, U5 and U6atac snRNAs) spliceosomes (Matera and Wang, 2014) (**Figure 1-2**). Zinc finger CCCH-type, RNA binding motif and serine / arginine Rich 2 (ZRSR2) is thought to function in the U12 snRNP and be involved in the recognition of 3' splice sites of U12-type introns (~0.5% of all human introns).

In haematological malignancies, *ZRSR2* mutations occur frequently (MDS; 1 ~ 11% and CMML; 0.8 ~ 8%), but lesser as compared to *SF3B1, SRSF2* and *U2AF1* mutations (Zhang et al., 2021). *ZRSR2* is located on X chromosome (Xp22.1) and enriched for nonsense and frameshift mutations throughout its open reading frame in MDS, suggesting that the loss of function of *ZRSR2* contributes to developing MDS. *ZRSR2* mutations occur frequently in male MDS patients (Yoshida et al., 2011) and in a subtype of AML, blastic plasmacytoid dendritic cell neoplasms (Taylor et al., 2013). In a recent study, Inoue et al. have shown that increased retentions of minor (U12-type) introns in *LZTR1* genes, a regulator of RAS-related GTPases, are identified in MDS and AML cells harbouring *ZRSR2* mutations, which results in NMD and decreased expression level of mRNA and protein. They also have confirmed that restoration of *LZTR1* in Zrsr2-KO HSPCs significantly hindered their

self-renewal *in vivo*, suggesting the aberrantly spliced *LZTR1* caused by *ZRSR2* mutations as a potential leukaemia driver (Inoue et al., 2020, Inoue et al., 2021).

1.2.3 Epigenetic regulator gene mutations in MDS

Mutations in genes encoding epigenetic regulators are one of the most common molecular aberrations in MDS, along with molecular lesions that dysregulate the mRNA splicing process. 60% to 70% of MDS patients harbour this type of mutations such as *TET2*, *DNMT3A*, and *ASXL1* which are known as early driver-mutations as well (Grinfeld et al., 2018, Sato-Otsubo et al., 2012, Ley et al., 2010b). These mutations cause aberration of genes involved in DNA methylation control and histone modification control, which are thought to dysregulate the normal program of haematopoiesis through regulation of key genes in HSPCs, leading to establishment and persistence of MDS clones. In this chapter, we mainly focus on the epigenetic regulator genes including *TET2*, *ASXL1* and *DNMT3A*, the most frequently co-mutated genes with splicing factor gene mutations in MDS (Haferlach et al., 2014).

1.2.3.1 *TET2* mutations

TET2 gene is located on chromosome 4q24 and a member of tet methylcytosine dioxygenase family including *TET1* and *TET3*, and the function of *TET2* is to catalyse the conversion of the modified DNA base methylcytosine to 5-hydroxymethylcytosine (5hmC). Also, *TET2* is widely viewed as a tumour suppressor gene (Chiba, 2017). *TET2* mutations have no specific hot spots, being observed across the whole coding region, which are expected to lead to the loss of function.

During haematopoiesis, *TET2* is highly expressed in HSPCs and it plays crucial roles in cell commitment for differentiation and promotion of the self-renewal of HSPCs (Solary et al., 2014). For example, loss of *TET2* causes an increase in immature $cKit^{+}$ Lin^{-} cells, implying that *TET2* depletion may have an impact on HSPC differentiation and development (Feng et al., 2019, Figueroa et al., 2010). Itzykson et al. also reported that *TET2* deletion promotes monocyte expansion in $CD34^{+}CD38^{+}$ cells (Itzykson and Solary, 2013, Itzykson et al., 2013).

In MDS, mutations of *TET2* were identified up to 30% and observed in other myeloid malignancies (CMML; up to 50%, AML; up to 25%) (Nagata and Maciejewski, 2019, Feng et al., 2019). A recent study showed that the haematologic phenotype of myeloid malignancies is associated with deletion of *Tet2* in a mice model of MDS and CMML (Li et al., 2011, Muto et al., 2013). In the study, *Tet2* $^{-/-}$ mice developed multiple characteristics of myeloid malignancy within 4 months compared to *Tet2* WT mice. Interestingly, recent studies demonstrated that vitamin C treatment mimics effects of *Tet2* / *Tet3* function, resulting in normal haematopoiesis from HSPCs in *Tet2*-deficient mouse (Cimmino et al., 2017, Huang et al., 2022), suggesting a potential therapy to block leukemogenesis.

1.2.3.2 *ASXL1* mutations

A chromatin modulator, *Additional sex combs-like 1* (*ASXL1*), is a member of *ASXL* family genes including *ASXL2* and *ASXL3* which are the mammalian homologs of *Dorosophila* Additional sex combs (Asx) (Nagata and Maciejewski, 2019). They are involved in both activation and repression of Hox genes and known to regulate the balance between polycomb and trithorax functions (Park et al., 2011). The human *ASXL1* gene is located on

the chromosome 20q11.21 and encodes a 1,541 amino acid protein (Fisher et al., 2003). It plays a role in H3K4 methylation through O-linked N-acetylglucosamine transferase (OGT), histone H2A at lysine 119 ubiquitination through PRC1, and histone H3 at lysine 27 methylation through polycomb repressive complex 2 (Asada and Kitamura, 2019, Inoue et al., 2018).

In MDS, *ASXL1* mutations occur in approximately 20% of the patients and are also identified frequently in other myeloid malignancies, including especially CMML (up to 50%), occasionally myeloproliferative neoplasms (MPN) (up to 10%), juvenile myelomonocytic leukaemia (JMML) (up to 8%) and AML (up to 10%) (Medina et al., 2022). *ASXL1* mutations are indicators of poorer prognosis and more frequently observed in patients with high-risk MDS. Majority of *ASXL1* mutations are frameshift or nonsense mutations in the last exon (exon 12), which suggest that the mutant transcripts are expected to escape from NMD, generating C-terminally truncated protein (Kitamura, 2018). A recent study demonstrated that mutant *ASXL1* cooperates with *BAP1* to cause aberrant myeloid differentiation and MDS-like diseases, upregulating posterior *HOXA* genes and *IRF8* (Asada et al., 2018). *SETBP1* mutations frequently co-occur in MDS patients harbouring *ASXL1* mutations. Inoue et al. have reported that *SETBP1* mutations rapidly drive leukemic transformation of MDS patients harbouring *ASXL1* mutations both in patients and in a mouse model (Inoue et al., 2015, Makishima et al., 2013).

1.2.3.3 *DNMT3A* mutations

DNA methyltransferase 3A (DNMT3A) is a member of mammalian DNA methyltransferases (DNMT3A, DNMT3B, and DNMT1) that catalyse 5-methylcytosine

methylation in CpG dinucleotides. *DNMT3A* consists of 23 exons and locates on the human chromosome 2p23. DNMT3A is involved in initiating *de novo* DNA methylation, allowing interactions with histones and histone methyltransferases to regulate gene transcription (Yang et al., 2015). *DNMT3A* and *DNMT3B* play an important role in HSC differentiation. Challen et at. have shown that loss of *Dnmt3a* in HSCs results in markedly increased HSC self-renewal and block of differentiation ability, implying critical roles of DNA methyltransferases in haematopoiesis (Challen et al., 2012).

Mutations in *DNMT3A* occur less frequently in MDS (up to 10%) than in primary AML (up to 23%) (Ogawa, 2019). *DNMT3A* mutations are heterozygous, mostly missense or nonsense. The mutation hotspot is located at the arginine 882 (R882) residue in the methyltransferase domain in approximately 60% of AML patients. In other myeloid malignancies including MDS, MPN and MDS / MPN (CMML) also, R882 is known as the most frequently mutated spot (Williams et al., 2021, Yang et al., 2015). AML patients harbouring *DNMT3A* mutations showed a poorer outcome when they were treated with standard-dose-based induction chemotherapy (Ley et al., 2010a). Also, a recent study has revealed an important impact of *DNMT3A*^{R882} mutations on AML chemoresistance. They identified attenuated recruitment of histone chaperone SPT-16 to chromatin in cells expressing *DNMT3A*^{R882}, which leads to impaired chromatin remodelling and histone eviction in response to anthracycline treatment (Guryanova et al., 2016). Among the splicing factor genes, *SF3B1* mutation is the most frequently coexist with *DNMT3A*. Interestingly, Song et al. reported that patients with *SF3B1 / DNMT3A* double mutations show better clinical prognosis than patients with *DNMT3A* mutation alone in myeloid malignancies (Song et al., 2019, Song et al., 2020).

1.3 Therapeutic approaches for MDS with splicing factor mutations

Due to the heterogeneous nature of MDS, it is imperative to develop personalised and comprehensive treatments. Over the last decades, several therapeutic approaches have been developed but met with little success. To date, one potentially curative treatment for MDS patients is allogeneic haematopoietic stem cell transplant (HSCT), but only a small number of MDS patients fit the criteria to receive such treatment.

Several preclinical works have been contributing to developing a range of potential therapies for MDS harbouring splicing factor mutations (Obeng et al., 2016, Seiler et al., 2018, Lee et al., 2016b, Steensma et al., 2021, Folco et al., 2011). In this chapter, we focus on these approaches to target splicing mutations and its products.

1.3.1 Small molecule splicing modulators

Some small molecules such as spliceostatin A, sudemycin, pladienolide, E7107, H3B-8800 and herboxidiene have been identified as potential drugs to modulate splicing process (Hasegawa et al., 2011, Kotake et al., 2007, Fan et al., 2011, Kaida et al., 2007). They were discovered in bacterial fermentation products and known to be anti-proliferative or pro-apoptotic via interaction with SF3b complex. These drugs bind to HEAT repeat domain of *SF3B1* and inhibit its interactions with the branch site adenosine on mRNA during E/A phase of splicing. However, how they act as antitumour compounds instead of promoting splicing inhibition or toxicity is not fully understood.

E7107 treatment has entered clinical trials (phase I) for patients with solid tumours who didn't respond to existing therapies in 2007 (NCT00499499 and NCT00459823) (Bonnal et al., 2020). However, unexpected ocular toxicity has been observed and precluded any further clinical evaluations for this drug. In the clinical phase I trial of H3B-8800, orally bioavailable analogue of E7107 for haematological neoplasm patients harbouring splicing factor mutations, sufficient therapeutic responses have not been achieved, with only approximately 15% of patients showing decreased red cell transfusion requirements (Steensma et al., 2021).

1.3.2 Splice-switching oligonucleotide (SSO)

SSO is a type of antisense oligonucleotide (ASO) that can modulate splicing, disrupting RNA-RNA base paring or RNA-protein interactions during pre-mRNA splicing process (Havens and Hastings, 2016). These approaches have succeeded clinically for the treatment of non-cancer diseases such as spinal muscular atrophy (SMA) and Duchenne muscular dystrophy (Finkel et al., 2017, McDonald et al., 2021). For example, SSO targeting *SMN2*, called nusinersen, promotes inclusion of exon 7 in *SMN2* gene, which leads to the production of survival motor neuron (SMN) protein in SMA patients (95% ~ 98%) harbouring the homozygous deletion or mutations in the *SMN1* gene. *SMN2* protein containing the exon 7 is very similar to the protein encoded by *SMN1* gene (Finkel et al., 2017). Some studies have also shown that SSO could be a potential therapeutic approach in breast or prostate cancer (Mercatante et al., 2001, Taylor et al., 1999). Inoue et al. have demonstrated that aberrantly splicing events of *BRD9* in melanoma cells harbouring *SF3B1* mutations were restored using SSOs, suggesting the potential as a cancer therapy both *in*

vitro and *in vivo* (Inoue et al., 2019). However, there are two major challenges in SSO treatments for cancer patients with splicing factor mutations - 1) malignant cells with splicing factor mutations usually have at least a few hundreds of aberrant splicing events (Pellagatti et al., 2018), and 2) it is difficult to deliver SSOs to the target sites or cells systemically.

1.3.3 Protein arginine methyltransferase (PRMT) inhibitors

Post-translational modifications (PTM) of splicing factor proteins are important to assemble spliceosome and establish protein-protein interactions. It is also required to localise subcellular molecules (Stanley and Abdel-Wahab, 2022, Patel and Bellini, 2008). PTMs is known to be responsible for regulation of splicing via SR protein kinases and arginine methylation (Musiani et al., 2019). Fong et al. have reported that the treatment of PRMT5 inhibitor and type 1 PRMT enzymes diminished splicing fidelity and gave rise to preferential killing of leukaemia cells harbouring splicing factor mutations, compared to wild-type counterparts (Fong et al., 2019). PRMT5 inhibitors such as PRT543 and JNJ-64619178 have entered clinical trials of phase I / II in advanced MDS/AML (NCT03614728) and patients with low-risk MDS (NCT03573310), respectively.

1.3.4 Targeting downstream pathway and cellular process

A recent study has shown that *U2AF1* mutations cause IRAK4-L isoforms which are essential for leukemic cell growth in myeloid malignancies (Smith et al., 2019). In the studies by Choudhary et al. and Smith et al. the treatment of IRAK4 kinase inhibitor CA-4948, called Emavusertib, was evaluated, showing restoration of differentiation in

preclinical models of MDS and AML harbouring *U2AF1* or *SF3B1* mutations (Choudhary et al., 2022, Smith et al., 2019). Currently, CA-4948 is being tested in clinical trials in MDS, AML and lymphomas (NCT04278768 and NCT03328078).

In several studies, the accumulation of R-loops (RNA-DNA hybrids with a displaced single stranded DNA) was observed in leukaemia cells harbouring splicing factor mutations (*SF3B1*, *U2AF1* and *SRSF2*) (Singh et al., 2020, Nguyen et al., 2018, Chen et al., 2018). This phenomenon is involved in replication stress and activation of the ATR pathway, implying that inhibitors of ATR pathway can be one of the therapeutic approaches for MDS patients with splicing factor mutations. This approach is now being evaluated in a phase 1B clinical trial of the ATR inhibitor, called AZD6738, for the patients with MDS or CMML (NCT03770429).

1.4 Aims

The aim of this thesis is to deepen the understanding of the molecular pathogenesis of MDS harbouring splicing factor mutations so as to contribute to the identification of potential therapeutic targets. This thesis is focused on three main parts as shown below.

1. The first focus is on aberrantly spliced key target gene in MDS with splicing factor mutations to understand dysregulated pathways and cellular processes (chapter 3). We started with an investigation of the molecular mechanisms underlying ineffective erythropoiesis in MDS with splicing factor mutations. To this end, we have identified an aberrant splicing event of *STRAP* in *SRSF2* mutant cases, leading to impaired erythropoiesis. Functional studies demonstrated that knockdown of

STRAP led to inactivation of p38, a member of MAPK pathway, and downregulation of CSDE1-bound transcripts.

2. The second focus is on the global landscape of mRNA splicing alterations in MDS caused by splicing factor mutations (chapter 4). Our global splicing analysis confirmed that alternative splicing is cell-type dependent and revealed that splicing factor (*SF3B1* and *SRSF2*) mutations alter patterns of mRNA splicing within bone marrow subpopulations, showing similar patterns of splicing changes depending on genotypes. This result suggests that there may be a potential mechanism underlying regularity of the altered splicing patterns among complex factors when mRNA splicing process is perturbed.

3. My last focus is to reveal the transcriptional landscape of MDS harbouring *SRSF2* mutations with co-mutations of *TET2* and/or *ASXL1*, at single-cell resolution, in MDS HSPCs (chapter 5). Our results showed that there are distinct clonal heterogeneities within MDS HSPC populations harbouring *SRSF2* mutations, depending on the combination of the co-mutations. This transcriptomic heterogeneity within MDS cells implies the possibility to further define sub-groups of MDS and delineate different therapeutic approaches according to the genotypes.

2 Chapter Two - Materials and Methods

2.1 Cell culture

2.1.1 Cell lines

K562 cells were purchased from Horizon Discovery. K562 cells were cultured in RPMI 1640 medium (Sigma-Aldrich, St. Louis, Missouri, United States) with 10% heat-inactivated foetal bovine serum (FBS) (Life Technologies, Carlsbad, California, United States), 100 units/mL penicillin and 100 µg/mL streptomycin (Sigma-Aldrich) at 37 °C and 5% CO^2.

HEK 293T cells were purchased from ATCC and cultured in Dulbecco's Modified Eagle's medium (DMEM) (Life Technologies) supplemented containing 10% FBS (Life Technologies), 100 units/mL penicillin and 100 µg/mL streptomycin (Sigma- Aldrich).

2.1.2 Human primary $CD34^+$ culture and differentiation toward erythroid lineage

Bone marrow $CD34^+$ cells from healthy controls (Stemcell Technologies, Vancouver, Canada) were cultured for 14 days to generate erythroblasts in StemSpan™ Serum-Free Expansion Medium (Stemcell Technologies) plus 0.5 U/mL recombinant human erythropoietin (EPO) (Roche, Basel, Switzerland), 100 ng/mL stem cell factor (SCF) (Miltenyi Biotec, Bergisch Gladbach, Germany), 10 ng/mL interleukin 3 (IL-3) (Miltenyi Biotec), 40 µg/mL Lipids Cholesterol Rich from adult bovine serum (Sigma- Aldrich, St. Louis, Missouri, United States), 2 mM L-glutamine (Sigma- Aldrich), and 1% penicillin-streptomycin solution (Sigma Aldrich). On day 7, there was increase in concentration of

EPO to 3 U/ml during the second week. Medium was changed every second day to maintain the same cell concentration.

2.1.3 Viral transduction

To create retrovirus stock with high-titers used for downstream transduction, MISSION lentiviral packaging mix (Sigma-Aldrich) was co-transfected cells along with lentiviral transfer vector into HEK293T. Spinoculation was performed with 8 μg/ml polybrene (Sigma-Aldrich) at 800 ×g and 32°C for 2 hours. Following incubation overnight, transduced cells were selected using medium with 0.65 μg/ml puromycin (Thermo Fisher Scientific). Lentiviral constructs containing shRNAs targeting *STRAP* in the pLKO.1 vector (TRCN0000060465 and TRCN0000060466) were purchased from Sigma Mission.

2.2 Cell growth assay

Transduced cells were planted in 96 well plates ($0.5*10^5$ cells) on day 7 of culture. Each sample was duplicated. Viable count was performed using haemocytometer after staining with trypan blue from day 8 to day 14 of differentiation.

2.3 DNA and RNA extraction

2.3.1 Genomic DNA extraction

DNA extraction from cells was extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

2.3.2 RNA extraction

RNA was extracted from cells using TRIzol™ Plus RNA Purification Kit (Ambion, Austin, Texas, United States) according to the manufacturer's protocol.

2.3.3 Preparation of RNA samples prior to RT-PCR and qRT-PCR

Total RNA extracted was treated with DNase I (Invitrogen) ahead of reverse transcription. 1µg RNA was added into 10ul of reaction buffer containing 1µl 10X DNase I Reaction Buffer (Invitrogen), 1µl DNase I 1U/µl (Invitrogen) and DEPC-treated water to 10µl. The reaction mixture was incubated for 15 min at room temperature. To inactivate the DNAse I, 1µl of 25mM EDTA solution (Invitrogen) was added and then the reaction mixture was incubated at 65°C for 10 minutes. The concentration was measured by spectrophotometry using a Nanodrop spectrophotometer (Thermo Scientific, UK).

2.3.4 cDNA synthesis

For cDNA synthesis the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Waltham, Massachusetts, United) was used. 1µg of RNA were reverse transcribed according to the manufacturer's instructions. 1µl of RNase inhibitor was added to a 20µl of reverse transcription reaction. The quantification was determined by using a Nanodrop spectrophotometer (Thermo Scientific).

2.4 Real-time quantitative polymerase chain reaction

Quantitative Real Time Polymerase Chain Reaction (qRT-PCR) were performed in triplicate duplex reactions using LightCycler SYBR Green I Master Mix (Roche) on a Roche LightCycler 96 instrument (Roche Diagnostics). No template controls (NTC) were added as general controls to check extraneous nucleic acid contamination for each gene expression assay. For the reactions, the 5µl of cDNA was added to the reaction mixture containing 2X LightCycler SYBR Green I Master, 1µl of forward primer, 1µl of reverse primer (10µM) and 3µl of Nuclease-free water. Specificity of each primer set was assessed by the melt curve analysis checking additional peaks and visualising PCR amplicons on agarose gels (**Table 2-2**). The qRT-PCR reaction condition was as follows:

Table 2-1. qRT-PCR program condition

Step	Temperature	Duration	Cycles
Initial denaturation	95°C	300 sec	1
Denaturation	95°C	10 sec	
Annealing	60°C	20 sec	45
Extension	72°C	20 sec	
	95°C	5 sec	
Melting	65°C	60 sec	1
	97°C	1 sec	

The LightCycler 96 software version 1.1.0.1320 (Roche Diagnostics) was used to analyse raw qRT-PCR fluorescence data to compute cycle threshold (CT) for each sample with primer sets. The CT values of the genes of interest were normalised to endogenous control

β2-microglobulin (*B2M*) measured in the same reaction. Relative gene expression level was measured using the ΔΔCT method. The following calculation was used:

ΔCT = CT(Gene) – CT(B2M),

ΔΔCT = ΔCT(sample) - ΔCT(control).

The concentration of cDNA in each sample was determined using a NanoDrop (Thermo Fisher Scientific).

Table 2-2. Primers of qRT-PCR (SYBR Green) used for relative mRNA expression level

Gene		Sequence (5' > 3')	Application
STRAP	Forward	CCTACAAGGGCAACTTTGGTCCTA	qRT-PCR for *STRAP*
	Reverse	CTAGCTCCTCTTCTGTTGTCTCTGG	qRT-PCR for *STRAP*
PABPC1	Forward	TAATCAACCCCTACCAGCCAG	qRT-PCR for CSDE1 bound transcripts
	Reverse	AGGATAGTATGCAGCACGGT	qRT-PCR for CSDE1 bound transcripts
PABPC4	Forward	TCTGCCGCGATATGATCACC	qRT-PCR for CSDE1 bound transcripts
	Reverse	ATGCGGATTGGCTTTCCCTT	qRT-PCR for CSDE1 bound transcripts
RANBP1	Forward	ACCATGACCCTCAGTTTGAGCC	qRT-PCR for CSDE1 bound transcripts
	Reverse	AGTGCCTCGCTCCTTCCATTCT	qRT-PCR for CSDE1 bound transcripts
THOC5	Forward	AAGCCAAAGTGACAACTGCC	qRT-PCR for CSDE1 bound transcripts
	Reverse	CTGATTGGCTGGATTCGGAGT	qRT-PCR for CSDE1 bound transcripts
HMBS	Forward	ACCCTAGAAACCCTGCCAGA	qRT-PCR for CSDE1 bound transcripts
	Reverse	GGTGTTGAGGTTTCCCCGAA	qRT-PCR for CSDE1 bound transcripts
OSTF1	Forward	GCTGAGCAGGCAGAATCCAT	qRT-PCR for CSDE1 bound transcripts
	Reverse	CCGTGGCAAGCCCAGTATAA	qRT-PCR for CSDE1 bound transcripts
B2M	Forward	CTCCGTGGCCTTAGCTGTG	Endogenous control
	Reverse	TTTGGAGTACGCTGGATAGCCT	Endogenous control

2.5 Colony-forming cell assay

Colony-forming cell assay (CFU) was carried out using methylcellulose (MethoCult H4434 Classic, Stemcell Technologies) containing 0.65µg/ml puromycin (Thermo Fisher Scientific) according to the manufacturer's protocols. 3000 transduced cells were seeded onto the methylcellulose on day 7 of erythroid differentiation. To prevent the culture from being dried-up, the plates were incubated in a humidified incubator at 37°C with 5% CO2 in air. Colonies' identification and the counting were performed after 14 days of culture.

2.6 Protein expression analysis

2.6.1 Preparation of cell lysates

Cell lysates were prepared using Cell Lysis Buffer(10X) (Cell Signaling Technology) according to manufacturer's instruction. Briefly the cells were washed with PBS to remove residual medium. 10^7 cells pellets were resuspended in 400µl of 1X Cell Lysis Buffer and incubated on ice for 5 mins. The lysates were centrifuged for removal of the debris. 200X of Phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich) and Protease/Phosphatase Inhibitor Cocktail (100X) (Cell Signaling Technology) were added.

2.6.2 Determination of protein concentration

The amount and concentration of protein in each cell lysates were determined using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instruction prior to Western blotting.

2.6.3 Western blotting

Equal amounts (10μg for p38, 20ug for phosphorylated p38) of extracted proteins used for western blot analysis. Proteins were separated using NuPAGE 4-12% Bis-Tris Protein Gels (Invitrogen) and transferred to the nitrocellulose membranes. The membranes were incubated with blocking buffer, containing 5% bovine serum albumin (BSA) in Tris-buffered saline with Tween-20 detergent (TBST) (50 mM Tris-HCl, 150 mM NaCl and 0.05% Tween 20, pH 7.6) for one hour. The membranes were then washed with TBST and incubated with primary antibodies diluted 1:1000 in the blocking buffer at overnight 4°C. Following the incubation with primary antibodies, membranes were washed repeatedly by TBST and then incubated with HRP-conjugated secondary antibody for one hour at room temperature. After the washing step, the membranes were developed using ECL substrate (ThermoFisher Scientific) according to the manufacturer's instruction and then were visualised as images using gel documentation system (G: Box Chemi XRQ, Syngene, Cambridge, United Kingdom). β-actin was used as a loading control. p38 MAPK Antibody rabbit antibody and Phospho-p38 MAPK (Thr180/Tyr182) antibody were purchased from Cell signalling Technology (Danvers, Massachusetts, United States). Anti-β-actin -HRP was purchased from Abcam (Cambridge, UK).

2.7 Flow cytometry

To assess differentiation toward the erythroid lineage, flow cytometry experiments were performed on a BD LSRII instrument (BD Bioscience; Franklin Lakes, NJ, USA). To determine the levels of compensation, BD CompBeads Set Anti-Mouse Ig (Becton Dickinson UK) stained with each antibody in our panel and were used as single-stained samples. Fluorescence Minus One controls (FMOs) were included in all the experiments

to determine the boundary between background fluorescence caused by other fluorochromes and positive signal, and thus to settle the gates for distinguishing the positive and negative populations. Unstained live cells and dead cells stained with DAPI (Sigma Aldrich) were used to set the gates for the viable cell population. On day 11 and day 14 of erythroid differentiation, cells were collected and then stained with anti-CD36-PE (clone 5-271, BioLegend), anti-CD71-FITC (clone OKT9, eBioscience), and anti–CD235a-APC (clone HIR2, BioLegend) antibodies. DAPI was used for viability of cells. The cells were incubated with the antibodies for 30mins on ice in dark, washed with FACS buffer containing PBS with 1% Bovine serum albumin and then resuspended in the FACS buffer. All flow cytometry data were analysed by using FlowJO VX software.

2.8 RNA sequencing and data analysis for chapter 4

2.8.1 Patient cohorts for analysis of aberrant splicing

The clinical information on patient cohorts (**Table 2-3** and **Table 2-4**) and bulk RNA-seq dataset discussed for aberrant splicing analysis in this thesis (chapter 4) were generated for our lab's previous project (Pellagatti et al., 2018) and re-investigated for this thesis. The RNA-seq data sets have been stored in the NCBI's Gene Expression Omnibus (GEO) repository (GEO accession number: GSE114922).

Table 2-3. Clinical information on 84 MDS patient cohorts with splicing factor mutations

* This data is adapted from Pellagatti et al. (Pellagatti et al., 2018)

ID	SF3B1	SRSF2	U2AF1	ZRSR2	Disease Subtype	Dead or alive at last observation	Days from sample to last observation	IPSS	Karyotype details	Sex	Age	Hb	WBC	Neut	Platelet	BM blast %	Transfusion dependency
A112	1	0	0	0	RARS	ALIVE	1758	Low	46XX	F	61	9.5	3.84	2.2272	226	1	no
A114	0	0	1	0	RA	ALIVE	1254	Na	46XX	F	49	na	na	na	na	na	na
A116	1	0	0	0	RARS	ALIVE	0	Low	46XY	M	71	7.9	7.1	5.538	430	1	yes
A117	0	0	0	0	RAEB2	ALIVE	258	Int-1	46XX,del(5q)	F	74	8.8	na	0.9968	228	na	no
A118	1	0	0	0	RA	ALIVE	1056	Low	46XX,del(5q)	F	65	10	6.34	2.7262	347	3	yes
A119	0	1	0	0	RAEB1	DEAD	168	Int-1	46XY	M	52	11.9	3.81	2.5908	85	7	no
A120	0	0	0	0	RARS	ALIVE	300	Int-1	46XY	M	77	8.1	1.93	0.91	52	4	no
A121	0	0	0	0	RA	ALIVE	36	Int-1	46XY	M	87	10.4	3.25	1.3	89	2	no
A122	0	1	0	0	RAEB2	ALIVE	1040	Int-2	46XX	F	79	9.3	2	0.54	63	11	no
A123	1	0	0	0	RARS	ALIVE	1596	Low	46XX	F	83	9.3	4.41	2.8665	227	2	no
A124	0	0	0	0	RA	ALIVE	1125	Low	46XX	F	76	10.1	5.38	2.5286	389	3	no
A125	1	0	0	0	RARS	ALIVE	264	Int-1	46XY	M	66	7.9	2.7	1.62	114	1	no
A126	0	0	0	0	RA	ALIVE	1986	Low	46XX	F	85	9.4	4	2.8	224	1	yes
A127	1	0	0	0	RARS	ALIVE	1097	Low	46XX	F	59	9	5.5	2.64	305	1	no
A128	1	0	0	0	RARS	ALIVE	975	Int-1	46XX,del(20q)	F	65	9.6	na	1.019	139	4	yes
A129	0	0	0	0	RAEB1	ALIVE	2549	Int-1	46,XX,del(5)(q14q34)[20]; 46,XX[5]	F	56	9	3.43	2.13	349	8	na
A130	0	0	0	0	RAEB2	ALIVE	184	Int-2	46XX	F	34	10.5	1.5	0.39	87	17	no
A131	0	0	0	0	RAEB1	ALIVE	147	Int-1	46XY	M	55	8.6	4.46	1.87	196	5	no
A132	0	0	0	0	RA	ALIVE	2130	Low	46XX	F	40	10.7	na	2.6	77	1	no
A134	0	0	0	0	RAEB2	DEAD	988	Int-2	46,XY	M	67	10.4	na	0.8	336	12	no
A136	0	0	0	0	RAEB2	DEAD	228	Int-2	46XY	M	57	5.4	1.9	0.95	45	12	yes
A138	0	0	0	0	RARS	DEAD	458	Int-1	46XY,del(17p)	M	62	6.6	3.25	1.0725	278	3	yes
A139	0	0	0	0	RARS	DEAD	1348	Int-1	46XY;del(11q),del(17p)	M	63	9.5	4.68	2.9952	218	3	no
A140	0	0	0	0	RA	ALIVE	1350	Int-1	46,XX,del(5q)/9[25]	F	67	12.1	na	1.5	85	2	yes
A141	0	0	0	0	RA	ALIVE	2116	Int-1	complex	M	74	na	na	na	na	1	no
A142	1	0	0	0	RARS	ALIVE	1231	Low	46XY	M	66	10.7	2.5	0.96	171	0	yes
A143	0	0	0	0	RA	ALIVE	520	Low	46XY	M	71	10.3	5.4	2.43	318	2	no
A145	0	0	0	0	RA	ALIVE	981	Int-1	46XX, del(12p), del(5q)	F	74	8.2	na	2.16	292	3	no
A147	1	0	0	0	RARS	ALIVE	1158	Low	46XX	F	53	7	na	3.55	577	2	yes
A148	0	0	0	0	RAEB2	DEAD	539	High	46,XX	F	69	11.1	na	1.7	10	14	yes
A149	1	0	0	0	RARS	ALIVE	693	Int-1	46XY,del(7q),del(20q)	M	75	8.7	5.2	3.016	251	4	no
A150	1	0	0	0	RARS	ALIVE	525	Int-1	46XY,del(9q),del(13q)	M	62	9.5	na	2.855	180	4	no
A151	1	0	0	0	RARS	ALIVE	409	Low	46XY	M	71	10.4	4.74	3.0336	227	3	no

A153	1	0	0	0	RARS	ALIVE	1316	Low	46XY	M	69	11.2	na	3.022	461	2	no
A154	1	0	0	0	RARS	DEAD	1119	Int-1	46XY	M	57	8.4	na	1.005	305	0	no
A155	0	0	0	0	RA	ALIVE	1352	Int-1	46XY	M	26	11	4	0.16	51	2	no
A156	0	1	0	0	RAEB1	DEAD	444	Int-2	47XY,+8	M	73	8.6	1.7	0.44	148	7	no
A158	1	0	0	0	RARS	ALIVE	1534	Low	46XX	F	48	10.1	7.91	4.9042	543	3	no
A161	0	0	1	0	RAEB1	ALIVE	21	Int-1	46XY	M	66	7.4	9.39	1.6902	18	8	yes
A163	0	0	0	0	CMML	na	na		46XX,del(5q)	F	82	na	na	na	232	0.2	na
A164	0	0	0	0	RA	ALIVE	410	Int-1	46XY	M	64	9.7	2.79	1.4508	105	3	yes
A165	0	0	0	0	RAEB2	ALIVE	313	Int-2	46XY	M	57	4.5	1.22	0.1708	43	15	no
A166	1	0	0	0	RARS	ALIVE	2233	Int-1	46XX,del(12p)	F	79	8.1	3.83	2.298	594	1	no
A167	1	0	0	0	RA	ALIVE	1643	Int-1	47XY,+19	M	74	8.5	5.8	2.7	160	0	yes
A168	0	0	0	0	RAEB2	ALIVE	430	Low	hyperdyploidy	F	59	8.8	2	0.88	77	11	no
A170	1	0	0	0	RARS	ALIVE	1625	Low	46XY	M	65	9.8	6.57	3.4821	104	3	no
A171	0	0	0	0	RAEB1	ALIVE	22	Int-2	complex	M	74	8.3	2.88	1.7568	16	6	yes
A172	1	0	0	0	RAEB1	DEAD	407	Int-1	47XX,+8, del17p	F	67	8.6	6.87	3.7098	538	6	yes
A173	0	0	0	0	CMML	ALIVE	1690	Int-1	46,XY	M	35	11	na	8	36	15.5	no
A175	0	0	0	0	RAEB1	DEAD	185	Int-2	46XY,del(12p)	M	71	9.3	6.47	2.98	93	6	yes
A176	0	0	0	0	CMML	DEAD	663		46,XX	F	56	11.4	na	na	336	6	no
A177	0	1	0	0	RARS	ALIVE	25	Low	46XY,del(20q)	M	60	11.5	2.67	1.36	187	2	no
A178	1	0	0	0	RARS	ALIVE	1440	Int-1	46XX	F	81	10.7	7	4.8	504	2	no
A179	0	0	0	0	RA	ALIVE	14	Int-2	46XX, -7	F	72	13	na	1.4098	86	3	no
A180	0	0	1	0	RAEB1	ALIVE	49	Int-1	46XY	M	77	7.5	3.3	1.58	20	6	no
A181	1	0	0	0	RARS	ALIVE	1107	Low	46XX,del(20q)	F	78	9.1	3.63	2.4684	166	1	no
A182	0	0	0	0	RAEB1	ALIVE	19	Int-1	46XY	M	69	12	2.25	0.4275	42	8	no
A183	1	0	0	0	RARS	DEAD	747	Int-1	46XX	F	69	9.5	5.6	3.7	362	4.5	no
A184	0	0	0	0	RAEB1	DEAD	1635	Int-1	46 XY	M	67	10	0.4	0.1	83	7	yes
A185	1	0	0	0	RARS	ALIVE	223	Int-1	48XY,+8,+21	M	58	6.9	na	2.854	252	0	yes
A186	0	1	0	0	CMML	DEAD	930	Na	46,XY	M	58	11.1	na	15.2	45	10.5	no
A187	0	0	1	0	RAEB2	ALIVE	1755	Int-2	46 XY	M	76	9.9	2.6	0.5	103	13.5	yes
A188	0	0	0	0	RARS	DEAD	1057	Low	46,XY	M	64	7.9	6.5	4.4	244	2	yes
A189	0	0	0	0	RA	ALIVE	1833	Low	46XX	F	65	7.4	5.8	2.9	382	4	yes
A190	1	0	0	0	RARS	ALIVE	1603	Low	46XY	M	69	10.8	4.7	2.6	349	2	no
A191	0	0	1	0	RAEB1	ALIVE	80	Int-1	46XX	F	46	11.9	2.8	1.204	120	7	no
A192	0	1	0	0	RAEB2	ALIVE	154	Int-2	47XY,+21	M	61	8.3	5.1	2.4	314	13	yes
A193	0	0	0	0	RA	DEAD	412	Low	46XY	M	77	10.1	4.3	1.4	183	3	no
A194	0	0	0	0	RA	ALIVE	1650	Int-1	46XX, del(5)(q13q33)	F	66	9.5	9.1	6.1	305	4	no
A195	0	0	0	0	RAEB2	ALIVE	0	Int-2	46XY	M	63	9.6	1.82	0.8372	34	16	no
A196	1	0	0	0	RARS	ALIVE	1221	Low	46,XY	M	70	10.8	na	2.4	348	3	no

A197	1	0	0	0	RA	na	na	Low	46XX	F	58	10	na	2.5	500	2	no
A198	0	0	0	0	RAEB1	DEAD	695	Int-1	46XX	F	72	9.9	5.5	3.025	465	7	yes
A199	1	0	0	0	RARS	DEAD	124	Int-1	47XY,+8	M	72	8.2	6.33	2.7852	300	1	yes
A200	0	0	1	0	RAEB2	ALIVE	63	High	46XY;46XY,t(X;3),del(17p)	M	78	8.3	5.51	2.1	63	18	no
A203	0	0	0	0	RAEB2	DEAD	366	High	5q-+ complex (X;12, del5, der12)	M	73	13.4	2.82	1.02	91	16	na
A204	0	0	0	0	RAEB1	DEAD	1040	Int-1	46XX	F	77	13.1	3.1	0.94	100	8	no
A205	1	0	0	0	RAEB1	ALIVE	466	Int-1	46XX,t(3;3),t(4;6)	F	81	8.8	na	3.6	496	6	no
A206	0	0	0	0	RAEB1	DEAD	953	Int-1	46XX, riarr(1p)	F	76	11.9	2.74	0.7672	120	9	no
A207	0	1	0	0	RAEB2	ALIVE	193	Int-2	47XY,+8	M	63	12.1	30	20.1	59	11	no
A208	0	1	0	0	RARS	ALIVE	1832	Low	46XY	M	62	7.5	11.73	4.8093	106	4	yes
A209	0	0	0	0	RA	DEAD	2057	Low	46XY,del(5q)	M	63	10.3	na	3.9	99	1	No

Table 2-4. Mutational spots in each splicing factor mutation in the cohort

* This data is adapted from Pellagatti et al. (Pellagatti et al., 2018)

ID	Splicing factor Mutations in the cohort			
	SF3B1	SRSF2	U2AF1	ZRSR2
A153	E622D			
A142	E622D			
A190	E622D			
A178	E622D			
A166	H662Q			
A123	H662Q			
A205	K666N			
A172	K700E			
A199	K700E			
A118	K700E			
A149	K700E			
A125	K700E			
A127	K700E			
A116	K700E			
A150	K700E			
A185	K700E			

A128	K700E		
A147	K700E		
A181	K700E		
A170	K700E		
A158	K700E		
A197	K700E		
A112	K700E		
A196	K700E		
A167	K700E		
A183	K700E		
A154	R625L		
A151	R625L		
A119		P95H	
A192		P95_R102delPPDSHHSR	
A156		P95H	
A207		P95H	
A208		P95H	
A122		P95H	
A186		P95H	
A177		P95L	
A180			Q157P
A161			Q157P
A200			Q157R
A187			R156H
A191			S34F
A114			S34F
A160		P95H	H191Y
A174		P95R	V253fs*36

2.8.2 Detection of aberrant splicing events

rMATS v3.2.2beta (Shen et al., 2014, Dolatshad et al., 2016, Yip et al., 2017) was used to detect aberrant splicing events associated with each splicing factor mutation following alignment by HISAT2. (False discovery rate (FDR) < 0.05 and |ΔPSI| > 0.1)

2.8.3 Gene Ontology (GO) Analysis

In R, clusterProfiler 4.0 was used to perform GO analysis and visualise the functional enrichment results on gene-concept network plots and dot plots (Wu et al., 2021). To reduce redundant GO terms, generated by overlap between parent and child terms, the simplify function was applied.

2.8.4 Principal component analysis (PCA) of splicing events

PSI values were extracted from rMATS ourtputs and used to perform PCA dimension reduction analysis using MARVEL (Wen et al., 2022). With the help of Sean Wen at Weatherall Institute of Molecular Medicine (WIMM) in University of Oxford, the MARVEL object was generated.

2.9 Single-cell RNA-sequencing

2.9.1 Patient cohorts for single-cell analysis

Table 2-5. Mutational status on 15 MDS patient with *SRSF2* mutation for single-cell analysis

ID	Gene	Reference	Genotype	VAF	c.	p.
PV1506	ASXL1	-	A	0.4878	c.1772dupA	p.Y591_Q592delinsX
	ETV6	T	-	0.4555	c.614delT	p.L205fs
	NRAS	C	G	0.1336	c.G34C	p.G12R
	PTPN11	A	G	0.034	c.A182G	p.D61G
	RUNX1	GG	-	0.4167	c.1151_1152del	p.384_384del
	SETBP1	C	A	0.0507	c.C2607A	p.S869R
	SRSF2	G	T	0.5352	c.C284A	p.P95H
PV1553	SRSF2	G	C	0.491	c.C284G	p.P95R
	TET2	G	C	0.4649	c.G4021C	p.A1341P
	TET2	T	-	0.469	c.4285delT	p.F1429fs
PV1488	SRSF2	G	A	0.5281	c.C284T	p.P95L
	TET2	C	T	0.5046	c.C5500T	p.Q1834X
543	SRSF2	G	A	46.8	c.284C>T	p.Pro95Leu
	ASXL1	AG	A	42.2	c.2201delG	p.Arg734Leufs*9
	RUNX1	G	A	40.3	c.529C>T	p.Arg177Ter
	STAG2	C	T	36.2	c.775C>T	p.Arg259Ter
	BCOR	A	AC	32.4	c.675dupG	p.Ser226Valfs*75
973	SRSF2	n/a	n/a	n/a	c.284C>T	p.Pro95Leu
	TET2	TC	T	44.4	c.2044del	p.His682Ilefs*18
	ASXL1	A	AG	ns	c.1934dup	p.Gly646Trpfs*12
PV2210	ASXL1	-	G	0.4529	c.1927dupG	p.G642fs
	IDH1	C	T	0.4895	c.G395A	p.R132H
	SRSF2	GGCGGCTGTGGTGTGAGTCCGGGG	-	0.6101	c.284_307del	p.95_103del
PV1253	ASXL1	-	G	0.3886	c.1927dupG	p.G642fs
	SRSF2	G	C	0.4491	c.C284G	p.P95R
	TET2	AG	-	0.4491	c.2273_2274del	p.758_758del
	TET2	-	TT	0.4383	c.3385_3386insTT	p.D1129fs
PV593	CBL	T	C	0.073668	c.T1139C	p.L380P
	JAK2	G	T	0.102167	c.G1849T	p.V617F
	SRSF2	G	T	0.329386	c.C284A	p.P95H
	TET2	G	A	0.464427	c.G3632A	p.C1211Y

	TET2	A	-	0.363401	c.4786delA	p.N1596fs
PV244	CBL	G	A	0.24689	c.G1259A	p.R420Q
	SRSF2	G	A	0.423742	c.C284T	p.P95L
	TET2	AT	-	0.403593	c.2171_2172 del	p.724_724del
	TET2	C	T	0.449845	c.C2887T	p.Q963X
115	SRSF2	CGGCGGCTGTGGT GTGAGTCCGGGG	CGGCGGC TGTGGT GTGAGTC CGGGA	40.49	c.284C>T	p.Pro95Leu
	TET2	G	T	48.57	c.184G>T	p.Gly62Ter
	ASXL1	CGGTGGACAAG	C	42.77	c.3085_3094 delGTGGAC AAGG	p.Val1029Metfs *15
	SETBP1	G	G/A	45.27	c.2609G>A	p.Gly870Asp
PV492	ASXL1	-	G	0.288071	c.1927dupG	p.G642fs
	IDH1	C	T	0.372134	c.G395A	p.R132H
	SRSF2	G	T	0.396985	c.C284A	p.P95H
PV1424	CUX1	G	T	0.102845	c.G2596T	p.E866X
	SRSF2	G	T	0.403561	c.C284A	p.P95H
	TET2	C	T	0.613596	c.C3412T	p.Q1138X
PV1539	RUNX1	G	A	0.0235	c.C622T	p.Q208X
	RUNX1	G	C	0.4346	c.C908G	p.S303X
	SRSF2	G	C	0.5268	c.C284G	p.P95R
	TET2	C	T	0.4962	c.C2272T	p.Q758X
666	SRSF2	G	T	59.0	c.284C>A	p.Pro95His
	TET2	C	CT	21.5	c.2562dup	p.Ala855Cysfs* 17
	TET2	C	T	49.5	c.4621C>T	p.Gln1541Ter
	ASXL1	ACTCC	A	49.9	c.3637_3640 del	p.Leu1213Ilefs* 3
	CBL	G	A	23.4	c.1259G>A	p.Arg420Gln
930	SRSF2	G	T	90.1	c.284C>A	p.Pro95His
	TET2	C	T	14.8	c.4165C>T	p.Gln1389Ter
	ASXL1	ns	ns	47	c.1900_1922 del	p.Glu635Argfs* 15
	NRAS	C	A	4.1	c.35G>T	p.Gly12Val

2.9.2 Multiplex Samples with Hashtags, Sample staining and Fluorescent activated cell sorting (FACS)

Bone marrow mononuclear cells from MDS patient samples were hashed using TotalSeq™-B hashtag reagents (**Table 2-6**) (BioLegend, San Diego, California, United States) according to the manufacturer's protocol and then the hashed cells were stained with fluorochrome conjugated antibodies (**Table 2-7**) to enrich HSPCs (Lineage$^-$ CD34$^+$). Briefly 1:200 dilution of Human Fc-block (BD) added in each sample and incubated for 5 minutes on ice before staining. The hashing and fluorochrome antibodies were added to each sample tube and then incubated for 25 minutes on ice. After that, the cells were washed once with 1 mL of FACS buffer with EDTA (2mM), resuspended with 500 µL of FACS Buffer with EDTA and 7-AAD was added to set the gates for the viable cell population.

Table 2-6 Hashtag reagents used for multiplex samples.

Name	Isotype	Barcode Sequence	Company
TotalSeq™-B0251 anti-human Hashtag 1	Mouse IgG1, κ (all clones)	GTCAACTCTTTAGCG	BioLegend
TotalSeq™-B0252 anti-human Hashtag 2	Mouse IgG1, κ (all clones)	TGATGGCCTATTGGG	BioLegend
TotalSeq™-B0253 anti-human Hashtag 3	Mouse IgG1, κ (all clones)	TTCCGCCTCTCTTTG	BioLegend

Table 2-7. Antibodies used for enrichment of Lineage$^-$ CD34$^+$ cells by FACS.

Antibody-Fluorochrome	Clone	Dilution	Company
CD8-FITC (Lineage)	Clone: RPA-T8	1:100	BioLegend
CD20-FITC (Lineage)	Clone: 2H7	1:150	BioLegend
CD66b-FITC (Lineage)	Clone: G10F5	1:15	BioLegend
CD10-FITC (Lineage)	Clone: HI10a	1:30	BioLegend
CD127-FITC (Lineage)	Clone eBioRDR5	1:30	eBioscience
Human Hematopoietic Lineage Cocktail-FITC (Lineage)	NA	1:15	eBioscience
CD34-APC-eF780	Clone: 4H11	1:150	eBioscience

To determine initial voltages and compensations, BD CompBeads Set Anti-Mouse Ig (BD) used as single-stained samples. FMOs were applied in all the experiments to set the boundary between background fluorescence caused by other fluorochromes and positive signal, and thus to settle the gates for distinguishing the positive and negative populations.

FACS was performed using BD Fusion I instrument (BD) with the assistance of Dr. Sally-Ann Clark at the Flow Cytometry Facility of Weatherall Institute of Molecular Medicine. Lineage$^-$ CD34$^+$ cells were collected in 1.5ml microfuge tube with FACS buffer. The number of cells were counted with FACS instrument and haematocytometer.

2.9.3 Single-cell transcriptome library preparation and sequencing

Each sample was prepared for single-cell transcriptome library right after enrichment by FACS. The libraries were generated at Oxford Genomics Centre using the standard 10X

Genomics platform (Chromium Next GEM Single Cell 3'Reagent Kits v3.1; Pleasanton, California, United States) according to manufacturer's recommendations.

Individual libraries were normalised using Qubit (Invitrogen), and the profile of size was measured on the 2200 or 4200 TapeStation (Agilent Technologies, Santa Clara, California, United States). Barcoded libraries were normalised and pooled together. The pooled libraries were diluted to ~10 nM and then denatured and further diluted before loading on the sequencer. Paired end sequencing was carried out using a NovaSeq6000 platform (Illumina, NovaSeq 6000 S2/S4 reagent kit v1.5, 300 cycles; San Diego, CA, USA)

2.9.4 Single cell data analysis

2.9.4.1 Sample processing and library preparation

For each sequencing run, two pools of samples were included, each pool consisted of three donors. In total three sequencing runs were performed, and 18 donors were sequenced.

2.9.4.2 Data pre-processing

Raw sequencing reads in FASTQ format were aligned to the human reference genome (GRCh38) using the cellranger count function of cellranger (v6.1.2) module. The option --include-introns was specified so that reads mapping to the intronic regions were included for downstream analysis.

We used two approaches to demultiplex each pool of samples into their respective donor. The first approach was based on the hashtag oligos (HTOs) expression. Because we know a priori the specific HTO tagged to a given donor, we were able to assign each cell to its

corresponding donor based on the highest expressing HTO. To this end, the Seurat R package was used to assign each cell to its corresponding donor (Satija et al., 2015). The second approach was based on unsupervised clustering of the cells based on transcriptome-wide single nucleotide polymorphisms (SNPs). To this end, Souporcell was used to determine 3 clusters of cells, whereby each cluster represented one of the three donors pooled (Heaton et al., 2020). Because Souporcell is agnostic to the donor ID, we assigned each cluster to their respective donor based on the majority cell identity inferred from Seurat. For example, if majority of cells from Souporcell cluster 1 were identified as originating from donor A, then all cells within cluster 1 will be assigned as being originated from donor A. The benefit of performing Souporcell on top of HTO-based demultiplexing is to rescue cells with low HTO expression.

Once the cells have been assigned to their respective donor, the input files required for downstream quality control were generated, namely the barcode (barcodes.tsv), feature (features.tsv), and matrix (matrix.mtx) files.

2.9.4.3 Quality control

For each donor, the SingCellaR R package was used for quality control of the single cells (Wang et al., 2022). For all donors, only cells with <10% mitochondria reads, >1,000 unique molecular identifier (UMI) counts, and >500 genes detected were retained. The upper threshold for UMI counts and genes detected, below which the cells were retained, were determined for each donor separately. Filtering of cells based on the aforementioned thresholds was performed using the filter_cells_and_genes function. The SingCellaR object for each donor after quality control was subsequently saved.

2.9.4.4 Cell type annotation

Each individual SingCellaR object representing each donor were integrated using the preprocess_integration function. The raw gene expression counts were normalised using the normalize_UMIs function. Specifically, the raw count for a given gene in a given cell was divided by that cell's total gene count, and then multiplied by 10,000. Highly variable genes were then determined using get_variable_genes_by_fitting_GLM_model function, and these genes were brought forward for dimension reduction analysis. We first performed linear dimension reduction analysis using principal component analysis (PCA) with the runPCA function. Based on the Jackstraw plot plotted using the plot_PCA_Elbowplot function, we brought forward the first 30 principal components (PC) for further non-linear dimension reduction analysis. To this end, we performed Uniform Manifold Approximation and Projection (UMAP) using the runUMAP function.

We observed the cells clustered according to their sequencing batch and donor ID on the UMAP embeddings. Therefore, we removed these batch effects using the runHarmony function. The resulting UMAP revealed cells to cluster by their cell lineage, i.e., erythroid, myeloid, lymphoid, and megakaryocyte lineages, and no longer by their sequencing batch and donor ID. This indicated that our batch correction was successful in mitigating the batch effect attributed to sequencing batch and donor ID.

Next, we performed Louvain clustering on our batch-corrected UMAP embeddings using the identifyClusters function with the options knn.metric="euclidean" and n.neighbors=30. In total, 21 Louvain clusters were identified. The top marker genes for each cluster were subsequently identified using the findMarkerGenes function. Furthermore, for each cluster,

differential gene expression analysis was performed against all other clusters using the identifyGSEAPreRankedGenes_for_all_clusters function to generate a list of genes ranked from the most significantly up-regulated to most significantly down-regulated. These ranked lists of genes were subsequently used for gene set enrichment analysis (GSEA) using the Run_fGSEA_for_multiple_comparisons function to identify the cell lineage signature gene sets most enriched within each cluster. The cell population identity for each cluster was determined based on the most enriched cell lineage signature gene sets. Additionally, the expression of canonical marker genes, e.g., AVP for HSC, were investigated for each putative cell population as a sanity check. Based on GSEA and canonical marker genes, we collapsed the original 21 clusters into 15 cell populations.

2.9.4.5 Differential cell abundance analysis

Based on our cell population assignment above, we next investigated if the different cell populations differed in their abundance (proportions) across the different genotype groups, i.e., *SRSF2/TET2*, *SRSF2/ASXL1*, *SRSF2/TET2/ASXL1*, and HD (healthy donors). The differential cell abundance analysis was performed using our in-house SingCellaRplus R package. SingCellaRplus is an R package which provides adjunct R functions to extend the existing functionalities of both SingCellaR and Seurat. We first down-sampled the cells for each genotype group to the number of cells of the genotype group with the lowest number of cells using the DownsampleCellsbyGroup function. Next, contour plot for each genotype group was generated using the PlotCellFraction.Contour function to enable us to observe cell populations that are explicitly expanded or depleted relative to one another. To enable more quantitative assessment of the cell abundance across the different genotypes, we assessed the average proportion of cells of a given cell population across

the different genotype groups using the PlotCellFraction.Boxplot function. Lastly, we scrutinised the proportion of cells for all cell populations in each donor in the form of a barplot using the PlotCellFraction.Barplot function.

2.9.4.6 Differential gene expression (DE) analysis

Differential gene expression analysis was performed for *SRSF2/TET2, SRSF2/ASXL1, SRSF2/TET2/ASXL1* vs HD using the identifyDifferentialGenes function from SingCellaR. Only genes expressed in at least 5% in either cell group were included for analysis. Differentially expressed genes were defined as false discovery rate (FDR) <1% and absolute log2 fold change >1. Custom R scripts were used to plot the results of the differential gene expression analysis on volcano plots (x-axis: log2 fold change; y-axis: FDR) to identify top differentially expressed genes. Furthermore, Venn Diagram was used to identify *SRSF2/TET2, SRSF2/ASXL1, SRSF2/TET2/ASXL1*-specified genes. Lastly, pathway enrichment analysis was performed using the cluserProfiler R package to identify pathways enriched among the differentially expressed genes (Wu et al., 2021). Top 15 pathways from each comparison were plotted side-by-side in the form of a dot plot to reveal pathways shared or specific to *SRSF2/TET2, SRSF2/ASXL1, SRSF2/TET2/ASXL1*.

3 Chapter Three – Investigation of the molecular mechanisms underlying ineffective erythropoiesis in MDS with splicing factor mutations

3.1 Introduction

Anaemia is widely known to be the most common manifestation of MDS, which is associated with ineffective development of erythroid precursor cells, leading to chronic dependency on transfusions of red blood cells (Platzbecker et al., 2012). Previous studies have demonstrated that the scarcities of erythrocyte subpopulations in MDS were more noticeable than other myeloid subpopulations (Flores-Figueroa et al., 1999), which emphasize the prevalence of ineffective erythroid maturation in MDS. Impaired erythroid maturation in MDS refers to the interruption of differentiation, morphological alteration, and increased apoptosis of erythroid precursor cells which result from a range of mechanism including genetic pathogenic factors in HSCs, dysregulation of steady-state cytokines and signalling pathways (Cazzola, 2020, Bhagat et al., 2013, Pellagatti et al., 2018, Yip et al., 2017). Recent investigations also have identified activation of the TGF-β signalling in HSPCs which leads to inhibition of erythropoiesis in MDS in SMAD2/3 dependent manner (Zhou et al., 2008, Zhou et al., 2011, Bhagat et al., 2013). Dolatshad et al. also identified the association between the morphological alteration of erythroblasts called ring sideroblasts (RS) and aberrant splicing of the iron transporter, *ABCB7*, caused

by *SF3B1* mutations (Dolatshad et al., 2016). However, other potential mechanisms causing defective erythropoiesis still remain elusive.

We have previously shown that expression of the *U2AF1*^{S34F} mutation in human haematopoietic progenitors impairs erythroid differentiation and skews granulomonocytic differentiation toward granulocytes (Yip et al., 2017). RNA sequencing of erythroid and granulomonocytic colonies revealed that *U2AF1*^{S34F} altered mRNA splicing of many transcripts in a lineage-specific manner. *U2AF1*^{S34F} induced skipping of exon 2 of the *STRAP* gene preferentially in erythroid colonies (**Figure 3-1**), giving rise to a premature stop codon. Resultantly, the mRNA transcript is targeted by NMD (Lewis et al., 2003), leading to the downregulation of the *STRAP* gene. Knockdown of *STRAP* in human bone marrow CD34^{+} cells resulted in impaired erythroid differentiation (Yip et al., 2017). Furthermore, Yip *et al.* have shown that overexpression of *STRAP* rescues the erythroid differentiation defect in *U2AF1*^{S34F} MDS cells (Yip et al., 2017). However, the mechanism by which down-regulation of *STRAP* results in impaired erythropoiesis remains to be defined.

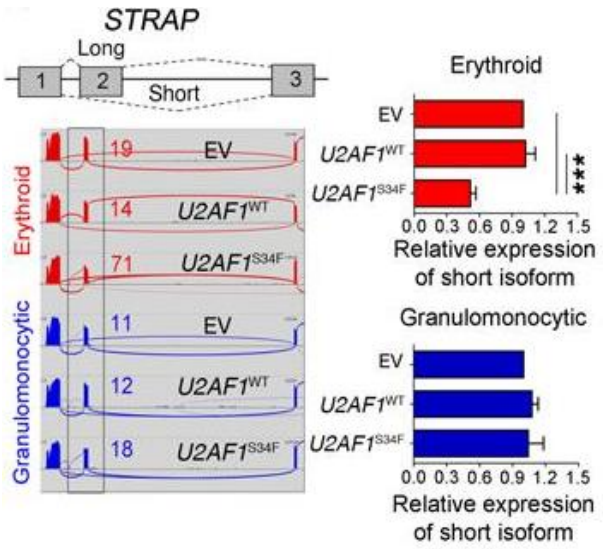


Figure 3-1. Confirmation of lineage-specific splicing alterations in *U2AF1*^{S34F} mutant erythroid and granulo-monocytic cells.

* This figure is adapted from Yip et al., JCI 2017

In this study, we identified down-regulation of *STRAP* expression in primary bone marrow CD34^{+} cells of MDS and found an aberrant splicing event of *STRAP* in erythroid lineage cells of MDS patients harbouring *SRSF2* mutations. We demonstrated that these dysregulations of *STRAP* cause impaired erythroid maturation in our *in vitro* studies. The functional studies showed that knockdown of *STRAP* led to inactivation of p38, a member of MAPK pathway, and downregulation of CSDE1 bound transcripts. These *in vitro* results suggest that the underlying mechanism of ineffective erythropoiesis in MDS with low expression of *STRAP* is caused by *SRSF2* mutations.

3.2 Results

3.2.1 *STRAP* is downregulated in bone marrow CD34$^+$ cells in patients with MDS

Firstly, we analysed the gene expression levels of *STRAP* in our large RNA-seq gene expression dataset on CD34$^+$ cells derived from the bone marrow samples of 82 MDS patients and 8 healthy controls. Of the 82 MDS cases, 28 were *SF3B1* mutant (*SF3B1*mut), 8 *SRSF2* mutant (*SRSF2*mut), 6 *U2AF1* mutant (*U2AF1*mut) and 40 had no mutations in splicing factor genes (SFwt), as determined by targeted next-generation sequencing data (Pellagatti et al., 2018) (Table S 1 and Table S 2 11).

We found that *STRAP* expression levels were significantly decreased in MDS samples when compared to healthy controls (**Figure 3-2**). *STRAP* was one of the most significant differentially expressed genes in MDS and was markedly reduced in some patients. Correlation of *STRAP* levels with clinical parameters did not reveal any associations with age, sex, or cytogenetics, suggesting it to be a pervasive finding in MDS (**Figure 3-3**).

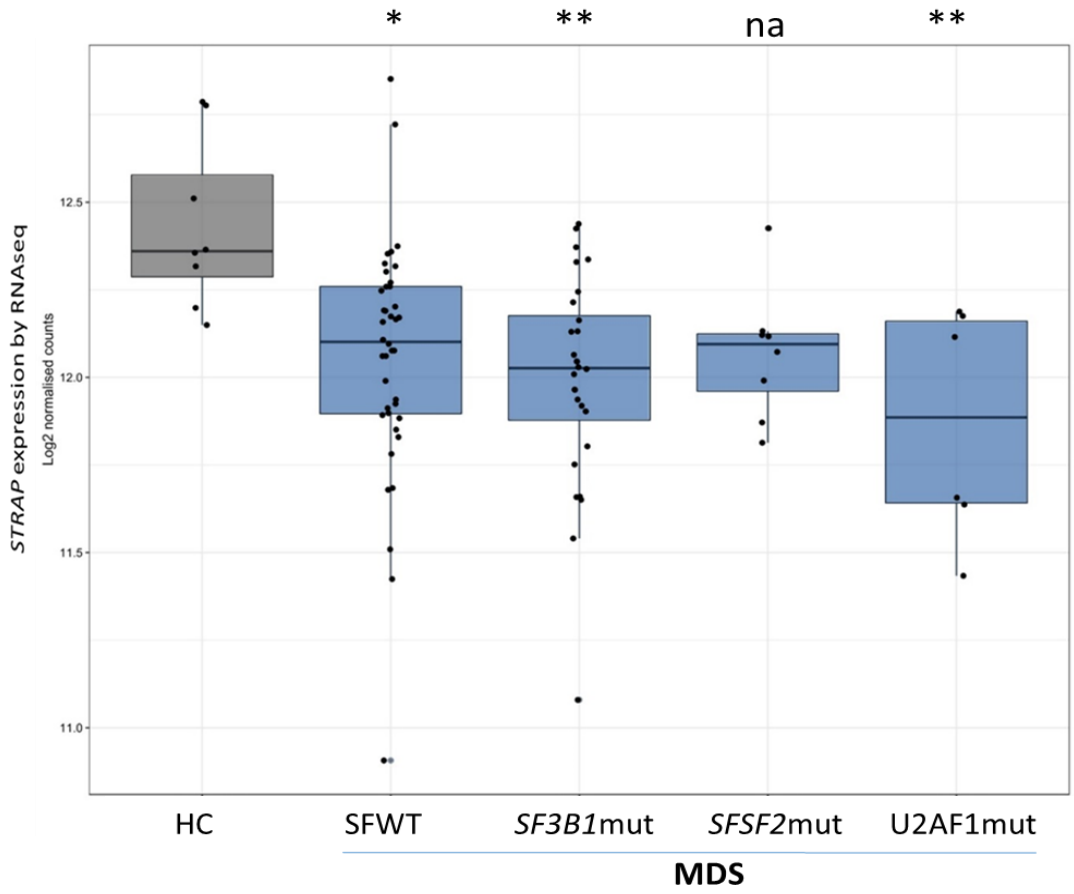


Figure 3-2. A box plot showing mRNA expression of *STRAP* which is significantly decreased in $CD34^{+}$ cells from MDS patient samples compared to healthy control. P-values in panel were calculated by repeated-measures 1-way ANOVA with Bonferroni's multiple comparison test *P < 0.05 **P < 0.01 ***P < 0.001

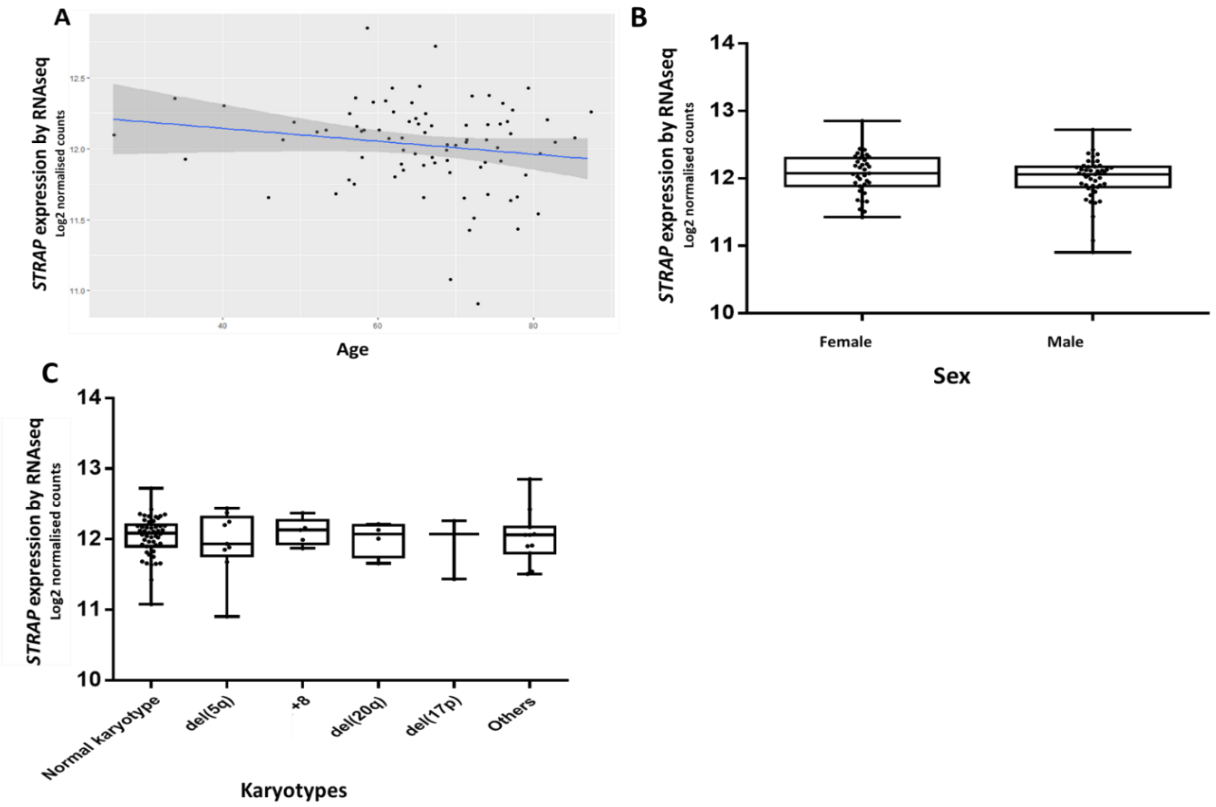


Figure 3-3. Plots showing no significant correlation of *STRAP* expression levels with (A) age, (B) sex, (C) cytogeneticsin bone marrow CD34^{+} cells from MDS patients.

(A) Scatter plot. Estimate: -0.004558 (age), P value: 0.1333 p-value: 0.3408. These figures were calculated using linear regression model in R. (B) Box plots. P value: 0.3408. this figure was calculated by unpaired t test. (C) Box plots. P value: 0.9370. This figure was calculated by 1 way ANOVA**.**

3.2.2 Aberrant splicing of *STRAP* in erythroid precursors of *SRSF2*mut MDS

Our group has previously shown that *U2AF1*^{S34F} mutation induces skipping of exon 2 of *STRAP*, predominantly in erythroid cells (Yip et al., 2017). In order to determine whether aberrant splicing of *STRAP* occurs in association with splicing factor mutations other than *U2AF1*, we analysed the aberrant splicing events from the RNA-seq analysis of bone marrow monocytic, granulocytic and erythroid precursors obtained from *SF3B1*mut (n=7) or *SRSF2*mut (n=8) MDS patient samples compared to healthy controls (n=4) (**Table 3-1**) (Pellagatti et al., 2018).

Interestingly, we found an aberrant splicing event in *STRAP*, which is identical to the one caused by the *U2AF1*^{S34F} mutation, in the erythroid precursors of *SRSF2* mutant MDS cases (**Figure 3-4**, **Table 3-1** and **Table 3-2**). These results suggest that this *STRAP* splicing alteration by *SRSF2*mut, as well as *U2AF1*mut, gives rise to premature termination codons which can lead to degradation of *STRAP* mRNA by NMD (Lewis et al., 2003), in erythroid cells of MDS patients.

Table 3-1. Clinical details and mutational status of the 11 patients used in the study of bone marrow monocytic, granulocytic and erythroid precursors

			Cell population present (1=yes, 0 =no)			Mutation status (1=yes, 0 =no)								
ID	Disease Subtype	Age	MON	GRA	ERY	SF3B1	SRSF2	TET2	CBL	IDH2	ASXL1	JAK2	PHF6	NPM1
MDS142	RA	76	1	0	0	1	0	1	0	0	0	0	0	0
MDS152	RARS	64	1	0	1	0	1	0	1	1	0	0	0	0
MDS155	MDS/AML	72	1	1	1	1	0	0	0	0	1	1	0	0
MDS163	MDS/AML	76	1	1	1	1	0	0	0	0	0	0	0	0
MDS166	RAEB2	44	1	1	1	1	0	0	0	0	0	0	1	0
MDS168	MDS/AML	71	1	0	1	0	1	0	0	1	0	0	0	1
MDS177	RAEB2	81	1	0	1	1	0	0	0	0	0	0	0	0
MDS178	RARS	63	1	1	1	1	0	1	0	0	0	0	0	0
MDS189	RARS	81	1	1	1	1	0	0	0	0	0	0	0	0
MDS191	RA	83	1	1	0	0	1	0	0	0	0	0	0	0
MDS218	RA	87	1	1	1	0	1	1	0	0	1	0	0	0

Table 3-2. Aberrant splicing of *STRAP* in erythroid precursors from *SRSF2* mutant MDS patient samples.

The table provides details of the output generated by the rMATS bioinformatics pipeline (Dolatshad et al., 2016, Shen et al., 2014, Yip et al., 2017) for the *STRAP* aberrant splicing event. SE= exon skipping.

Event ID	*Symbol*	*Event Type*	*Chr*	*Strand*	*Start - End position*	*IncLevelDifference*	*P-value*	*FDR*
24374	**STRAP**	**SE**	**12**	**+**	**16036474-16036610**	**0.409**	**3.78663767009e-10**	**4.8999091451e-10**

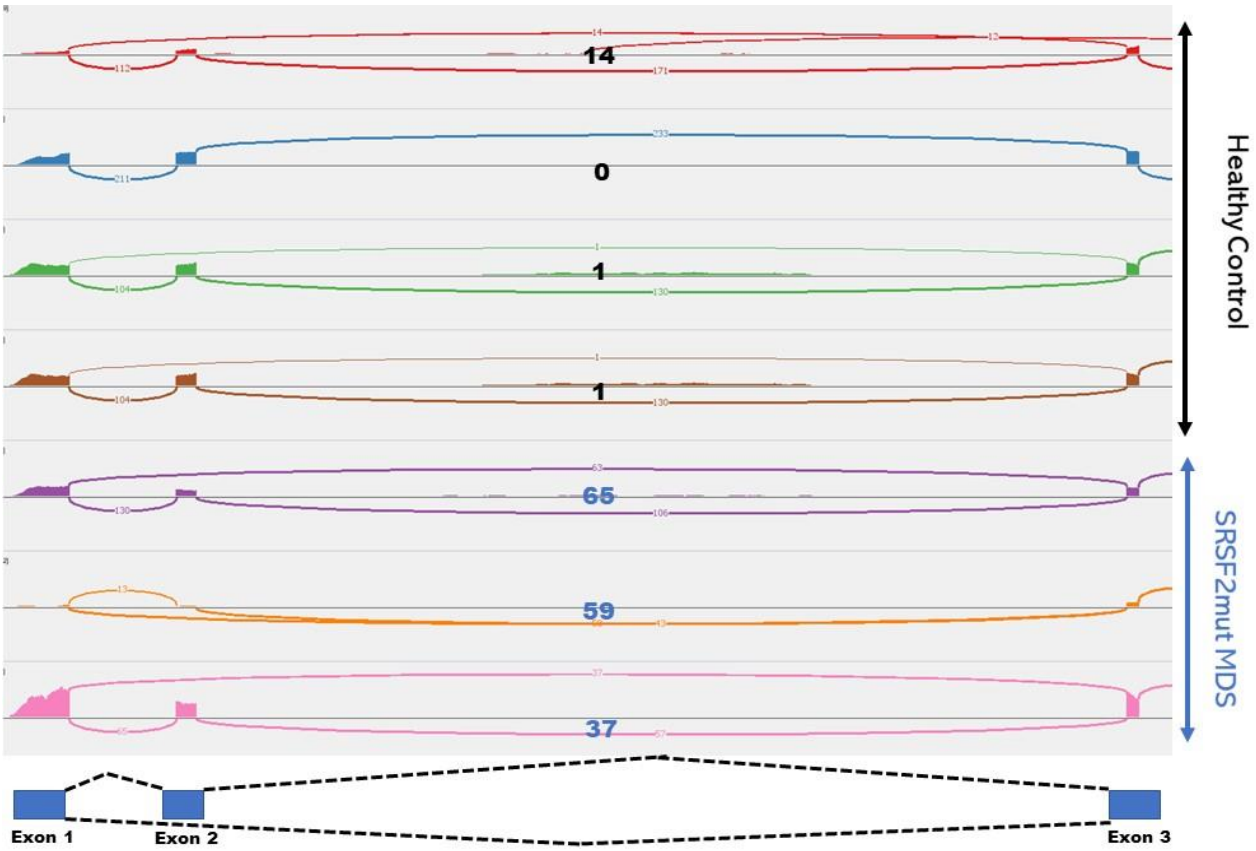


Figure 3-4. Sashimi plot showing the *STRAP* aberrant splicing event (skipped exon 2) in erythroid precursors from *SRSF2* mutant MDS patient samples.

The number on each plot refers to the number of sequencing reads skipping exon2.

3.2.3 Reduction of *STRAP* expression in CD34$^+$ cells leads to decreased cell growth and impaired erythroid differentiation

To investigate the effect of *STRAP* down-regulation on human bone marrow CD34$^+$ cells and erythroid differentiation, we performed *STRAP* knockdown in bone marrow CD34$^+$ cells from healthy controls using lentivirally-delivered shRNAs. Firstly, the shRNAs were tested in the K562 cell line: after lentiviral transduction, cells were selected in liquid media supplemented with puromycin and *STRAP* expression levels measured by real-time quantitative PCR (qRT-PCR). The two shRNAs (sh65, sh66) successfully downregulated the expression of *STRAP* in K562 cells by approximately 60-80% compared to a scramble control (**Figure 3-5**. A). These two shRNAs were then used to knock down *STRAP* in CD34$^+$ cells from healthy controls, and a 40-90% decrease was obtained compared to the scramble control (**Figure 3-5**. B).

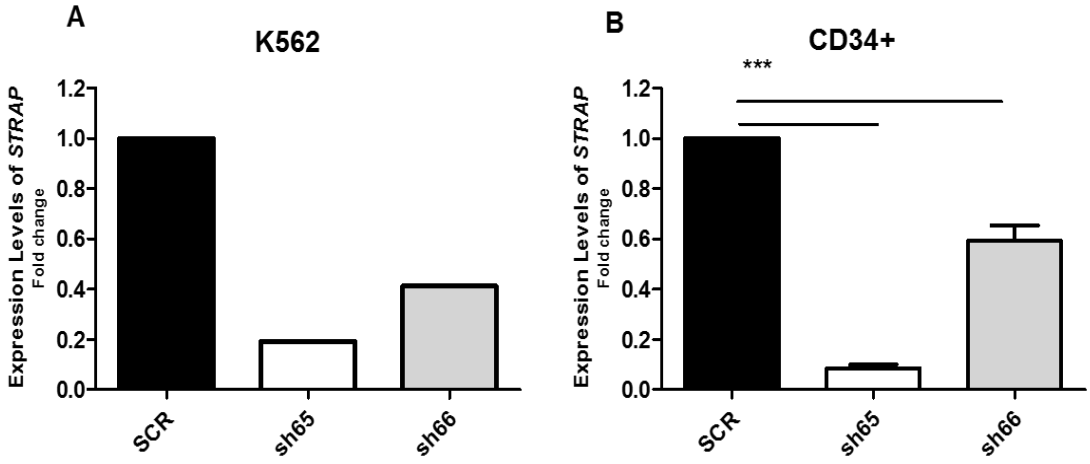


Figure 3-5. mRNA expression levels of *STRAP* determined using qRT-PCR in (A)K562 and (B)erythroid cells derived from bone marrow CD34$^+$ with *STRAP* knockdown.

sh65 and sh66 were significantly decreased compared to scramble control. Bar graphs (**B**) show mean +S.E.M.of 3 biologically independent experiments . P-values in panel (**B**) were calculated by repeated-measures 1-way ANOVA with Bonferroni's multiple comparison test. ***P <0.001.

$CD34^+$ cells with *STRAP* knockdown were differentiated towards the erythroid lineage to study the effects of the knockdown on erythropoiesis. Erythroid cells with *STRAP* knockdown showed impaired growth (**Figure 3-6**.A) compared to the scramble control. In addition, knockdown of *STRAP* caused a significant decrease in the intermediate erythroid cell population ($CD71^+$ / $CD235a^+$ and $CD36^+$ / $CD235a^+$) on day 11 of differentiation (**Figure 3-6**. B, C), compared to the scramble control.

Erythropoiesis is a multistep process involving the commitment and then sequential differentiation and proliferation of a multipotent progenitor [colony-forming unit granulocyte erythroid macrophage mixed (CFU-GEMM)]. CFU-GEMM further differentiates into two subpopulations of erythroid progenitors, namely burst-forming unit erythroid (BFU-E) and colony-forming unit erythroid (CFU-E), which are characterized by their *in vitro* pattern of colony formation in semisolid culture (Gregory and Eaves, 1978).

To determine the effect of *STRAP* down-regulation on erythroid differentiation, bone marrow $CD34^+$ cells from healthy donors (Stemcell Technologies) were cultured in erythroid differentiation media with cytokines (Erythropoietin, IL-3, SCF, Dexamethasone and lipids). On day 3 of differentiation, lentiviral transduction was performed. After the cells were selected in liquid media with puromycin, colony-forming cell assay was performed by plating transduced cells (4,000 cells) on day 7 of differentiation on methylcellulose (MethoCult H4434 Classic, Stemcell Technologies) containing 0.65 µg/ml puromycin.

Transduced progenitors with *STRAP* knockdown produced a markedly lower number of erythroid colonies (BFU-E and CFU-E) compared to progenitors transduced with the

scramble control (**Figure 3-6**. F and G) in colony-forming cell assays. It is interesting to note that the number of CFU-E colonies with *STRAP* knockdown (sh66) decreased markedly in comparison to the scramble control (**Figure 3-6**. F and G). These data suggest that *STRAP* downregulation impacts the development and differentiation of erythroid progenitor cells.

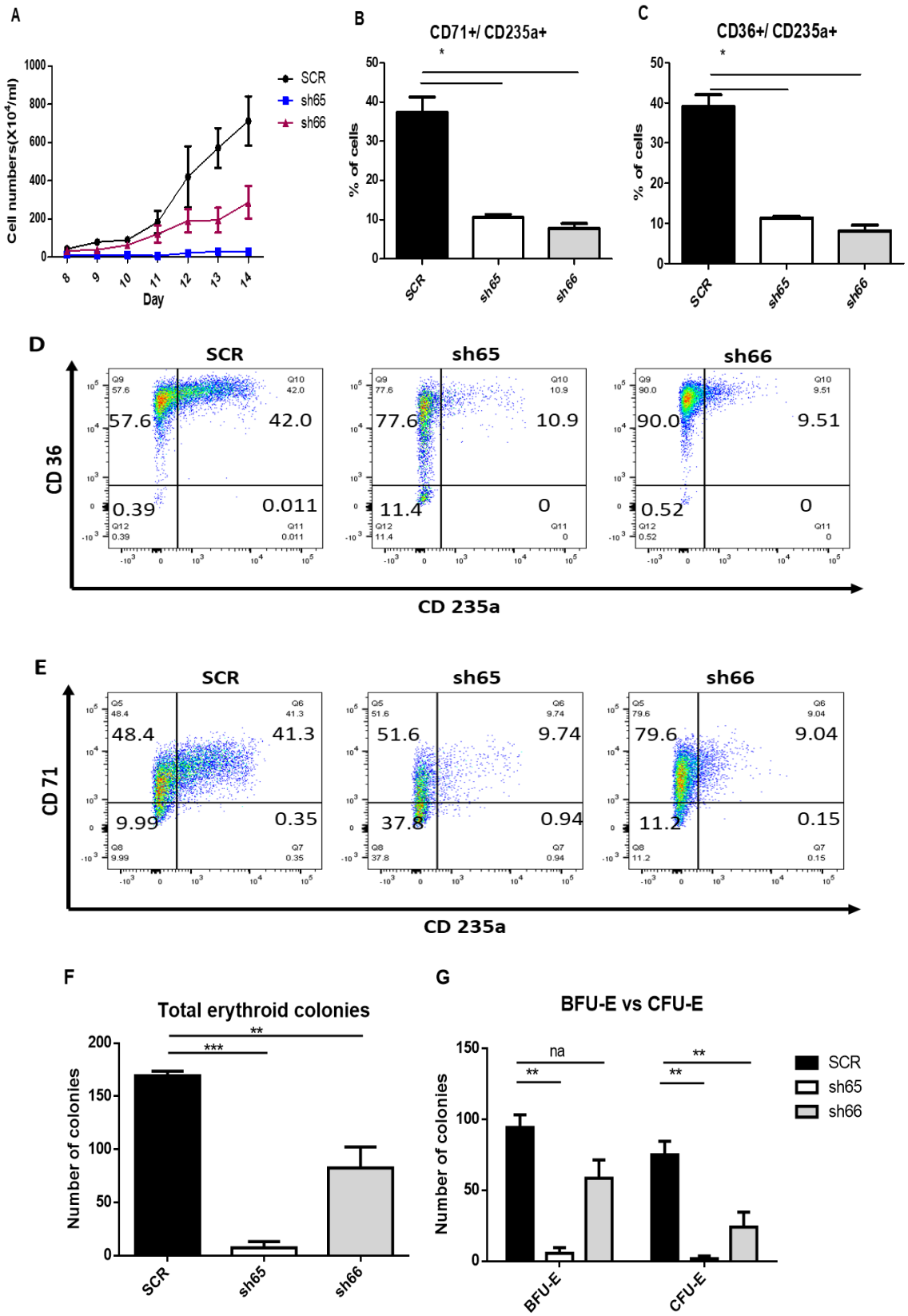

A
Cell numbers(X10⁴/ml)
Day
SCR
sh65
sh66
B
CD71+/ CD235a+
% of cells
C
CD36+/ CD235a+
% of cells
D
SCR
sh65
sh66
CD 36
CD 235a
57.6
42.0
0.39
0.011
77.6
10.9
11.4
0
90.0
9.51
0.52
0
E
SCR
sh65
sh66
CD 71
CD 235a
48.4
41.3
9.99
0.35
51.6
9.74
37.8
0.94
79.6
9.04
11.2
0.15
F
Total erythroid colonies
Number of colonies
G
BFU-E vs CFU-E
Number of colonies
na
BFU-E
CFU-E

Figure 3-6. Effects of *STRAP* knockdown on differentiation of erythroid cells derived from bone marrow CD34$^+$ cells.

(A) Growth curves for erythroid cells with knockdown of *STRAP*. (B, C) Flow cytometry-based quantification of erythroid differentiation. Percentage of CD71$^{+/}$CD235a$^+$ and CD36$^+$ /CD235a$^+$ cells (intermediate erythroid cell population) in erythroid cultures with knockdown of *STRAP* on day 11. (D, E) Representative flow cytometric plots showing expression of CD36, CD71 and CD235a on day 11 of erythroid culture. (F) Total erythroid colonies. (G) Comparison between BFU-E and CFU-E. Bar graphs show mean +S.E.M of biologically 3 independent experiments. (A, B, F and G) show mean +S.E.M of biologically 3 independent experiments. P-values in panel B and C were calculated by repeated-measures 1-way ANOVA with Bonferroni's multiple comparison test. *P <0.05.

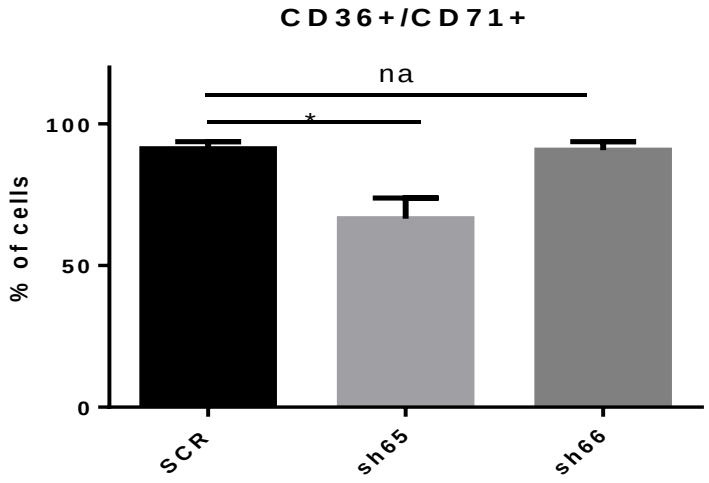


Figure 3-7. Effects of *STRAP* knockdown on differentiation at the early erythroid cells

3.2.4 *STRAP* knockdown in erythroid lineage cells results in inactivation of p38

Recent studies have demonstrated that STRAP can affect the regulation of p38 MAPK pathway via intracellular interacting partners including ASK1, TAK and MPK38, and it plays dual roles depending on cellular contexts (Manoharan et al., 2018, Reiner and Datta, 2011). The role of p38 MAPK (Mitogen-activated protein kinase) in primary human erythroid cells to coordinate differentiation, proliferation, and apoptosis has been reported, suggesting that p38 activity is required for erythropoiesis, especially at the terminal differentiation stage including enucleation (Wang et al., 2018b, Schultze et al., 2012, Somervaille et al., 2003, Uddin et al., 2004). These findings indicated that *STRAP* knockdown cells may result in dysregulation of p38 (Qin et al., 2021), which may lead to defective erythroid development in human primary cells. Thus, we investigated whether p38 is activated in human primary $CD34^+$ cells with *STRAP* knockdown, differentiating towards the erythroid lineage.

To measure p38 activation in erythroid progenitor cells with *STRAP* knockdown, we examined phosphorylation of p38 by using western blot analysis with anti-p38 and anti-phospho-p38 antibodies. Erythroblasts with *STRAP* knockdown showed significantly decreased phosphorylation of p38 compared to the scrambled control (**Figure 3-8**. B), with no significant difference in protein expression level of total p38 between the *STRAP* knockdown and the scrambled control (**Figure 3-8**. A). These results indicate that p38 is inactivated in erythroid progenitor cells with *STRAP* knockdown.

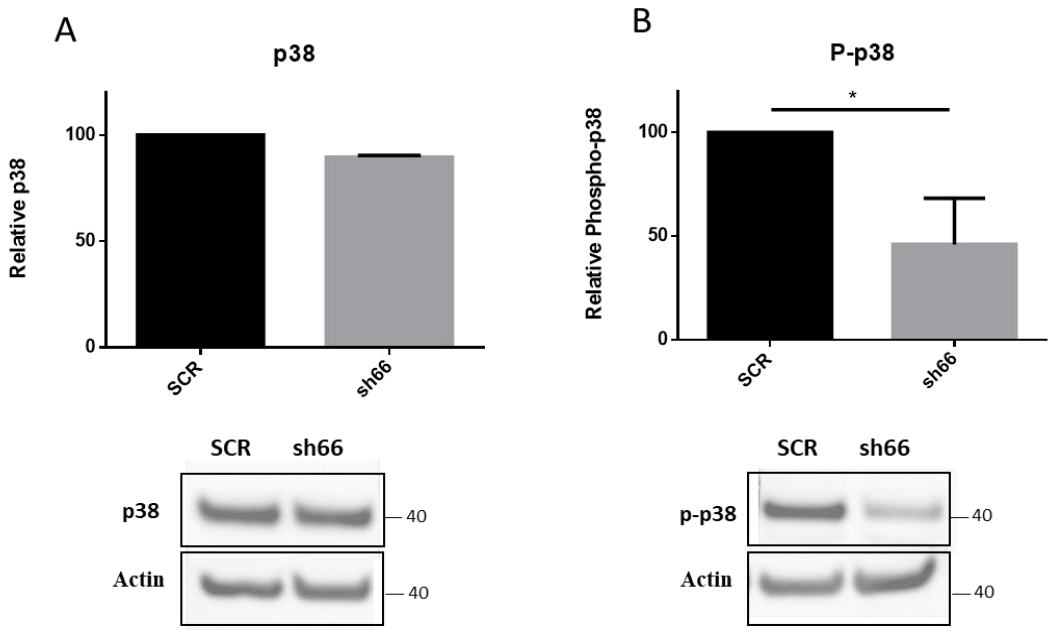


Figure 3-8. Protein expression of p38 and phosphorylated p38 in erythroid cells derived from bone marrow CD34$^+$ cells.

(A) Total p38 and (B) phosphorylated p38 expression were measured by western blot. Intensity of each band was measured with ImageJ. Statistical significance was determined by paired t test from 3 independent experiments. *P <0.05

3.2.5 *STRAP* knockdown reduced mRNA expression of CSDE1-bound transcripts.

Moor *et al*. have previously shown that Strap physically associates with Cold shock domain protein e1 (Csde1) by using strong affinity of streptavidin-biotin interaction in Murine erythroleukemia (MEL, mouse erythroblasts transformed with Friend virus) and the association between Strap and Csde1 is almost twice as much as compared to second highest enriched protein (Znfx1) (Moore et al., 2018a). Csde1 is an RNA-binding protein which is crucial for erythroid maturation and is highly expressed in erythroblasts (Moore and von Lindern, 2018, Moore et al., 2018b). These studies also showed that Csde1 are affected by *Strap* knockdown, changing mRNA and/or protein expressions of Csde1-bound transcripts in erythroblasts of mouse (Moore et al., 2018a). However, the dysregulations of CSDE1-bound transcripts by *STRAP* knockdown have not been investigated in human erythroid progenitor cells. Therefore, we performed gene expression analysis of CSDE1-bound transcripts by using qRT-PCR to determine whether the dysregulation of CSDE1-bound transcripts take place in human primary erythroid progenitor cells with low level of *STRAP* expression.

Of CSDE1-bound transcripts, we have selected 6 genes which play important roles, especially in the process of erythroid differentiation (Moore et al., 2018b, Moore et al., 2018a), including Poly(A) Binding Protein Cytoplasmic 1 (*PABPC1*), Poly(A) Binding Protein Cytoplasmic 4 (*PABPC4*), RAN Binding Protein 1 (*RANBP1*), Hydroxymethylbilane Synthase (*HMBS*), Osteoclast Stimulating Factor 1 (*OSTF1*) and THO Complex 5 (*THOC5)*.

We observed a statistically significant decrease in mRNA expression of 6 CSDE1-bound transcripts in the erythroid lineage cells with approximately 90% of *STRAP* knockdown (sh65) compared to a scramble control (**Figure 3-9**. A-F), which was in concordance with a recent study (Moore et al., 2018a) in MEL cells with *Strap* knockdown. Notably, among these 6 genes, there were significant reductions of *HMBS* and *OSTF1* in the erythroid lineage cells with approximately 50% of *STRAP* knockdown (sh66), a level similar to that observed in MDS erythroid cells(Yip et al., 2017), compared to a scramble control (**Figure 3-9**. E and F).

These results indicate that down-regulation of *STRAP* leads to deregulated gene expression of CSDE1-bound transcripts in human erythroid progenitor cells, suggesting that these dysregulations may contribute to inadequate production of erythroblasts in MDS patients.

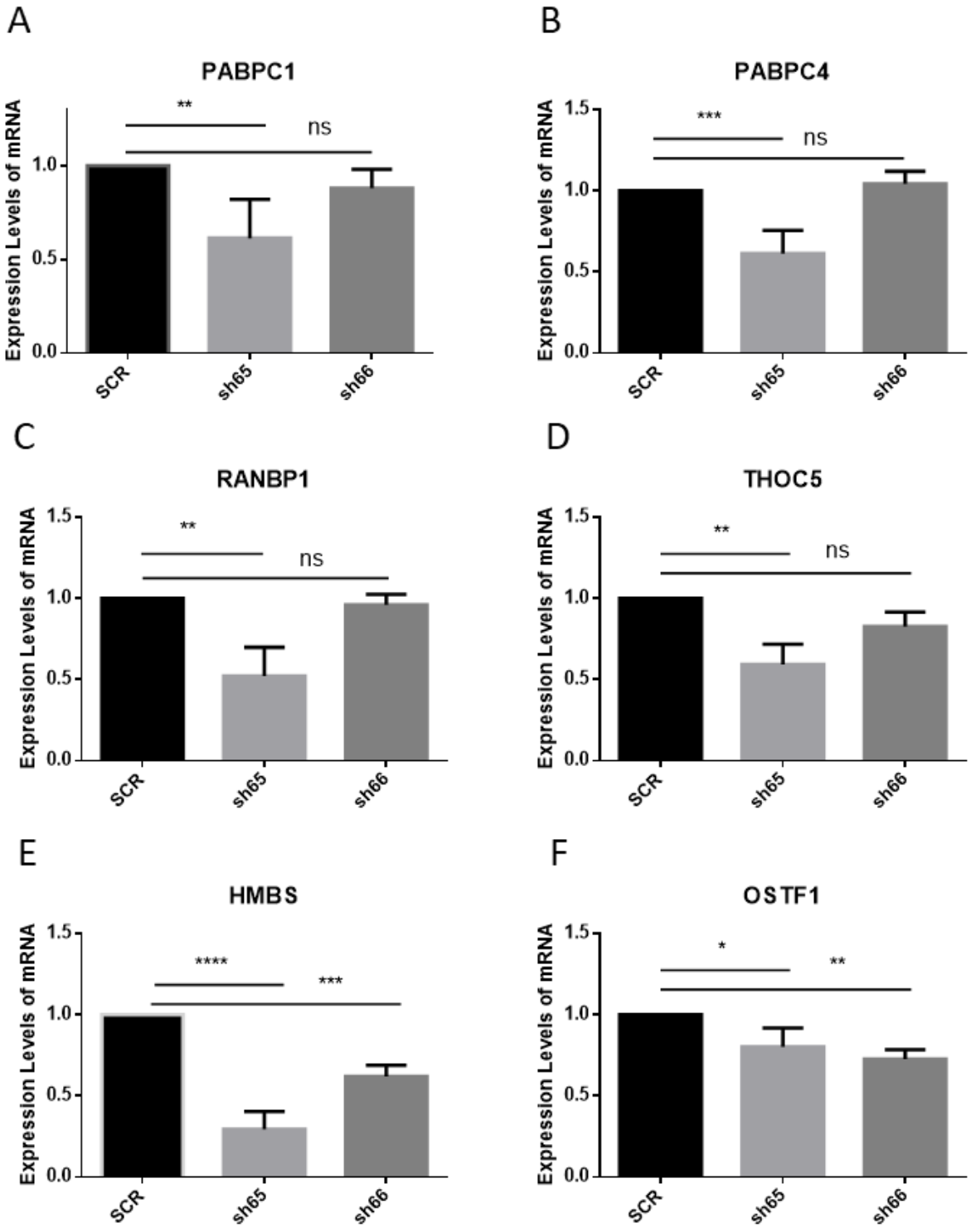

A
PABPC1
Expression Levels of mRNA
**
ns
1.0
0.5
0.0
SCR
sh65
sh66
B
PABPC4
Expression Levels of mRNA
***
ns
1.5
1.0
0.5
0.0
SCR
sh65
sh66
C
RANBP1
Expression Levels of mRNA
**
ns
1.5
1.0
0.5
0.0
SCR
sh65
sh66
D
THOC5
Expression Levels of mRNA
**
ns
1.5
1.0
0.5
0.0
SCR
sh65
sh66
E
HMBS
Expression Levels of mRNA
****
***
1.5
1.0
0.5
0.0
SCR
sh65
sh66
F
OSTF1
Expression Levels of mRNA
*
**
1.5
1.0
0.5
0.0
SCR
sh65
sh66

Figure 3-9. mRNA expression levels of CSDE1-bound transcripts.

Expression levels of 6 CSDE1-bound transcripts including *PABPC1* (**A**), *PABPC4* (**B**), *RANBP1* (**C**), *THOC5* (**D**), *HMBS* (**E**) and *OSTF1* (**F**) were determined using qRT-PCR in erythroid cells derived from human bone marrow $CD34^{+}$ with *STRAP* knockdown. The bar graphs show mean +S.E.M.of 4 biologically independent experiments. P-values were calculated by repeated-measures 1-way ANOVA with Tukey's multiple comparison test. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

3.3 Discussion

The underlying mechanisms by which splicing factor mutations lead to impaired erythropoiesis in MDS are incompletely understood. A previous study from our group has shown that the *U2AF1*^{S34F} mutation induces aberrant splicing of *STRAP* (Yip et al., 2017), predominantly in the erythroid lineage, resulting in impaired erythropoiesis. In this study, we identified down-regulation of *STRAP* expression in primary bone marrow CD34^{+} cells of MDS and also found an aberrant splicing event of *STRAP* in erythroid lineage cells of MDS patients harbouring *SRSF2* mutations as well. We demonstrated that these dysregulation of *STRAP* causes impaired erythroid maturation in our *in vitro* studies. The functional studies showed that knockdown of *STRAP* led to inactivation of p38, a member of MAPK pathway, and downregulation of CSDE1 bound transcripts. These *in vitro* results thus unveil the underlying mechanisms of ineffective erythropoiesis in MDS with low expression of *STRAP* caused by *SRSF2* mutations.

We analysed a large dataset of RNA-seq, which allowed us to demonstrate the reduction of *STRAP* expression in bone marrow CD34^{+} cells of MDS patients. It is intriguing that the expression level of *STRAP* is significantly reduced in most cases, even in cases without splicing factor mutations. Recent studies have shown deregulated miRNAs, which cause downregulation of specific key genes in MDS, leading to impaired haematopoiesis (Fang et al., 2012, Muench et al., 2018, Bhagat et al., 2013). However, no miRNA, thus far, has been associated with regulation of *STRAP*. Little is known in terms of molecular interactions of *STRAP* that lead to a MDS phenotype during erythropoiesis, thus there is a need for in-depth characterisation of the role of *STRAP* in MDS pathogenesis.

We found abnormal splicing of *STRAP* in the erythroid precursors of MDS with *SRSF2* mutations, as well as in MDS with *U2AF1*^{S34F} mutation, from our rMATS bioinformatics pipeline. This abnormally spliced isoform of *STRAP* in MDS patient with *SRSF2* mutation is expected to be degraded in the same way of *U2AF1*^{S34F} mutations cases in MDS (Yip et al., 2017) since the aberrant transcripts contain premature stop codon, leading to NMD. Considering the significant role of STRAP during erythropoiesis, identification that aberrant splicing of *STRAP* occurs in two major splicing factor mutant MDS (*SRSF2* and *U2AF1*) is important (Moore et al., 2018b, Yip et al., 2017). Also, it would be very interesting to speculate about potential splicing mechanism between *SRSF2* and *U2AF1* mutations given that they result in aberrant splicing of *STRAP* identically. Interestingly, a recent study showed that *SRSF2* mutation causes "splicing-cascade" via altering other RNA processing and splicing genes in haematopoietic cells, which implicate functional association between splicing factor genes to regulate global mRNA splicing (Liang et al., 2018).

Reduction of p38 activation has been observed in human primary erythroid cells with *STRAP* knockdown in our study. Recent studies revealed that STRAP, being associated with p38 MAPK pathway, plays dual roles depending on cellular contexts (Manoharan et al., 2018, Reiner and Datta, 2011). Also, it has been well documented that p38 activation is required for terminal erythroid maturation in response to EPO (Somervaille et al., 2003, Uddin et al., 2004, Schultze et al., 2012, Wang et al., 2018b). More recent studies revealed dysregulated p38 pathway as the pathogenesis of MDS. However, conflicting findings have been reported about the activation of p38 in MDS (Smith et al., 2019, Navas et al., 2006, Peng et al., 2012, Spinelli et al., 2012). Notably, several studies have shown that p38 is over-activated in low-risk MDS, interrupting proliferation of $CD34^+$ progenitors and

suggested that over-activation of p38 may contribute to apoptosis of HSCs (Navas et al., 2006, Peng et al., 2012). It is noted that these results were based upon data from $CD34^+$ progenitor cells and focused on the relation between stem-cell survival and p38 activity. On the other hand, Spinelli et al. showed that the statistical analysis did not represent significant differences of basal p38 activation in distinct subpopulations ($CD45^+$ cells and $CD71^+/CD45^-$ cells) of MDS bone marrow, although the activation of p38 were observed in $CD34^+$ cells of a few MDS patients (n = 11) among the total number of cases (n = 54). This study suggests that the p38 activation is highly heterogeneous, dependent on disease staging and the subpopulations in MDS. We found that negative effects of knockdown of *STRAP* on erythropoiesis have appeared at the relatively late stage ($CD71^{+/}CD235^+$) (**Figure 3-6**. B) of erythroid cells compared to cells at the early stage ($CD36^+/CD71^+$) (**Figure 3-7**), consistent with results of the studies about the role of p38 in terminal erythropoiesis. In summary, given that the association between STRAP and p38 pathway and the role of p38 at the terminal stage of erythropoiesis (Schultze et al., 2012), our findings revealed that the inactivation of p38 in the cells at the late erythropoiesis is hindered by *STRAP* down-regulations which may be representative of the underlying mechanism of impairment of erythroid maturation in MDS with *SRSF2* and *U2AF1* mutation.

Csde1 is an RNA-binding protein initially called Unr (upstream of N-ras) and it is indispensable for erythroblast proliferation and differentiation in mouse erythroblasts (Triqueneaux et al., 1999, Moore et al., 2018b, Moore et al., 2018a, Horos et al., 2012). A recent study showed that Csde1 is strongly associated with Strap in MEL cells and the low expression of Strap altered some Csde1-bound transcripts (Moore et al., 2018a). In this study, we have identified the reduction of 6 CSDE1-bound transcripts which are required

for the regulation of terminal erythropoiesis and the hypoxic response. *PABPC1* and *PABPC4* play crucial roles in stable expression of a subset of erythroid mRNAs via binding to the 3'poly(A) tails (Kini et al., 2014). *RANBP1* is highly expressed in erythroid progenitor cells and mediate the cell cycle via nuclear export reactions (Fujishima et al., 2004, Tedeschi et al., 2007, Kehlenbach et al., 2001). *THOC5* is known for functioning as a nuclear exporter of transcripts which is vital for of haematopoietic differentiation and proliferation (Mancini et al., 2010, Tran et al., 2014). HMBS is an important enzyme that is associated with catalysation in heme biosynthetic pathway and is dysregulated under hypoxic conditions, maintaining heme homeostasis (Vargas et al., 2008, Dailey and Meissner, 2013), and *OSTF1* is regulated by hypoxia (Day et al., 2017, Fuhrmann et al., 2015, Clees et al., 2022). Thus, our results suggest that dysregulations of these CSDE-bound transcripts may lead to ineffective erythropoiesis and hypoxic response in MDS with low expression of *STRAP* caused by *SRSF2* mutations.

The intracellular signalling in HSPCs must be regulated precisely to adjust systemic, cell-intrinsic and microenvironment factors for optimal generation of different haematopoietic-cell lineages. These signal interactions during haematopoiesis also are very diverse and differed depending on haematopoietic lineages. STRAP contains seven WD40 domains, which function as an indiscriminate adaptor and contribute to its regulatory functions mainly in several cellular signalling processes, including signal transduction, cell cycle progression, transcription regulation, RNA processing, and vesicular trafficking (Kashikar et al., 2010, Buess et al., 2004, Manoharan et al., 2018, Yuan et al., 2016, Reiner and Datta, 2011, Jin and Datta, 2014, Huh et al., 2016). Given that the role of STRAP as a signaling adaptor for protein interaction in a variety of pathways, it would be of value to investigate

further other dysregulated signaling pathways to characterise the role of downregulated *STRAP* in MDS pathogenesis.

This study sheds light on the underlying mechanism of impairments of erythropoiesis of MDS harbouring the common spliceosome mutations in erythroid progenitor cells. It is crucial for elucidating how the mutations hinder erythroid maturations and for identifying potential therapeutic treatments for anaemia of MDS.

4 Chapter Four - Global landscape of mRNA splicing alterations in MDS harbouring splicing factor mutations

4.1 Introduction

With the emergence of genome-wide sequencing technology, alternative splicing has become an field of interest in developmental biology and disease pathologies (Pan et al., 2008). mRNA splicing regulates the production of various isoforms from a single genomic region, dramatically increasing the diversity of the functional proteins (Blencowe, 2017). Transcriptomic studies have determined that alternative splicing is prevalent across eukaryotes. Particularly, approximately over 90% of human genes undergo alternative splicing (Wang et al., 2008).

Systematic transcriptomic analysis revealed that several cancers utilise dysregulated pre-mRNA splicing to initiate or maintain the tumorigenicity (Supek et al., 2014). In many cases, splicing-associated mutations cause altered gene expression (Karni et al., 2007) and impair the function of core and/or accessory spliceosome units (Yoshida et al., 2011, Papaemmanuil, 2014). Splicing factor genes are the most mutated genes in myeloid malignancies (Yoshida et al., 2011, Haferlach et al., 2014, Bonnal et al., 2020). The most frequently mutated genes encoding splicing factors are *SF3B1, SRSF2, U2AF1*, and *ZRSR2* in myeloid malignancies. The spliceosome proteins translated from these genes are all

involved in 3’ splice sites recognition during the pre-mRNA splicing process (Yoshida et al., 2011, Sperling et al., 2017).

Haematopoiesis refers to the process of commitment and differentiation, which result in the production of mature blood and immune cells from HSCs (Birbrair and Frenette, 2016). Alternative splicing programme plays a crucial role in the regulation of haematopoietic cell differentiation and development (Li et al., 2021, de Jong, 2019, Gao et al., 2018, Chen et al., 2014, Goldstein et al., 2017) such as erythropoiesis (Pimentel et al., 2014), granulopoiesis (Wong et al., 2013), megakaryopoiesis (Edwards et al., 2016), and monocyte-to-macrophage differentiation (Liu et al., 2018).

Initially, the research in the 'splicing code' (a set of splicing regulatory rules) was focused on *cis*-acting sequence features within the pre-mRNA (Barash et al., 2010) and a variety of regulatory *trans*-acting factors (Long and Caceres, 2009, Krecic and Swanson, 1999). Enhancer and silencer motifs within exon and intron locus promote or inhibit splicing, combined by RNA-binding proteins (RBP). More recently, many studies on splicing have taken up a macroscopic view, considering cell type specificity (Chepelev and Chen, 2013, Zhang et al., 2016), epigenetic regulations of chromatin (Schwartz et al., 2009, Dvinge, 2018, Clough et al., 2022, Yoshimi et al., 2019), transcription efficiency and RNA pol II progressivity (Carrillo Oesterreich et al., 2016), and structural features of mRNA (Soemedi et al., 2017) (**Figure 4-16**), to delineate the far-reaching nature of the mechanism (Hayashi et al., 2018, Wen et al., 2020, Gaiti et al., 2022, Wang et al., 2018a, Jenkins and Kielkopf, 2017, Kesarwani et al., 2017). Most of the elements affecting splicing outcomes are intricately linked, and therefore the final splicing fate will be determined by the combinatorial effect of multiple factors. Concomitantly, these studies have also highlighted the complexity of the ‘splicing code’ and the difficulties to fully decipher it.

In MDS with *U2AF1*^{S34F} mutation, Yip *et al.* have shown that abnormal splicing events occurs in a lineage-specific manner (Yip et al., 2017). One plausible explanation can be that tissue-specific alternative splicing takes place during haematopoiesis, which is regulated by lineage-specific and ubiquitously expressed RNA-biding proteins and other aforementioned factors (Wang et al., 2008, Hou et al., 2002, Yamamoto et al., 2009, Liu et al., 2011). Furthermore, a recent study has shown that *SRSF2* mutations globally affect splicing by targeting other splicing factors in Human Erythroid Leukaemia (HEL) cells with inducible expression of mutant *SRSF2*^{P95H} (Liang et al., 2018). Undoubtedly, these findings add another layer of complexity over the research on aberrantly spliced key target genes in MDS. However, these studies only focused on a few target genes or were conducted in an individual cell line. To better understand the abnormal splicing alterations in MDS, a comprehensive analysis in global context on splicing outcomes is pertinent.

In this chapter, we show that splicing factor mutations (*SF3B1* and *SRSF2*) reshape the mRNA splicing landscape in MDS. This study showed that the majority of abnormal splicing events take place in a cell-type specific manner in *SRSF2*mut and *SF3B1*mut MDS. Using GO, we identified that *SF3B1* and *SRSF2* mutations affects mRNA splicing process in all the bone marrow subpopulations analysed, causing aberrant splicing in other splicing factor genes. Our splicing analysis in the whole transcriptomes confirmed that alternative splicing is cell-type dependent and revealed that splicing factor (*SF3B1* and *SRSF2*) mutations alter patterns of whole mRNA splicing within bone marrow subpopulation, showing similar patterns of splicing changes depending on genotypes. This result suggests that there might be a potential mechanism underlying regularity of the altered splicing patterns among complex factors when mRNA splicing process is perturbed.

4.2 Results

4.2.1 Abnormal splicing events by splicing factor mutations in MDS occur in a lineage-specific manner

To characterise and identify global aberrant splicing events in bone marrow subpopulations of MDS patients harbouring splicing factor mutations (*SF3B1* and *SRSF2)* we analysed our large RNA-seq dataset of BM subpopulations in MDS patients including $CD34^+$, monocytic, granulocytic, and erythroid precursor cells from 84 patients with MDS and 8 healthy control individuals (Pellagatti et al., 2018). In brief, 28 cases of *SF3B1* mutant MDS, 8 cases of *SRSF2* mutant MDS and 8 healthy controls accounted for the data set for $CD34^+$ populations (Pellagatti et al., 2018). The data of other subpopulations including granulocytic, monocytic, and erythroid precursors were from the BM samples of 11 patients with MDS (7 *SF3B1* mutant and 4 *SRSF2* mutant) compared with 5 healthy control individuals (Pellagatti et al., 2018). To determine alternative splicing events we used the rMATS bioinformatics pipeline (Shen et al., 2014, Dolatshad et al., 2016, Yip et al., 2017) classifying them as alternative 3' splice site (A3SS) usage, alternative 5' splice site (A5SS) usage, exon skipping (SE), mutually exclusive exons (MXE), or retained introns (RI).

Pertaining to the *SF3B1* mutation cases (**Figure 4-1**. A and **Table 4-2**), we identified a total of 1034 alternative splicing events which are significantly different compared to healthy controls, including the five alternative splicing types in the $CD34^+$ cell populations. SE was the most common alternative splicing event (529), accounting for 51.16% of the total alternative splicing events in the $CD34^+$ cell population. RI was the second most prevalent alternative splicing event (31.04%). A3SS and A5SS account for much lower proportion of the alternative splicing events recorded (10.25% and 5.32% respectively),

while MXE is a rare event (2.22%). As for the alternative splicing events in the monocytic population, a total of 483 events were identified. SE was the most common alternative splicing event (250), reaching a percentage 51.76%, and RI was shown as the second most frequent alternative splicing event (26.09%). Lastly, the alternative splicing events in granulocytic and erythrocytic cells where a total of 397 and 172 of events were identified respectively, SE was once again the most common alternative splicing event (50.88% and 52.91%, respectively) (**Figure 4-15**. A).

On the other hand, a total of 998 alternative splicing events which are significantly different compared to healthy control, in the case of $CD34^+$, were recorded for *SRSF2* mutation cases (**Figure 4-2**. A and **Table 4-3**). SE was identified as the most common alternative splicing event (634), occurring at the highest frequency of 63.53% of the total alternative splicing events of $CD34^+$ cells. RI was ranked second most prevalent (20.44%), followed by A3SS and A5SS accounting for 8.22% and 5.01% respectively, with MXE ranked last (2.81%). As for granulocytic cell populations, a total of 723 alternative splicing events were identified. SE was the most common alternative splicing event (395), reaching a percentage 54.63%, and RI was indicated as the second most prevalent alternative splicing event (21.99%). Regarding the alternative splicing events in monocytic and erythrocytic cells, 475 and 325 events were identified, respectively and SE in both subpopulations were ranked as the most common alternative splicing event (62.11%,65.23%, respectively) (**Figure 4-15**. B).

A **SF3B1mut**

Rank	AS events	Schema of structure	CD34+	MON	GRA	ERY
1	SE		529	250	202	91
2	RI		321	126	107	52
3	A3SS		106	62	54	17
4	A5SS		55	26	23	7
5	MXE		23	19	11	5
	Total		1034	483	397	172

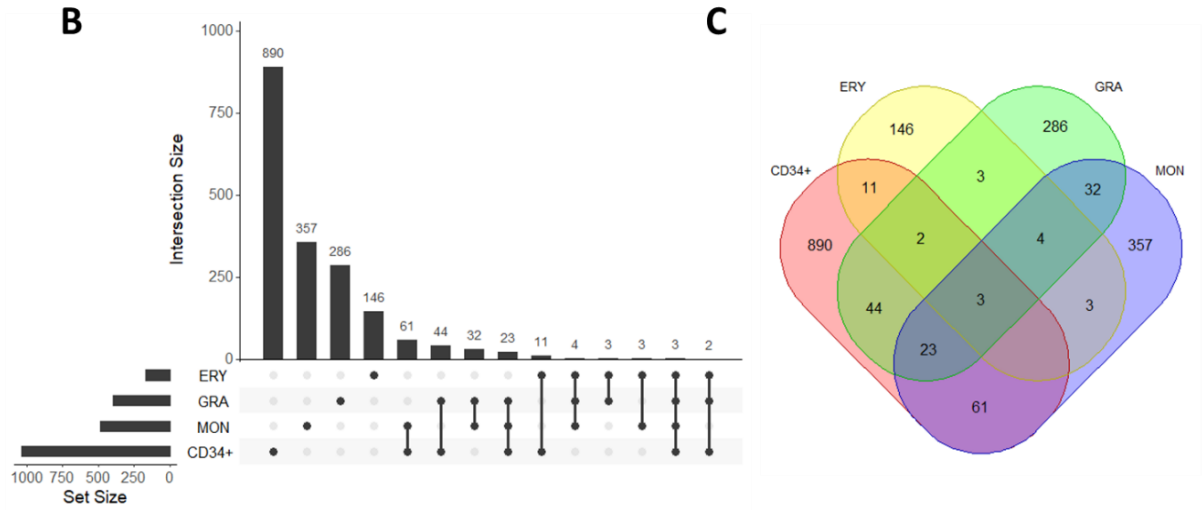


Figure 4-1. Aberrant splicing events in BM subpopulations of *SF3B1* mutations.

Number of splicing events detected and altered across BM subpopulations for comparisons (A) between healthy control and *SF3B1*-mutant MDS patients. The UpSet plots and Venn diagram showing the overlap of alternative splicing events in $CD34^+$ cells and in monocyte (MON), granulocyte (GRA) and erythroid (ERY) precursor cells isolated from MDS patient samples with (B, C) *SF3B1* mutations compared to healthy control. These figures were made with the data by using rMATS bioinformatics pipeline ($|\Delta PSI|>0.1$ and $FDR<0.05$).

A **SRSF2mut**

Rank	AS events	Schema of structure	CD34+	MON	GRA	ERY
1	SE		634	295	395	212
2	RI		204	103	159	69
3	A3SS		82	37	66	18
4	A5SS		50	21	36	19
5	MXE		28	19	67	7
	Total		998	475	723	325

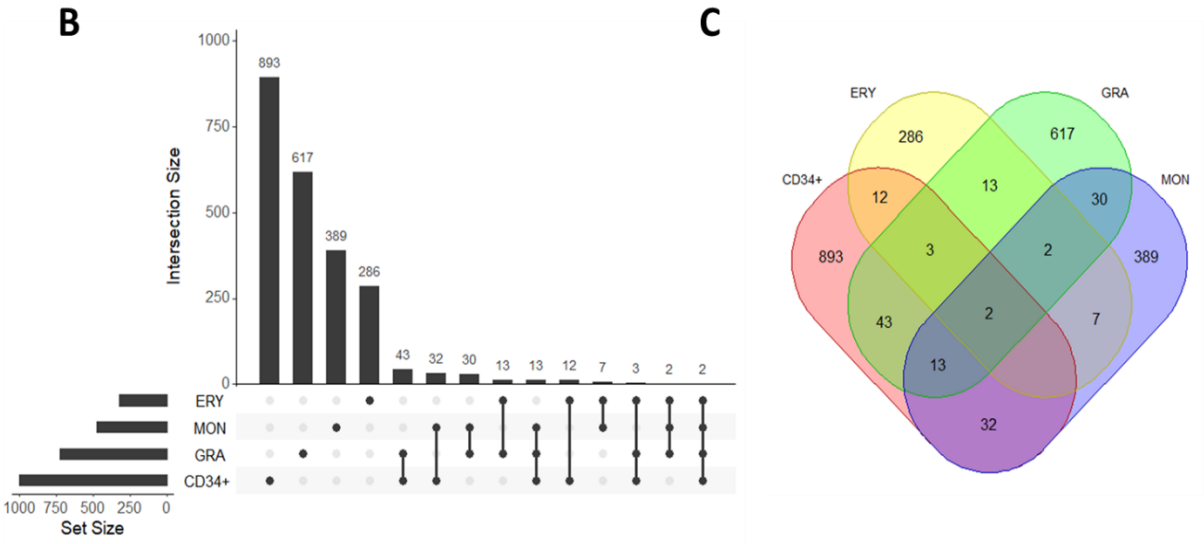


Figure 4-2. Aberrant splicing events in BM subpopulations of *SRSF2* mutations.

Number of splicing events detected and altered across BM subpopulations for comparisons (A) between healthy control and *SRSF2*-mutant MDS patients. The UpSet plots and Venn diagram showing the overlap of alternative splicing events in $CD34^+$ cells and in monocyte (MON), granulocyte (GRA) and erythroid (ERY) precursor cells isolated from MDS patient samples with (B, C) *SRSF2* mutations compared to healthy control. These figures were made with the data by using rMATS bioinformatics pipeline ($|\Delta PSI|>0.1$ and FDR<0.05).

Table 4-1. **The number of alternative splicing genes in the list of alternative splicing events detected by rMATS pipeline** (|ΔPSI|>0.1 and FDR<0.05)

Experiments	BM subpopulations			
	CD34$^+$	**MON**	**GRA**	**ERY**
Healthy control vs SF3B1mut	779/1034	402/483	323/397	145/172
Healthy control vs SRSF2mut	782/998	377/475	542/723	265/325

The results showed that the frequency of RI and A3SS events in *SF3B1* mutations cases are more commonly observed than ones in *SRSF2* mutation cases. Many studies including ours have shown that *SF3B1* mutations in myeloid malignancies are associated with altered selection of 3'splice sites of transcripts (DeBoever et al., 2015, Dolatshad et al., 2015, Dolatshad et al., 2016, Kesarwani et al., 2017, Obeng et al., 2016). In addition, the results showed that *SRSF2* mutations result in more common frequency of SE events, compared to *SF3B1* mutations. During splicing, SRSF2 has a RRM and thus promotes exon recognition by binding to ESE motifs in pre-mRNA through its RBD, allowing exon inclusion or exon skipping. Several groups have demonstrated that SRSF2 protein is responsible for either activating exon inclusion or exon skipping (Pandit et al., 2013, Moon et al., 2019).

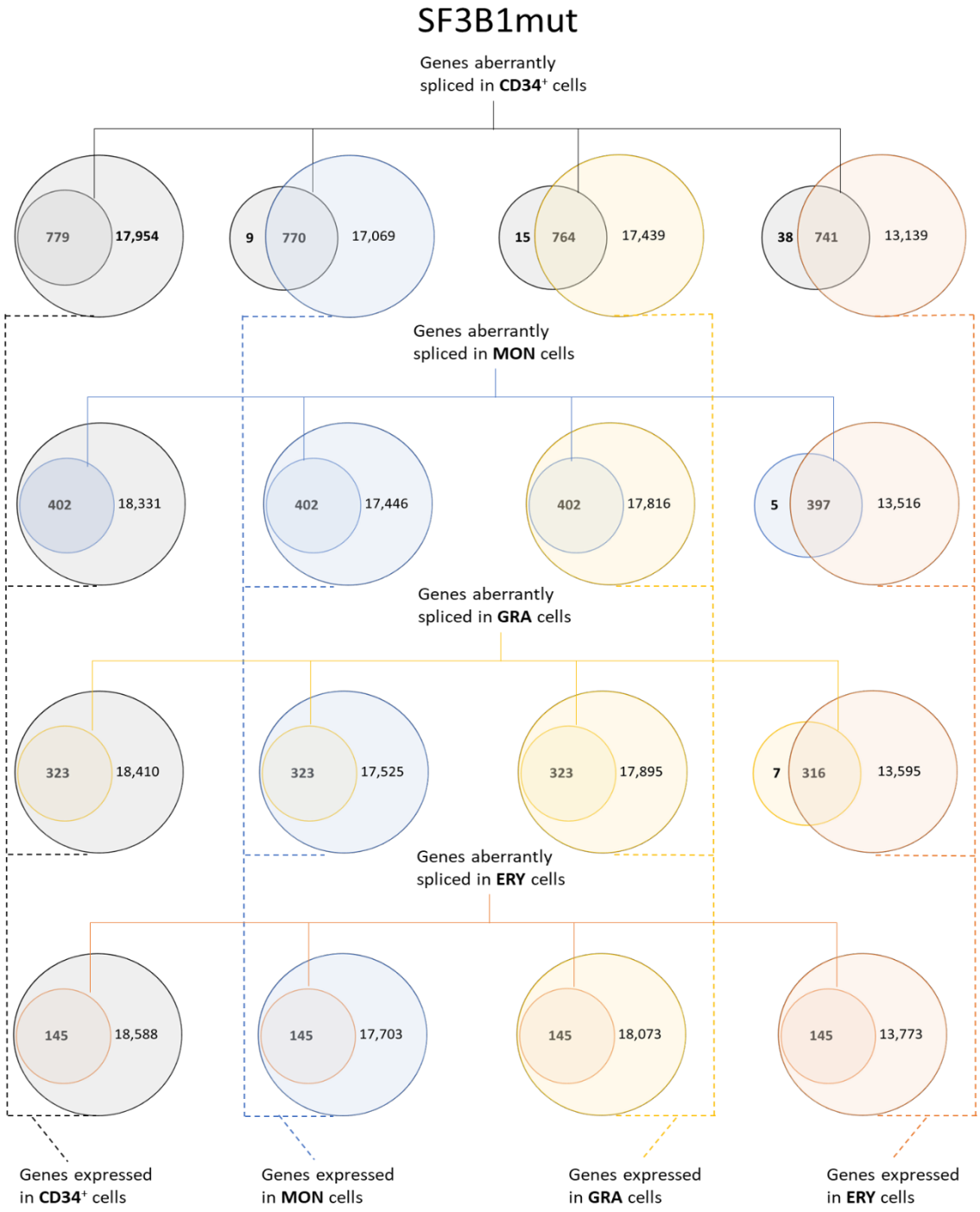


Figure 4-3. Overlap between genes showing aberrant splicing in the BM subpopulations and genes expressed in the BM subpopulations of MDS with *SF3B1* mutation.

The majority of genes aberrantly spliced in the BM subpopulations were also expressed in the other BM subpopulations.

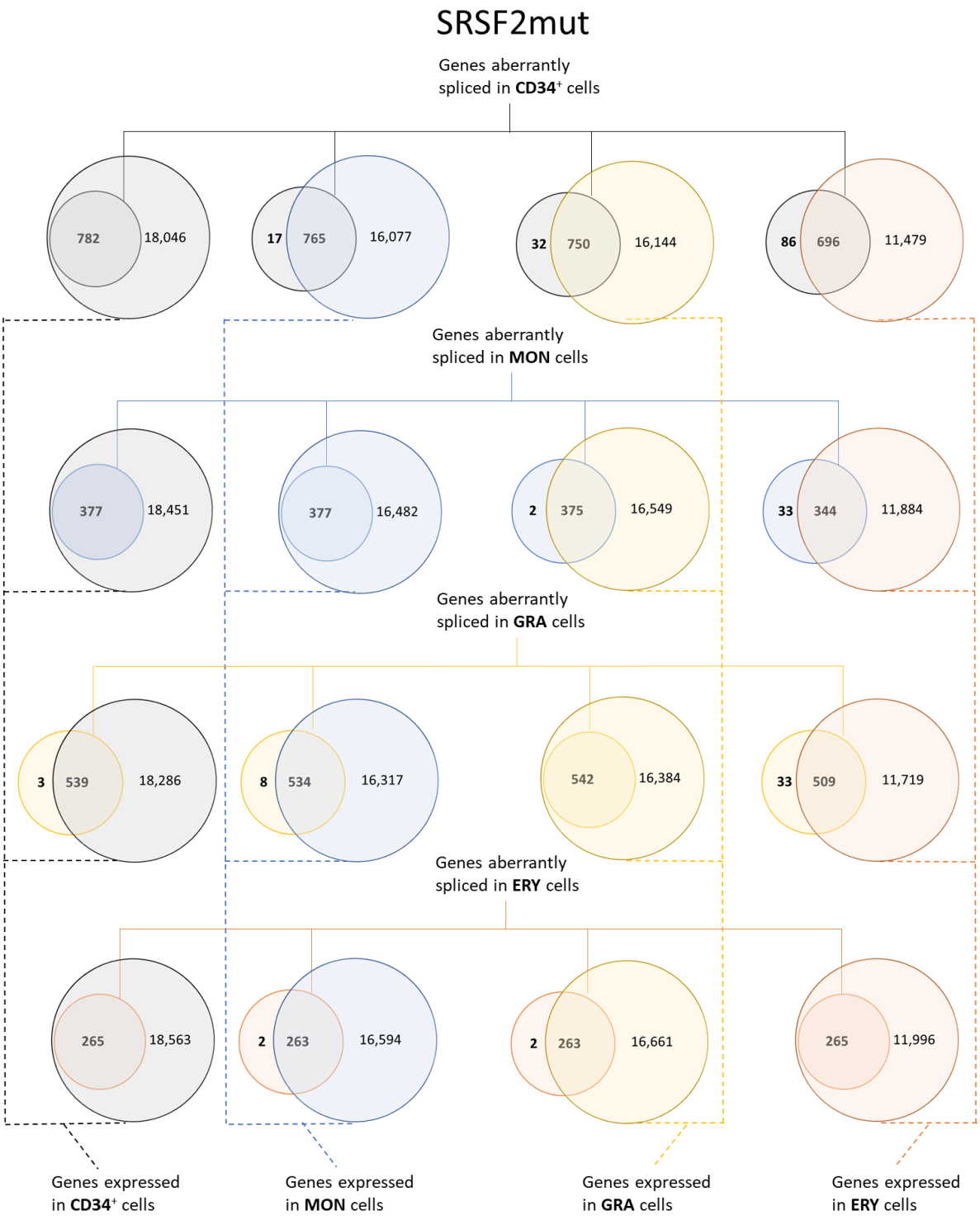


Figure 4-4. Overlap between genes showing aberrant splicing in the BM subpopulations and genes expressed in the BM subpopulations of MDS with *SRSF2* mutation.

The majority of genes aberrantly spliced in the BM subpopulations were also expressed in the other BM subpopulations.

To identify key target genes aberrantly spliced across the BM subpopulations, we investigated the overlaps of alternative splicing events in each subpopulation of *SF3B1* mutation cases and *SRSF2* mutation cases (Pellagatti et al., 2018). Surprisingly, only 3 common events (*SEPT2*, *DDX24*, and *DYNLL1*) and 2 common events (*RIMKLB* and *SNHG6*) were aberrantly spliced in common to all the BM subpopulations for *SF3B1* mutations or *SRSF2* mutations, respectively (**Figure 4-1**. B and **Figure 4-2**. C). Interestingly, most abnormal splicing events occur in a cell-type specific manner in MDS with *SRSF2* and *SF3B1* mutation. In *SF3B1* mutation cases, we identified that a considerable number of cell-type specific events occurred in the 4 BM subpopulations; $CD34^+$ (890; 86.07%), monocytic (357;73.91%), granulocytic (286;72.04%), erythrocytic (146;84.88%) cells. As for *SRSF2* mutation, the cell-type alternative splicing events in all the BM subpopulations occurred at a high frequency as well; $CD34^+$(893;89.48%), monocytic (389;81.89%), granulocytic (617;85.34%), erythrocytic (286;88.00%) cells. To determine whether this phenomenon is due to cell-type specific transcriptomics, we investigated the overlap between genes showing aberrant splicing in the BM subpopulations and genes expressed in the subpopulations of MDS with *SF3B1* and *SRSF2* mutation (**Figure 4-3 and Figure 4-4**) The results showed that large majority of genes aberrantly spliced in the BM subpopulations were also expressed in the other BM subpopulations though at different levels. Therefore, the results suggest that the cell-type specific manner of abnormal splicing events is not due to the differences in the types of genes expressed between the BM subpopulations.

4.2.2 *SFSF2* and *SF3B1* mutations cause aberrant splicing of transcripts associated with RNA splicing process in MDS

We next investigated which biological functions the aberrant spliced transcripts were involved in to better understand the biology of *SF3B1* and *SRSF2* mutations in MDS. We applied functional annotation tools, Gene ontology (GO) (Ashburner et al., 2000), to the aberrantly spliced genes (**Table 4-1**) identified across the BM subpopulations. The GO analysis revealed that biological processes enriched for each subpopulations in *SF3B1* mutant MDS include "double-strand break repair (GO:0006302)", "response to oxidative stress (GO:0006979)", "positive regulation of chromosome separation (GO:1905820)" and "positive regulation of translation (GO:0045727)" for $CD34^+$, monocytic, granulocytic and erythrocytic subpopulations, respectively (**Figure 4-5~8** and **Table 8-1~4**). For *SRSF2* mutant MDS, biological processes were identified in each subpopulation (**Figure 4-9~12** and **Table 8-5~8**) namely "DNA recombination (GO:0006310)", "positive regulation of proteolysis (GO:0045862)", "proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)" and "energy derivation by oxidation of organic compounds (GO:0015980)" for $CD34^+$, monocytic, granulocytic and erythrocytic subpopulations, respectively. Interestingly, the biological process commonly shared in all the BM subpopulation are linked with mRNA processing such as "RNA modification (GO:0009451)", regulation of RNA stability (GO:0043487)", regulation of mRNA metabolic process (GO:1903311)" and "RNA splicing (GO:0008380)" (**Figure 4-13**). Collectively, these results suggested that whilst mutated splicing factor (*SF3B1* or *SRSR2*) in MDS largely affects splicing of different genes in different bone marrow cell populations, these genes converge in overlapping gene ontology themes, "RNA splicing (GO:0008380)", suggesting a common perturbed cellular process.

Several studies have shown that RBPs play an important role in execution of sophisticated splicing process (Witten and Ule, 2011, Yang et al., 2016, Li et al., 2017, Mukherjee et al., 2018, Kosti et al., 2012, Sternburg and Karginov, 2020). In addition, recent studies have suggested that these various abnormal isoforms of splicing factor genes such as heterogeneous nuclear ribonucleoproteins (hnRNPs) and SR proteins, are able to distort the overall splicing process by hindering RBP-RBP and RBP-RNA interactions (Cerasuolo et al., 2020, Liang et al., 2018, Quattrone and Dassi, 2019, Castello et al., 2012, Geuens et al., 2016, Gautrey et al., 2015). Taken together, these findings demonstrate that *SF3B1* and *SRSF2* mutations affect mRNA splicing process in BM cell populations of MDS causing aberrant splicing in other splicing factor genes, indicating an exacerbation of aberrant splicing in MDS.

SF3B1mut / CD34^{+}

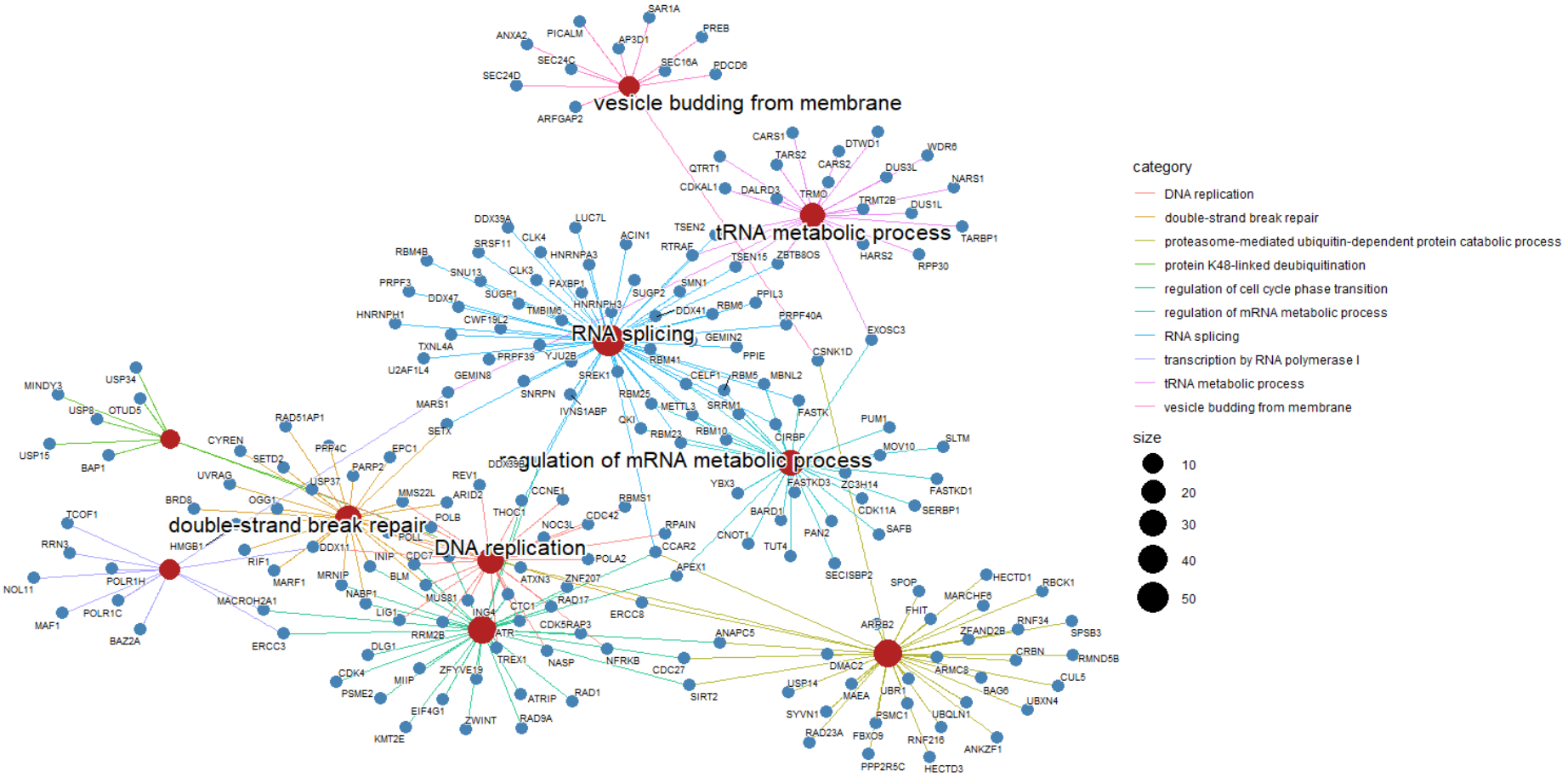


Figure 4-5. Gene Ontology analysis for aberrantly spliced genes in CD34^{+} cell population of MDS with *SF3B1* mutations.
The Gene-Concept Network showing the linkages of genes and biological concepts as networks

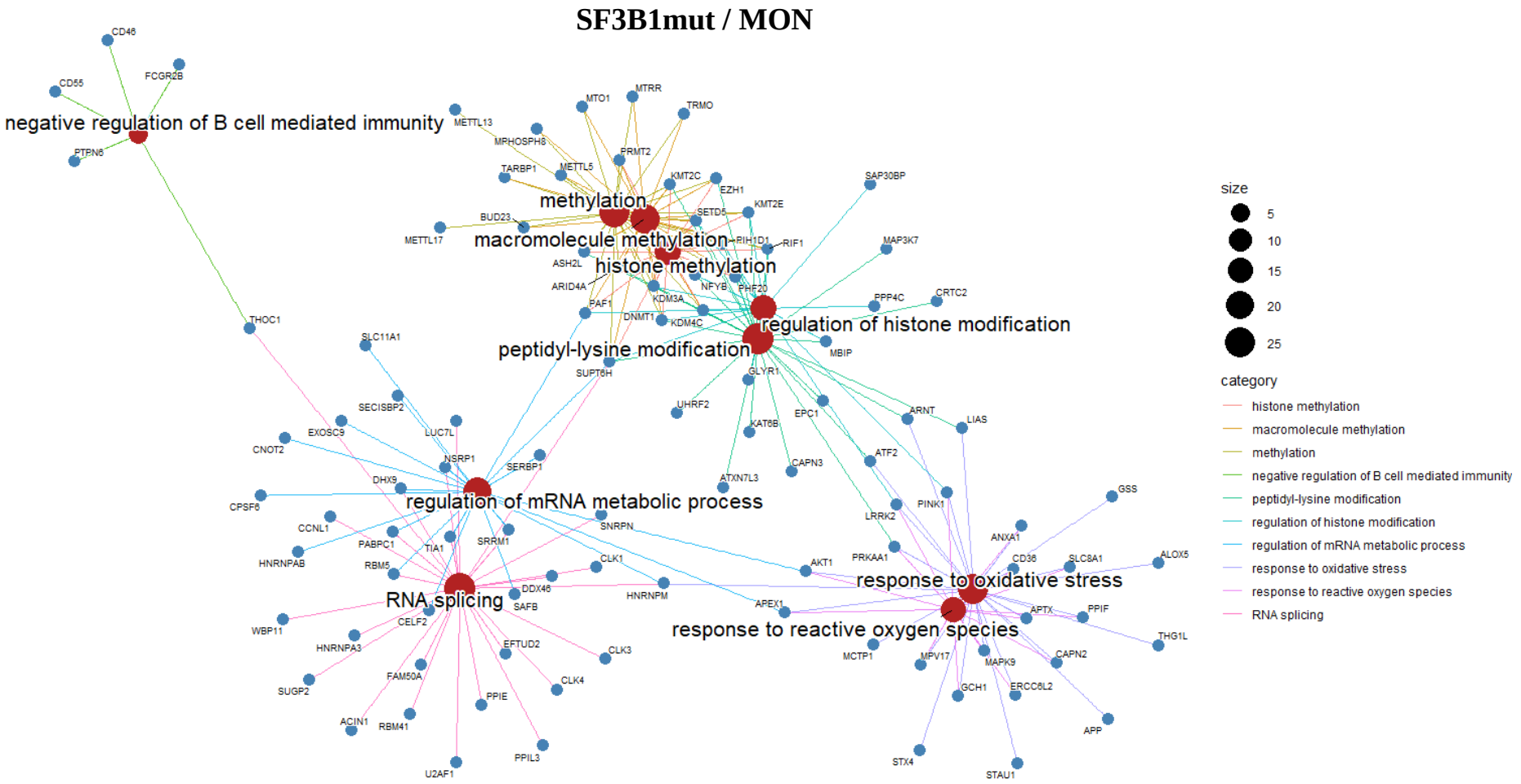


Figure 4-6. Gene Ontology analysis for aberrantly spliced genes in monocytic cell population of MDS with *SF3B1* mutations.
The Gene-Concept Network showing the linkages of genes and biological concepts as networks

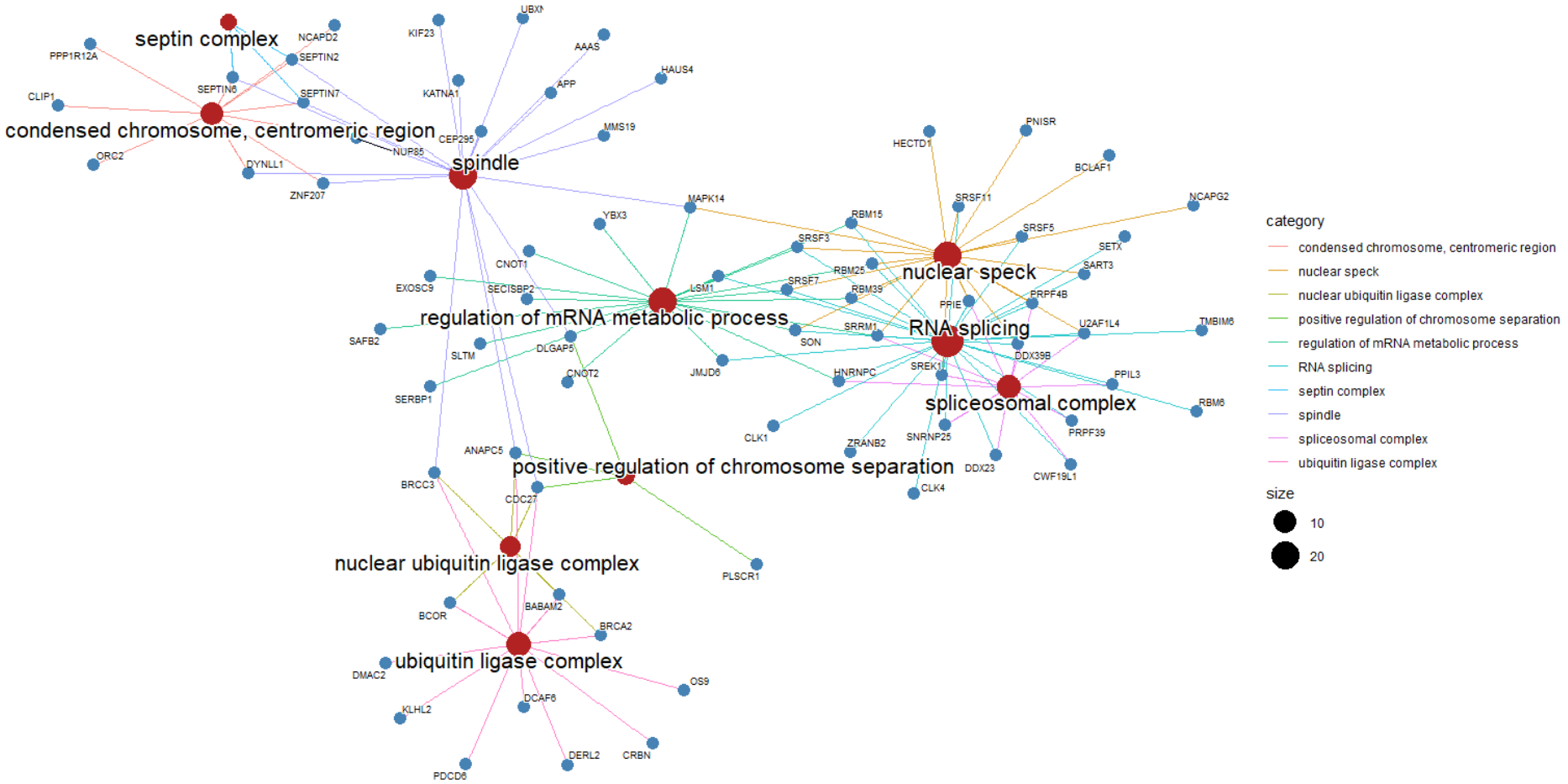


Figure 4-7. Gene Ontology analysis for aberrantly spliced genes in erythrocytic cell population of MDS with *SF3B1* mutations.

The Gene-Concept Network showing the linkages of genes and biological concepts as networks

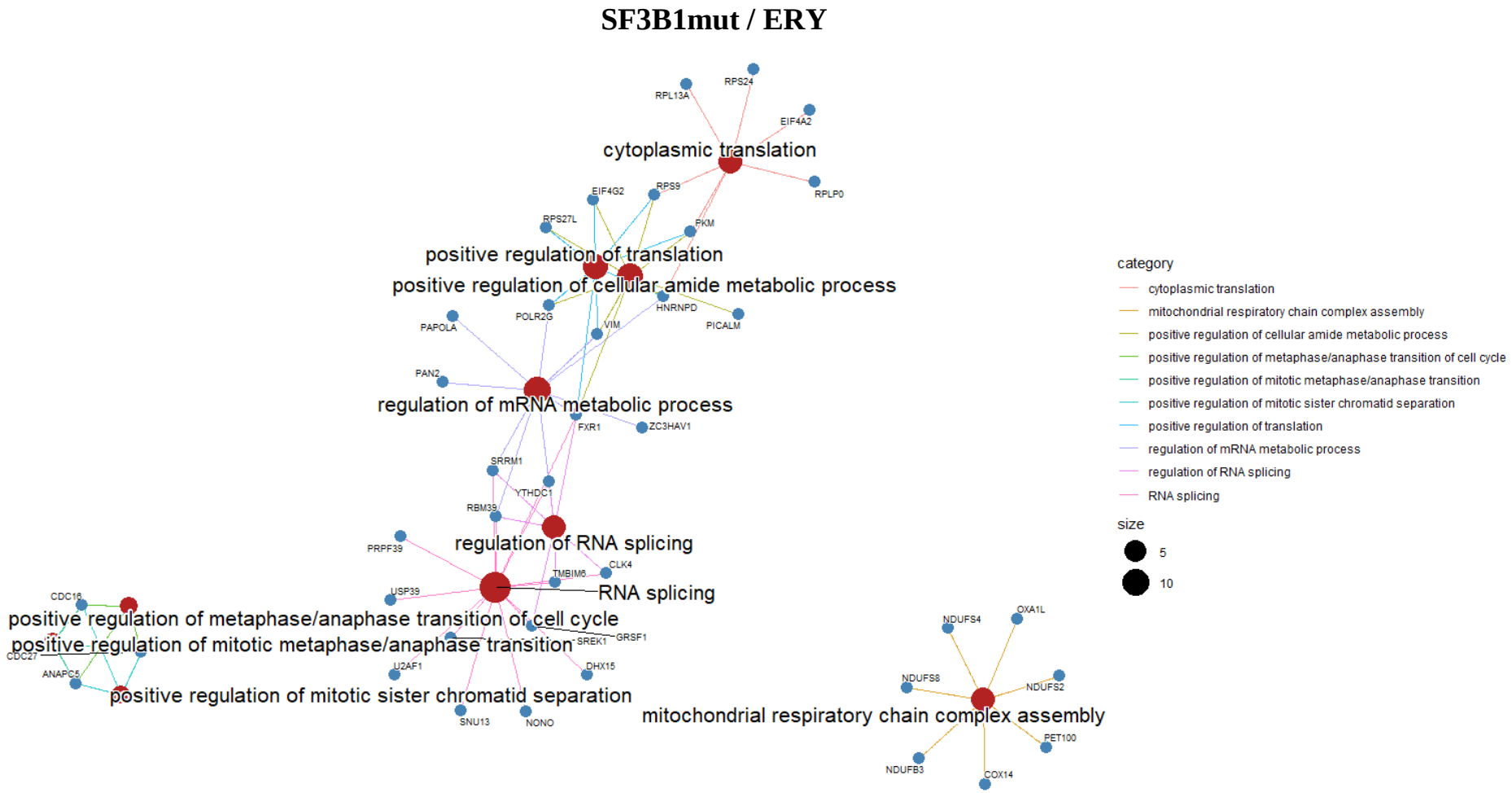


Figure 4-8. Gene Ontology analysis for aberrantly spliced genes in erythrocytic cell population of MDS with *SF3B1* mutations.
The Gene-Concept Network showing the linkages of genes and biological concepts as networks

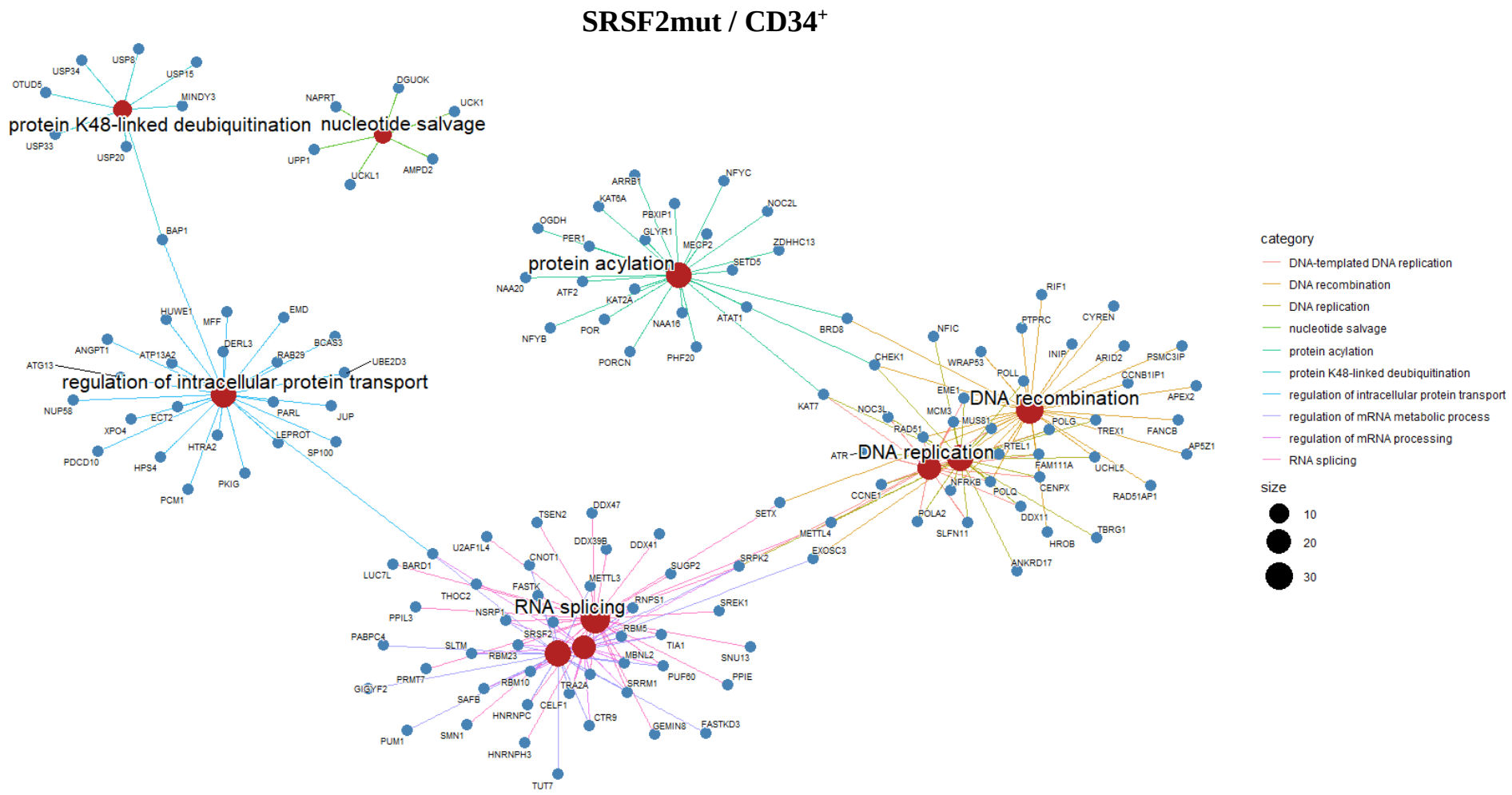


Figure 4-9. Gene Ontology analysis for aberrantly spliced genes in CD34$^+$ cell population of MDS with *SRSF2* mutations.
The Gene-Concept Network showing the linkages of genes and biological concepts as networks

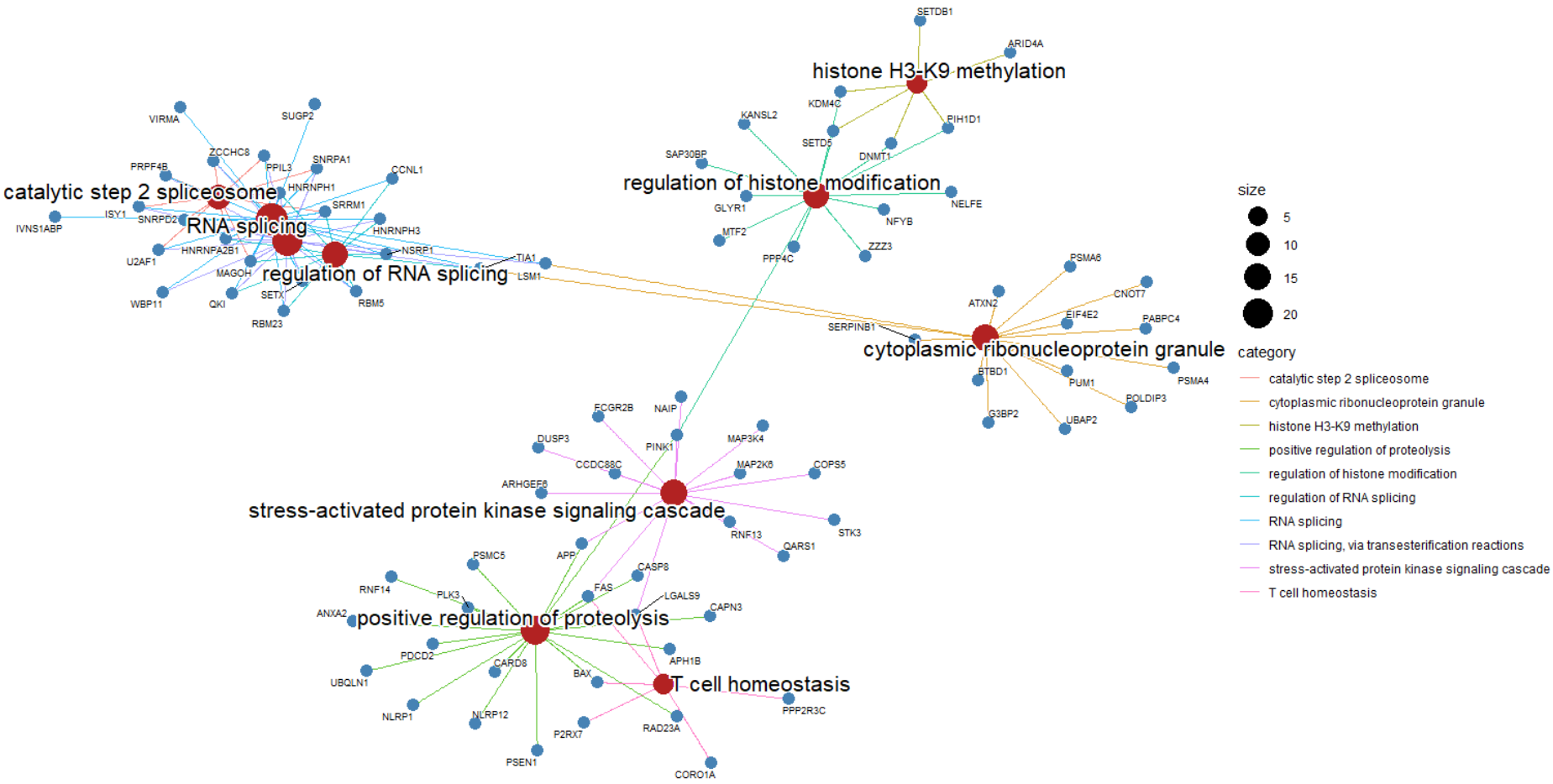


Figure 4-10. Gene Ontology analysis for aberrantly spliced genes in monocytic cell population of MDS with *SRSF2* mutations.
The Gene-Concept Network showing the linkages of genes and biological concepts as networks

SRSF2mut / GRA

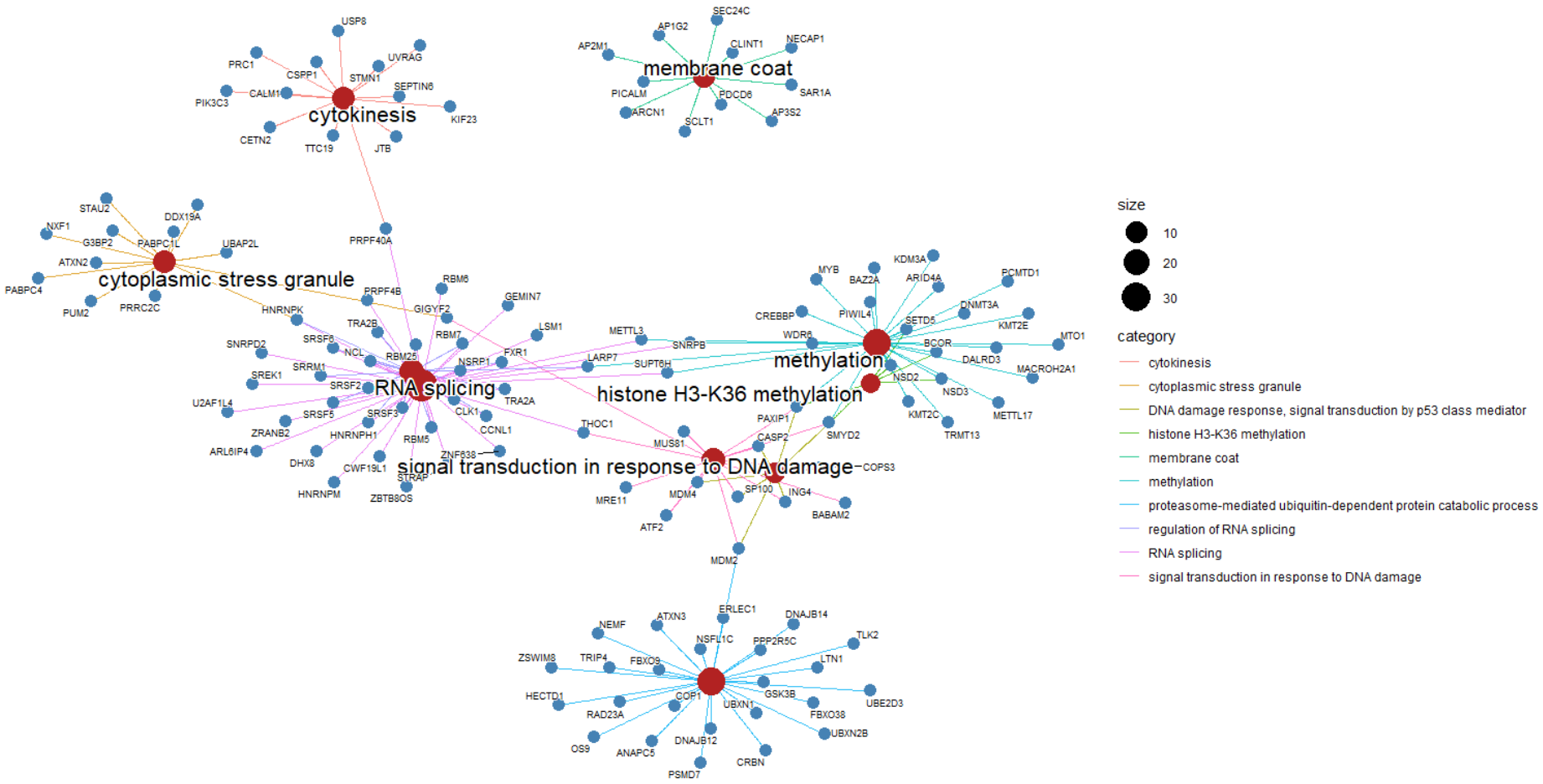


Figure 4-11. Gene Ontology analysis for aberrantly spliced genes in granulocytic cell population of MDS with *SRSF2* mutations.
The Gene-Concept Network showing the linkages of genes and biological concepts as networks

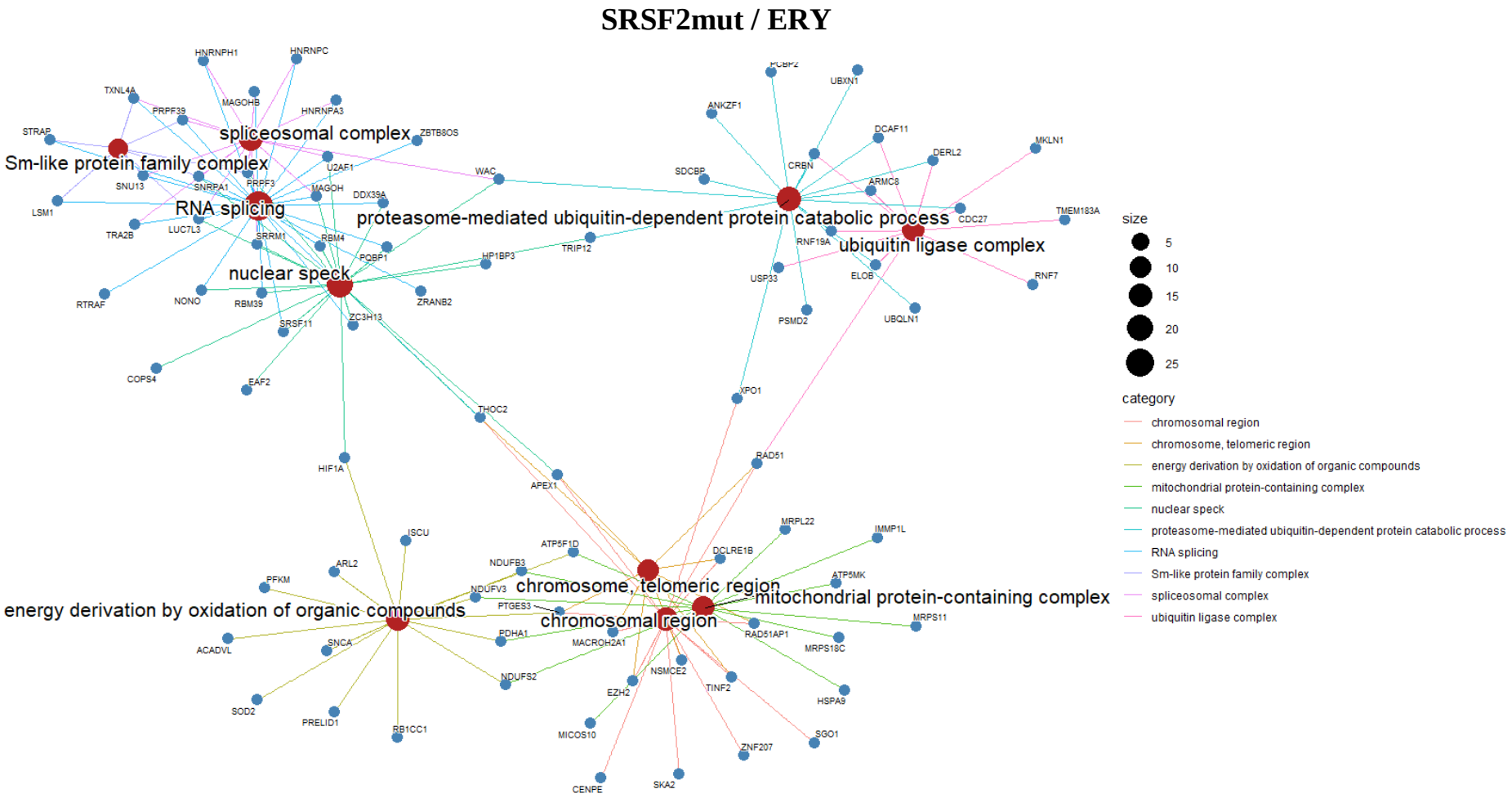


Figure 4-12. Gene Ontology analysis for aberrantly spliced genes in erythrocytic cell population of MDS with *SRSF2* mutations.
The Gene-Concept Network showing the linkages of genes and biological concepts as networks

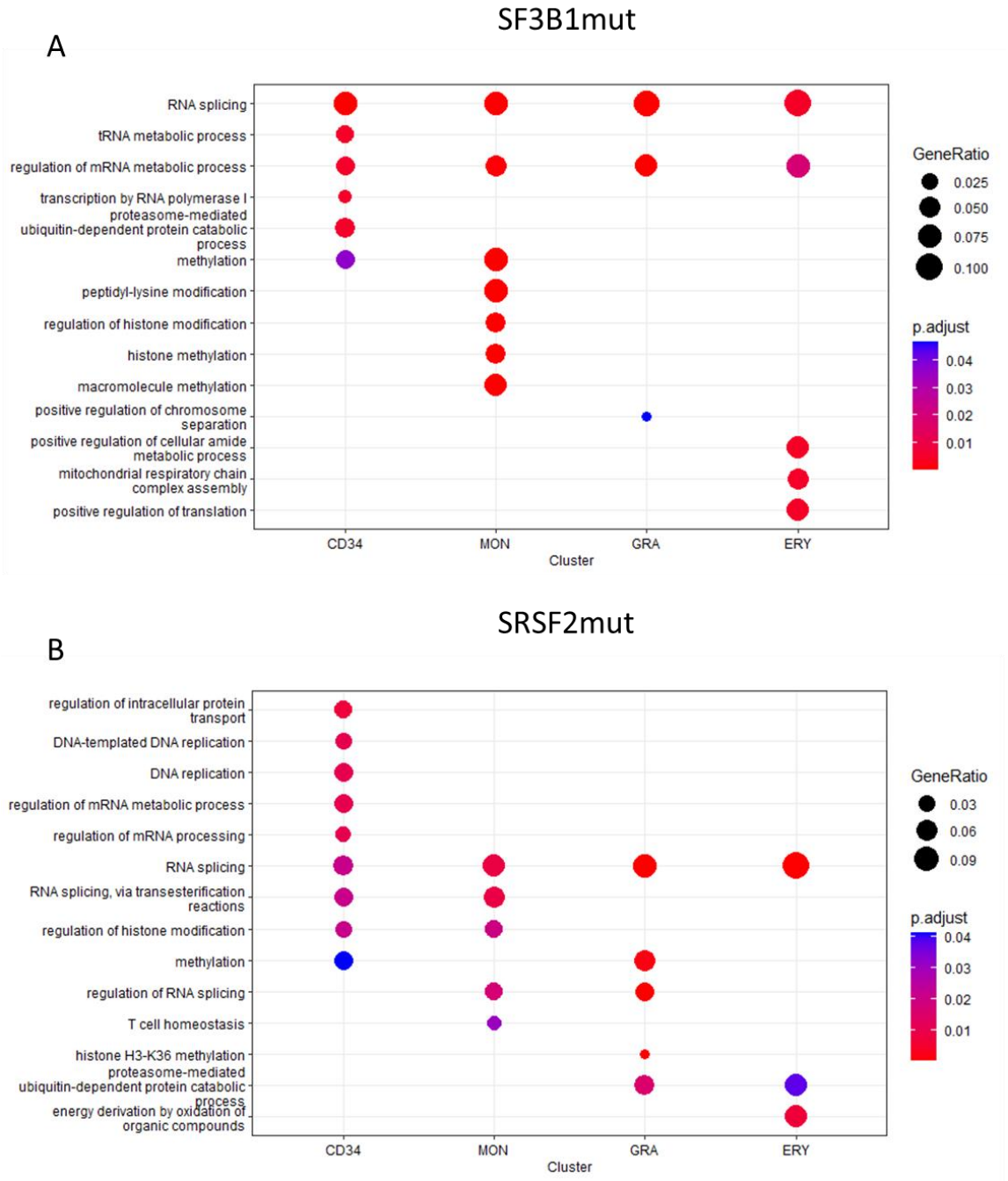


Figure 4-13. Biological theme comparison of GO enrichment analysis across the BM subpopulations in (A) *SF3B1* and (B) *SRSF2* mutant MDS.

Significant enrichment for "RNA splicing" was observed in the 4 BM subpopulation. To reduce redundant GO terms, generated by overlap between parent and child terms, "simplify" method of clusterProfiler was applied.

4.2.3 Global patterns of the aberrant splicing associated with each splicing factor mutation in all four subpopulations of MDS patients

The analysis in all the bone marrow populations suggested that mutations in *SF3B1* and *SRSF2* led to alteration in other splicing factors in the context of the splicing process. This indicates that a great number of abnormal splicing events may be implicated, either directly or indirectly, due to mutated *SF3B1* or *SRSF2*, resulting in dysregulation of splicing process at a global level in MDS. Thus, we sought to explore how global pattern of abnormal splicing events are altered by splicing factor mutations in each subpopulation using maser and MARVEL, which are bioinformatic tools used to analyse alteration of splicing based on the output of rMATS and PSI value. PSI values generated by rMATS outputs across the bone marrow populations, were integrated and subjected to dimension reduction analysis. Individual samples including *SF3B1* mutant and *SRSF2* mutant cases, and healthy controls in the 4 bone marrow populations were inspected using principal component analysis (PCA) based on PSI values (**Figure 4-14**). In concordance with previous reports (Zhang et al., 2016, Gao et al., 2018, Mazin et al., 2021, Baralle and Giudice, 2017), samples were strongly clustered by cell types regardless of mutational status (**Figure 4-14** A and B). Interestingly, this result suggests that each cell population in human bone marrow has unique signatures of alternative splicing pattern and that the alteration of mRNA splicing caused by *SF3B1* and *SRSF2* mutations do not changed cell-type intrinsic patterns of splicing. To further uncover patterns of splicing alteration, affected by *SF3B1* and *SRSF2* mutations within each bone marrow population, we re-performed PCA using the data in each cell type (**Figure 4-14** C-F). Notably, this analysis revealed that samples harbouring *SF3B1* mutation and *SRSF2* mutation in bone marrow

cell populations of MDS were clustered distinctly based on genotype, indicating that there are similarities of whole splicing alteration patterns.

Next, we analysed distribution of the five splicing event types in all the bone marrow populations. Interestingly, consistent distribution of the aberrant splicing types associated with each splicing factor mutations in all four subpopulations of MDS patients were observed (**Figure 4-15**. A-B, **Table 4-2** and **Table 4-3**). Inclusion of the cassette exon is regulated by *cis*-elements including exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE), and intronic splicing silencer (ISS) and these sequence sites are recognized and activated by *trans*-acting splicing factors such as SR proteins, hnRNPs, and other RPBs to promote or inhibit inclusion of exons (Wang and Burge, 2008, Cerasuolo et al., 2020, Kim et al., 2018, Baralle and Baralle, 2018). Many abnormal spliced genes in the list of "RNA splicing (GO:0008380)" of GO term, observed in our MDS samples are associated with splicing enhancers and silencers. For instance, we identified aberrantly spliced genes, encoding proteins interacting with the splicing enhancers or silencer sites, promoting skipping or inclusion of exons, such as the spliceosomal small nuclear ribonucleoproteins (*SNRPA1*, *SNRPB, SNRPD2, SNRPN* and *SNRNP25*), RNA-binding motif proteins (*RBM4, RBM4B, RBM5, RBM6, RBM7, RBM10, RBM15, RBM23, RBM25, RBM39* and *RBM41*), Serine/Arginine-rich (SR) splicing factors (*SRSF3, SRSF6, SRSF5, SRSF7*, and *SRSF11*), heterogeneous nuclear ribonucleoproteins (*HNRNPA2B1, HNRNPA3, HNRNPC, HNRNPH1, HNRNPH3, HNRNPK* and *HNRNPM*) (**Table 8-1~8**). This may explain in part why the main event type in each sample group is SE event. To evaluate further the effect of these genes on SE events, we analysed the SE event separately. There are two types of SE events which we can identify using rMATS: exon inclusion and exon exclusion (**Figure 4-15**. C). Interestingly, the bar chart showed

that the ratios of the 2 types of SE events, associated with each splicing factor mutations, are consistent in all four cell subpopulations of MDS patients (**Figure 4-15**. D and E).

In summary, our splicing analysis in the whole transcriptomes confirmed that alternative splicing is cell-type dependent and revealed that splicing factor (*SF3B1* and *SRSF2*) mutations altered patterns of whole mRNA splicing within each BM subpopulation, showing similar patterns of alterations depending on genotypes. This indicates there might be a potential mechanism underlying regularity of the altered splicing patterns among complex factors when mRNA splicing process is perturbed.

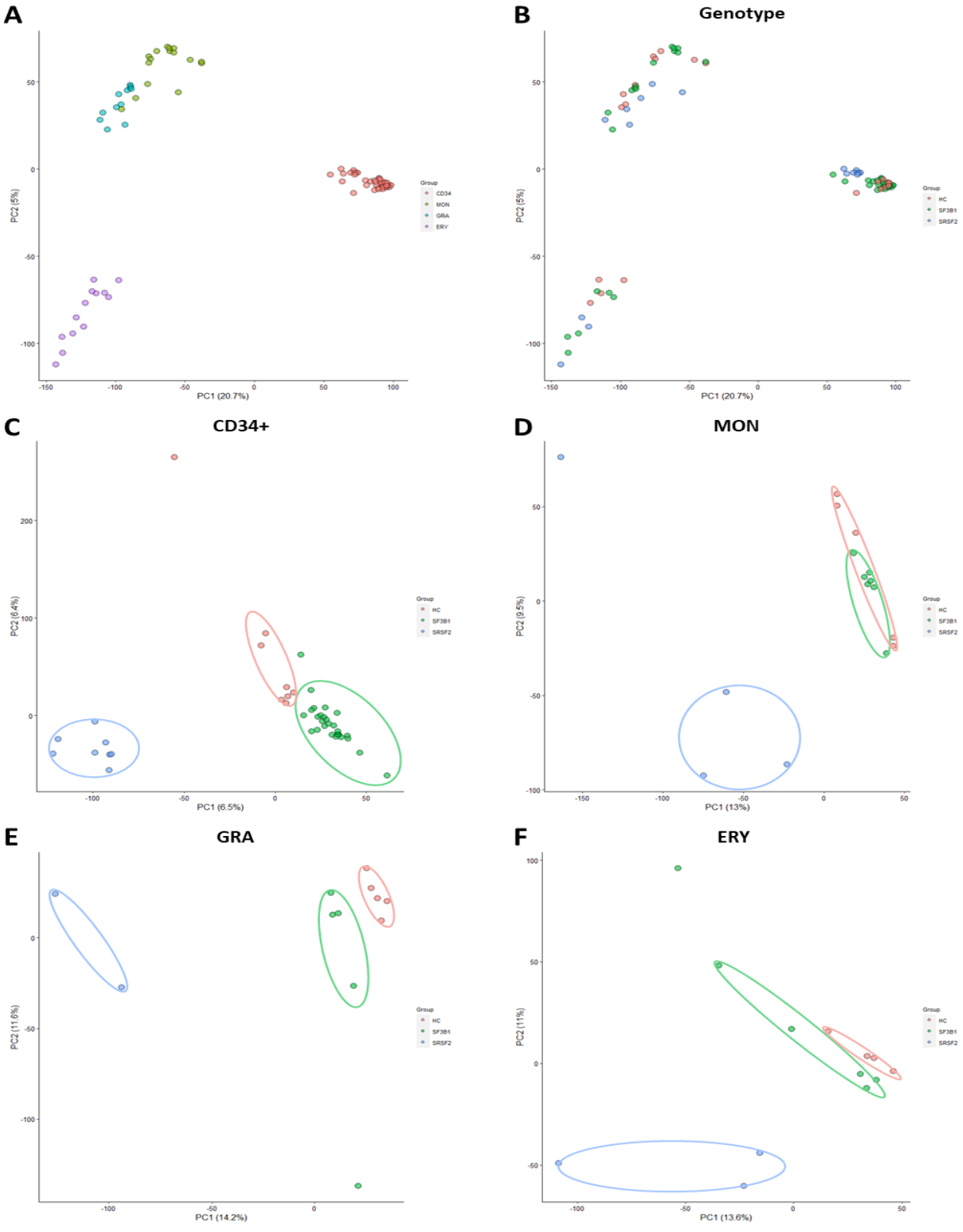

A
PC2 (5%)
PC1 (20.7%)
Group
CD34
MON
GRA
ERY
B
Genotype
PC2 (5%)
PC1 (20.7%)
Group
HC
SF3B1
SRSF2
C
CD34+
PC2 (6.4%)
PC1 (6.5%)
Group
HC
SF3B1
SRSF2
D
MON
PC2 (9.5%)
PC1 (13%)
Group
HC
SF3B1
SRSF2
E
GRA
PC2 (11.6%)
PC1 (14.2%)
Group
HC
SF3B1
SRSF2
F
ERY
PC2 (11%)
PC1 (13.6%)
Group
HC
SF3B1
SRSF2

Figure 4-14. Principal component analysis (PCA) of splicing events in bone marrow cell populations of *SF3B1* mutant and *SRSF2* mutant MDS.

PSI values are used to perform dimension reduction analysis. Individual samples are plotted on 2D space (PC1 and PC2) and labelled by (A) cell types and (B) genotypes. The PCA plot showing strong clustering by cell type but not by genotype. Within each cell populations (C-F), PCAs representing distinct clusters depending on genotype indicating their similarities of splicing alterations. All PSI values were extracted from rMATS outputs and PCA were performed using MARVEL.

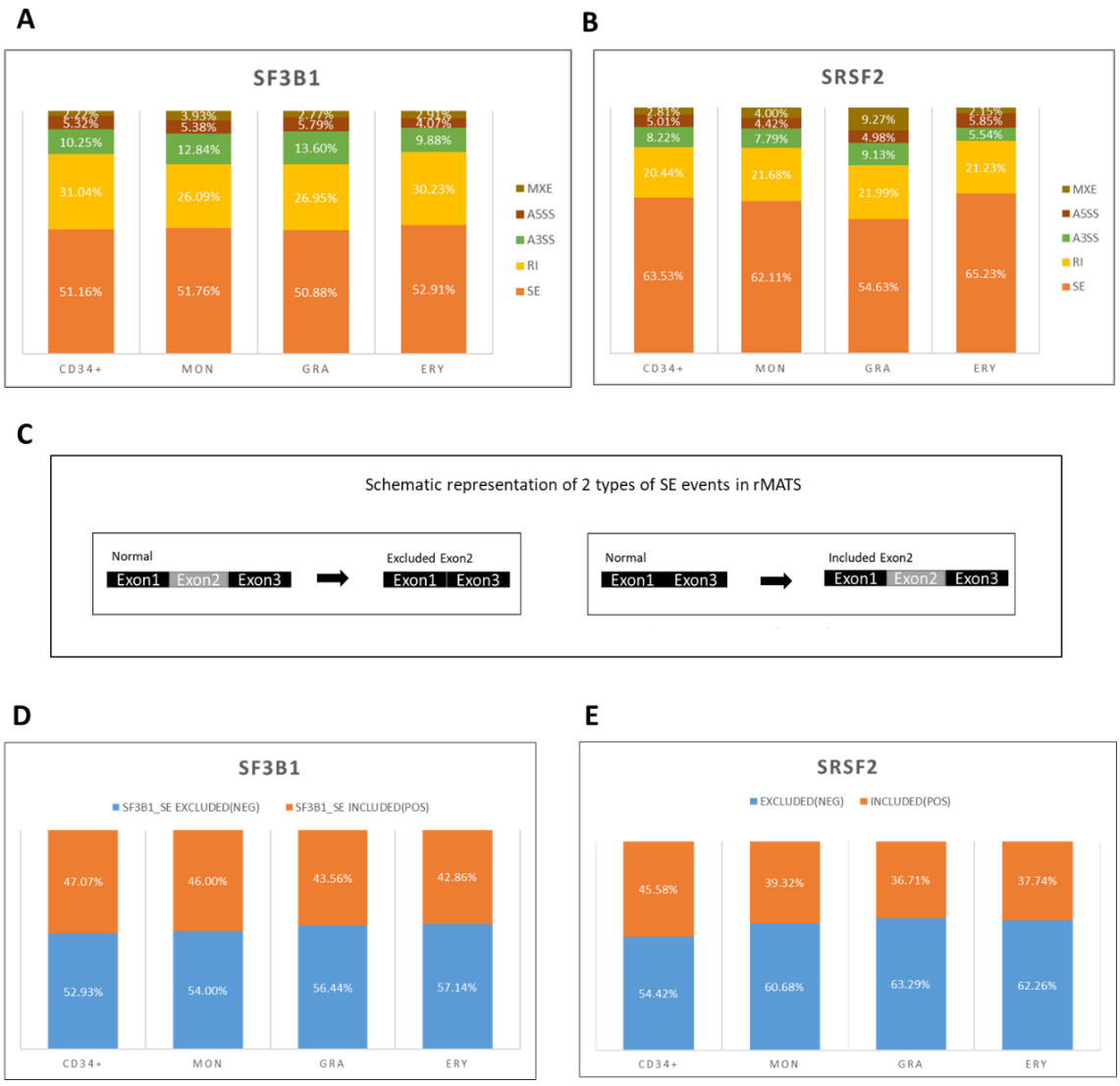


Figure 4-15. Distribution of aberrant splicing events found in bone marrow populations of MDS harbouring *SF3B1* and *SRSF2* mutations.

The bar charts showing the ratio of aberrant splicing events identified in (A) *SF3B1* and (B) *SRSF2* mutant MDS cohort. (C) Schematic of 2 types of SE events in rMATS. The bar graphs showing the ratio of 2 types of SE events identified in (D) *SF3B1* and (E) *SRSF2* mutant MDS.

Table 4-2. Comparison of alternative splicing frequencies across four BM subpopulations for the five main alternative splicing types in MDS patient with *SF3B1* mutation

Event types	Healthy control vs *SF3B1mut*			
	CD34$^+$	MON	GRA	ERY
SE	51.16%	51.76%	50.88%	52.91%
RI	31.04%	26.09%	26.95%	30.23%
A3SS	10.25%	12.84%	13.60%	9.88%
A5SS	5.32%	5.38%	5.79%	4.07%
MXE	2.22%	3.93%	2.77%	2.91%

Table 4-3. Comparison of alternative splicing frequencies across four BM subpopulations for the five main alternative splicing types in MDS patient with *SRSF2* mutation

Event types	Healthy control vs *SRSF2mut*			
	CD34$^+$	MON	GRA	ERY
SE	63.53%	62.11%	54.63%	65.23%
RI	20.44%	21.68%	21.99%	21.23%
A3SS	8.22%	7.79%	9.13%	5.54%
A5SS	5.01%	4.42%	4.98%	5.85%
MXE	2.81%	4.00%	9.27%	2.15%

4.3 Discussion

Here, we presented the global landscape of mRNA splicing alterations in MDS harbouring splicing factor mutations (*SF3B1* and *SRSF2*). This study showed that the majority of abnormal splicing events take place in a cell-type specific manner in *SRSF2*mut and *SF3B1*mut MDS. Using GO, we identified that *SF3B1* and *SRSF2* mutations affect mRNA splicing process in all BM subpopulations studied, causing aberrant splicing in other splicing factor genes. Our findings revealed that alternative splicing is cell-type dependent and splicing factor (*SF3B1* and *SRSF2*) mutations alter the patterns of whole mRNA splicing within each bone marrow subpopulation, showing similar patterns of alterations depending on genotypes. This result suggests that there may be a potential mechanism underlying regularity of the altered splicing patterns among complex factors when mRNA splicing process is perturbed.

We analysed alteration of mRNA splicing in our patient cohort (84 MDS patients) of RNA-seq, which allowed us to identify aberrantly spliced key target genes (Pellagatti et al., 2018) in 4 bone marrow subpopulations, and to demonstrate that most of the key events happen in a cell-type specific manner in *SF3B1*mut and *SRSF2*mut MDS cases. Many studies in myeloid neoplasms harbouring splicing factor mutations have focused on key target genes and downstream implications, how the aberrantly spliced target genes dysregulate pathways and cellular processes involved in impairment of differentiation and proliferation or phenotype related to myeloid neoplasms (Dolatshad et al., 2016, Kim et al., 2015, Smith et al., 2019, Clough et al., 2022). For example, Dolatshad et al. demonstrated that the *SF3B1* mutation leads to aberrant splicing in *ABCB7* and downregulation of the mRNA expression via NMD in human myeloid cancer cells (Dolatshad et al., 2016). Recent data uncovered that *U2AF1* mutations lead to usage of specific isoforms of *GNAS* and *IRAK4*,

which may result in malignant transformation (Smith et al., 2019, Wheeler et al., 2022). Also, *SRSF2* mutations lead to an *EZH2* isoform with inclusion of a poison exon that causes NMD, giving rise to impaired haematopoiesis (Kim et al., 2015). In fact, the inclusion of the poison exon of *EZH2* in our data is observed only within $CD34^+$ population as a significant event, but not in the other three bone marrow cell populations. Our results warrant a more in-depth approach for cell-type specific mis-splicing to study the impact of key target genes on haematological malignancies.

Our GO analysis for aberrant spliced genes in MDS harbouring splicing factor mutations (*SF3B1* and *SRSF2*) showed that each mutated splicing factor extensively affects genes involved in splicing, targeting different genes in different bone marrow cell populations, but these genes converge in overlapping gene ontology themes (RNA splicing; GO:0008380), indicating a common dysregulated cellular process. In our previous work (Pellagatti et al., 2018), which formed the basis of this study, this phenomenon was shown only in $CD34^+$ populations of MDS but not in other lineages. Interestingly, the target genes of splicing factors in the category of "RNA splicing"; GO:0008380) are different depending on cell types but the biological process, commonly affected in all the BM subpopulation, is mRNA splicing. This result indicates several possibilities: (1) there may be networks between the splicing factors in the context of the splicing process, (2) there may be "splicing-cascade" in cell-type dependant manner, and (3) these may result in exacerbation of aberrant splicing in MDS. A recent research (Liang et al., 2018) showed findings that support the above possibilities in *SRSF2*^{P95H} mutant HEL cells.

Using PCA, we confirm that alternative splicing is cell-type dependent in bone marrow cell populations. This suggested that global pre-mRNA splicing pattern in a certain cell-type can be considered as an intrinsic signature to distinguish from other cell-types during

haematopoiesis. Several examples of cell-type dependent splicing were demonstrated in recent studies during development in many types of tissues (Chepelev and Chen, 2013, Zhang et al., 2016, Gao et al., 2018, Mazin et al., 2021, Barash et al., 2010), suggesting that splicing plays an important role in organ differentiation and development. We also showed that *SF3B1* mutations and *SRSF2* mutations lead to alteration of splicing patterns in the whole transcriptomes, displaying similar patterns of splicing changes within each bone marrow population, depending on genotypes. Each stage of haematopoiesis should be accurately governed by cell type-specific differential splicing. However, our data showed that splicing factor mutations in MDS distort patterns of global splicing within each population, implying the splicing alteration affects differentiation of haematological cells in MDS.

Research in mRNA splicing regulation has achieved great advances over last few decades. Initially research was focused on *cis*-acting sequencing elements and trans-acting factors, which provided the important knowledge on most of the mechanisms (Sharp, 1994, Lerner et al., 1980, Breathnach et al., 1978). To decipher the 'splicing code' (a set of splicing regulatory rules), some groups have used computational analysis to understand the mechanism underlying, between multiple splicing factors and splicing regulatory elements including ESE, ESS, ISE and ISS, enabling prediction of splicing upshot from sequence information in the pre-mRNA (Wang et al., 2004, Barash et al., 2010, Jha et al., 2019). More recently, many studies on splicing have been conducted at a single cell resolution, identifying the widespread nature of the mechanism (Hayashi et al., 2018, Wen et al., 2020, Gaiti et al., 2022, Wang et al., 2018a, Jenkins and Kielkopf, 2017, Kesarwani et al., 2017). These studies have also highlighted the complexity of ‘splicing code’ and the difficulties to be fully decipher it. Furthermore, the final fate of splicing outcome during the mRNA

splicing process is multifactorial dependent (**Figure 4-16**). As such, the involvements of other cellular processes and elements in splicing further adds to the complexity. Hence, there is a need to decipher the 'splicing codes' at a greater resolution. Beside *cis*-acting elements, trans-acting factors and cell types, chromatin structure changes can impact transcription efficiency and RNA pol II progressivity which consequently affect the selection of splice sites (Schwartz et al., 2009, Spies et al., 2009, Kfir et al., 2015, Dvinge, 2018, Carrillo Oesterreich et al., 2016). Structural features of mRNA are also indispensable elements given its canonical role in regulating the accessibilities around splice sites (Soemedi et al., 2017, Kesarwani et al., 2017, Warf and Berglund, 2010). In addition, interaction profiles of RBP networks should be considered as an important element as well (Sternburg and Karginov, 2020). Although these elements consisting of the 'splicing code' are correlative and extremely complex, similar patterns and consistencies of splicing alterations caused by splicing factor mutations were observed in our data analysis. This suggests that there may be a potential mechanism underlying the regularity of the patterns of alterations when mRNA splicing process is perturbed.

Advances in RNA-seq have been allowing us to identify splicing outcomes, which is shaping nearly every aspect of our knowledge on splicing (Stark et al., 2019). The emerging long-read sequencing technologies (Sharon et al., 2013, Cartolano et al., 2016) provide opportunities to observe an accurate splicing outcome, as each read covers a full transcript (Clark et al., 2020). We also anticipate that single-cell analysis and multi-omics approaches will further help decipher the 'splicing code', enabling us to find the associations of cell-type specificity and chromatin accessibility with splicing alteration (Wu et al., 2022).

Our study is the first to show the landscape of global splicing patterns in MDS patients harbouring *SF3B1* mutations and *SRSF2* mutations across 4 bone marrow populations. The results of this study may inform future efforts to uncover the mechanism underlying the altered splicing patterns when mRNA splicing process is perturbed. The approaches to understand pattern and regularity of aberrant splicing caused by mutated splicing factors should shed light on the 'splicing code'.

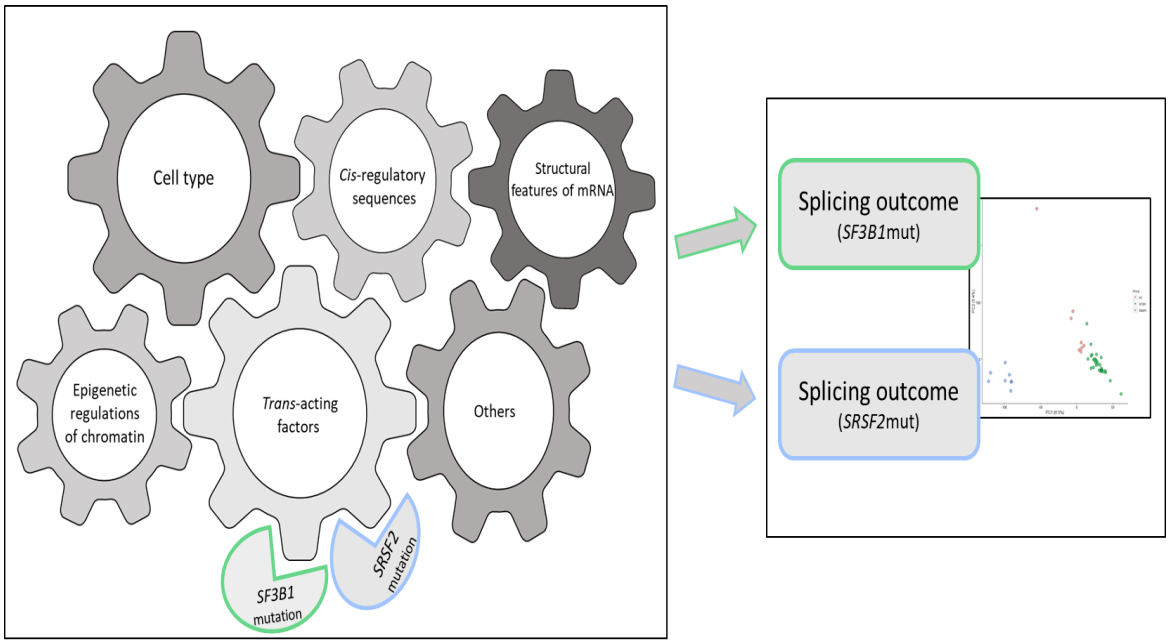


Figure 4-16. Schematic representations of elements affecting splicing fate, their combinatorial control on splicing and different splicing outcomes depending on types of splicing factor mutations.

5 Chapter Five – Single-cell analysis reveals transcriptional landscapes of MDS HSPC harbouring *SRSF2* mutations with co-mutations of *TET2* and/or *ASXL1*

5.1 Introduction

With advances in genomics, there have been many studies elucidating the driver mutations of MDS, which confirmed that the median number of driver mutations is two or three per patient (Ogawa, 2019, Cazzola, 2020, Haferlach et al., 2014). Approximately 60% of MDS patients harbouring *SRSF2* mutation also have other driver mutations such as *TET2* or/and *ASXL1* (**Figure 5-1**) (Todisco et al., 2021). However, despite the frequent overlap of mutations in the splicing factor gene (*SRSF2*) and epigenetic regulation genes (*TET2* and *ASXL1)*, little is known about the combinatory effects of the driver mutations on leukemogenesis.

Many cancers are propagated by rare self-renewing cancer stem cells which display genetic and phenotypic heterogeneity, and typically show resistance to current treatments (Eun et al., 2017, Tefferi et al., 2022, Clevers, 2011). A major obstacle in the study of genetic and phenotypic heterogeneity of leukemic cells has been the inability to exclude confounders and noises introduced by normal cells and their mixed characteristics in bone marrow niche (Li and Wang, 2021). Single-cell RNA sequencing (scRNA-seq) is a powerful tool that enables precise characterisation of individual cells in a myeloid neoplasm, providing the

ability to distinguish between normal and malignant cells, which would be normally obscured in bulk analysis (Lawson et al., 2015, Madaci et al., 2021).

Herein, we generated a single-cell transcriptome atlas encompassing 33,982 HSPCs harbouring *SRSF2* mutation, categorising them depending on the genotype of the co-mutations (*TET2* and/or *ASXL1*) from 18 donors (15 MDS patients and 3 healthy donors). We analysed cell-type distributions and cell population composition in *SRSF2* mutant MDS HSPCs and demonstrated that there are distinct differences in the cell population composition depending on the genotype of the co-mutations. Also, differential gene expression analysis revealed genotype-specific dysregulated pathways and genes in primitive HSPC populations of *SRSF2* mutant MDS. This study may provide novel insight into the biological features and clinical use for *SRSF2* mutant MDS at the single-cell level.

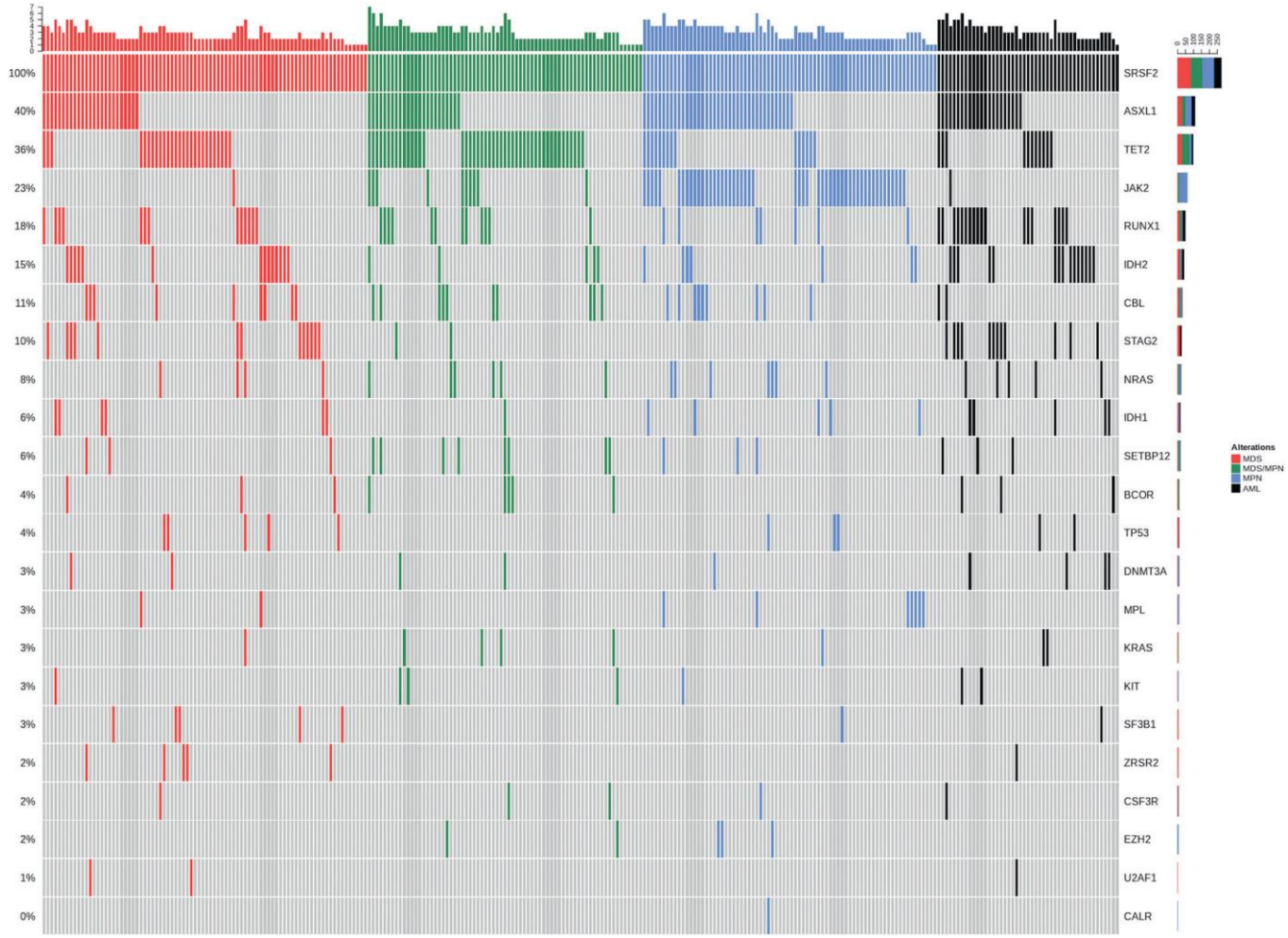


Figure 5-1. Co-mutation plot for somatically mutated genes in *SRSF2*^{P95}-mutated neoplasms.
Diagnosis of each sample according to WHO criteria is shown by indicated colours.

* This figure adapted from Todisco et al. (Todisco et al., 2021)

5.2 Results

5.2.1 Generation of single-cell transcriptomic atlas of MDS HSPCs harbouring *SRSF2* mutations with co-mutations (*TET2* and/or *ASXL1*)

To identify the genetic landscape of MDS harbouring *SRSF2* mutations with the co-mutations, we analysed 15 MDS patients with *SRSF2* mutations using bulk-level next generation sequencing (**Table 2-5**). We confirmed $SRSF2^{P95}$ mutations (15/15, VAF $\geq$ 40%) and other co-occurring mutations in all the patient samples. The mean number of mutations per sample is 4 (between 2 and 7). Among the co-occurring mutations with *SRSF2* mutations, *TET2* mutations (11 / 15; 73%) or *ASXL1* mutations (10 / 15; 60%) were present in all the samples. We then categorised the patient samples into three groups ; (1) *SRSF2 / TET2* mutations (n = 6), (2) *SRSF2 / ASXL1* mutations (n = 4) and (3) *SRSF2 / TET2 / ASXL1* mutations (n = 5), based on the genotype of the co-mutations (**Figure 5-2**. A).

To investigate transcriptome heterogeneity in bone marrow MDS HSPCs harbouring *SRSF2* mutations with co-mutations (*TET2* and/or *ASXL1*), we performed scRNA-seq of bone marrow Lin^-CD34^+ HSPCs enriched by FACS, using 15 MDS patients harbouring *SRSF2* mutations and 3 age-matched healthy donors (HD). Quality checks were performed to ensure that the cells were suitable for downstream analysis including integration, batch correction, dimension reduction and cell-type annotation (**Table 5-1**). A total of 33,982 cells were captured to generate the *SRSF2* mutant MDS HSPC atlas (**Figure 5-2**. B) on Uniform manifold approximation and projection (UMAP). According to the gene expression patterns and canonical gene markers, we identified 15 cell populations within

the $Lin^{-}CD34^{+}$ cell population, namely HSPCs; hematopoietic stem progenitor cells, Myeloid-primed; myeloid-primed progenitors, IMP; immature myeloid progenitors, CMP; common myeloid progenitors, EMBP; eosinophil / mast cell / basophil progenitors, NP; neutrophil progenitors, Mono/DC; monocytic/dendritic cell progenitors, CD14 Mono; CD14 monocytic progenitors, Pre-B; precursors B cells, MEP; megakaryocytic-erythroid progenitors, MkP; megakaryocytic progenitors and EP; erythroid progenitors. The HSPCs are divided into three clusters, HSPC-1, HSPC-2, and HSPC-3. The related canonical gene markers are shown in **Figure 5-2**. C and D.

Table 5-1. The result of quality control to proceed downstream analysis.

* Excluded from downstream analysis due to too low cell number.

Sample ID	Donor ID	HTO ID	No. of cells pooled	No. of cells sequenced (cellranger)	No. of cells successfully demultiplexed (HTO + Souporcell)	No. of cells passed no. of genes, no. of UMIs, % MT QC (SingleCellarR)
1	PV1506	HTO1	17,500		4,543	4,372
2	PV1553	HTO2	1,400	7,153 (30%)	467	371
3	MAN973	HTO3	5,250		1,754	1,689
4	PV1488	HTO1	3,500		300	239
5	NOC161	HTO2	12,250	2,757 (8%)	827	750
6	MAN543	HTO3	17.500		1,433	1,364
7	PV1253	HTO1	1,200		267	167*
8	PV593	HTO2	6,400	4,999 (27%)	1,618	1,573
9	NOC124	HTO3	11,000		2,906	2,838
10	MAN115	HTO1	11,000		3,788	3,658
11	PV244	HTO2	150	5,447 (34%)	-	-*
12	PV2210	HTO3	4,760		1,507	1,341
13	MAN930	HTO1	17,500		8,236	7,923
14	PV1539	HTO2	950	11,963 (42%)	184	29 *
15	NOC126	HTO3	10,000		2,780	2.689
16	MAN666	HTO1	5,000		1,888	1,823
17	PV1424	HTO2	3,000	7,318 (29%)	774	610
18	PV492	HTO3	17,500		4,032	2,742

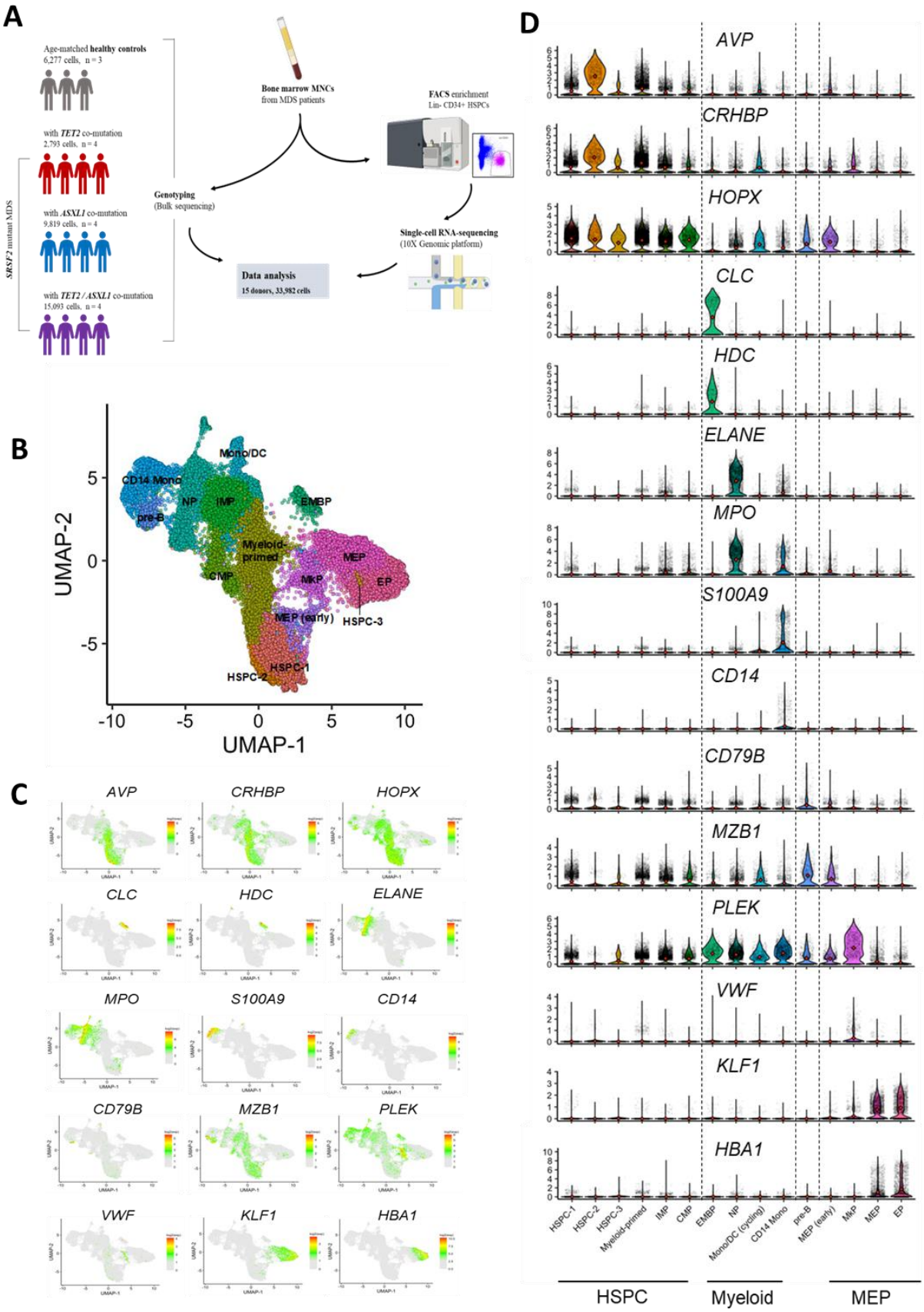

A
Age-matched healthy controls
6,277 cells, n = 3
with TET2 co-mutation
2,793 cells, n = 4
with ASXL1 co-mutation
9,819 cells, n = 4
with TET2 / ASXL1 co-mutation
15,093 cells, n = 4
SRSF2 mutant MDS
Bone marrow MNCs
from MDS patients
FACS enrichment
Lin- CD34+ HSPCs
Genotyping
(Bulk sequencing)
Single-cell RNA-sequencing
(10X Genomic platform)
Data analysis
15 donors, 33,982 cells
B
Mono/DC
CD14 Mono
NP
IMP
EMBP
pre-B
Myeloid-primed
MEP
CMP
MkP
EP
MEP (early)
HSPC-3
HSPC-1
HSPC-2
UMAP-2
UMAP-1
C
AVP
CRHBP
HOPX
CLC
HDC
ELANE
MPO
S100A9
CD14
CD79B
MZB1
PLEK
VWF
KLF1
HBA1
D
AVP
CRHBP
HOPX
CLC
HDC
ELANE
MPO
S100A9
CD14
CD79B
MZB1
PLEK
VWF
KLF1
HBA1
HSPC-1
HSPC-2
HSPC-3
Myeloid-primed
IMP
CMP
EMBP
NP
Mono/DC (cycling)
CD14 Mono
pre-B
MEP (early)
MkP
MEP
EP
HSPC
Myeloid
MEP

Figure 5-2. Single-cell transcriptomic atlas of MDS HSPCs harbouring SRSF2 mutations with co-mutations (*TET2* and/or *ASXL1*).

(A) Schematic overview of the experimental workflow in the study. (B) UMAP of Lin^- $CD34^+$ cells (n = 34,149 cells) from MDS patient samples with *SRSF2* mutations (n = 12 individuals) and healthy donors (n = 3 individuals), overlaid with cluster cell-type assignments. HSPC; hematopoietic stem progenitor cells, Myeloid-primed; myeloid-primed progenitors, IMP; immature myeloid progenitors, CMP; common myeloid progenitors, EMBP; eosinophil / mast cell / basophil progenitors, NP; neutrophil progenitors, Mono/DC; monocytic/dendritic cell progenitors, CD14 Mono; CD14 monocytic progenitors, Pre-B; precursors B cells, MEP; megakaryocytic-erythroid progenitors, MkP; megakaryocytic progenitors and EP; erythroid progenitors. (C) Violin plots and (D) dot plots showing the expression of the canonical marker genes in the 15 populations on UMAP.

5.2.2 Distinct differences in cell population composition of MDS HSPCs harbouring *SRSF2* mutations depending on the co-mutation types

A number of studies have sought to investigate clonal heterogeneity in MDS HSPC population, focusing on a type of driver mutations (Gaiti et al., 2022, Huang et al., 2022, Mian et al., 2015, Dutta et al., 2020). However, it is important to study the combinatory effects of driver mutations given that the mean number of driver mutations is two or three per MDS patient. After generating the transcriptomic atlas of MDS HSPCs harbouring *SRSF2* mutations, we performed analysis of cell-type distributions and cell population composition to characterise clonal heterogeneity depending on the co-mutations. Our results revealed that among the four groups with the different genotypes, each group showed distinct differences in composition of cell populations depending on the co-mutation types (**Figure 5-3**. A).

In all the MDS patients harbouring *SRSF2* mutations, we identified decreased erythroid, megakaryocytes, and lymphoid outputs, consistent with MDS clinical phenotype – pancytopenia (Cazzola, 2020, Arber et al., 2016, Montalban-Bravo and Garcia-Manero, 2018, Fenaux et al., 2021), compared to HD samples (**Figure 5-3**. B and C). Notably, as for *SRSF2 / TET2* mutation cases, marked increase in $CD14^{+}$ monocytes, which has not been reported yet in other MDS and decrease in HSPC-2 and myeloid-primed cells were observed (**Figure 5-3**. C). In *SRSF2*/*ASXL1* mutant samples, regardless of *TET2* mutations, we observed increase in primitive HSPCs such as HSPC-1, HSPC-3, myeloid-primed HSPCs and immature myeloid progenitors. Moreover, it is interesting to note that *SRSF2*/*ASXL1* with or without *TET2* mutant cases have very similar distributions of cellular composition compared to other genotype groups (**Figure 5-3**. A).

While performing FACS to enrich Lin^- $CD34^+$ cells, we found that the frequencies of the Lin^- $CD34^+$ cells are different in *SRSF2* mutant MDS depending on the genotype of the co-mutations. In the FACS experiments, the means of the frequencies of Lin^- $CD34^+$ HSPCs were 0.24%, 0.079%, 0.12% and 0.15% in *SRSF2*/*ASXL1* mutant, *SRSF2*/*TET2* mutant, *SRSF2*/*ASXL1*/*TET2* mutant and HD samples respectively. Interestingly, the samples harbouring *ASXL1* mutations had higher frequencies of Lin^- $CD34^+$ cells than HD, while the samples with *TET2* mutations showed lower frequencies of Lin^- $CD34^+$ cells. The analysis of cell population composition (**Figure 5-3**. C) suggests that the increased number of the primitive cell populations such as HSPC-1, Myeloid-primed and IMP may contribute to the higher frequencies of Lin^- $CD34^+$ cells in *SRSF2*/*ASXL1* mutant MDS samples.

Altogether, our results revealed that the cell population composition of MDS HSPCs with *SRSF2* mutations are distinctly different depending on the co-mutation types, suggesting that the combination of the co-mutations with *SRSF2* mutation in MDS can affect HSPC population composition.

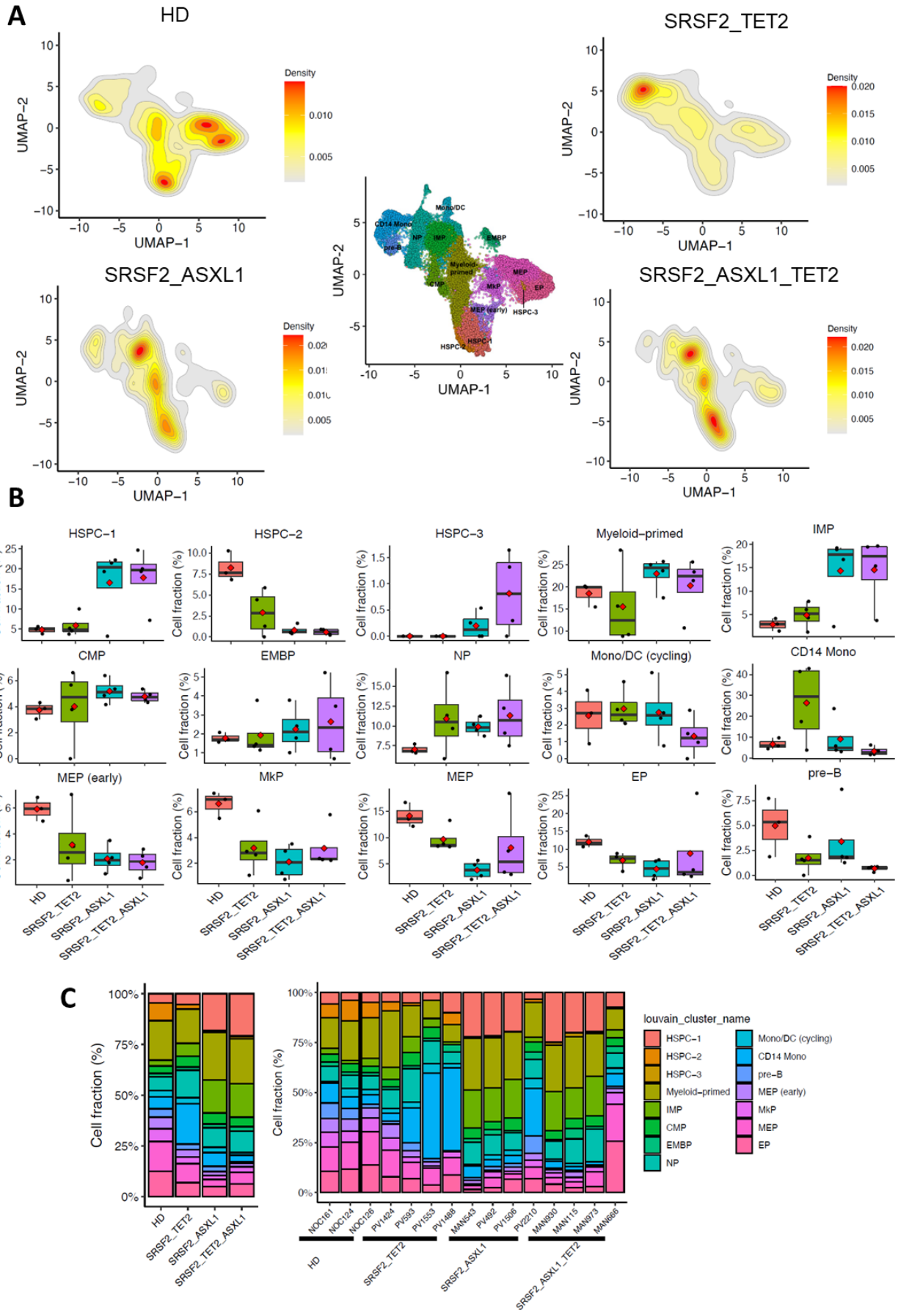

A
HD
SRSF2_TET2
SRSF2_ASXL1
SRSF2_ASXL1_TET2
UMAP-1
UMAP-2
Density
Mono/DC
CD14 Mono
NP
IMP
EMBP
pre-B
Myeloid-primed
MEP
CMP
MkP
EP
MEP (early)
HSPC-3
HSPC-1
HSPC-2
B
HSPC-1
HSPC-2
HSPC-3
Myeloid-primed
IMP
CMP
EMBP
NP
Mono/DC (cycling)
CD14 Mono
MEP (early)
MkP
MEP
EP
pre-B
Cell fraction (%)
HD
SRSF2_TET2
SRSF2_ASXL1
SRSF2_TET2_ASXL1
C
Cell fraction (%)
louvain_cluster_name
HSPC-1
HSPC-2
HSPC-3
Myeloid-primed
IMP
CMP
EMBP
NP
Mono/DC (cycling)
CD14 Mono
pre-B
MEP (early)
MkP
MEP
EP
NOC161
NOC124
NOC126
PV1424
PV593
PV1553
PV1488
MAN543
PV492
PV1506
PV2210
MAN930
MAN115
MAN973
MAN666

Figure 5-3. Distinct differences in cellular composition of MDS HSPCs harbouring *SRSF2* mutations depending on the co-mutation types.

(A) The distribution maps showing distinct cell-populations depending on the combination of 3 major driver mutations in MDS. (B) Box plots representing the proportion of each cell fraction. (C) Normalised ratio of each cell fraction.

5.2.3 Transcriptional profile differences across *SRSF2* mutant MDS cases harbouring different co-mutations

We next sought to investigate dysregulated cellular processes, performing differential gene expression (DE) analysis of *SRSF2* mutant MDS cases harbouring different co-mutations, compared to HD samples. Although DE analysis has been conducted for all the cell populations, here, we focused on HSPC-1, HSPC-2 and IMP because ; 1) they are the primitive populations and therefore likely to be the disease-propagating cells (Woll et al., 2014, Shastri et al., 2017) and 2) these populations showed different abundance in the cell population composition analysis compared to HD (**Figure 5-4**. A, **Figure 5-5**. A and **Figure 5-6**. A).

In HSPC-1 population, we identified 648 up-regulated/343 down-regulated genes, 1,669 up-regulated/285 down-regulated genes and 1,304 up-regulated/890 down-regulated genes in *SRSF2* mutant MDS harbouring co-mutations of *TET2*, *ASXL1* and *TET2/ASXL1* respectively, compared to HD (**Figure 5-4**. B and C). In all the *SRSF2* mutant cases, GO analysis of up-regulated genes revealed the enrichment of genes related to GTPase activities, such as "regulation of small GTPase mediated signal transduction", "positive regulation of GTPase activity" and "regulation of GTPase activity" in all the *SRSF2* mutant cases (**Figure 5-4**. D and E (top)). For example, we found significantly higher expression of genes associated with GTPase activities such as *NF1*, *RALGAPA1* and *RALGAPA2* in HSPC-1 cluster compared to HD samples (**Figure 5-4**. E). This result implies that there are common dysregulated genes and pathways in all the *SRSF2* mutant cases regardless of the differences of the co-mutations. In addition, we identified genotype-specific dysregulated genes and pathways in HSPC-1 population (**Figure 5-4**. C). For example, the enrichment of the genes associated with "Ras protein signal transduction" was observed

only in *SRSF2*/*TET2* mutant MDS while the gene enrichments of "histone modification" and "covalent chromatin modification" were identified only in *SRSF2*/*ASXL1* mutant cases with or without *TET2* mutations (**Figure 5-4**. D). These findings are further supported by violin plots presenting single-cell expression of the genes such as *ASXL1*, *JAK2* and *KMT2A* (all up-regulated in *SRSF2*/*ASXL1* with or without *TET2* cases compared to HD) (**Figure 5-4**. E (bottom)). In contrast, GO analysis of down-regulated genes in HSPC-1 cluster showed that dysregulations of biological process and pathways are barely shared across the 3 genotype groups (**Figure 5-4**. F). For instance, the dysregulation of "Mitochondrial respiratory chain complex assembly" including the genes such as *COA3*, *COA4*, *NDUFS3*, *TIMMDC1*, *UQCC2* and *UQCR10* was observed solely in *SRSF2/ASXL1/TET2* mutant MDS cases, compared to other groups (**Figure 5-4**. F and G).

Next, in HSPC-2 population, our DE analysis found 278 up-regulated/84 down-regulated genes, 347 up-regulated/126 down-regulated genes and 537 up-regulated/195 down-regulated genes in *SRSF2* mutant MDS with co-mutations of *TET2*, *ASXL1* and *TET2/ASXL1* respectively, compared to HD (**Figure 5-5**. B and C). In all the *SRSF2* mutant cases, we identified the enrichment of genes associated with "covalent chromatin modification" and "histone modification" (**Figure 5-5**. D and E (top)). Interestingly, in the GO analysis of up-regulated genes, there were several dysregulated biological processes and pathways, commonly identified in both the *SRSF2*/*TET2* and the *SRSF2*/*TET2*/*ASXL1* mutant cases unlike in the analysis of HSPC-1 cluster. For example, significantly higher expression of genes associated with "RNA splicing" and "G1/S transition of mitotic cell cycle" were observed in both the *SRSF2*/*TET2* and the *SRSF2*/*TET2*/*ASXL1* mutant groups (**Figure 5-5**. D and E (bottom)).

In IMP population, the DE analysis identified 254 up-regulated / 1,162 down-regulated genes, 538 up-regulated/ 1,088 down-regulated genes and 489 up-regulated/ 1,837 down-regulated genes in *SRSF2* mutant cases with *TET2, ASXL1* and *TET2/ASXL1* mutations respectively, compared to HD (**Figure 5-6**. B and C). In the GO analysis of up-regulated genes in IMP populations, genotype-specific dysregulated genes and pathways were observed (**Figure 5-6**. D). For instance, the result showed significantly up-regulated expression of genes associated with "neutrophil degranulation" and "cellular response to interferon−gamma" in *SRSF*/*TET2* and *SRSF2*/*ASXL1* mutant cases respectively (**Figure 5-6**. E). In contrast, GO analysis of down-regulated genes in IMP population showed that several dysregulated pathways were shared in all the *SRSF2* mutant groups, such as "DNA replication", "nuclear division", "mitotic nuclear division" and "signal transduction by p53 class mediator" (**Figure 5-6**. F). Notably, in *SRSF2*/*ASXL1* mutant cases with or without *TET2* mutations, most of the dysregulated pathways were shared such as "chromosome segregation", "nuclear chromosome segregation" and "DNA conformation change" (**Figure 5-6**. F and G).

In summary, our DE data and GO analysis elucidated key characteristics of transcriptional profiles in MDS HSPCs harbouring *SRSF2* mutations and co-mutation of *TET2* and/or *ASXL1* at single-cell resolution. Firstly, we identified the dysregulated genes and pathways specific to the co-mutation genotype groups as well as the ones commonly shared in all the *SRSF2* mutant cases. Secondly, our results showed that the pathways associated with epigenetic regulations (DNA methylation and histone modification) were dysregulated, which is consistent with genetic mutations in *TET2* and *ASXL1* (Ogawa, 2019). Thirdly, we identified similar transcriptional profiles in both cases of *SRSF2*/*ASXL1* mutant groups with and without *TET2* mutation, compared to other genotype groups (**Figure 5-4**.C,

Figure 5-5. C and **Figure 5-6**. C). This may also provide the explanation on the similarities of cell population compositions between the two groups (**Figure 5-3**. A).

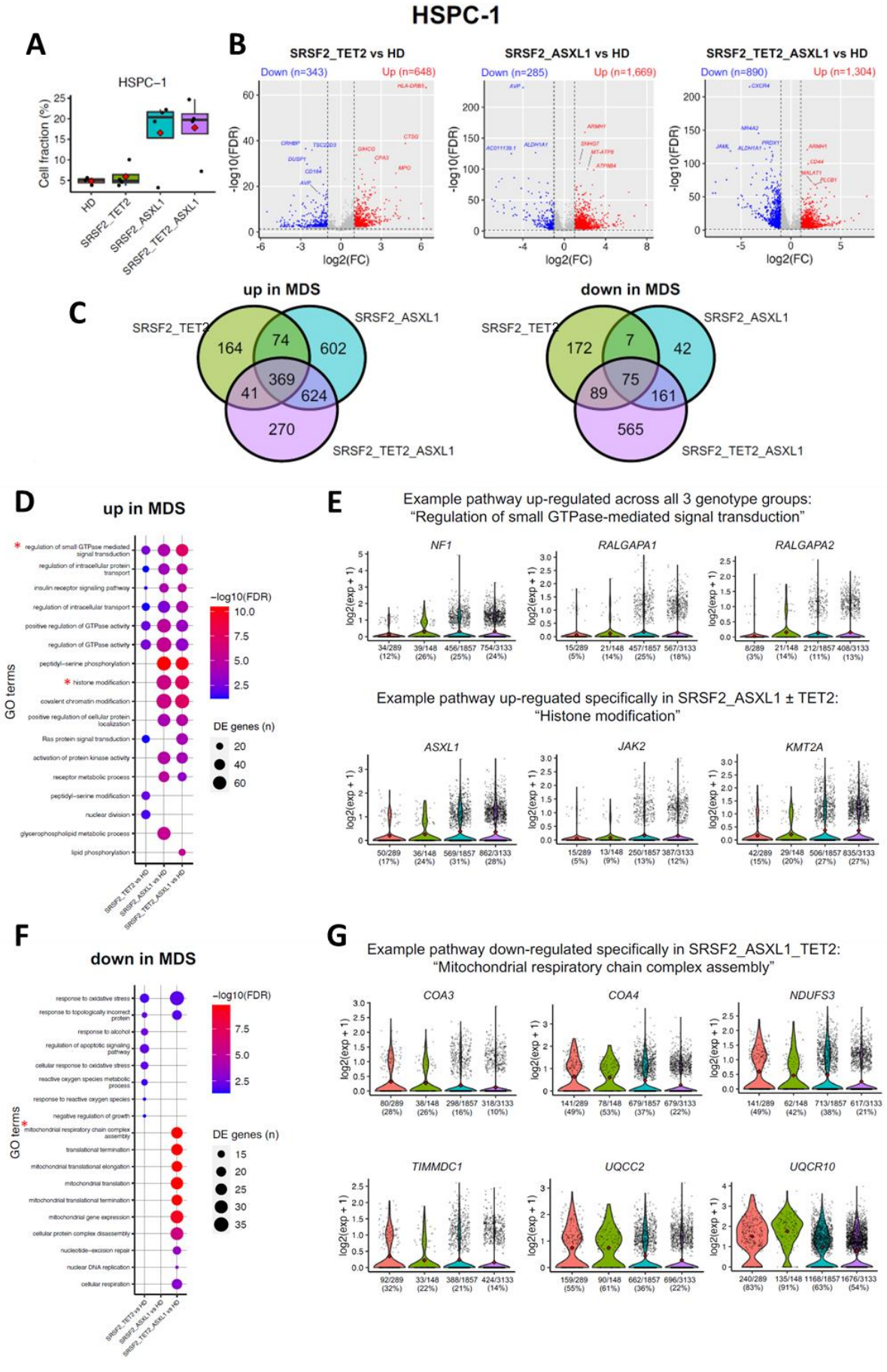

HSPC-1
A
HSPC–1
Cell fraction (%)
HD
SRSF2_TET2
SRSF2_ASXL1
SRSF2_TET2_ASXL1
B
SRSF2_TET2 vs HD
Down (n=343)
Up (n=648)
SRSF2_ASXL1 vs HD
Down (n=285)
Up (n=1,669)
SRSF2_TET2_ASXL1 vs HD
Down (n=890)
Up (n=1,304)
-log10(FDR)
log2(FC)
C
up in MDS
SRSF2_TET2
SRSF2_ASXL1
SRSF2_TET2_ASXL1
164
74
602
369
41
624
270
down in MDS
172
7
42
75
89
161
565
D
up in MDS
GO terms
regulation of small GTPase mediated signal transduction
regulation of intracellular protein transport
insulin receptor signaling pathway
regulation of intracellular transport
positive regulation of GTPase activity
regulation of GTPase activity
peptidyl−serine phosphorylation
histone modification
covalent chromatin modification
positive regulation of cellular protein localization
Ras protein signal transduction
activation of protein kinase activity
receptor metabolic process
peptidyl−serine modification
nuclear division
glycerophospholipid metabolic process
lipid phosphorylation
−log10(FDR)
DE genes (n)
E
Example pathway up-regulated across all 3 genotype groups:
"Regulation of small GTPase-mediated signal transduction"
NF1
RALGAPA1
RALGAPA2
log2(exp + 1)
Example pathway up-reguated specifically in SRSF2_ASXL1 ± TET2:
"Histone modification"
ASXL1
JAK2
KMT2A
F
down in MDS
GO terms
response to oxidative stress
response to topologically incorrect protein
response to alcohol
regulation of apoptotic signaling pathway
cellular response to oxidative stress
reactive oxygen species metabolic process
response to reactive oxygen species
negative regulation of growth
mitochondrial respiratory chain complex assembly
translational termination
mitochondrial translational elongation
mitochondrial translation
mitochondrial translational termination
mitochondrial gene expression
cellular protein complex disassembly
nucleotide−excision repair
nuclear DNA replication
cellular respiration
−log10(FDR)
DE genes (n)
G
Example pathway down-regulated specifically in SRSF2_ASXL1_TET2:
"Mitochondrial respiratory chain complex assembly"
COA3
COA4
NDUFS3
TIMMDC1
UQCC2
UQCR10
log2(exp + 1)

Figure 5-4. Transcriptional profile differences across *SRSF2* mutant MDS cases harbouring different co-mutations in the HSPC-1 population

(A) A box plot showing the ratio of HSPC-1 compartment (HD, *SRSF2/TET2* mutant, *SRSF2/ASXL1* mutant and *SRSF2/TET2/ASXL1* mutant). (B) Volcano plots representing differentially expressed genes in the three *SRSF2* mutant groups (left*; SRSF2/TET2*, middle; *SRSF2/ASXL1*, right; *SRSF2/TET2/ASXL1*) compared to HD. (C) Venn diagrams showing the relations of differentially expressed genes in the three groups (left; up-regulated genes, right; down-regulated genes). (D) A bubble plot showing GO analysis on up-regulated genes. (E) The relative expression level of selected genes in the list of GO terms (top; regulation of small GTPase-mediated signal transduction, bottom; histone modification) (F) A bubble plot showing GO analysis on down-regulated genes. (G) The relative expression level of selected genes in the list of GO terms (Mitochondrial respiratory chain complex assembly).

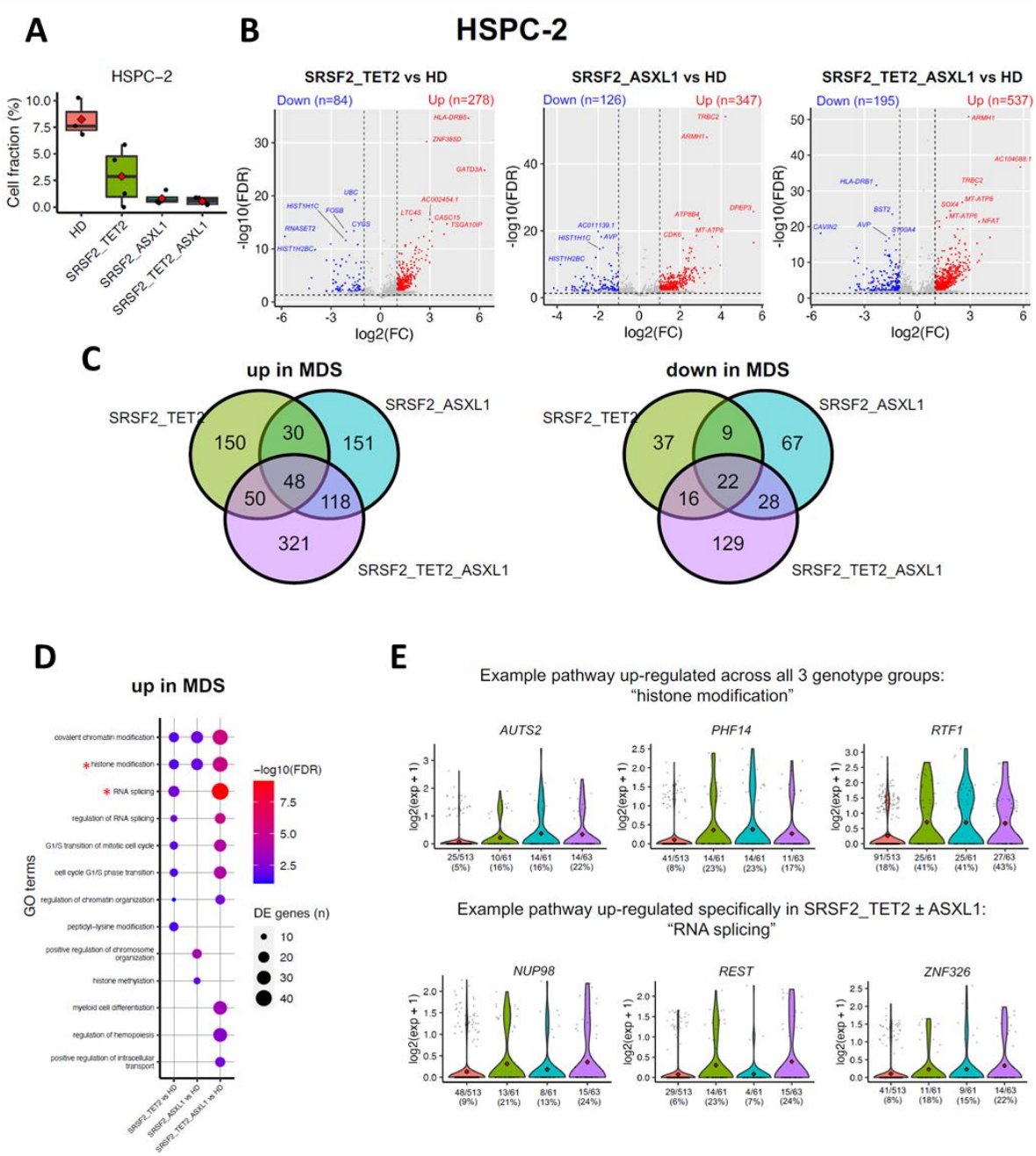
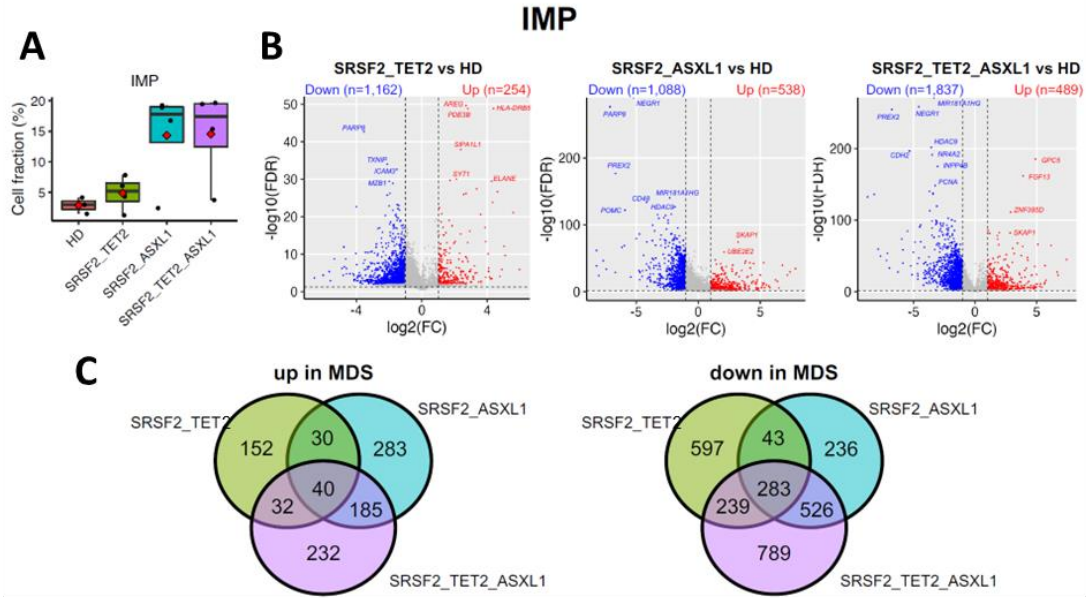
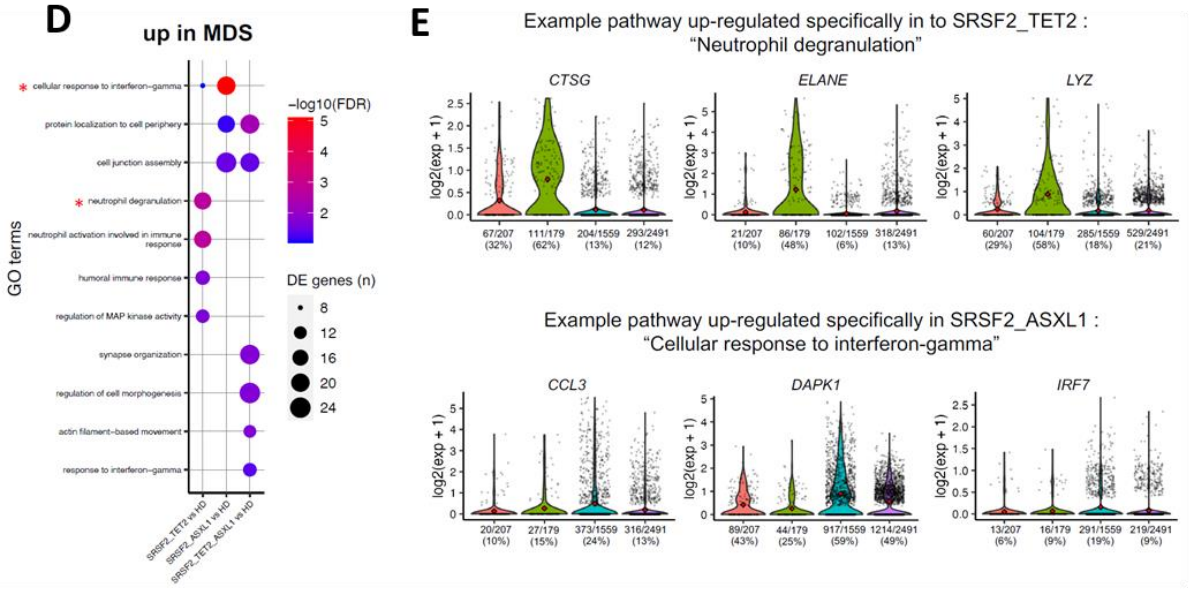

A
HSPC−2
Cell fraction (%)
HD
SRSF2_TET2
SRSF2_ASXL1
SRSF2_TET2_ASXL1
B
HSPC-2
SRSF2_TET2 vs HD
Down (n=84)
Up (n=278)
SRSF2_ASXL1 vs HD
Down (n=126)
Up (n=347)
SRSF2_TET2_ASXL1 vs HD
Down (n=195)
Up (n=537)
-log10(FDR)
log2(FC)
C
up in MDS
SRSF2_TET2
SRSF2_ASXL1
SRSF2_TET2_ASXL1
150
30
151
48
50
118
321
down in MDS
37
9
67
22
16
28
129
D
up in MDS
GO terms
covalent chromatin modification
histone modification
RNA splicing
regulation of RNA splicing
G1/S transition of mitotic cell cycle
cell cycle G1/S phase transition
regulation of chromatin organization
peptidyl−lysine modification
positive regulation of chromosome organization
histone methylation
myeloid cell differentiation
regulation of hemopoiesis
positive regulation of intracellular transport
−log10(FDR)
DE genes (n)
E
Example pathway up-regulated across all 3 genotype groups:
“histone modification”
AUTS2
PHF14
RTF1
log2(exp + 1)
Example pathway up-regulated specifically in SRSF2_TET2 ± ASXL1:
“RNA splicing”
NUP98
REST
ZNF326

Figure 5-5. Transcriptional profile differences across *SRSF2* mutant MDS cases harbouring different co-mutations in the HSPC-2 population

(A) A box plot showing the ratio of HSPC-2 compartment in the four groups (HD, *SRSF2/TET2* mutant, *SRSF2/ASXL1* mutant and *SRSF2/TET2/ASXL1* mutant). (B) Volcano plots representing differentially expressed genes in the three *SRSF2* mutant groups (left; *SRSF2/TET2*, middle; *SRSF2/ASXL1*, right; *SRSF2/TET2/ASXL1*) compared to HD. (C) Venn diagrams showing the relations of differentially expressed genes in the three groups (left; up-regulated genes, right; down-regulated genes). (D) A bubble plot showing GO analysis on up-regulated genes. (E) The relative expression level of selected genes in the list of GO terms (top; histone modification, down; RNA splicing).


IMP
A
IMP
Cell fraction (%)
HD
SRSF2_TET2
SRSF2_ASXL1
SRSF2_TET2_ASXL1
B
SRSF2_TET2 vs HD
Down (n=1,162)
Up (n=254)
SRSF2_ASXL1 vs HD
Down (n=1,088)
Up (n=538)
SRSF2_TET2_ASXL1 vs HD
Down (n=1,837)
Up (n=489)
-log10(FDR)
log2(FC)
C
up in MDS
SRSF2_TET2
SRSF2_ASXL1
SRSF2_TET2_ASXL1
152
30
283
40
32
185
232
down in MDS
597
43
236
283
239
526
789


D
up in MDS
GO terms
cellular response to interferon−gamma
protein localization to cell periphery
cell junction assembly
neutrophil degranulation
neutrophil activation involved in immune response
humoral immune response
regulation of MAP kinase activity
synapse organization
regulation of cell morphogenesis
actin filament−based movement
response to interferon−gamma
−log10(FDR)
DE genes (n)
E
Example pathway up-regulated specifically in to SRSF2_TET2 :
“Neutrophil degranulation”
CTSG
ELANE
LYZ
log2(exp + 1)
Example pathway up-regulated specifically in SRSF2_ASXL1 :
“Cellular response to interferon-gamma”
CCL3
DAPK1
IRF7

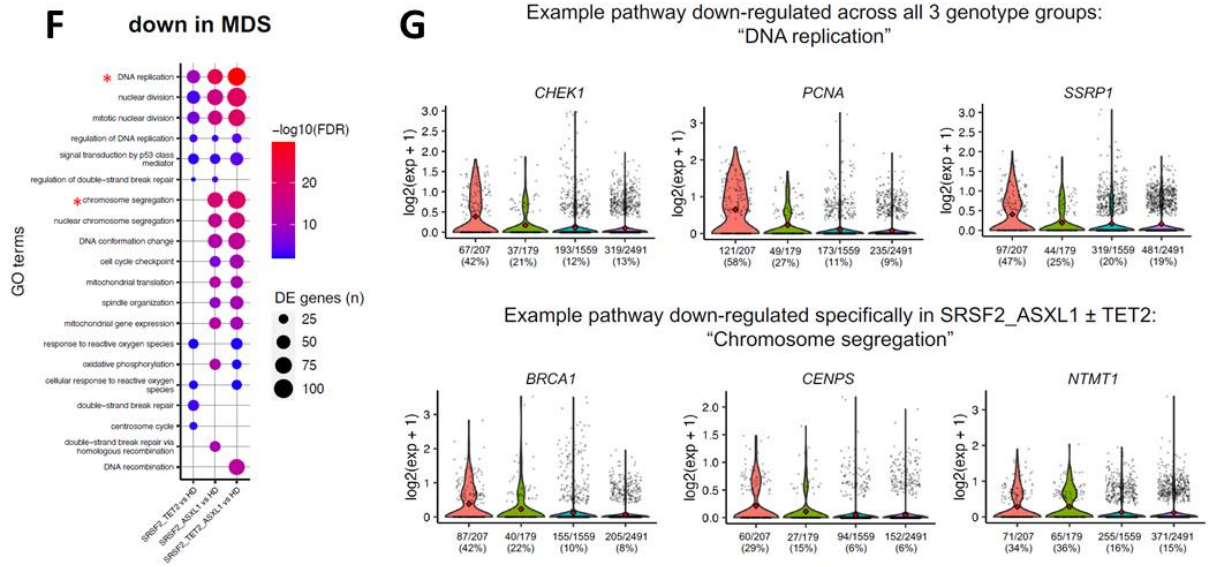

F
down in MDS
GO terms
DNA replication
nuclear division
mitotic nuclear division
regulation of DNA replication
signal transduction by p53 class mediator
regulation of double−strand break repair
chromosome segregation
nuclear chromosome segregation
DNA conformation change
cell cycle checkpoint
mitochondrial translation
spindle organization
mitochondrial gene expression
response to reactive oxygen species
oxidative phosphorylation
cellular response to reactive oxygen species
double−strand break repair
centrosome cycle
double−strand break repair via homologous recombination
DNA recombination
−log10(FDR)
DE genes (n)
G
Example pathway down-regulated across all 3 genotype groups:
“DNA replication”
CHEK1
PCNA
SSRP1
log2(exp + 1)
Example pathway down-regulated specifically in SRSF2_ASXL1 ± TET2:
“Chromosome segregation”
BRCA1
CENPS
NTMT1

Figure 5-6. Transcriptional profile differences across *SRSF2* mutant MDS cases harbouring different co-mutations in the IMP population

(A) A box plot showing the ratio of IMP compartment in the four groups (HD, *SRSF2/TET2* mutant, *SRSF2/ASXL1* mutant and *SRSF2/TET2/ASXL1* mutant). (B) Volcano plots representing differentially expressed genes in the three *SRSF2* mutant groups (left; *SRSF2/TET2*, middle; *SRSF2/ASXL1*, right; *SRSF2/TET2/ASXL1*) compared to HD. (C) Venn diagrams showing the relations of differentially expressed genes in the three groups (left; up-regulated gene, right; down-regulated genes). (D) A bubble plot showing GO analysis on up-regulated genes. (E) The relative expression level of selected genes in the list of GO terms (top; neutrophil degranulation, down; cellular response to interferon-gamma). (F) A bubble plot showing GO analysis on down-regulated genes. (G) The relative expression level of selected genes in the list of GO terms (top; DNA replication, down; chromosome segregation).

5.3 Discussion

In this study, we characterised the transcriptome of MDS HSPCs harbouring *SRSF2* mutations with co-mutations of *TET2* and/or *ASXL1*, at single-cell resolution. Our analysis revealed that distinct differences in cell population composition of MDS HSPCs harbouring *SRSF2* mutations depending on the co-mutation genotypes. Also, differential gene expression analysis unveiled genotype-specific gene expression signatures and dysregulated pathways.

Within myeloid neoplasm patients harbouring *SRSF2* mutation, over 60% of patients carry co-mutations of *TET2* and/or *ASXL1* (Todisco et al., 2021). We generated a single-cell atlas, categorising the samples into 4 sub-groups (*SRSF2*/*TET2* mutant, *SRSF2*/*ASXL1* mutant, *SRSF2*/*TET2*/*ASXL1* mutant and HD) depending on the genotypes. We identified 15 distinct cell populations based on GSEA and canonical marker genes. Our MDS single-cell atlas is the first to reveal the genetic landscape of MDS harbouring *SRSF2* mutations with co-mutations of *TET2* and/or *ASXL1* in MDS HSPCs, at the single-cell level. This allows to trace relevant associations between genotype and phenotype and also provides insight on unexplainable clinical heterogeneity within current MDS classification categories.

We confirmed distinct differences in the cell population composition of MDS HSPCs with *SRSF2* mutations depending on the co-mutation genotypes. Interestingly, in both patient groups (*SRSF2*/*ASXL1* mutant groups with or without *TET2* mutation), more similar characteristics of the cell population composition and higher frequencies of primitive HSPCs were identified, compared to other groups. This suggests that the presence of *ASXL1* mutations in *SRSF2* mutant MDS may confer growth-advantages to the primitive

HSPCs, which leads to clonal dominance of mutant stem cells. This observation is supported by other studies on the relations between *ASXL1* mutations and dysregulated proliferation (Fujino et al., 2021, Wu et al., 2015). For example, Fujino et al. have reported that *ASXL1* mutation confers clonal advantage on LT-HSCs in genetic mosaic mouse model, activating Akt/mTOR pathway in an epigenetics-independent manner. This results in aberrant expansion of LT-HSCs to occupy the HSC compartment. In contrast, we found the lower frequencies of Lin^{-} $CD34^{+}$ cells during FACS enrichment in *SRSF2*/*TET2* mutant groups. *TET2* is highly expressed in HSPCs and plays crucial roles in cell commitment for differentiation and promotion of the self-renewal of HSPCs (Solary et al., 2014). For example, loss of *TET2* causes an increase in immature $cKit^{+}$ Lin^{-} cells, implying that *TET2* depletion may have an impact on HSPC differentiation and development (Feng et al., 2019, Figueroa et al., 2010). A recent study also has reported that loss of *Tet2* in *Srsf2* mutant mouse model leads to less self-renewal capacity compared to others in the study by Yoshimi et al. (Yoshimi et al., 2019). In addition, the cell population composition analysis showed marked increase of the CD14 MONO cell population in *SRSF2*/*TET2* mutant groups. A recent study by Itzykson et al. reported that *TET2* deletion promotes monocyte expansion in $CD34^{+}CD38^{+}$ cells (Itzykson and Solary, 2013, Itzykson et al., 2013).

Moreover, this study unveiled transcriptional profiles in MDS HSPCs harbouring *SRSF2* mutations and co-mutation of *TET2* and/or *ASXL1* at single-cell resolution, performing DE data and GO analysis. We identified the dysregulated genes and pathways specific to the co-mutation genotype groups as well as the ones commonly shared in all the *SRSF2* mutant cases. The results implies that there are distinct biological characteristics depending on the combination of the driver mutations in MDS HSPCs.

Despite the high-resolution of gene expression level achieved upon our scRNA profiling, it should be noted that 3' RNA sequencing methods only provide a macro-overview on transcription level without detailed characteristics of the whole transcripts, such as splicing variant and mutations. We are currently generating full-length TARGET-seq (Rodriguez-Meira et al., 2019) library to study on clonal hierarchy and aberrantly splicing in MDS at the single-cell resolution. This approach may potentially contribute to unveil the combinatory effect of mutations on the alteration of splicing process and identify new key aberrantly spliced target genes.

In conclusion, we believe that these high-resolution transcriptional profiles at single-cell level will provide insights to understand characteristics of *SRSF2* mutant MDS with major co-mutations and shed light on identifying novel and specific targets/treatments for MDS patients.

6 Chapter Six - General discussion

MDS are a group of myeloid malignancies characterised by ineffective haematopoiesis and blood cytopenias. Genes involved in the splicing process are most frequently mutated, as observed in over half of MDS patients. Splicing factor gene mutations play a key role in the pathogenesis of MDS, causing aberrant splicing and dysregulated expression of many target genes. However, the molecular mechanisms through which these mutations drive the MDS phenotype are not fully understood.

In the study of myeloid malignancies harbouring splicing factor mutations, one of the main aims is to identify the key aberrantly spliced target genes and to investigate the underlying mechanism including how these key target genes result in impaired haematopoiesis, and cellular process that are dysregulated. This approach has been shown to be effective and practical to find potential targets and drugs for myeloid cancer therapy (Zhou et al., 2011, Smith et al., 2019, Yoshimi et al., 2019, Choudhary et al., 2022). A previous study in our lab has shown that *U2AF1*^{S34F} induced skipping of exon 2 of the *STRAP* gene preferentially in erythroid colonies. Here, we identified the same splicing event of *STRAP* in erythroid precursor cells of *SRSF2* mutant MDS, highlighting the correlation between the low level of *STRAP* mRNA and the presence of *SRSF2* mutations in erythroid progenitor cells. Our *in vitro* studies revealed that knockdown of *STRAP* inactivated p38 and dysregulated CSDE1-bound transcripts. The role of p38 MAPK in primary human erythroid cells to coordinate differentiation, proliferation, and apoptosis has been reported, suggesting that p38 activity is required for erythropoiesis, especially at the terminal differentiation stage including enucleation (Wang et al., 2018b, Schultze et al., 2012, Somervaille et al., 2003, Uddin et al., 2004). CSDE1 is an RNA-binding protein which is crucial for erythroid

maturation and is highly expressed in erythroblasts (Moore and von Lindern, 2018, Moore et al., 2018b). Given the significant role of STRAP during erythropoiesis, the finding that aberrant splicing of *STRAP* occurs in MDS with *SRSF2* mutations, as well as in MDS with *U2AF1*^{S34F} mutation is important (Moore et al., 2018b, Yip et al., 2017). These observations shed light on the underlying mechanism of impairments of erythropoiesis in MDS with *SRSF2* mutations. Future work to investigate hidden mechanism of aberrant splicing of *STRAP* shared in *SRSF2* and *U2AF1* mutant cases would deepen our understanding of mRNA splicing process. Interestingly, a recent study suggested that *SRSF2* mutation causes "splicing-cascade" via affecting other RNA processing and splicing genes in haematopoietic cells, which implies association between splicing factor genes to govern mRNA splicing process (Liang et al., 2018).

The final splicing fate is determined by the combinatorial effect of multiple factors and these factors interrelate each other (Chepelev and Chen, 2013, Schwartz et al., 2009, Zhang et al., 2016, Clough et al., 2022, Carrillo Oesterreich et al., 2016, Soemedi et al., 2017). In addition, malignant cells with splicing factor mutations usually have at least a few hundreds of aberrant splicing events (Pellagatti et al., 2018). Thus, it can be difficult to identify the key aberrantly spliced target genes and to fully understand the mechanism underlying splicing alterations in MDS. Against this backdrop, analysis on final splicing outcomes in a global context is pertinent. We hypothesised that a pattern of splicing alteration in global context might exist, caused by a type of splicing factor mutation. Our analysis identified that splicing factor mutations in MDS distort patterns of global splicing within each bone marrow cell population. This approach may potentially kickstart future efforts to uncover the mechanism underlying the regularity of the altered splicing patterns when the mRNA splicing process is perturbed. Further work to characterise the

combinational effect of multiple factors including splicing factors, epigenetic regulators and cell-type specificity would be of value to decipher the 'splicing code'.

With advances in genomics and transcriptomics, our knowledge on the association between the genetic landscape and disease biology in MDS has been expanding into new areas such as the characteristics of driver mutations and their co-occurrence (Ogawa, 2019, Yoshida et al., 2011, Haferlach et al., 2014). Recently, the 2022 WHO classification and ICC for myeloid neoplasms were released (Khoury et al., 2022, Arber et al., 2022). However, the criteria for the classifications varied and were mostly still based on clinical phenotypes rather than the presence of genetic lesions. We investigated the transcriptome of MDS HSPCs harbouring *SRSF2* mutations with co-mutations of *TET2* and/or *ASXL1*, at single-cell resolution. We revealed that there are distinct transcriptional profiles within MDS HSPC populations harbouring *SRSF2* mutations, depending on the combination of the co-mutations. Single-cell based studies may help to genetically define sub-groups of MDS and delineate different therapeutic approaches according to the characteristics of genotype. Future work could include performing single-cell analysis in MDS cohorts with different combinations of driver mutations, such as *SF3B1* mutant cases with co-mutations of *TET2* and/or *DNMT3A*.

The Illumina short-read sequencing technology has furthered our understanding of biology, by unveiling the regulation of gene expression and mRNA splicing process. Both wet-lab and computational developments have driven this technological evolution, allowing a richer and less biases of transcriptomic data. However, it can be difficult to assign short-sequencing reads to specific mRNA isoforms. A significant number of short-reads map ambiguously on reference data when the reads span exon-exon junctions being shared between isoforms, which is a hurdle in the study of mRNA splicing and the relevant

diseases. Emerging long-read sequencing platforms like Pacific Bioscience and Oxford Nanopore can overcome the mapping issue caused from short-read sequencing by removing the demands to assemble fragmented reads, providing complete capture of full-length mRNAs. Single-cell multi-omics platform are also being refined and this approach will provide comprehensive information associated with mRNA splicing, enabling us to integrate splicing alteration and chromatin accessibility data in specific cell populations, at single-cell resolution.

In conclusion, the results presented in this thesis deepen our understanding of the molecular pathogenesis of MDS harbouring splicing factor mutations and may help identifying new therapeutic targets.

7 Bibliography

ALIEVA, A. K., SHADRINA, M. I., FILATOVA, E. V., KARABANOV, A. V., ILLARIOSHKIN, S. N., LIMBORSKA, S. A. & SLOMINSKY, P. A. 2014. Involvement of Endocytosis and Alternative Splicing in the Formation of the Pathological Process in the Early Stages of Parkinson's Disease. *BioMed Research International,* 2014**,** 718732.

ARBER, D. A., ORAZI, A., HASSERJIAN, R., THIELE, J., BOROWITZ, M. J., LE BEAU, M. M., BLOOMFIELD, C. D., CAZZOLA, M. & VARDIMAN, J. W. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood,* 127**,** 2391-2405.

ARBER, D. A., ORAZI, A., HASSERJIAN, R. P., BOROWITZ, M. J., CALVO, K. R., KVASNICKA, H. M., WANG, S. A., BAGG, A., BARBUI, T., BRANFORD, S., BUESO-RAMOS, C. E., CORTES, J. E., DAL CIN, P., DINARDO, C. D., DOMBRET, H., DUNCAVAGE, E. J., EBERT, B. L., ESTEY, E. H., FACCHETTI, F., FOUCAR, K., GANGAT, N., GIANELLI, U., GODLEY, L. A., GÖKBUGET, N., GOTLIB, J., HELLSTRÖM-LINDBERG, E., HOBBS, G. S., HOFFMAN, R., JABBOUR, E. J., KILADJIAN, J. J., LARSON, R. A., LE BEAU, M. M., LOH, M. L., LÖWENBERG, B., MACINTYRE, E., MALCOVATI, L., MULLIGHAN, C. G., NIEMEYER, C., ODENIKE, O. M., OGAWA, S., ORFAO, A., PAPAEMMANUIL, E., PASSAMONTI, F., PORKKA, K., PUI, C. H., RADICH, J. P., REITER, A., ROZMAN, M., RUDELIUS, M., SAVONA, M. R., SCHIFFER, C. A., SCHMITT-GRAEFF, A., SHIMAMURA, A., SIERRA, J., STOCK, W. A., STONE, R. M., TALLMAN, M. S., THIELE, J., TIEN, H. F., TZANKOV, A., VANNUCCHI, A. M., VYAS, P., WEI, A. H., WEINBERG, O. K., WIERZBOWSKA, A., CAZZOLA, M., DÖHNER, H. & TEFFERI, A. 2022. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood,* 140**,** 1200-1228.

ASADA, S., GOYAMA, S., INOUE, D., SHIKATA, S., TAKEDA, R., FUKUSHIMA, T., YONEZAWA, T., FUJINO, T., HAYASHI, Y., KAWABATA, K. C., FUKUYAMA, T., TANAKA, Y., YOKOYAMA, A., YAMAZAKI, S., KOZUKA-HATA, H., OYAMA, M., KOJIMA, S., KAWAZU, M., MANO, H. & KITAMURA, T. 2018. Mutant ASXL1 cooperates with BAP1 to promote myeloid leukaemogenesis. *Nat Commun,* 9**,** 2733.

ASADA, S. & KITAMURA, T. 2019. Aberrant histone modifications induced by mutant ASXL1 in myeloid neoplasms. *Int J Hematol,* 110**,** 179-186.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D.

P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics,* 25**,** 25-29.

BARALLE, F. E. & GIUDICE, J. 2017. Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology,* 18**,** 437-451.

BARALLE, M. & BARALLE, F. E. 2018. The splicing code. *Biosystems,* 164**,** 39-48.

BARASH, Y., CALARCO, J. A., GAO, W., PAN, Q., WANG, X., SHAI, O., BLENCOWE, B. J. & FREY, B. J. 2010. Deciphering the splicing code. *Nature,* 465**,** 53.

BENNETT, J. M., CATOVSKY, D., DANIEL, M. T., FLANDRIN, G., GALTON, D. A., GRALNICK, H. R. & SULTAN, C. 1982. Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol,* 51**,** 189-99.

BHAGAT, T. D., ZHOU, L., SOKOL, L., KESSEL, R., CACERES, G., GUNDABOLU, K., TAMARI, R., GORDON, S., MANTZARIS, I., JODLOWSKI, T., YU, Y., JING, X., POLINENI, R., BHATIA, K., PELLAGATTI, A., BOULTWOOD, J., KAMBHAMPATI, S., STEIDL, U., STEIN, C., JU, W., LIU, G., KENNY, P., LIST, A., BITZER, M. & VERMA, A. 2013. miR-21 mediates hematopoietic suppression in MDS by activating TGF-beta signaling. *Blood,* 121**,** 2875-81.

BIRBRAIR, A. & FRENETTE, P. S. 2016. Niche heterogeneity in the bone marrow. *Annals of the New York Academy of Sciences,* 1370**,** 82-96.

BLENCOWE, B. J. 2017. The Relationship between Alternative Splicing and Proteomic Complexity. *Trends Biochem Sci,* 42**,** 407-408.

BONDU, S., ALARY, A.-S., LEFÈVRE, C., HOUY, A., JUNG, G., LEFEBVRE, T., ROMBAUT, D., BOUSSAID, I., BOUSTA, A., GUILLONNEAU, F., PERRIER, P., ALSAFADI, S., WASSEF, M., MARGUERON, R., ROUSSEAU, A., DROIN, N., CAGNARD, N., KALTENBACH, S., WINTER, S., KUBASCH, A.-S., BOUSCARY, D., SANTINI, V., TOMA, A., HUNAULT, M., STAMATOULLAS, A., GYAN, E., CLUZEAU, T., PLATZBECKER, U., ADÈS, L., PUY, H., STERN, M.-H., KARIM, Z., MAYEUX, P., NEMETH, E., PARK, S., GANZ, T., KAUTZ, L., KOSMIDER, O. & FONTENAY, M. 2019. A variant erythroferrone disrupts iron homeostasis in SF3B1-mutated myelodysplastic syndrome. *Science Translational Medicine,* 11**,** eaav5467.

BONNAL, S. C., LÓPEZ-OREJA, I. & VALCÁRCEL, J. 2020. Roles and mechanisms of alternative splicing in cancer — implications for care. *Nature Reviews Clinical Oncology*.

BREATHNACH, R., BENOIST, C., O'HARE, K., GANNON, F. & CHAMBON, P. 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proceedings of the National Academy of Sciences,* 75**,** 4853-4857.

BUESS, M., TERRACCIANO, L., REUTER, J., BALLABENI, P., BOULAY, J. L., LAFFER, U., METZGER, U., HERRMANN, R. & ROCHLITZ, C. 2004. STRAP is a strong predictive marker of adjuvant chemotherapy benefit in colorectal cancer. *Neoplasia,* 6**,** 813-20.

CARRILLO OESTERREICH, F., HERZEL, L., STRAUBE, K., HUJER, K., HOWARD, J. & NEUGEBAUER, KARLA M. 2016. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell,* 165**,** 372-381.

CARTOLANO, M., HUETTEL, B., HARTWIG, B., REINHARDT, R. & SCHNEEBERGER, K. 2016. cDNA Library Enrichment of Full Length Transcripts for SMRT Long Read Sequencing. *PLoS One,* 11**,** e0157779.

CASTELLO, A., FISCHER, B., EICHELBAUM, K., HOROS, R., BECKMANN, B. M., STREIN, C., DAVEY, N. E., HUMPHREYS, D. T., PREISS, T., STEINMETZ, L. M., KRIJGSVELD, J. & HENTZE, M. W. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell,* 149**,** 1393-406.

CAZZOLA, M. 2020. Myelodysplastic Syndromes. *N Engl J Med,* 383**,** 1358-1374.

CERASUOLO, A., BUONAGURO, L., BUONAGURO, F. M. & TORNESELLO, M. L. 2020. The Role of RNA Splicing Factors in Cancer: Regulation of Viral and Human Gene Expression in Human Papillomavirus-Related Cervical Cancer. *Frontiers in Cell and Developmental Biology,* 8.

CHALLEN, G. A., SUN, D., JEONG, M., LUO, M., JELINEK, J., BERG, J. S., BOCK, C., VASANTHAKUMAR, A., GU, H., XI, Y., LIANG, S., LU, Y., DARLINGTON, G. J., MEISSNER, A., ISSA, J.-P. J., GODLEY, L. A., LI, W. & GOODELL, M. A. 2012. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genetics,* 44**,** 23-31.

CHEN, J., KAO, Y.-R., SUN, D., TODOROVA, T. I., REYNOLDS, D., NARAYANAGARI, S.-R., MONTAGNA, C., WILL, B., VERMA, A. & STEIDL, U. 2019. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. *Nature Medicine,* 25**,** 103-110.

CHEN, L., CHEN, J.-Y., HUANG, Y.-J., GU, Y., QIU, J., QIAN, H., SHAO, C., ZHANG, X., HU, J., LI, H., HE, S., ZHOU, Y., ABDEL-WAHAB, O., ZHANG, D.-E. & FU, X.-D. 2018. The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. *Molecular Cell,* 69**,** 412-425.e6.

CHEN, L., KOSTADIMA, M., MARTENS, J. H. A., CANU, G., GARCIA, S. P., TURRO, E., DOWNES, K., MACAULAY, I. C., BIELCZYK-MACZYNSKA, E., COE, S., FARROW, S., POUDEL, P., BURDEN, F., JANSEN, S. B. G., ASTLE, W. J., ATTWOOD, A., BARIANA, T., DE BONO, B., BRESCHI, A., CHAMBERS, J. C., CONSORTIUM, B.,

CHOUDRY, F. A., CLARKE, L., COUPLAND, P., VAN DER ENT, M., ERBER, W. N., JANSEN, J. H., FAVIER, R., FENECH, M. E., FOAD, N., FRESON, K., VAN GEET, C., GOMEZ, K., GUIGO, R., HAMPSHIRE, D., KELLY, A. M., KERSTENS, H. H. D., KOONER, J. S., LAFFAN, M., LENTAIGNE, C., LABALETTE, C., MARTIN, T., MEACHAM, S., MUMFORD, A., NÜRNBERG, S., PALUMBO, E., VAN DER REIJDEN, B. A., RICHARDSON, D., SAMMUT, S. J., SLODKOWICZ, G., TAMURI, A. U., VASQUEZ, L., VOSS, K., WATT, S., WESTBURY, S., FLICEK, P., LOOS, R., GOLDMAN, N., BERTONE, P., READ, R. J., RICHARDSON, S., CVEJIC, A., SORANZO, N., OUWEHAND, W. H., STUNNENBERG, H. G., FRONTINI, M. & RENDON, A. 2014. Transcriptional diversity during lineage commitment of human blood progenitors. *Science,* 345**,** 1251033.

CHEN, S., BENBARCHE, S. & ABDEL-WAHAB, O. 2021. Splicing factor mutations in hematologic malignancies. *Blood,* 138**,** 599-612.

CHEPELEV, I. & CHEN, X. 2013. Alternative splicing switching in stem cell lineages. *Front Biol (Beijing),* 8**,** 50-59.

CHIBA, S. 2017. Dysregulation of TET2 in hematologic malignancies. *International Journal of Hematology,* 105**,** 17-22.

CHOUDHARY, G. S., PELLAGATTI, A., AGIANIAN, B., SMITH, M. A., BHAGAT, T. D., GORDON-MITCHELL, S., SAHU, S., PANDEY, S., SHAH, N., ALURI, S., AGGARWAL, R., AMINOV, S., SCHWARTZ, L., STEEPLES, V., BOOHER, R. N., RAMACHANDRA, M., SAMSON, M., CARBAJAL, M., PRADHAN, K., BOWMAN, T. V., PILLAI, M. M., WILL, B., WICKREMA, A., SHASTRI, A., BRADLEY, R. K., MARTELL, R. E., STEIDL, U. G., GAVATHIOTIS, E., BOULTWOOD, J., STARCZYNOWSKI, D. T. & VERMA, A. 2022. Activation of targetable inflammatory immune signaling is seen in myelodysplastic syndromes with SF3B1 mutations. *eLife,* 11**,** e78136.

CIMMINO, L., DOLGALEV, I., WANG, Y., YOSHIMI, A., MARTIN, G. H., WANG, J., NG, V., XIA, B., WITKOWSKI, M. T., MITCHELL-FLACK, M., GRILLO, I., BAKOGIANNI, S., NDIAYE-LOBRY, D., MARTÍN, M. T., GUILLAMOT, M., BANH, R. S., XU, M., FIGUEROA, M. E., DICKINS, R. A., ABDEL-WAHAB, O., PARK, C. Y., TSIRIGOS, A., NEEL, B. G. & AIFANTIS, I. 2017. Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell,* 170**,** 1079-1095.e20.

CLARK, M. B., WRZESINSKI, T., GARCIA, A. B., HALL, N. A. L., KLEINMAN, J. E., HYDE, T., WEINBERGER, D. R., HARRISON, P. J., HAERTY, W. & TUNBRIDGE, E. M. 2020.

Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene CACNA1C in human brain. *Molecular Psychiatry*, 25**,** 37-47.

CLEES, A.-S., STOLP, V., HÄUPL, B., FUHRMANN, D. C., WEMPE, F., SEIBERT, M., WEBER, S., BANNING, A., TIKKANEN, R., WILLIAMS, R., BRÜNE, B., SERVE, H., SCHNÜTGEN, F., VON METZLER, I. & KURRLE, N. 2022. Identification of the Cysteine Protease Legumain as a Potential Chronic Hypoxia-Specific Multiple Myeloma Target Gene. *Cells,* 11**,** 292.

CLEVERS, H. 2011. The cancer stem cell: premises, promises and challenges. *Nat Med,* 17**,** 313-9.

CLOUGH, C. A., PANGALLO, J., SARCHI, M., ILAGAN, J. O., NORTH, K., BERGANTINOS, R., STOLLA, M. C., NARU, J., NUGENT, P., KIM, E., STIREWALT, D. L., SUBRAMANIAM, A. R., ABDEL-WAHAB, O., ABKOWITZ, J. L., BRADLEY, R. K. & DOULATOV, S. 2022. Coordinated missplicing of TMEM14C and ABCB7 causes ring sideroblast formation in SF3B1-mutant myelodysplastic syndrome. *Blood,* 139**,** 2038-2049.

CRETU, C., SCHMITZOVÁ, J., PONCE-SALVATIERRA, A., DYBKOV, O., DE LAURENTIIS, E. I., SHARMA, K., WILL, C. L., URLAUB, H., LÜHRMANN, R. & PENA, V. 2016. Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. *Mol Cell,* 64**,** 307-319.

DAILEY, H. A. & MEISSNER, P. N. 2013. Erythroid heme biosynthesis and its disorders. *Cold Spring Harbor perspectives in medicine,* 3**,** a011676-a011676.

DAUBNER, G. M., CLÉRY, A., JAYNE, S., STEVENIN, J. & ALLAIN, F. H. 2012. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *Embo j,* 31**,** 162-74.

DAY, T. F., MEWANI, R. R., STARR, J., LI, X., CHAKRAVARTY, D., RESSOM, H., ZOU, X., EIDELMAN, O., POLLARD, H. B., SRIVASTAVA, M. & KASID, U. N. 2017. Transcriptome and Proteome Analyses of TNFAIP8 Knockdown Cancer Cells Reveal New Insights into Molecular Determinants of Cell Survival and Tumor Progression. *In:* KASID, U. & CLARKE, R. (eds.) *Cancer Gene Networks.* New York, NY: Springer New York.

DE JONG, J. L. O. 2019. Splicing up hematopoietic development. *Blood,* 133**,** 770-771.

DEBOEVER, C., GHIA, E. M., SHEPARD, P. J., RASSENTI, L., BARRETT, C. L., JEPSEN, K., JAMIESON, C. H., CARSON, D., KIPPS, T. J. & FRAZER, K. A. 2015. Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS Comput Biol,* 11**,** e1004105.

DEZERN, A. E., MALCOVATI, L. & EBERT, B. L. 2019. CHIP, CCUS, and Other Acronyms: Definition, Implications, and Impact on Practice. *American Society of Clinical Oncology Educational Book,* 400-410.

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. & BOULTWOOD, 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**,** 1092-103.

DOLATSHAD, H., PELLAGATTI, A., LIBERANTE, F. G., LLORIAN, M., REPAPI, E., STEEPLES, V., ROY, S., SCIFO, L., ARMSTRONG, R. N., SHAW, J., YIP, B. H., KILLICK, S., KUSEC, R., TAYLOR, S., MILLS, K. I., SAVAGE, K. I., SMITH, C. W. & BOULTWOOD, J. 2016. Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. *Leukemia,* 30**,** 2322-2331.

DUTTA, A., YANG, Y., LE, B. & MOHI, G. 2020. The RNA Splicing Factor U2AF1 Controls Hematopoietic Stem Cell Survival and Function. *Blood,* 136**,** 9-9.

DVINGE, H. 2018. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Letters,* 592**,** 2987-3006.

EDWARDS, C. R., RITCHIE, W., WONG, J. J., SCHMITZ, U., MIDDLETON, R., AN, X., MOHANDAS, N., RASKO, J. E. & BLOBEL, G. A. 2016. A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. *Blood,* 127**,** e24-e34.

EUN, K., HAM, S. W. & KIM, H. 2017. Cancer stem cell heterogeneity: origin and new perspectives on CSC targeting. *BMB Rep,* 50**,** 117-125.

FAN, L., LAGISETTI, C., EDWARDS, C. C., WEBB, T. R. & POTTER, P. M. 2011. Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. *ACS Chem Biol,* 6**,** 582-9.

FANG, J., VARNEY, M. & STARCZYNOWSKI, D. T. 2012. Implication of microRNAs in the pathogenesis of MDS. *Curr Pharm Des,* 18**,** 3170-9.

FENAUX, P., HAASE, D., SANTINI, V., SANZ, G. F., PLATZBECKER, U. & MEY, U. 2021. Myelodysplastic syndromes: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up^{†☆}. *Annals of Oncology,* 32**,** 142-156.

FENG, Y., LI, X., CASSADY, K., ZOU, Z. & ZHANG, X. 2019. TET2 Function in Hematopoietic Malignancies, Immune Regulation, and DNA Repair. *Frontiers in Oncology,* 9.

FIGUEROA, M. E., LUGTHART, S., LI, Y., ERPELINCK-VERSCHUEREN, C., DENG, X., CHRISTOS, P. J., SCHIFANO, E., BOOTH, J., VAN PUTTEN, W., SKRABANEK, L., CAMPAGNE, F., MAZUMDAR, M., GREALLY, J. M., VALK, P. J. M., LÖWENBERG, B., DELWEL, R. & MELNICK, A. 2010. DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia. *Cancer Cell,* 17**,** 13-27.

FINKEL, R. S., MERCURI, E., DARRAS, B. T., CONNOLLY, A. M., KUNTZ, N. L., KIRSCHNER, J., CHIRIBOGA, C. A., SAITO, K., SERVAIS, L., TIZZANO, E., TOPALOGLU, H., TULINIUS, M., MONTES, J., GLANZMAN, A. M., BISHOP, K., ZHONG, Z. J., GHEUENS, S., BENNETT, C. F., SCHNEIDER, E., FARWELL, W. & DE VIVO, D. C. 2017. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N Engl J Med,* 377**,** 1723-1732.

FISHER, C. L., BERGER, J., RANDAZZO, F. & BROCK, H. W. 2003. A human homolog of Additional sex combs, ADDITIONAL SEX COMBS-LIKE 1, maps to chromosome 20q11. *Gene,* 306**,** 115-26.

FLORES-FIGUEROA, E., GUTIÉRREZ-ESPINDOLA, G., GUERRERO-RIVERA, S., PIZZUTO-CHAVEZ, J. & MAYANI, H. 1999. Hematopoietic progenitor cells from patients with myelodysplastic syndromes: in vitro colony growth and long-term proliferation. *Leuk Res,* 23**,** 385-94.

FOLCO, E. G., COIL, K. E. & REED, R. 2011. The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. *Genes Dev,* 25**,** 440-4.

FONG, J. Y., PIGNATA, L., GOY, P.-A., KAWABATA, K. C., LEE, S. C.-W., KOH, C. M., MUSIANI, D., MASSIGNANI, E., KOTINI, A. G., PENSON, A., WUN, C. M., SHEN, Y., SCHWARZ, M., LOW, D. H. P., RIALDI, A., KI, M., WOLLMANN, H., MZOUGHI, S., GAY, F., THOMPSON, C., HART, T., BARBASH, O., LUCIANI, G. M., SZEWCZYK, M. M., WOUTERS, B. J., DELWEL, R., PAPAPETROU, E. P., BARSYTE-LOVEJOY, D., ARROWSMITH, C. H., MINDEN, M. D., JIN, J., MELNICK, A., BONALDI, T., ABDEL-WAHAB, O. & GUCCIONE, E. 2019. Therapeutic Targeting of RNA Splicing Catalysis through Inhibition of Protein Arginine Methylation. *Cancer Cell,* 36**,** 194-209.e9.

FUHRMANN, D. C., TAUSENDSCHÖN, M., WITTIG, I., STEGER, M., DING, M. G., SCHMID, T., DEHNE, N. & BRÜNE, B. 2015. Inactivation of Tristetraprolin in Chronic Hypoxia Provokes the Expression of Cathepsin B. *Molecular and Cellular Biology,* 35**,** 619-630.

FUJINO, T., GOYAMA, S., SUGIURA, Y., INOUE, D., ASADA, S., YAMASAKI, S., MATSUMOTO, A., YAMAGUCHI, K., ISOBE, Y., TSUCHIYA, A., SHIKATA, S.,

SATO, N., MORINAGA, H., FUKUYAMA, T., TANAKA, Y., FUKUSHIMA, T., TAKEDA, R., YAMAMOTO, K., HONDA, H., NISHIMURA, E. K., FURUKAWA, Y., SHIBATA, T., ABDEL-WAHAB, O., SUEMATSU, M. & KITAMURA, T. 2021. Mutant ASXL1 induces age-related expansion of phenotypic hematopoietic stem cells through activation of Akt/mTOR pathway. *Nature Communications,* 12**,** 1826.

FUJISHIMA, N., HIROKAWA, M., AIBA, N., ICHIKAWA, Y., FUJISHIMA, M., KOMATSUDA, A., SUZUKI, Y., KAWABATA, Y., MIURA, I. & SAWADA, K. 2004. Gene expression profiling of human erythroid progenitors by micro-serial analysis of gene expression. *Int J Hematol,* 80**,** 239-45.

GAITI, F., CHAMELY, P., HAWKINS, A. G., CORTÉS-LÓPEZ, M., SWETT, A. D., GANESAN, S., MOUHIEDDINE, T. H., DAI, X., KLUEGEL, L., CHEN, C., BATTA, K., BEAULAURIER, J., DRONG, A. W., HICKEY, S., DUSAJ, N., MULLOKANDOV, G., SU, J., CHALIGNÉ, R., JUUL, S., HARRINGTON, E., KNOWLES, D. A., WISEMAN, D. H., GHOBRIAL, I. M., TAYLOR, J., ABDEL-WAHAB, O. & LANDAU, D. A. 2022. Single-cell multi-omics defines the cell-type specific impact of splicing aberrations in human hematopoietic clonal outgrowths. *bioRxiv***,** 2022.06.08.495292.

GAO, K., MASUDA, A., MATSUURA, T. & OHNO, K. 2008. Human branch point consensus sequence is yUnAy. *Nucleic Acids Research,* 36**,** 2257-2267.

GAO, Y., VASIC, R. & HALENE, S. 2018. Role of alternative splicing in hematopoietic stem cells during development. *Stem Cell Investig,* 5**,** 26.

GAUTREY, H., JACKSON, C., DITTRICH, A. L., BROWELL, D., LENNARD, T. & TYSON-CAPPER, A. 2015. SRSF3 and hnRNP H1 regulate a splicing hotspot of HER2 in breast cancer cells. *RNA Biol,* 12**,** 1139-51.

GEUENS, T., BOUHY, D. & TIMMERMAN, V. 2016. The hnRNP family: insights into their role in health and disease. *Hum Genet,* 135**,** 851-67.

GOLDSTEIN, O., MEYER, K., GREENSHPAN, Y., BUJANOVER, N., FEIGIN, M., NER-GAON, H., SHAY, T. & GAZIT, R. 2017. Mapping Whole-Transcriptome Splicing in Mouse Hematopoietic Stem Cells. *Stem Cell Reports,* 8**,** 163-176.

GOZANI, O., POTASHKIN, J. & REED, R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol,* 18**,** 4752-60.

GRAVELEY, B. R. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics,* 17**,** 100-107.

GRAVELEY, B. R. & MANIATIS, T. 1998. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell,* 1**,** 765-71.

GREENBERG, P., COX, C., LEBEAU, M. M., FENAUX, P., MOREL, P., SANZ, G., SANZ, M., VALLESPI, T., HAMBLIN, T., OSCIER, D., OHYASHIKI, K., TOYAMA, K., AUL, C., MUFTI, G. & BENNETT, J. 1997. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood,* 89**,** 2079-88.

GREENBERG, P. L., TUECHLER, H., SCHANZ, J., SANZ, G., GARCIA-MANERO, G., SOLÉ, F., BENNETT, J. M., BOWEN, D., FENAUX, P., DREYFUS, F., KANTARJIAN, H., KUENDGEN, A., LEVIS, A., MALCOVATI, L., CAZZOLA, M., CERMAK, J., FONATSCH, C., LE BEAU, M. M., SLOVAK, M. L., KRIEGER, O., LUEBBERT, M., MACIEJEWSKI, J., MAGALHAES, S. M. M., MIYAZAKI, Y., PFEILSTÖCKER, M., SEKERES, M., SPERR, W. R., STAUDER, R., TAURO, S., VALENT, P., VALLESPI, T., VAN DE LOOSDRECHT, A. A., GERMING, U. & HAASE, D. 2012. Revised International Prognostic Scoring System for Myelodysplastic Syndromes. *Blood,* 120**,** 2454-2465.

GREGORY, C. J. & EAVES, A. C. 1978. Three stages of erythropoietic progenitor cell differentiation distinguished by a number of physical and biologic properties. *Blood,* 51**,** 527-37.

GRINFELD, J., NANGALIA, J., BAXTER, E. J., WEDGE, D. C., ANGELOPOULOS, N., CANTRILL, R., GODFREY, A. L., PAPAEMMANUIL, E., GUNDEM, G., MACLEAN, C., COOK, J., O'NEIL, L., O'MEARA, S., TEAGUE, J. W., BUTLER, A. P., MASSIE, C. E., WILLIAMS, N., NICE, F. L., ANDERSEN, C. L., HASSELBALCH, H. C., GUGLIELMELLI, P., MCMULLIN, M. F., VANNUCCHI, A. M., HARRISON, C. N., GERSTUNG, M., GREEN, A. R. & CAMPBELL, P. J. 2018. Classification and Personalized Prognosis in Myeloproliferative Neoplasms. *N Engl J Med,* 379**,** 1416-1430.

GURYANOVA, O. A., SHANK, K., SPITZER, B., LUCIANI, L., KOCHE, R. P., GARRETT-BAKELMAN, F. E., GANZEL, C., DURHAM, B. H., MOHANTY, A., HOERMANN, G., RIVERA, S. A., CHRAMIEC, A. G., PRONIER, E., BASTIAN, L., KELLER, M. D., TOVBIN, D., LOIZOU, E., WEINSTEIN, A. R., GONZALEZ, A. R., LIEU, Y. K., ROWE, J. M., PASTORE, F., MCKENNEY, A. S., KRIVTSOV, A. V., SPERR, W. R., CROSS, J. R., MASON, C. E., TALLMAN, M. S., ARCILA, M. E., ABDEL-WAHAB, O., ARMSTRONG, S. A., KUBICEK, S., STABER, P. B., GÖNEN, M., PAIETTA, E. M., MELNICK, A. M., NIMER, S. D., MUKHERJEE, S. & LEVINE, R. L. 2016. DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nature Medicine,* 22**,** 1488-1495.

HAFERLACH, T., NAGATA, Y., GROSSMANN, V., OKUNO, Y., BACHER, U., NAGAE, G., SCHNITTGER, S., SANADA, M., KON, A., ALPERMANN, T., YOSHIDA, K.,

ROLLER, A., NADARAJAH, N., SHIRAISHI, Y., SHIOZAWA, Y., CHIBA, K., TANAKA, H., KOEFFLER, H. P., KLEIN, H. U., DUGAS, M., ABURATANI, H., KOHLMANN, A., MIYANO, S., HAFERLACH, C., KERN, W. & OGAWA, S. 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia,* 28**,** 241-7.

HASEGAWA, M., MIURA, T., KUZUYA, K., INOUE, A., WON KI, S., HORINOUCHI, S., YOSHIDA, T., KUNOH, T., KOSEKI, K., MINO, K., SASAKI, R., YOSHIDA, M. & MIZUKAMI, T. 2011. Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. *ACS Chem Biol,* 6**,** 229-33.

HAVENS, M. A. & HASTINGS, M. L. 2016. Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res,* 44**,** 6549-63.

HAYASHI, T., OZAKI, H., SASAGAWA, Y., UMEDA, M., DANNO, H. & NIKAIDO, I. 2018. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nature Communications,* 9**,** 619.

HEATON, H., TALMAN, A. M., KNIGHTS, A., IMAZ, M., GAFFNEY, D. J., DURBIN, R., HEMBERG, M. & LAWNICZAK, M. K. N. 2020. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nature Methods,* 17**,** 615-620.

HOROS, R., IJSPEERT, H., POSPISILOVA, D., SENDTNER, R., ANDRIEU-SOLER, C., TASKESEN, E., NIERADKA, A., CMEJLA, R., SENDTNER, M., TOUW, I. P. & VON LINDERN, M. 2012. Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood,* 119**,** 262-272.

HOU, V. C., LERSCH, R., GEE, S. L., PONTHIER, J. L., LO, A. J., WU, M., TURCK, C. W., KOURY, M., KRAINER, A. R., MAYEDA, A. & CONBOY, J. G. 2002. Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch. *Embo j,* 21**,** 6195-204.

HUANG, F., SUN, J., CHEN, W., ZHANG, L., HE, X., DONG, H., WU, Y., WANG, H., LI, Z., BALL, B., KHALED, S., MARCUCCI, G. & LI, L. 2022. TET2 deficiency promotes MDS-associated leukemogenesis. *Blood Cancer Journal,* 12**,** 141.

HUH, H. D., RA, E. A., LEE, T. A., KANG, S., PARK, A., LEE, E., CHOI, J. L., JANG, E., LEE, J. E., LEE, S. & PARK, B. 2016. STRAP Acts as a Scaffolding Protein in Controlling the TLR2/4 Signaling Pathway. *Sci Rep,* 6**,** 38849.

ILAGAN, J. O., RAMAKRISHNAN, A., HAYES, B., MURPHY, M. E., ZEBARI, A. S., BRADLEY, P. & BRADLEY, R. K. 2015. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res,* 25**,** 14-26.

IMIELINSKI, M., BERGER, A. H., HAMMERMAN, P. S., HERNANDEZ, B., PUGH, T. J., HODIS, E., CHO, J., SUH, J., CAPELLETTI, M., SIVACHENKO, A., SOUGNEZ, C., AUCLAIR, D., LAWRENCE, M. S., STOJANOV, P., CIBULSKIS, K., CHOI, K., DE WAAL, L., SHARIFNIA, T., BROOKS, A., GREULICH, H., BANERJI, S., ZANDER, T., SEIDEL, D., LEENDERS, F., ANSÉN, S., LUDWIG, C., ENGEL-RIEDEL, W., STOELBEN, E., WOLF, J., GOPARJU, C., THOMPSON, K., WINCKLER, W., KWIATKOWSKI, D., JOHNSON, B. E., JÄNNE, P. A., MILLER, V. A., PAO, W., TRAVIS, W. D., PASS, H. I., GABRIEL, S. B., LANDER, E. S., THOMAS, R. K., GARRAWAY, L. A., GETZ, G. & MEYERSON, M. 2012. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell,* 150**,** 1107-20.

INOUE, D., CHEW, G. L., LIU, B., MICHEL, B. C., PANGALLO, J., D'AVINO, A. R., HITCHMAN, T., NORTH, K., LEE, S. C., BITNER, L., BLOCK, A., MOORE, A. R., YOSHIMI, A., ESCOBAR-HOYOS, L., CHO, H., PENSON, A., LU, S. X., TAYLOR, J., CHEN, Y., KADOCH, C., ABDEL-WAHAB, O. & BRADLEY, R. K. 2019. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature,* 574**,** 432-436.

INOUE, D., FUJINO, T. & KITAMURA, T. 2018. ASXL1 as a critical regulator of epigenetic marks and therapeutic potential of mutated cells. *Oncotarget,* 9**,** 35203-35204.

INOUE, D., KITAURA, J., MATSUI, H., HOU, H. A., CHOU, W. C., NAGAMACHI, A., KAWABATA, K. C., TOGAMI, K., NAGASE, R., HORIKAWA, S., SAIKA, M., MICOL, J. B., HAYASHI, Y., HARADA, Y., HARADA, H., INABA, T., TIEN, H. F., ABDEL-WAHAB, O. & KITAMURA, T. 2015. SETBP1 mutations drive leukemic transformation in ASXL1-mutated MDS. *Leukemia,* 29**,** 847-857.

INOUE, D., POLASKI, J. T., TAYLOR, J., CASTEL, P., CHEN, S., KOBAYASHI, S., HOGG, S. J., HAYASHI, Y., BELLO PINEDA, J. M., PENSON, A. V., MARABTI, E. E., ERICKSON, C., FUKUMOTO, M., YAMAZAKI, H., FUKUI, C., LU, S. X., DURHAM, B. H., LIU, B., MEHTA, S., ZAKHEIM, D., GARIPPA, R., CHEW, G.-L., MCCORMICK, F., BRADLEY, R. K. & ABDEL-WAHAB, O. 2020. ZRSR2 Mutation Induced Minor Intron Retention Drives MDS and Diverse Cancer Predisposition Via Aberrant Splicing of LZTR1. *Blood,* 136**,** 10-11.

INOUE, D., POLASKI, J. T., TAYLOR, J., CASTEL, P., CHEN, S., KOBAYASHI, S., HOGG, S. J., HAYASHI, Y., PINEDA, J. M. B., EL MARABTI, E., ERICKSON, C., KNORR, K., FUKUMOTO, M., YAMAZAKI, H., TANAKA, A., FUKUI, C., LU, S. X., DURHAM,

B. H., LIU, B., WANG, E., MEHTA, S., ZAKHEIM, D., GARIPPA, R., PENSON, A., CHEW, G.-L., MCCORMICK, F., BRADLEY, R. K. & ABDEL-WAHAB, O. 2021. Minor intron retention drives clonal hematopoietic disorders and diverse cancer predisposition. *Nature Genetics,* 53**,** 707-718.

ITZYKSON, R., KOSMIDER, O., RENNEVILLE, A., MORABITO, M., PREUDHOMME, C., BERTHON, C., ADÈS, L., FENAUX, P., PLATZBECKER, U., GAGEY, O., RAMEAU, P., MEURICE, G., ORÉAR, C., DELHOMMEAU, F., BERNARD, O. A., FONTENAY, M., VAINCHENKER, W., DROIN, N. & SOLARY, E. 2013. Clonal architecture of chronic myelomonocytic leukemias. *Blood,* 121**,** 2186-2198.

ITZYKSON, R. & SOLARY, E. 2013. An evolutionary perspective on chronic myelomonocytic leukemia. *Leukemia,* 27**,** 1441-1450.

JENKINS, J. L. & KIELKOPF, C. L. 2017. Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. *Trends in Genetics,* 33**,** 336-348.

JHA, A., AICHER, J. K., SINGH, D. & BARASH, Y. 2019. Improving interpretability of deep learning models: splicing codes as a case study. *bioRxiv***,** 700096.

JIN, L. & DATTA, P. K. 2014. Oncogenic STRAP functions as a novel negative regulator of E-cadherin and p21(Cip1) by modulating the transcription factor Sp1. *Cell Cycle,* 13**,** 3909-20.

KAIDA, D., MOTOYOSHI, H., TASHIRO, E., NOJIMA, T., HAGIWARA, M., ISHIGAMI, K., WATANABE, H., KITAHARA, T., YOSHIDA, T., NAKAJIMA, H., TANI, T., HORINOUCHI, S. & YOSHIDA, M. 2007. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol,* 3**,** 576-83.

KARNI, R., DE STANCHINA, E., LOWE, S. W., SINHA, R., MU, D. & KRAINER, A. R. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol,* 14**,** 185-93.

KASHIKAR, N. D., REINER, J., DATTA, A. & DATTA, P. K. 2010. Serine threonine receptor-associated protein (STRAP) plays a role in the maintenance of mesenchymal morphology. *Cell Signal,* 22**,** 138-49.

KATAOKA, N., MATSUMOTO, E. & MASAKI, S. 2021. Mechanistic Insights of Aberrant Splicing with Splicing Factor Mutations Found in Myelodysplastic Syndromes. *Int J Mol Sci,* 22.

KEHLENBACH, R. H., ASSHEUER, R., KEHLENBACH, A., BECKER, J. & GERACE, L. 2001. Stimulation of Nuclear Export and Inhibition of Nuclear Import by a Ran Mutant Deficient in Binding to Ran-binding Protein 1*. *Journal of Biological Chemistry,* 276**,** 14524-14531.

KESARWANI, A. K., RAMIREZ, O., GUPTA, A. K., YANG, X., MURTHY, T., MINELLA, A. C. & PILLAI, M. M. 2017. Cancer-associated SF3B1 mutants recognize otherwise inaccessible cryptic 3' splice sites within RNA secondary structures. *Oncogene,* 36**,** 1123-1133.

KFIR, N., LEV-MAOR, G., GLAICH, O., ALAJEM, A., DATTA, A., SZE, S. K., MESHORER, E. & AST, G. 2015. SF3B1 association with chromatin determines splicing outcomes. *Cell Rep,* 11**,** 618-29.

KHOURY, J. D., SOLARY, E., ABLA, O., AKKARI, Y., ALAGGIO, R., APPERLEY, J. F., BEJAR, R., BERTI, E., BUSQUE, L., CHAN, J. K. C., CHEN, W., CHEN, X., CHNG, W.-J., CHOI, J. K., COLMENERO, I., COUPLAND, S. E., CROSS, N. C. P., DE JONG, D., ELGHETANY, M. T., TAKAHASHI, E., EMILE, J.-F., FERRY, J., FOGELSTRAND, L., FONTENAY, M., GERMING, U., GUJRAL, S., HAFERLACH, T., HARRISON, C., HODGE, J. C., HU, S., JANSEN, J. H., KANAGAL-SHAMANNA, R., KANTARJIAN, H. M., KRATZ, C. P., LI, X.-Q., LIM, M. S., LOEB, K., LOGHAVI, S., MARCOGLIESE, A., MESHINCHI, S., MICHAELS, P., NARESH, K. N., NATKUNAM, Y., NEJATI, R., OTT, G., PADRON, E., PATEL, K. P., PATKAR, N., PICARSIC, J., PLATZBECKER, U., ROBERTS, I., SCHUH, A., SEWELL, W., SIEBERT, R., TEMBHARE, P., TYNER, J., VERSTOVSEK, S., WANG, W., WOOD, B., XIAO, W., YEUNG, C. & HOCHHAUS, A. 2022. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia,* 36**,** 1703-1719.

KIM, E., ILAGAN, J. O., LIANG, Y., DAUBNER, G. M., LEE, S. C., RAMAKRISHNAN, A., LI, Y., CHUNG, Y. R., MICOL, J. B., MURPHY, M. E., CHO, H., KIM, M. K., ZEBARI, A. S., AUMANN, S., PARK, C. Y., BUONAMICI, S., SMITH, P. G., DEEG, H. J., LOBRY, C., AIFANTIS, I., MODIS, Y., ALLAIN, F. H., HALENE, S., BRADLEY, R. K. & ABDEL-WAHAB, O. 2015. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell,* 27**,** 617-30.

KIM, H. K., PHAM, M. H. C., KO, K. S., RHEE, B. D. & HAN, J. 2018. Alternative splicing isoforms in health and disease. *Pflügers Archiv - European Journal of Physiology,* 470**,** 995-1016.

KINI, H. K., KONG, J. & LIEBHABER, S. A. 2014. Cytoplasmic poly(A) binding protein C4 serves a critical role in erythroid differentiation. *Molecular and cellular biology,* 34**,** 1300-1309.

KITAMURA, T. 2018. ASXL1 mutations gain a function. *Blood,* 131**,** 274-275.

KOSTI, I., RADIVOJAC, P. & MANDEL-GUTFREUND, Y. 2012. An Integrated Regulatory Network Reveals Pervasive Cross-Regulation among Transcription and Splicing Factors. *PLOS Computational Biology,* 8**,** e1002603.

KOTAKE, Y., SAGANE, K., OWA, T., MIMORI-KIYOSUE, Y., SHIMIZU, H., UESUGI, M., ISHIHAMA, Y., IWATA, M. & MIZUI, Y. 2007. Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat Chem Biol,* 3**,** 570-5.

KRECIC, A. M. & SWANSON, M. S. 1999. hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol,* 11**,** 363-71.

KWOK, B., HALL, J. M., WITTE, J. S., XU, Y., REDDY, P., LIN, K., FLAMHOLZ, R., DABBAS, B., YUNG, A., AL-HAFIDH, J., BALMERT, E., VAUPEL, C., EL HADER, C., MCGINNISS, M. J., NAHAS, S. A., KINES, J. & BEJAR, R. 2015. MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. *Blood,* 126**,** 2355-61.

LARA-PEZZI, E., GÓMEZ-SALINERO, J., GATTO, A. & GARCÍA-PAVÍA, P. 2013. The Alternative Heart: Impact of Alternative Splicing in Heart Disease. *Journal of Cardiovascular Translational Research,* 6**,** 945-955.

LAWSON, D. A., BHAKTA, N. R., KESSENBROCK, K., PRUMMEL, K. D., YU, Y., TAKAI, K., ZHOU, A., EYOB, H., BALAKRISHNAN, S., WANG, C.-Y., YASWEN, P., GOGA, A. & WERB, Z. 2015. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature,* 526**,** 131-135.

LEE, C., LOW, C. Y. B., FRANCIS, P. T., ATTEMS, J., WONG, P. T.-H., LAI, M. K. P. & TAN, M. G. K. 2016a. An isoform-specific role of FynT tyrosine kinase in Alzheimer's disease. *Journal of Neurochemistry,* 136**,** 637-650.

LEE, S. C.-W., DVINGE, H., KIM, E., CHO, H., MICOL, J.-B., CHUNG, Y. R., DURHAM, B. H., YOSHIMI, A., KIM, Y. J., THOMAS, M., LOBRY, C., CHEN, C.-W., PASTORE, A., TAYLOR, J., WANG, X., KRIVTSOV, A., ARMSTRONG, S. A., PALACINO, J., BUONAMICI, S., SMITH, P. G., BRADLEY, R. K. & ABDEL-WAHAB, O. 2016b. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nature Medicine,* 22**,** 672-678.

LEE, S. C.-W., NORTH, K., KIM, E., JANG, E., OBENG, E., LU, S. X., LIU, B., INOUE, D., YOSHIMI, A., KI, M., YEO, M., ZHANG, X. J., KIM, M. K., CHO, H., CHUNG, Y. R., TAYLOR, J., DURHAM, B. H., KIM, Y. J., PASTORE, A., MONETTE, S., PALACINO, J., SEILER, M., BUONAMICI, S., SMITH, P. G., EBERT, B. L., BRADLEY, R. K. & ABDEL-WAHAB, O. 2018. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. *Cancer Cell,* 34**,** 225-241.e8.

LERNER, M. R., BOYLE, J. A., MOUNT, S. M., WOLIN, S. L. & STEITZ, J. A. 1980. Are snRNPs involved in splicing? *Nature,* 283**,** 220-224.

LEWIS, B. P., GREEN, R. E. & BRENNER, S. E. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A,* 100**,** 189-92.

LEY, T. J., DING, L., WALTER, M. J., MCLELLAN, M. D., LAMPRECHT, T., LARSON, D. E., KANDOTH, C., PAYTON, J. E., BATY, J., WELCH, J., HARRIS, C. C., LICHTI, C. F., TOWNSEND, R. R., FULTON, R. S., DOOLING, D. J., KOBOLDT, D. C., SCHMIDT, H., ZHANG, Q., OSBORNE, J. R., LIN, L., O'LAUGHLIN, M., MCMICHAEL, J. F., DELEHAUNTY, K. D., MCGRATH, S. D., FULTON, L. A., MAGRINI, V. J., VICKERY, T. L., HUNDAL, J., COOK, L. L., CONYERS, J. J., SWIFT, G. W., REED, J. P., ALLDREDGE, P. A., WYLIE, T., WALKER, J., KALICKI, J., WATSON, M. A., HEATH, S., SHANNON, W. D., VARGHESE, N., NAGARAJAN, R., WESTERVELT, P., TOMASSON, M. H., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F., MARDIS, E. R. & WILSON, R. K. 2010a. DNMT3A Mutations in Acute Myeloid Leukemia. *New England Journal of Medicine,* 363**,** 2424-2433.

LEY, T. J., DING, L., WALTER, M. J., MCLELLAN, M. D., LAMPRECHT, T., LARSON, D. E., KANDOTH, C., PAYTON, J. E., BATY, J., WELCH, J., HARRIS, C. C., LICHTI, C. F., TOWNSEND, R. R., FULTON, R. S., DOOLING, D. J., KOBOLDT, D. C., SCHMIDT, H., ZHANG, Q., OSBORNE, J. R., LIN, L., O'LAUGHLIN, M., MCMICHAEL, J. F., DELEHAUNTY, K. D., MCGRATH, S. D., FULTON, L. A., MAGRINI, V. J., VICKERY, T. L., HUNDAL, J., COOK, L. L., CONYERS, J. J., SWIFT, G. W., REED, J. P., ALLDREDGE, P. A., WYLIE, T., WALKER, J., KALICKI, J., WATSON, M. A., HEATH, S., SHANNON, W. D., VARGHESE, N., NAGARAJAN, R., WESTERVELT, P., TOMASSON, M. H., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F., MARDIS, E. R. & WILSON, R. K. 2010b. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med,* 363**,** 2424-33.

LI, X. & WANG, C.-Y. 2021. From bulk, single-cell to spatial RNA sequencing. *International Journal of Oral Science,* 13**,** 36.

LI, Y., WANG, D., WANG, H., HUANG, X., WEN, Y., WANG, B., XU, C., GAO, J., LIU, J., TONG, J., WANG, M., SU, P., REN, S., MA, F., LI, H.-D., BRESNICK, E. H., ZHOU, J. & SHI, L. 2021. A splicing factor switch controls hematopoietic lineage specification of pluripotent stem cells. *EMBO reports,* 22**,** e50535.

LI, Y. E., XIAO, M., SHI, B., YANG, Y.-C. T., WANG, D., WANG, F., MARCIA, M. & LU, Z. J. 2017. Identification of high-confidence RNA regulatory elements by combinatorial classification of RNA–protein binding sites. *Genome Biology,* 18**,** 169.

LI, Z., CAI, X., CAI, C. L., WANG, J., ZHANG, W., PETERSEN, B. E., YANG, F. C. & XU, M. 2011. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood,* 118**,** 4509-18.

LIANG, Y., TEBALDI, T., REJESKI, K., JOSHI, P., STEFANI, G., TAYLOR, A., SONG, Y., VASIC, R., MAZIARZ, J., BALASUBRAMANIAN, K., ARDASHEVA, A., DING, A., QUATTRONE, A. & HALENE, S. 2018. SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. *Leukemia*.

LIU, H., LORENZINI, P. A., ZHANG, F., XU, S., WONG, M. S. M., ZHENG, J. & ROCA, X. 2018. Alternative splicing analysis in human monocytes and macrophages reveals MBNL1 as major regulator. *Nucleic Acids Res,* 46**,** 6069-6086.

LIU, H. X., CHEW, S. L., CARTEGNI, L., ZHANG, M. Q. & KRAINER, A. R. 2000. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol,* 20**,** 1063-71.

LIU, P., BARB, J., WOODHOUSE, K., JAMES G. TAYLOR, V., MUNSON, P. J. & RAGHAVACHARI, N. 2011. Transcriptome profiling and sequencing of differentiated human hematopoietic stem cells reveal lineage-specific expression and alternative splicing of genes. *Physiological Genomics,* 43**,** 1117-1134.

LONG, J. C. & CACERES, J. F. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J,* 417**,** 15-27.

MADACI, L., COLLE, J., VENTON, G., FARNAULT, L., LORIOD, B. & COSTELLO, R. 2021. The contribution of single-cell analysis of acute leukemia in the therapeutic strategy. *Biomarker Research,* 9**,** 50.

MAKISHIMA, H., YOSHIDA, K., NGUYEN, N., PRZYCHODZEN, B., SANADA, M., OKUNO, Y., NG, K. P., GUDMUNDSSON, K. O., VISHWAKARMA, B. A., JEREZ, A., GOMEZ-SEGUI, I., TAKAHASHI, M., SHIRAISHI, Y., NAGATA, Y., GUINTA, K., MORI, H., SEKERES, M. A., CHIBA, K., TANAKA, H., MURAMATSU, H., SAKAGUCHI, H., PAQUETTE, R. L., MCDEVITT, M. A., KOJIMA, S., SAUNTHARARAJAH, Y., MIYANO, S., SHIH, L. Y., DU, Y., OGAWA, S. & MACIEJEWSKI, J. P. 2013. Somatic SETBP1 mutations in myeloid malignancies. *Nat Genet,* 45**,** 942-6.

MAKISHIMA, H., YOSHIZATO, T., YOSHIDA, K., SEKERES, M. A., RADIVOYEVITCH, T., SUZUKI, H., PRZYCHODZEN, B., NAGATA, Y., MEGGENDORFER, M., SANADA,

M., OKUNO, Y., HIRSCH, C., KUZMANOVIC, T., SATO, Y., SATO-OTSUBO, A., LAFRAMBOISE, T., HOSONO, N., SHIRAISHI, Y., CHIBA, K., HAFERLACH, C., KERN, W., TANAKA, H., SHIOZAWA, Y., GÓMEZ-SEGUÍ, I., HUSSEINZADEH, H. D., THOTA, S., GUINTA, K. M., DIENES, B., NAKAMAKI, T., MIYAWAKI, S., SAUNTHARARAJAH, Y., CHIBA, S., MIYANO, S., SHIH, L.-Y., HAFERLACH, T., OGAWA, S. & MACIEJEWSKI, J. P. 2017. Dynamics of clonal evolution in myelodysplastic syndromes. *Nature Genetics,* 49**,** 204-212.

MALAKAR, P., CHARTARIFSKY, L., HIJA, A., LEIBOWITZ, G., GLASER, B., DOR, Y. & KARNI, R. 2016. Insulin receptor alternative splicing is regulated by insulin signaling and modulates beta cell survival. *Scientific Reports,* 6**,** 31222.

MALCOVATI, L., HELLSTRÖM-LINDBERG, E., BOWEN, D., ADÈS, L., CERMAK, J., DEL CAÑIZO, C., DELLA PORTA, M. G., FENAUX, P., GATTERMANN, N., GERMING, U., JANSEN, J. H., MITTELMAN, M., MUFTI, G., PLATZBECKER, U., SANZ, G. F., SELLESLAG, D., SKOV-HOLM, M., STAUDER, R., SYMEONIDIS, A., VAN DE LOOSDRECHT, A. A., DE WITTE, T., CAZZOLA, M. & EUROPEAN LEUKEMIA, N. 2013. Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. *Blood,* 122**,** 2943-2964.

MALCOVATI, L., STEVENSON, K., PAPAEMMANUIL, E., NEUBERG, D., BEJAR, R., BOULTWOOD, J., BOWEN, D. T., CAMPBELL, P. J., EBERT, B. L., FENAUX, P., HAFERLACH, T., HEUSER, M., JANSEN, J. H., KOMROKJI, R. S., MACIEJEWSKI, J. P., WALTER, M. J., FONTENAY, M., GARCIA-MANERO, G., GRAUBERT, T. A., KARSAN, A., MEGGENDORFER, M., PELLAGATTI, A., SALLMAN, D. A., SAVONA, M. R., SEKERES, M. A., STEENSMA, D. P., TAURO, S., THOL, F., VYAS, P., VAN DE LOOSDRECHT, A. A., HAASE, D., TÜCHLER, H., GREENBERG, P. L., OGAWA, S., HELLSTROM-LINDBERG, E. & CAZZOLA, M. 2020. SF3B1-mutant MDS as a distinct disease subtype: a proposal from the International Working Group for the Prognosis of MDS. *Blood,* 136**,** 157-170.

MANCINI, A., NIEMANN-SEYDE, S. C., PANKOW, R., EL BOUNKARI, O., KLEBBA-FÄRBER, S., KOCH, A., JAWORSKA, E., SPOONCER, E., GRUBER, A. D., WHETTON, A. D. & TAMURA, T. 2010. THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo. *BMC Biology,* 8**,** 1.

MANOHARAN, R., SEONG, H. A. & HA, H. 2018. Dual Roles of Serine-Threonine Kinase Receptor-Associated Protein (STRAP) in Redox-Sensitive Signaling Pathways Related to Cancer Development. *Oxid Med Cell Longev,* 2018**,** 5241524.

MATERA, A. G. & WANG, Z. 2014. A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology,* 15**,** 108-121.

MAZIN, P. V., KHAITOVICH, P., CARDOSO-MOREIRA, M. & KAESSMANN, H. 2021. Alternative splicing during mammalian organ development. *Nature Genetics,* 53**,** 925-934.

MCDONALD, C. M., SHIEH, P. B., ABDEL-HAMID, H. Z., CONNOLLY, A. M., CIAFALONI, E., WAGNER, K. R., GOEMANS, N., MERCURI, E., KHAN, N., KOENIG, E., MALHOTRA, J., ZHANG, W., HAN, B. & MENDELL, J. R. 2021. Open-Label Evaluation of Eteplirsen in Patients with Duchenne Muscular Dystrophy Amenable to Exon 51 Skipping: PROMOVI Trial. *J Neuromuscul Dis,* 8**,** 989-1001.

MEDINA, E. A., DELMA, C. R. & YANG, F.-C. 2022. ASXL1/2 mutations and myeloid malignancies. *Journal of Hematology & Oncology,* 15**,** 127.

MERCATANTE, D. R., BORTNER, C. D., CIDLOWSKI, J. A. & KOLE, R. 2001. Modification of alternative splicing of Bcl-x pre-mRNA in prostate and breast cancer cells. analysis of apoptosis and cell death. *J Biol Chem,* 276**,** 16411-7.

MIAN, S. A., ROUAULT-PIERRE, K., SMITH, A. E., SEIDL, T., PIZZITOLA, I., KIZILORS, A., KULASEKARARAJ, A. G., BONNET, D. & MUFTI, G. J. 2015. SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment. *Nature Communications,* 6**,** 10004.

MONTALBAN-BRAVO, G. & GARCIA-MANERO, G. 2018. Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. *Am J Hematol,* 93**,** 129-147.

MOON, H., JANG, H. N., LIU, Y., CHOI, N., OH, J., HA, J., ZHENG, X. & SHEN, H. 2019. Activation of Cryptic 3' Splice-Sites by SRSF2 Contributes to Cassette Exon Skipping. *Cells,* 8**,** 696.

MOORE, K. S. & VON LINDERN, M. 2018. RNA Binding Proteins and Regulation of mRNA Translation in Erythropoiesis. *Front Physiol,* 9**,** 910.

MOORE, K. S., YAGCI, N., VAN ALPHEN, F., MEIJER, A. B., T HOEN, P. A. C. & VON LINDERN, M. 2018a. Strap associates with Csde1 and affects expression of select Csde1-bound transcripts. *PLoS One,* 13**,** e0201690.

MOORE, K. S., YAGCI, N., VAN ALPHEN, F., PAOLINI, N. A., HOROS, R., HELD, N. M., HOUTKOOPER, R. H., VAN DEN AKKER, E., MEIJER, A. B., T HOEN, P. A. C. & VON LINDERN, M. 2018b. Csde1 binds transcripts involved in protein homeostasis and controls their expression in an erythroid cell line. *Sci Rep,* 8**,** 2628.

MUENCH, D. E., FERCHEN, K., VELU, C. S., PRADHAN, K., CHETAL, K., CHEN, X., WEIRAUCH, M. T., COLMENARES, C., VERMA, A., SALOMONIS, N. & GRIMES,

H. L. 2018. SKI controls MDS-associated chronic TGF-beta signaling, aberrant splicing, and stem cell fitness. *Blood,* 132**,** e24-e34.

MUKHERJEE, N., WESSELS, H.-H., LEBEDEVA, S., SAJEK, M., GHANBARI, M., GARZIA, A., MUNTEANU, A., YUSUF, D., FARAZI, T., HOELL, J. I., AKAT, K. M., AKALIN, A., TUSCHL, T. & OHLER, U. 2018. Deciphering human ribonucleoprotein regulatory networks. *Nucleic Acids Research,* 47**,** 570-581.

MUSIANI, D., BOK, J., MASSIGNANI, E., WU, L., TABAGLIO, T., IPPOLITO, M. R., CUOMO, A., OZBEK, U., ZORGATI, H., GHOSHDASTIDER, U., ROBINSON, R. C., GUCCIONE, E. & BONALDI, T. 2019. Proteomics profiling of arginine methylation defines PRMT5 substrate specificity. *Science Signaling,* 12**,** eaat8388.

MUTO, T., SASHIDA, G., OSHIMA, M., WENDT, G. R., MOCHIZUKI-KASHIO, M., NAGATA, Y., SANADA, M., MIYAGI, S., SARAYA, A., KAMIO, A., NAGAE, G., NAKASEKO, C., YOKOTE, K., SHIMODA, K., KOSEKI, H., SUZUKI, Y., SUGANO, S., ABURATANI, H., OGAWA, S. & IWAMA, A. 2013. Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. *J Exp Med,* 210**,** 2627-39.

NAGATA, Y. & MACIEJEWSKI, J. P. 2019. The functional mechanisms of mutations in myelodysplastic syndrome. *Leukemia,* 33**,** 2779-2794.

NAVAS, T. A., MOHINDRU, M., ESTES, M., MA, J. Y., SOKOL, L., PAHANISH, P., PARMAR, S., HAGHNAZARI, E., ZHOU, L., COLLINS, R., KERR, I., NGUYEN, A. N., XU, Y., PLATANIAS, L. C., LIST, A. A., HIGGINS, L. S. & VERMA, A. 2006. Inhibition of overactivated p38 MAPK can restore hematopoiesis in myelodysplastic syndrome progenitors. *Blood,* 108**,** 4170-7.

NGUYEN, H. D., LEONG, W. Y., LI, W., REDDY, P. N. G., SULLIVAN, J. D., WALTER, M. J., ZOU, L. & GRAUBERT, T. A. 2018. Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. *Cancer Research,* 78**,** 5363-5374.

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. & 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.

OGAWA, S. 2019. Genetics of MDS. *Blood,* 133**,** 1049-1059.

PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics,* 40**,** 1413-1415.

PANDIT, S., ZHOU, Y., SHIUE, L., COUTINHO-MANSFIELD, G., LI, H., QIU, J., HUANG, J., YEO, G. W., ARES, M., JR. & FU, X. D. 2013. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell,* 50**,** 223-35.

PAPAEMMANUIL, E. 2014. Somatic Mutations in Myelodysplastic Syndrome. *Blood,* 124**,** SCI-22-SCI-22.

PARK, U. H., YOON, S. K., PARK, T., KIM, E. J. & UM, S. J. 2011. Additional sex comb-like (ASXL) proteins 1 and 2 play opposite roles in adipogenesis via reciprocal regulation of peroxisome proliferator-activated receptor {gamma}. *J Biol Chem,* 286**,** 1354-63.

PATEL, S. B. & BELLINI, M. 2008. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res,* 36**,** 6482-93.

PELLAGATTI, A., ARMSTRONG, R. N., STEEPLES, V., SHARMA, E., REPAPI, E., SINGH, S., SANCHI, A., RADUJKOVIC, A., HORN, P., DOLATSHAD, H., ROY, S., BROXHOLME, J., LOCKSTONE, H., TAYLOR, S., GIAGOUNIDIS, A., VYAS, P., SCHUH, A., HAMBLIN, A., PAPAEMMANUIL, E., KILLICK, S., MALCOVATI, L., HENNRICH, M. L., GAVIN, A. C., HO, A. D., LUFT, T., HELLSTROM-LINDBERG, E., CAZZOLA, M., SMITH, C. W. J., SMITH, S. & BOULTWOOD, J. 2018. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood,* 132**,** 1225-1240.

PENG, H., WEN, J., ZHANG, L., LI, H., CHANG, C. C., ZU, Y. & ZHOU, X. 2012. A systematic modeling study on the pathogenic role of p38 MAPK activation in myelodysplastic syndromes. *Mol Biosyst,* 8**,** 1366-74.

PIMENTEL, H., PARRA, M., GEE, S., GHANEM, D., AN, X., LI, J., MOHANDAS, N., PACHTER, L. & CONBOY, J. G. 2014. A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res,* 42**,** 4031-42.

PLATZBECKER, U., HOFBAUER, L. C., EHNINGER, G. & HÖLIG, K. 2012. The clinical, quality of life, and economic consequences of chronic anemia and transfusion support in patients with myelodysplastic syndromes. *Leuk Res,* 36**,** 525-36.

QIN, P., PANG, Y., HOU, W., FU, R., ZHANG, Y., WANG, X., MENG, G., LIU, Q., ZHU, X., HONG, N., CHENG, T. & JIN, W. 2021. Integrated decoding hematopoiesis and leukemogenesis using single-cell sequencing and its medical implication. *Cell Discovery,* 7**,** 2.

QUATTRONE, A. & DASSI, E. 2019. The Architecture of the Human RNA-Binding Protein Regulatory Network. *iScience,* 21**,** 706-719.

REINER, J. E. & DATTA, P. K. 2011. TGF-beta-dependent and -independent roles of STRAP in cancer. *Front Biosci (Landmark Ed),* 16**,** 105-15.

RODRIGUEZ-MEIRA, A., BUCK, G., CLARK, S.-A., POVINELLI, B. J., ALCOLEA, V., LOUKA, E., MCGOWAN, S., HAMBLIN, A., SOUSOS, N., BARKAS, N., GIUSTACCHINI, A., PSAILA, B., JACOBSEN, S. E. W., THONGJUEA, S. & MEAD, A. J. 2019. Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Molecular Cell,* 73**,** 1292-1305.e8.

SAEZ, B., WALTER, M. J. & GRAUBERT, T. A. 2017. Splicing factor gene mutations in hematologic malignancies. *Blood,* 129**,** 1260-1269.

SATIJA, R., FARRELL, J. A., GENNERT, D., SCHIER, A. F. & REGEV, A. 2015. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology,* 33**,** 495-502.

SATO-OTSUBO, A., SANADA, M. & OGAWA, S. 2012. Single-nucleotide polymorphism array karyotyping in clinical practice: where, when, and how? *Semin Oncol,* 39**,** 13-25.

SCHULTZE, S. M., MAIRHOFER, A., LI, D., CEN, J., BEUG, H., WAGNER, E. F. & HUI, L. 2012. p38α controls erythroblast enucleation and Rb signaling in stress erythropoiesis. *Cell Research,* 22**,** 539-550.

SCHWARTZ, S., MESHORER, E. & AST, G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol,* 16**,** 990-5.

SEILER, M., YOSHIMI, A., DARMAN, R., CHAN, B., KEANEY, G., THOMAS, M., AGRAWAL, A. A., CALEB, B., CSIBI, A., SEAN, E., FEKKES, P., KARR, C., KLIMEK, V., LAI, G., LEE, L., KUMAR, P., LEE, S. C.-W., LIU, X., MACKENZIE, C., MEESKE, C., MIZUI, Y., PADRON, E., PARK, E., PAZOLLI, E., PENG, S., PRAJAPATI, S., TAYLOR, J., TENG, T., WANG, J., WARMUTH, M., YAO, H., YU, L., ZHU, P., ABDEL-WAHAB, O., SMITH, P. G. & BUONAMICI, S. 2018. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nature Medicine,* 24**,** 497-504.

SHARON, D., TILGNER, H., GRUBERT, F. & SNYDER, M. 2013. A single-molecule long-read survey of the human transcriptome. *Nature Biotechnology,* 31**,** 1009-1014.

SHARP, P. A. 1994. Split genes and RNA splicing. *Cell,* 77**,** 805-815.

SHASTRI, A., WILL, B., STEIDL, U. & VERMA, A. 2017. Stem and progenitor cell alterations in myelodysplastic syndromes. *Blood,* 129**,** 1586-1594.

SHEN, S., PARK, J. W., LU, Z. X., LIN, L., HENRY, M. D., WU, Y. N., ZHOU, Q. & XING, Y. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A,* 111**,** E5593-601.

SINGH, S., AHMED, D., DOLATSHAD, H., TATWAVEDI, D., SCHULZE, U., SANCHI, A., RYLEY, S., DHIR, A., CARPENTER, L., WATT, S. M., ROBERTS, D. J., ABDEL-AAL, A. M., SAYED, S. K., MOHAMED, S. A., SCHUH, A., VYAS, P., KILLICK, S., KOTINI, A. G., PAPAPETROU, E. P., WISEMAN, D. H., PELLAGATTI, A. & BOULTWOOD, J. 2020. SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications. *Leukemia,* 34**,** 2525-2530.

SMITH, M. A., CHOUDHARY, G. S., PELLAGATTI, A., CHOI, K., BOLANOS, L. C., BHAGAT, T. D., GORDON-MITCHELL, S., VON AHRENS, D., PRADHAN, K., STEEPLES, V., KIM, S., STEIDL, U., WALTER, M., FRASER, I. D. C., KULKARNI, A., SALOMONIS, N., KOMUROV, K., BOULTWOOD, J., VERMA, A. & STARCZYNOWSKI, D. T. 2019. U2AF1 mutations induce oncogenic IRAK4 isoforms and activate innate immune pathways in myeloid malignancies. *Nature Cell Biology,* 21**,** 640-650.

SOEMEDI, R., CYGAN, K. J., RHINE, C. L., GLIDDEN, D. T., TAGGART, A. J., LIN, C. L., FREDERICKS, A. M. & FAIRBROTHER, W. G. 2017. The effects of structure on pre-mRNA processing and stability. *Methods,* 125**,** 36-44.

SOLARY, E., BERNARD, O. A., TEFFERI, A., FUKS, F. & VAINCHENKER, W. 2014. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia,* 28**,** 485-496.

SOMERVAILLE, T. C. P., LINCH, D. C. & KHWAJA, A. 2003. Different levels of p38 MAP kinase activity mediate distinct biological effects in primary human erythroid progenitors. *British Journal of Haematology,* 120**,** 876-886.

SONG, J., HUSSAINI, M., MOSCINSKI, L. C., ZHANG, L., ZHANG, X., SHAO, H. & ZHANG, H. 2019. SF3B1/DNMT3A Double Mutations Are Associated with Better Prognosis Than DNMT3A Mutation Alone in Myeloid Neopasms. *Blood,* 134**,** 5029.

SONG, J., HUSSAINI, M., QIN, D., ZHANG, X., SHAO, H., ZHANG, L., GAJZER, D., BASRA, P., MOSCINSKI, L. & ZHANG, H. 2020. Comparison of SF3B1/DNMT3A Comutations With DNMT3A or SF3B1 Mutation Alone in Myelodysplastic Syndrome and Clonal Cytopenia of Undetermined Significance. *American Journal of Clinical Pathology,* 154**,** 48-56.

SPERLING, A. S., GIBSON, C. J. & EBERT, B. L. 2017. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer,* 17**,** 5-19.

SPIES, N., NIELSEN, C. B., PADGETT, R. A. & BURGE, C. B. 2009. Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell,* 36**,** 245-54.

SPINELLI, E., CAPORALE, R., BUCHI, F., MASALA, E., GOZZINI, A., SANNA, A., SASSOLINI, F., VALENCIA, A., BOSI, A. & SANTINI, V. 2012. Distinct signal transduction abnormalities and erythropoietin response in bone marrow hematopoietic cell subpopulations of myelodysplastic syndrome patients. *Clin Cancer Res,* 18**,** 3079-89.

STANLEY, R. F. & ABDEL-WAHAB, O. 2022. Dysregulation and therapeutic targeting of RNA splicing in cancer. *Nature Cancer,* 3**,** 536-546.

STARK, R., GRZELAK, M. & HADFIELD, J. 2019. RNA sequencing: the teenage years. *Nature Reviews Genetics,* 20**,** 631-656.

STEENSMA, D. P., WERMKE, M., KLIMEK, V. M., GREENBERG, P. L., FONT, P., KOMROKJI, R. S., YANG, J., BRUNNER, A. M., CARRAWAY, H. E., ADES, L., AL-KALI, A., ALONSO-DOMINGUEZ, J. M., ALFONSO-PIÉROLA, A., COOMBS, C. C., DEEG, H. J., FLINN, I., FORAN, J. M., GARCIA-MANERO, G., MARIS, M. B., MCMASTERS, M., MICOL, J.-B., DE OTEYZA, J. P., THOL, F., WANG, E. S., WATTS, J. M., TAYLOR, J., STONE, R., GOURINENI, V., MARINO, A. J., YAO, H., DESTENAVES, B., YUAN, X., YU, K., DAR, S., OHANJANIAN, L., KUIDA, K., XIAO, J., SCHOLZ, C., GUALBERTO, A. & PLATZBECKER, U. 2021. Phase I First-in-Human Dose Escalation Study of the oral SF3B1 modulator H3B-8800 in myeloid neoplasms. *Leukemia,* 35**,** 3542-3550.

STERNBURG, E. L. & KARGINOV, F. V. 2020. Global Approaches in Studying RNA-Binding Protein Interaction Networks. *Trends Biochem Sci,* 45**,** 593-603.

SUPEK, F., MIÑANA, B., VALCÁRCEL, J., GABALDÓN, T. & LEHNER, B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell,* 156**,** 1324-1335.

TAYLOR, J., KIM, S. S., STEVENSON, K. E., YODA, A., KOPP, N., LOUISSAINT, A., LEE HARRIS, N., HOCHBERG, E. P., CHEN, Y.-B., LOVITCH, S. B., DEANGELO, D. J., WADLEIGH, M., STEENSMA, D. P., MOTYCKOVA, G., STONE, R. M., NEUBERG, D. S., JARDIN, F., PICCALUGA, P. P., WEINSTOCK, D. M. & LANE, A. A. 2013. Loss-Of-Function Mutations In The Splicing Factor ZRSR2 Are Common In Blastic Plasmacytoid Dendritic Cell Neoplasm and Have Male Predominance. *Blood,* 122**,** 741.

TAYLOR, J. K., ZHANG, Q. Q., WYATT, J. R. & DEAN, N. M. 1999. Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides. *Nat Biotechnol,* 17**,** 1097-100.

TEDESCHI, A., CICIARELLO, M., MANGIACASALE, R., ROSCIOLI, E., RENSEN, W. M. & LAVIA, P. 2007. RANBP1 localizes a subset of mitotic regulatory factors on spindle microtubules and regulates chromosome segregation in human cells. *Journal of Cell Science,* 120**,** 3748-3761.

TEFFERI, A., GANGAT, N., PARDANANI, A. & CRISPINO, J. D. 2022. Myelofibrosis: Genetic Characteristics and the Emerging Therapeutic Landscape. *Cancer Research,* 82**,** 749-763.

TENG, T., TSAI, J. H., PUYANG, X., SEILER, M., PENG, S., PRAJAPATI, S., AIRD, D., BUONAMICI, S., CALEB, B., CHAN, B., CORSON, L., FEALA, J., FEKKES, P., GERARD, B., KARR, C., KORPAL, M., LIU, X., J, T. L., MIZUI, Y., PALACINO, J., PARK, E., SMITH, P. G., SUBRAMANIAN, V., WU, Z. J., ZOU, J., YU, L., CHICAS, A., WARMUTH, M., LARSEN, N. & ZHU, P. 2017. Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. *Nat Commun,* 8**,** 15522.

TODISCO, G., CREIGNOU, M., GALLÌ, A., GUGLIELMELLI, P., RUMI, E., RONCADOR, M., RIZZO, E., NANNYA, Y., PIETRA, D., ELENA, C., BONO, E., MOLTENI, E., ROSTI, V., CATRICALÁ, S., SARCHI, M., DIMITRIOU, M., UNGERSTEDT, J., VANNUCCHI, A. M., HELLSTRÖM-LINDBERG, E., OGAWA, S., CAZZOLA, M. & MALCOVATI, L. 2021. Co-mutation pattern, clonal hierarchy, and clone size concur to determine disease phenotype of SRSF2P95-mutated neoplasms. *Leukemia,* 35**,** 2371-2381.

TRAN, D. D. H., KOCH, A. & TAMURA, T. 2014. THOC5, a member of the mRNA export complex: a novel link between mRNA export machinery and signal transduction pathways in cell proliferation and differentiation. *Cell communication and signaling : CCS,* 12**,** 3-3.

TRIQUENEAUX, G., VELTEN, M., FRANZON, P., DAUTRY, F. & JACQUEMIN-SABLON, H. 1999. RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Research,* 27**,** 1926-1934.

TURUNEN, J. J., NIEMELÄ, E. H., VERMA, B. & FRILANDER, M. J. 2013. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA,* 4**,** 61-76.

UDDIN, S., AH-KANG, J., ULASZEK, J., MAHMUD, D. & WICKREMA, A. 2004. Differentiation stage-specific activation of p38 mitogen-activated protein kinase isoforms in primary human erythroid cells. *Proceedings of the National Academy of Sciences,* 101**,** 147-152.

VALENT, P. 2019. ICUS, IDUS, CHIP and CCUS: Diagnostic Criteria, Separation from MDS and Clinical Implications. *Pathobiology,* 86**,** 30-38.

VARGAS, P. D., FURUYAMA, K., SASSA, S. & SHIBAHARA, S. 2008. Hypoxia decreases the expression of the two enzymes responsible for producing linear and cyclic tetrapyrroles in the heme biosynthetic pathway. *The FEBS Journal,* 275**,** 5947-5959.

WAHL, M. C., WILL, C. L. & LÜHRMANN, R. 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell,* 136**,** 701-718.

WANG, E. T., SANDBERG, R., LUO, S., KHREBTUKOVA, I., ZHANG, L., MAYR, C., KINGSMORE, S. F., SCHROTH, G. P. & BURGE, C. B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature,* 456**,** 470-6.

WANG, G., WEN, W. X., MEAD, A. J., ROY, A., PSAILA, B. & THONGJUEA, S. 2022. Processing single-cell RNA-seq datasets using SingCellaR. *STAR Protocols,* 3**,** 101266.

WANG, K., WU, D., ZHANG, H., DAS, A., BASU, M., MALIN, J., CAO, K. & HANNENHALLI, S. 2018a. Comprehensive map of age-associated splicing changes across human tissues and their contributions to age-associated diseases. *Scientific Reports,* 8**,** 10929.

WANG, L., WU, F., SONG, Y., DUAN, Y. & JIN, Z. 2018b. Erythropoietin induces the osteogenesis of periodontal mesenchymal stem cells from healthy and periodontitis sources via activation of the p38 MAPK pathway. *Int J Mol Med,* 41**,** 829-835.

WANG, Z. & BURGE, C. B. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna,* 14**,** 802-13.

WANG, Z., ROLISH, M. E., YEO, G., TUNG, V., MAWSON, M. & BURGE, C. B. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell,* 119**,** 831-45.

WARF, M. B. & BERGLUND, J. A. 2010. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci,* 35**,** 169-78.

WEN, W. X., MEAD, A. J. & THONGJUEA, S. 2020. Technological advances and computational approaches for alternative splicing analysis in single cells. *Computational and Structural Biotechnology Journal,* 18**,** 332-343.

WEN, W. X., MEAD, A. J. & THONGJUEA, S. 2022. MARVEL: An integrated alternative splicing analysis platform for single-cell RNA sequencing data. *bioRxiv***,** 2022.08.25.505258.

WHEELER, E. C., VORA, S., MAYER, D., KOTINI, A. G., OLSZEWSKA, M., PARK, S. S., GUCCIONE, E., TERUYA-FELDSTEIN, J., SILVERMAN, L., SUNAHARA, R. K., YEO, G. W. & PAPAPETROU, E. P. 2022. Integrative RNA-omics Discovers GNAS Alternative Splicing as a Phenotypic Driver of Splicing Factor–Mutant Neoplasms. *Cancer Discovery,* 12**,** 836-855.

WILLIAMS, L., DOUCETTE, K., KARP, J. E. & LAI, C. 2021. Genetics of donor cell leukemia in acute myelogenous leukemia and myelodysplastic syndrome. *Bone Marrow Transplantation,* 56**,** 1535-1549.

WITTEN, J. T. & ULE, J. 2011. Understanding splicing regulation through RNA splicing maps. *Trends in Genetics,* 27**,** 89-97.

WOLL, P. S., KJÄLLQUIST, U., CHOWDHURY, O., DOOLITTLE, H., WEDGE, D. C., THONGJUEA, S., ERLANDSSON, R., NGARA, M., ANDERSON, K., DENG, Q., MEAD, A. J., STENSON, L., GIUSTACCHINI, A., DUARTE, S., GIANNOULATOU, E., TAYLOR, S., KARIMI, M., SCHARENBERG, C., MORTERA-BLANCO, T., MACAULAY, I. C., CLARK, S. A., DYBEDAL, I., JOSEFSEN, D., FENAUX, P., HOKLAND, P., HOLM, M. S., CAZZOLA, M., MALCOVATI, L., TAURO, S., BOWEN, D., BOULTWOOD, J., PELLAGATTI, A., PIMANDA, J. E., UNNIKRISHNAN, A., VYAS, P., GÖHRING, G., SCHLEGELBERGER, B., TOBIASSON, M., KVALHEIM, G., CONSTANTINESCU, S. N., NERLOV, C., NILSSON, L., CAMPBELL, P. J., SANDBERG, R., PAPAEMMANUIL, E., HELLSTRÖM-LINDBERG, E., LINNARSSON, S. & JACOBSEN, S. E. 2014. Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell,* 25**,** 794-808.

WONG, J. J., RITCHIE, W., EBNER, O. A., SELBACH, M., WONG, J. W., HUANG, Y., GAO, D., PINELLO, N., GONZALEZ, M., BAIDYA, K., THOENG, A., KHOO, T. L., BAILEY, C. G., HOLST, J. & RASKO, J. E. 2013. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell,* 154**,** 583-95.

WU, Q., FENG, L., WANG, Y., MAO, Y., DI, X., ZHANG, K., CHENG, S. & XIAO, T. 2022. Multi-omics analysis reveals RNA splicing alterations and their biological and clinical implications in lung adenocarcinoma. *Signal Transduction and Targeted Therapy,* 7**,** 270.

WU, S., ROMFO, C. M., NILSEN, T. W. & GREEN, M. R. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature,* 402**,** 832-5.

WU, T., HU, E., XU, S., CHEN, M., GUO, P., DAI, Z., FENG, T., ZHOU, L., TANG, W., ZHAN, L., FU, X., LIU, S., BO, X. & YU, G. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation,* 2**,** 100141.

WU, X., BEKKER-JENSEN, I. H., CHRISTENSEN, J., RASMUSSEN, K. D., SIDOLI, S., QI, Y., KONG, Y., WANG, X., CUI, Y., XIAO, Z., XU, G., WILLIAMS, K., RAPPSILBER, J., SØNDERBY, C. K., WINTHER, O., JENSEN, O. N. & HELIN, K. 2015. Tumor suppressor ASXL1 is essential for the activation of INK4B expression in response to oncogene activity and anti-proliferative signals. *Cell Research,* 25**,** 1205-1218.

YAMAMOTO, M. L., CLARK, T. A., GEE, S. L., KANG, J. A., SCHWEITZER, A. C., WICKREMA, A. & CONBOY, J. G. 2009. Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis. *Blood,* 113**,** 3363-70.

YANG, L., RAU, R. & GOODELL, M. A. 2015. DNMT3A in haematological malignancies. *Nature Reviews Cancer,* 15**,** 152-165.

YANG, X., COULOMBE-HUNTINGTON, J., KANG, S., SHEYNKMAN, GLORIA M., HAO, T., RICHARDSON, A., SUN, S., YANG, F., SHEN, YUN A., MURRAY, RYAN R., SPIROHN, K., BEGG, BRIDGET E., DURAN-FRIGOLA, M., MACWILLIAMS, A., PEVZNER, SAMUEL J., ZHONG, Q., TRIGG, SHELLY A., TAM, S., GHAMSARI, L., SAHNI, N., YI, S., RODRIGUEZ, MARIA D., BALCHA, D., TAN, G., COSTANZO, M., ANDREWS, B., BOONE, C., ZHOU, XIANGHONG J., SALEHI-ASHTIANI, K., CHARLOTEAUX, B., CHEN, ALYCE A., CALDERWOOD, MICHAEL A., ALOY, P., ROTH, FREDERICK P., HILL, DAVID E., IAKOUCHEVA, LILIA M., XIA, Y. & VIDAL, M. 2016. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell,* 164**,** 805-817.

YIP, B. H., STEEPLES, V., REPAPI, E., ARMSTRONG, R. N., LLORIAN, M., ROY, S., SHAW, J., DOLATSHAD, H., TAYLOR, S., VERMA, A., BARTENSTEIN, M., VYAS, P., CROSS, N. C., MALCOVATI, L., CAZZOLA, M., HELLSTROM-LINDBERG, E., OGAWA, S., SMITH, C. W., PELLAGATTI, A. & BOULTWOOD, J. 2017. The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. *J Clin Invest,* 127**,** 2206-2221.

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. & OGAWA, S. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature,* 478**,** 64-9.

YOSHIMI, A., LIN, K.-T., WISEMAN, D. H., RAHMAN, M. A., PASTORE, A., WANG, B., LEE, S. C.-W., MICOL, J.-B., ZHANG, X. J., DE BOTTON, S., PENARD-LACRONIQUE, V., STEIN, E. M., CHO, H., MILES, R. E., INOUE, D., ALBRECHT, T. R., SOMERVAILLE, T. C. P., BATTA, K., AMARAL, F., SIMEONI, F., WILKS, D. P., CARGO, C., INTLEKOFER, A. M., LEVINE, R. L., DVINGE, H., BRADLEY, R. K., WAGNER, E. J., KRAINER, A. R. & ABDEL-WAHAB, O. 2019. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. *Nature,* 574**,** 273-277.

YUAN, G., ZHANG, B., YANG, S., JIN, L., DATTA, A., BAE, S., CHEN, X. & DATTA, P. K. 2016. Novel role of STRAP in progression and metastasis of colorectal cancer through Wnt/beta-catenin signaling. *Oncotarget,* 7**,** 16023-37.

ZEIDAN, A. M., BEWERSDORF, J. P., BUCKSTEIN, R., SEKERES, M. A., STEENSMA, D. P., PLATZBECKER, U., LOGHAVI, S., BOULTWOOD, J., BEJAR, R., BENNETT, J. M., BORATE, U., BRUNNER, A. M., CARRAWAY, H., CHURPEK, J. E., DAVER, N. G., DELLA PORTA, M., DEZERN, A. E., EFFICACE, F., FENAUX, P., FIGUEROA, M. E., GREENBERG, P., GRIFFITHS, E. A., HALENE, S., HASSERJIAN, R. P., HOURIGAN, C. S., KIM, N., KIM, T. K., KOMROKJI, R. S., KUTCHROO, V., LIST, A. F., LITTLE, R. F., MAJETI, R., NAZHA, A., NIMER, S. D., ODENIKE, O., PADRON, E., PATNAIK, M. M., ROBOZ, G. J., SALLMAN, D. A., SANZ, G., STAHL, M., STARCZYNOWSKI, D. T., TAYLOR, J., XIE, Z., XU, M., SAVONA, M. R., WEI, A. H., ABDEL-WAHAB, O. & SANTINI, V. 2022. Finding consistency in classifications of myeloid neoplasms: a perspective on behalf of the International Workshop for Myelodysplastic Syndromes. *Leukemia*.

ZHANG, X., CHEN, M. H., WU, X., KODANI, A., FAN, J., DOAN, R., OZAWA, M., MA, J., YOSHIDA, N., REITER, J. F., BLACK, D. L., KHARCHENKO, P. V., SHARP, P. A. & WALSH, C. A. 2016. Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell,* 166**,** 1147-1162.e15.

ZHANG, Y., QIAN, J., GU, C. & YANG, Y. 2021. Alternative splicing and cancer: a systematic review. *Signal Transduction and Targeted Therapy,* 6**,** 78.

ZHAO, Y., CAI, W., HUA, Y., YANG, X. & ZHOU, J. 2022. The Biological and Clinical Consequences of RNA Splicing Factor U2AF1 Mutation in Myeloid Malignancies. *Cancers,* 14**,** 4406.

ZHOU, L., MCMAHON, C., BHAGAT, T., ALENCAR, C., YU, Y., FAZZARI, M., SOHAL, D., HEUCK, C., GUNDABOLU, K., NG, C., MO, Y., SHEN, W., WICKREMA, A., KONG, G., FRIEDMAN, E., SOKOL, L., MANTZARIS, I., PELLAGATTI, A., BOULTWOOD, J., PLATANIAS, L. C., STEIDL, U., YAN, L., YINGLING, J. M., LAHN, M. M., LIST, A., BITZER, M. & VERMA, A. 2011. Reduced SMAD7 leads to overactivation of TGF-beta signaling in MDS that can be reversed by a specific inhibitor of TGF-beta receptor I kinase. *Cancer Res,* 71**,** 955-63.

ZHOU, L., NGUYEN, A. N., SOHAL, D., YING MA, J., PAHANISH, P., GUNDABOLU, K., HAYMAN, J., CHUBAK, A., MO, Y., BHAGAT, T. D., DAS, B., KAPOUN, A. M., NAVAS, T. A., PARMAR, S., KAMBHAMPATI, S., PELLAGATTI, A., BRAUNCHWEIG, I., ZHANG, Y., WICKREMA, A., MEDICHERLA, S.,

BOULTWOOD, J., PLATANIAS, L. C., HIGGINS, L. S., LIST, A. F., BITZER, M. & VERMA, A. 2008. Inhibition of the TGF-beta receptor I kinase promotes hematopoiesis in MDS. *Blood,* 112**,** 3434-43.

8 Appendix

Table 8-1. Gene Ontology analysis for aberrantly spliced genes in CD34^{+} cell population in MDS with *SF3B1* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0008380	RNA splicing	53/716	497/20973	2.10E-13	9.11E-10	8.68E-10	SNRPN/THOC1/SUGP1/ZBTB8OS/RBM23/SREK1/QKI/DDX41/CIRBP/PPIE/HNRNPH3/CLK4/PAXBP1/TSEN2/LUC7L/RBM25/RBM6/HNRNPH1/DDX39A/SUGP2/U2AF1L4/RBM5/CCAR2/DDX47/METTL3/DDX39B/FASTK/PRPF39/YJU2B/SNU13/PRPF40A/ACIN1/PRPF3/TMBIM6/HNRNPA3/RBM4B/CLK3/SRSF11/SRRM1/SETX/SMN1/GEMIN2/PPIL3/GEMIN8/RTRAF/IVNS1ABP/RBM10/CWF19L2/MBNL2/RBM41/TSEN15/CELF1/TXNL4A	53
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	35/716	364/20973	3.96E-08	5.65E-05	5.38E-05	SNRPN/RBM23/QKI/DDX41/CIRBP/PPIE/HNRNPH3/PAXBP1/LUC7L/RBM25/RBM6/HNRNPH1/DDX39A/U2AF1L4/RBM5/METTL3/DDX39B/PRPF39/SNU13/PRPF40A/PRPF3/HNRNPA3/RBM4B/SRRM1/SETX/SMN1/GEMIN2/PPIL3/GEMIN8/RBM10/CWF19L2/MBNL2/RBM41/CELF1/TXNL4A	35
BP	GO:0000398	mRNA splicing, via spliceosome	35/716	364/20973	3.96E-08	5.65E-05	5.38E-05	SNRPN/RBM23/QKI/DDX41/CIRBP/PPIE/HNRNPH3/PAXBP1/LUC7L/RBM25/RBM6/HNRNPH1/DDX39A/U2AF1L4/RBM5/METTL3/DDX39B/PRPF39/SNU13/PRPF40A/PRPF3/HNRNPA3/RBM4B/SRRM1/SETX/SMN1/GEMIN2/PPIL3/GEMIN8/RBM10/CWF19L2/MBNL2/RBM41/CELF1/TXNL4A	35
BP	GO:0000375	RNA splicing, via transesterification reactions	35/716	368/20973	5.20E-08	5.65E-05	5.38E-05	SNRPN/RBM23/QKI/DDX41/CIRBP/PPIE/HNRNPH3/PAXBP1/LUC7L/RBM25/RBM6/HNRNPH1/DDX39A/U2AF1L4/RBM5/METTL3/DDX39B/PRPF39/SNU13/PRPF40A/PRPF3/HNRNPA3/RBM4B/SRRM1/SETX/SMN1/GEMIN2/PPIL3/GEMIN8/RBM10/CWF19L2/MBNL2/RBM41/CELF1/TXNL4A	35
BP	GO:0006399	tRNA metabolic process	22/716	216/20973	5.20E-06	0.004523	0.004308	ZBTB8OS/CARS2/MARS1/RPP30/DALRD3/TSEN2/QTRT1/NARS1/DUS1L/CARS1/WDR6/DUS3L/TARBP1/DTWD1/HARS2/TRMT2B/RTRAF/EXOSC3/CDKAL1/TRMO/TSEN15/TARS2	22
BP	GO:1903311	regulation of mRNA metabolic process	28/716	323/20973	6.76E-06	0.004896	0.004664	SERBP1/FASTKD3/RBM23/QKI/SAFB/TUT4/CIRBP/CNOT1/RBM25/RBM5/ZC3H14/METTL3/FASTK/MOV10/YBX3/PAN2/APEX1/PUM1/SLTM/SRRM1/BARD1/SECISBP2/FASTKD1/EXOSC3/RBM10/CDK11A/MBNL2/CELF1	28
BP	GO:0006360	transcription by RNA polymerase I	11/716	63/20973	8.30E-06	0.005053	0.004813	MARS1/ERCC3/TCOF1/MAF1/POLR1C/NOL11/BAZ2A/DDX11/RRN3/MACROH2A1/POLR1H	11
BP	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	35/716	460/20973	9.30E-06	0.005053	0.004813	CDC27/ANAPC5/FHIT/RNF216/SIRT2/RMND5B/RAD23A/CCAR2/PSMC1/USP14/UBR1/MAEA/HECTD3/UBXN4/MARCHF6/SPSB3/ANKZF1/CSNK1D/SYVN1/UBQLN1/FBXO9/RBCK1/RNF34/ARRB2/ATXN3/DMAC2/PPP2R5C/ZFAND2B/CRBN/HECTD1/ARMC8/CUL5/BAG6/SPOP/ERCC8	35
BP	GO:0034470	ncRNA processing	35/716	472/20973	1.61E-05	0.006994	0.006662	METTL5/ZBTB8OS/TUT4/RPL7A/RPP30/NOL8/DALRD3/ADAR/TSEN2/QTRT1/DCAF13/NGDN/DDX47/METTL3/DUS1L/NOP2/WDR55/NOL11/WDR6/NSUN5P1/DUS3L/SNU13/TARBP1/LAS1L/DTWD1/PUM1/TRMT2B/NSUN4/C1D/RTRAF/EXOSC3/INTS14/CDKAL1/TRMO/TSEN15	35
BP	GO:0006302	double-strand break repair	25/716	285/20973	1.70E-05	0.006994	0.006662	CYREN/POLL/PARP2/PPP4C/MUS81/MARF1/MRNIP/MMS22L/EPC1/DDX11/RAD51AP1/INIP/OGG1/ARID2/HMGB1/SETX/RIF1/BRD8/BLM/POLB/NABP1/CDC7/SETD2/ERCC8/UVRAG	25
BP	GO:0006900	vesicle budding from membrane	11/716	68/20973	1.77E-05	0.006994	0.006662	ARFGAP2/ANXA2/SEC16A/PDCD6/SEC24C/CSNK1D/AP3D1/PREB/PICALM/SAR1A/SEC24D	11
BP	GO:1901987	regulation of cell cycle phase transition	34/716	467/20973	3.09E-05	0.011174	0.010644	CDC27/THOC1/ANAPC5/RAD9A/SIRT2/RAD17/RRM2B/PSME2/ZWINT/ERCC3/CDK5RAP3/CCAR2/TREX1/MUS81/DDX39B/MRNIP/MIIP/ING4/CTC1/APEX1/DLG1/CDK4/RAD1/ZFYVE19/INIP/MACROH2A1/ZNF207/EIF4G1/ATR/ATRIP/BLM/KMT2E/NABP1/CDC7	34
BP	GO:0071108	protein K48-linked deubiquitination	8/716	38/20973	3.50E-05	0.011694	0.01114	USP8/USP37/USP34/OTUD5/BAP1/ATXN3/USP15/MINDY3	8
BP	GO:0006260	DNA replication	25/716	302/20973	4.48E-05	0.013894	0.013235	THOC1/POLA2/NFRKB/LIG1/POLL/USP37/RAD17/RRM2B/NOC3L/RPAIN/TREX1/MUS81/REV1/ING4/MMS22L/CTC1/DDX11/CCNE1/ATR/BLM/POLB/CDC42/RBMS1/CDC7/NASP	25
BP	GO:0006901	vesicle coating	6/716	21/20973	5.42E-05	0.015223	0.014501	ARFGAP2/SEC16A/PDCD6/CSNK1D/PREB/SAR1A	6
BP	GO:0048193	Golgi vesicle transport	26/716	325/20973	5.64E-05	0.015223	0.014501	ERGIC3/ARFGAP2/ANK1/RAB6A/SEC16A/AP1G2/CLN3/SCYL1/PDCD6/GGA3/HTT/SEC24C/SNAP23/CSNK1D/AP3D1/PREB/VPS52/SEC22C/COPB2/PITPNB/TRAPPC2/EXOC1/SAR1A/SCAMP1/SEC24D/UVRAG	26
BP	GO:0006310	DNA recombination	27/716	345/20973	5.96E-05	0.015223	0.014501	THOC1/NFRKB/CYREN/POLL/RPAIN/PPP4C/TREX1/MUS81/MRNIP/MMS22L/EPC1/APEX1/RAD51AP1/INIP/ARID2/HMGB1/SETX/RIF1/BRD8/EXOSC3/BLM/POLB/APEX2/NABP1/PTPRC/CDC7/SETD2	27
BP	GO:0044839	cell cycle G2/M phase transition	17/716	168/20973	6.51E-05	0.015273	0.014549	PKMYT1/RAD17/ARPP19/RRM2B/CDK5RAP3/MRNIP/MIIP/ING4/CTC1/CDK4/ZFYVE19/INIP/MACROH2A1/BLM/NABP1/CDC7/CALM2	17
BP	GO:1902749	regulation of cell cycle G2/M phase transition	14/716	121/20973	6.68E-05	0.015273	0.014549	RAD17/RRM2B/CDK5RAP3/MRNIP/MIIP/ING4/CTC1/CDK4/ZFYVE19/INIP/MACROH2A1/BLM/NABP1/CDC7	14
BP	GO:0009451	RNA modification	17/716	171/20973	8.12E-05	0.017638	0.016801	METTL5/CMTR1/DALRD3/ADAR/QTRT1/METTL3/DUS1L/NOP2/WDR6/NSUN5P1/DUS3L/TARBP1/DTWD1/TRMT2B/NSUN4/CDKAL1/TRMO	17
BP	GO:0061013	regulation of mRNA catabolic process	18/716	189/20973	8.85E-05	0.017946	0.017095	SERBP1/FASTKD3/TUT4/CIRBP/CNOT1/ZC3H14/METTL3/FASTK/MOV10/YBX3/PAN2/APEX1/PUM1/SECISBP2/FASTKD1/EXOSC3/RBM10/CELF1	18
BP	GO:0043488	regulation of mRNA stability	17/716	174/20973	0.000101	0.017946	0.017095	SERBP1/FASTKD3/TUT4/CIRBP/CNOT1/ZC3H14/METTL3/FASTK/MOV10/YBX3/PAN2/APEX1/PUM1/FASTKD1/EXOSC3/RBM10/CELF1	17
BP	GO:0043487	regulation of RNA stability	18/716	191/20973	0.000101	0.017946	0.017095	SERBP1/FASTKD3/TUT4/CIRBP/CNOT1/ZC3H14/METTL3/FASTK/MOV10/YBX3/DXO/PAN2/APEX1/PUM1/FASTKD1/EXOSC3/RBM10/CELF1	18
BP	GO:0090114	COPII-coated vesicle budding	7/716	33/20973	0.000103	0.017946	0.017095	SEC16A/PDCD6/SEC24C/CSNK1D/PREB/SAR1A/SEC24D	7
BP	GO:0010452	histone H3-K36 methylation	5/716	15/20973	0.000103	0.017946	0.017095	BCOR/NSD3/SETD5/SMYD2/SETD2	5
BP	GO:0006997	nucleus organization	15/716	143/20973	0.000115	0.018519	0.017641	SERBP1/NUP107/SIRT2/BAZ2A/NUMA1/EMD/WDR73/DDX11/RRN3/PITPNB/PARP11/NEMP1/PAFAH1B1/POLR1B/USPL1	15
BP	GO:0008033	tRNA processing	15/716	143/20973	0.000115	0.018519	0.017641	ZBTB8OS/RPP30/DALRD3/TSEN2/QTRT1/DUS1L/WDR6/DUS3L/TARBP1/DTWD1/TRMT2B/RTRAF/CDKAL1/TRMO/TSEN15	15
BP	GO:0048199	vesicle targeting, to, from or within Golgi	6/716	24/20973	0.000123	0.018718	0.01783	ARFGAP2/SEC16A/PDCD6/CSNK1D/PREB/SAR1A	6
BP	GO:0006903	vesicle targeting	8/716	45/20973	0.000125	0.018718	0.01783	ARFGAP2/SEC16A/WDR11/PDCD6/SNAP23/CSNK1D/PREB/SAR1A	8
BP	GO:0001510	RNA methylation	11/716	85/20973	0.000146	0.021115	0.020113	METTL5/CMTR1/DALRD3/METTL3/NOP2/WDR6/NSUN5P1/TARBP1/TRMT2B/NSUN4/TRMO	11
BP	GO:0000910	cytokinesis	17/716	181/20973	0.000163	0.022858	0.021774	SEPTIN6/SEPTIN2/USP8/BECN1/PRC1/TTC19/NUP62/CNTROB/PRPF40A/ZFYVE19/ECT2/SEPTIN7/CDC42/SETD2/KIF23/UVRAG/CALM2	17
BP	GO:0043484	regulation of RNA splicing	16/716	165/20973	0.000174	0.022915	0.021829	RBM23/QKI/CIRBP/HNRNPH3/CLK4/RBM25/HNRNPH1/RBM5/FASTK/TMBIM6/CLK3/SRRM1/SETX/RBM10/MBNL2/CELF1	16
BP	GO:0051236	establishment of RNA localization	17/716	182/20973	0.000174	0.022915	0.021829	THOC1/NUP107/QKI/ZC3H11A/DDX39A/FLOT1/NUP85/DDX39B/NUP62/HNRNPA3/ATXN2/TSC1/ATR/RTRAF/PARP11/IGF2BP2/SETD2	17
BP	GO:0000075	cell cycle checkpoint signaling	18/716	200/20973	0.000182	0.022915	0.021829	THOC1/RAD9A/RAD17/ZWINT/CDK5RAP3/CCAR2/TREX1/MUS81/DDX39B/MRNIP/RAD1/ZFYVE19/INIP/ZNF207/ATR/ATRIP/BLM/NABP1	18

BP	GO:1905508	protein localization to microtubule organizing center	7/716	36/20973	0.000185	0.022915	0.021829	NUMA1/CCDC14/CSNK1D/NUP62/CEP131/WRAP73/PCM1	7
BP	GO:0007000	nucleolus organization	5/716	17/20973	0.000201	0.022985	0.021895	SIRT2/BAZ2A/DDX11/RRN3/POLR1B	5
BP	GO:0048207	vesicle targeting, rough ER to cis-Golgi	5/716	17/20973	0.000201	0.022985	0.021895	SEC16A/PDCD6/CSNK1D/PREB/SAR1A	5
BP	GO:0048208	COPII vesicle coating	5/716	17/20973	0.000201	0.022985	0.021895	SEC16A/PDCD6/CSNK1D/PREB/SAR1A	5
BP	GO:0000086	G2/M transition of mitotic cell cycle	15/716	152/20973	0.000227	0.025305	0.024105	PKMYT1/RAD17/ARPP19/RRM2B/CDK5RAP3/MRNIP/MIIP/CTC1/CDK4/ZFYVE19/ININP/BLM/NABP1/CDC7/CALM2	15
BP	GO:0090068	positive regulation of cell cycle process	22/716	278/20973	0.000239	0.025789	0.024566	CDC27/THOC1/ANAPC5/SIRT2/RRM2B/BECN1/NUMA1/DDX39B/NUP62/RANBP1/APEX1/PLSCR1/CDK4/DDX11/RAD51AP1/MACROH2A1/EIF4G1/ECT2/KMT2E/CDC42/CDC7/KIF23	22
BP	GO:0015931	nucleobase-containing compound transport	20/716	241/20973	0.000243	0.025789	0.024566	THOC1/ABCC5/NUP107/QKI/ZC3H11A/SLC17A9/SLC35B1/DDX39A/FLOT1/NUP85/DDX39B/NUP62/HNRNPA3/ATXN2/TSC1/RTRAF/PARP11/IGF2BP2/SETD2/CD47	20
BP	GO:0007059	chromosome segregation	27/716	380/20973	0.000291	0.029444	0.028047	CDC27/ANAPC5/SIRT2/TOP3B/ZWINT/BECN1/CIAO1/MUS81/NUMA1/PRC1/NUP62/WRAP73/PLSCR1/STAG2/DDX11/PUM1/MACROH2A1/ZNF207/HJURP/CCNE1/ECT2/BAG6/BEX4/CDC42/KIF23/DSN1/UVRAG	27
BP	GO:0043414	macromolecule methylation	26/716	360/20973	0.000291	0.029444	0.028047	METTL5/BCOR/CMTR1/NSD3/DALRD3/PAXBP1/PHF20/RAB6A/NDUFAF7/METTL3/NOP2/SETD5/SMYD2/BAZ2A/WDR6/NSUN5P1/TARBP1/CXXC1/MACROH2A1/CREBBP/TRMT2B/RIF1/NSUN4/TRMO/KMT2E/SETD2	26
BP	GO:0010389	regulation of G2/M transition of mitotic cell cycle	12/716	109/20973	0.000347	0.034301	0.032674	RAD17/RRM2B/CDK5RAP3/MRNIP/MIIP/CTC1/CDK4/ZFYVE19/ININP/BLM/NABP1/CDC7	12
BP	GO:0042790	nucleolar large rRNA transcription by RNA polymerase I	5/716	19/20973	0.000357	0.034461	0.032826	MARS1/TCOF1/NOL11/DDX11/MACROH2A1	5
BP	GO:0060368	regulation of Fc receptor mediated stimulatory signaling pathway	4/716	11/20973	0.000367	0.034658	0.033014	PLSCR1/RABGEF1/PTPRC/CD47	4
BP	GO:0032259	methylation	28/716	408/20973	0.000397	0.036677	0.034937	METTL5/BCOR/CMTR1/NSD3/DALRD3/PAXBP1/PHF20/RAB6A/NDUFAF7/METTL3/NOP2/METTL17/SETD5/SMYD2/BAZ2A/WDR6/NSUN5P1/DPH5/TARBP1/CXXC1/MACROH2A1/CREBBP/TRMT2B/RIF1/NSUN4/TRMO/KMT2E/SETD2	28
BP	GO:0044772	mitotic cell cycle phase transition	31/716	471/20973	0.000408	0.036677	0.034937	CDC27/ANAPC5/PKMYT1/SIRT2/USP37/RAD17/ARPP19/RRM2B/PSME2/ZWINT/ERCC3/CDK5RAP3/TREX1/MRNIP/MIIP/CTC1/APEX1/DLG1/CDK4/ZFYVE19/ININP/ZNF207/EIF4G1/CCNE1/CUL5/BLM/KMT2E/NABP1/CDC7/NASP/CALM2	31
BP	GO:0032465	regulation of cytokinesis	11/716	96/20973	0.00043	0.036677	0.034937	BECN1/PRC1/NUP62/PRPF40A/ZFYVE19/ECT2/CDC42/SETD2/KIF23/UVRAG/CALM2	11
BP	GO:1905515	non-motile cilium assembly	9/716	67/20973	0.000436	0.036677	0.034937	TBC1D32/CSNK1D/CEP131/WRAP73/TTC8/IFT122/MKS1/SEPTIN7/PCM1	9
BP	GO:0050657	nucleic acid transport	16/716	179/20973	0.000439	0.036677	0.034937	THOC1/NUP107/QKI/ZC3H11A/DDX39A/FLOT1/NUP85/DDX39B/NUP62/HNRNPA3/ATXN2/TSC1/RTRAF/PARP11/IGF2BP2/SETD2	16
BP	GO:0050658	RNA transport	16/716	179/20973	0.000439	0.036677	0.034937	THOC1/NUP107/QKI/ZC3H11A/DDX39A/FLOT1/NUP85/DDX39B/NUP62/HNRNPA3/ATXN2/TSC1/RTRAF/PARP11/IGF2BP2/SETD2	16
BP	GO:2001020	regulation of response to DNA damage stimulus	21/716	272/20973	0.000456	0.037398	0.035625	THOC1/RAD9A/CYREN/CD44/CCAR2/PPP4C/SMYD2/DDX39B/MRNIP/ING4/EPC1/DDX11/RAD51AP1/OGG1/ARID2/HMGB1/RIF1/ATR/BRD8/SETD2/ERCC8	21
BP	GO:0006402	mRNA catabolic process	21/716	274/20973	0.000502	0.040418	0.038501	SERBP1/FASTKD3/TUT4/SKIV2L/CIRBP/CNOT1/ZC3H14/METTL3/EDC4/FASTK/MOV10/YBX3/DXO/PAN2/APEX1/PUM1/SECISBP2/FASTKD1/EXOSC3/RBM10/CELF1	21
BP	GO:0090305	nucleic acid phosphodiester bond hydrolysis	21/716	275/20973	0.000527	0.041623	0.039649	RAD9A/CNOT1/RPP30/TSEN2/ERCC3/EDC4/CSTF2/DXO/PAN2/APEX1/PLSCR1/LAS1L/RAD1/REXO2/OGG1/POLR1H/FIP1L1/EXOSC3/POLB/APEX2/TSEN15	21
BP	GO:0006284	base-excision repair	7/716	43/20973	0.000579	0.044952	0.04282	LIG1/POLL/PARP2/APEX1/OGG1/POLB/APEX2	7
BP	GO:0010212	response to ionizing radiation	14/716	149/20973	0.000598	0.045566	0.043405	RAD9A/TREX1/MRNIP/RAD1/RAD51AP1/ININP/BARD1/ECT2/ATR/BLM/POLB/NABP1/PTPRC/ERCC8	14
CC	GO:0005819	spindle	39/727	434/21993	1.81E-08	1.06E-05	9.38E-06	CDC27/SEPTIN6/SEPTIN2/ANAPC5/TAF1D/PTPN7/SIRT2/TUBGCP4/DYNLL1/CIAO1/CCAR2/ATAT1/NUP85/HAUS7/MAEA/NUMA1/CSNK1D/PRC1/EMD/NUP62/WDR73/WRAP73/TUBGCP6/STAG2/DDX11/YPEL5/ZNF207/RIF1/ECT2/TFDP2/RTRAF/POLB/BEX4/SEPTIN7/CDC42/CDC7/PAFAH1B1/KIF23/CALM2	39
CC	GO:0005681	spliceosomal complex	26/727	229/21993	4.68E-08	1.37E-05	1.21E-05	SNRPN/SUGP1/SREK1/DDX41/CIRBP/PPIE/HNRNPH3/ADAR/LUC7L/HNRNPH1/U2AF1L4/RBM5/DDX39B/PRPF39/YJU2B/SNU13/PRPF40A/PRPF3/HNRNPA3/SRRM1/GEMIN2/PPIL3/IVNS1ABP/CWF19L2/RBM41/TXNL4A	26
CC	GO:0072686	mitotic spindle	19/727	178/21993	7.39E-06	0.001365	0.001204	CDC27/TAF1D/PTPN7/SIRT2/DYNLL1/ATAT1/HAUS7/NUMA1/PRC1/NUP62/WRAP73/STAG2/YPEL5/ECT2/TFDP2/RTRAF/CDC42/CDC7/KIF23	19
CC	GO:0034399	nuclear periphery	16/727	134/21993	9.32E-06	0.001365	0.001204	THOC1/NUP107/MATR3/MBD1/NARF/MAEA/NUMA1/PRPF40A/STAG2/CXXC1/ATXN3/OGG1/SRRM1/BLM/TINF2/ERCC8	16
CC	GO:0016363	nuclear matrix	14/727	113/21993	2.19E-05	0.002287	0.002017	THOC1/MATR3/MBD1/MAEA/NUMA1/PRPF40A/STAG2/CXXC1/ATXN3/OGG1/SRRM1/BLM/TINF2/ERCC8	14
CC	GO:0030496	midbody	20/727	210/21993	2.34E-05	0.002287	0.002017	SEPTIN6/SEPTIN2/SAFB/SIRT2/USP8/ANXA2/GNL3/PRC1/TTC19/NUP62/ANKRD54/DDX11/YPEL5/ZFYVE19/ECT2/EXOC1/SEPTIN7/CDC42/KIF23/UVRAG	20
CC	GO:0000151	ubiquitin ligase complex	25/727	319/21993	6.55E-05	0.005482	0.004835	CDC27/ANAPC5/BCOR/FBXO21/ABTB1/DCAF13/PDCD6/UBR1/MAEA/MARCHF6/SPSB3/SYVN1/FBXO9/RBCK1/TMEM183A/YPEL5/DMAC2/DCAF8/BARD1/CRBN/ARMC8/CUL5/RANBP10/SPOP/ERCC8	25
CC	GO:0098687	chromosomal region	29/727	405/21993	8.97E-05	0.006569	0.005794	SEPTIN6/THOC1/SEPTIN2/NUP107/SIRT2/PPP1R12A/RAD17/DYNLL1/ZWINT/NGDN/NUP85/CTC1/APEX1/STAG2/RAD51AP1/MACROH2A1/ZNF207/HJURP/PPP2R5C/SETX/RIF1/ATR/BLM/TINF2/SEPTIN7/NABP1/PAFAH1B1/DSN1/UVRAG	29
CC	GO:0035770	ribonucleoprotein granule	23/727	301/21993	0.000184	0.011965	0.010553	LARP4/FASTKD3/TUT4/CIRBP/CNOT1/EDC4/FASTK/MOV10/TIAL1/PAN2/SHFL/STAU2/HNRNPA3/PUM1/ATXN2/EIF4G1/SMN1/BARD1/FASTKD1/TRIM5/UBAP2L/CDC42/CELF1	23
CC	GO:0016607	nuclear speck	32/727	492/21993	0.000234	0.013716	0.012098	THOC1/SREK1/PPIE/PPP4R3A/RREB1/MBD1/RBM25/DDX39A/U2AF1L4/STK19/NOC3L/ZC3H14/METTL3/BAZ2A/DDX39B/TRIM22/APEX1/PRPF40A/ACIN1/PRPF3/RNF34/CXXC1/RBM4B/CLK3/SRSF11/OGG1/SRRM1/BARD1/HECTD1/RBM10/KMT2E/SPOP	32
CC	GO:0032153	cell division site	10/727	79/21993	0.000269	0.014354	0.01266	SEPTIN6/SEPTIN2/PSTPIP1/TUBGCP4/MAEA/WDR73/TUBGCP6/ZFYVE19/ECT2/SEPTIN7	10
CC	GO:0005963	magnesium-dependent protein serine/threonine phosphatase complex	6/727	30/21993	0.000385	0.018069	0.015937	PPP4R3A/PPP4C/PPP2R2A/PTPA/PPP2R5C/PPP2R1B	6
CC	GO:0010494	cytoplasmic stress granule	10/727	83/21993	0.000404	0.018069	0.015937	LARP4/CIRBP/MOV10/TIAL1/STAU2/PUM1/ATXN2/EIF4G1/UBAP2L/CELF1	10
CC	GO:0035869	ciliary transition zone	9/727	69/21993	0.000432	0.018069	0.015937	SEPTIN2/CFAP410/WDR19/CEP131/AHI1/TTC8/IFT122/MKS1/PCM1	9

CC	GO:0000793	condensed chromosome	20/727	266/21993	0.00058	0.022641	0.019969	SEPTIN6/SEPTIN2/NUP107/RAD9A/TOP3B/PPP1R12A/DYNLL1/ZWINT/NUP85/STAG2/RAD1/MACROH2A1/ZNF207/HJURP/HMGB1/RIF1/BLM/SEPTIN7/PAFAH1B1/DSN1	20
CC	GO:0070971	endoplasmic reticulum exit site	7/727	45/21993	0.000636	0.023311	0.02056	HLA-A/SEC16A/PDCD6/SEC24C/PREB/SAR1A/SEC24D	7
CC	GO:0030127	COPII vesicle coat	4/727	15/21993	0.001208	0.041637	0.036723	PDCD6/SEC24C/SAR1A/SEC24D	4
CC	GO:0036064	ciliary basal body	14/727	168/21993	0.001422	0.043774	0.038608	TTLL4/CFAP410/WDR11/CSNK1D/CEP131/WRAP73/AHI1/TTC8/IFT122/ERC1/MKS1/SDCCAG8/PCM1/ODF2L	14
CC	GO:0031965	nuclear membrane	22/727	329/21993	0.001488	0.043774	0.038608	NUP107/MLX/MATR3/PHF20/SMOX/NUP85/P2RX1/EMD/NUP62/EPC1/STAU2/CDK4/PHF11/TMEM120A/TMEM97/RIF1/MINDY3/NEMP1/PCM1/PAFAH1B1/OSBPL8/TXNL4A	22
CC	GO:0071004	U2-type prespliceosome	4/727	16/21993	0.001569	0.043774	0.038608	SNRPN/LUC7L/PRPF39/PRPF40A	4
CC	GO:0071010	prespliceosome	4/727	16/21993	0.001569	0.043774	0.038608	SNRPN/LUC7L/PRPF39/PRPF40A	4
MF	GO:0140098	catalytic activity, acting on RNA	47/730	453/20738	3.76E-11	2.84E-08	2.61E-08	METTL5/CARS2/DDX55/DDX41/TUT4/DDX31/SKIV2L/MARS1/CNOT1/CMTR1/RPP30/DALRD3/TSEN2/QTRT1/DDX39A/NARS1/DDX47/METTL3/DUS1L/RPAP1/NOP2/POLR1C/DDX39B/CARS1/MOV10/DDX24/DXO/PAN2/APEX1/DUS3L/TARBP1/DTWD1/HARS2/REXO2/DDX50/POLR1H/NT5C3A/TRMT2B/NSUN4/EXOSC3/CDKAL1/TRMO/DDX58/EIF4A2/TARS2/POLR1B/PTRH2	47
MF	GO:0140101	catalytic activity, acting on a tRNA	19/730	156/20738	2.56E-06	0.000853	0.000785	CARS2/MARS1/RPP30/DALRD3/TSEN2/QTRT1/NARS1/DUS1L/CARS1/DUS3L/TARBP1/DTWD1/HARS2/NT5C3A/TRMT2B/CDKAL1/TRMO/TARS2/PTRH2	19
MF	GO:0140097	catalytic activity, acting on DNA	23/730	223/20738	4.28E-06	0.000853	0.000785	HLTF/RAD9A/LIG1/POLL/TOP3B/RAD17/ALKBH2/ERCC3/TREX1/MUS81/REV1/APEX1/RAD1/DDX11/REXO2/SMARCA2/OGG1/CHD8/SETX/BLM/POLB/APEX2/BPTF	23
MF	GO:0004386	helicase activity	20/730	177/20738	4.52E-06	0.000853	0.000785	HLTF/DDX55/DDX41/DDX31/SKIV2L/DDX39A/ERCC3/DDX47/DDX39B/MOV10/DDX24/DDX11/DDX50/SMARCA2/CHD8/ERCC6L2/SETX/BLM/DDX58/EIF4A2	20
MF	GO:0008186	ATP-dependent activity, acting on RNA	13/730	100/20738	4.86E-05	0.007323	0.006738	DDX55/DDX41/DDX31/SKIV2L/DDX39A/DDX47/DDX39B/MOV10/DDX24/DDX11/DDX50/DDX58/EIF4A2	13
MF	GO:0003684	damaged DNA binding	10/730	68/20738	0.000126	0.014027	0.012905	ERCC3/RAD23A/REV1/APEX1/RAD1/OGG1/HMGB1/CREBBP/BLM/POLB	10
MF	GO:0008408	3'-5' exonuclease activity	9/730	57/20738	0.000156	0.014027	0.012905	RAD9A/CNOT1/TREX1/PAN2/APEX1/RAD1/REXO2/EXOSC3/APEX2	9
MF	GO:0003724	RNA helicase activity	12/730	98/20738	0.000168	0.014027	0.012905	DDX55/DDX41/DDX31/SKIV2L/DDX39A/DDX47/DDX39B/MOV10/DDX24/DDX50/DDX58/EIF4A2	12
MF	GO:0004529	exodeoxyribonuclease activity	6/730	25/20738	0.000186	0.014027	0.012905	RAD9A/TREX1/APEX1/RAD1/REXO2/APEX2	6
MF	GO:0016895	exodeoxyribonuclease activity, producing 5'-phosphomonoesters	6/730	25/20738	0.000186	0.014027	0.012905	RAD9A/TREX1/APEX1/RAD1/REXO2/APEX2	6
MF	GO:0016796	exonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters	9/730	61/20738	0.000266	0.018234	0.016775	RAD9A/CNOT1/TREX1/PAN2/APEX1/RAD1/REXO2/EXOSC3/APEX2	9
MF	GO:0003727	single-stranded RNA binding	11/730	91/20738	0.000349	0.021909	0.020157	LARP4/PPIE/HNRNPH1/ZC3H14/LONP1/ILF3/CNBP/DDX11/EIF4H/DDX58/RBMS1	11
MF	GO:0046975	histone methyltransferase activity (H3-K36 specific)	4/730	11/20738	0.000412	0.022206	0.020429	NSD3/SETD5/SMYD2/SETD2	4
MF	GO:0051880	G-quadruplex DNA binding	4/730	11/20738	0.000412	0.022206	0.020429	LONP1/CNBP/DDX11/BLM	4
MF	GO:0004843	cysteine-type deubiquitinase activity	12/730	110/20738	0.000497	0.024962	0.022965	USP8/USP21/USP37/USP11/USP34/USP14/OTUD5/BAP1/USP16/ATXN3/USP15/MINDY3	12
MF	GO:0003906	DNA-(apurinic or apyrimidinic site) endonuclease activity	4/730	12/20738	0.000601	0.028335	0.026069	APEX1/OGG1/POLB/APEX2	4
MF	GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	15/730	167/20738	0.000849	0.036274	0.033372	METTL5/CMTR1/NSD3/NDUFAF7/METTL3/NOP2/SETD5/SMYD2/DPH5/TARBP1/CXXC1/TRMT2B/NSUN4/TRMO/SETD2	15
MF	GO:0101005	deubiquitinase activity	12/730	117/20738	0.000866	0.036274	0.033372	USP8/USP21/USP37/USP11/USP34/USP14/OTUD5/BAP1/USP16/ATXN3/USP15/MINDY3	12
MF	GO:0004527	exonuclease activity	10/730	88/20738	0.001043	0.04127	0.037969	RAD9A/CNOT1/TREX1/DXO/PAN2/APEX1/RAD1/REXO2/EXOSC3/APEX2	10
MF	GO:0008168	methyltransferase activity	18/730	226/20738	0.001119	0.04127	0.037969	METTL5/CMTR1/NSD3/AMT/NDUFAF7/METTL3/NOP2/METTL17/SETD5/SMYD2/NSUN5P1/DPH5/TARBP1/CXXC1/TRMT2B/NSUN4/TRMO/SETD2	18
MF	GO:0008296	3'-5'-exodeoxyribonuclease activity	4/730	14/20738	0.001149	0.04127	0.037969	TREX1/APEX1/REXO2/APEX2	4
MF	GO:0140030	modification-dependent protein binding	16/730	193/20738	0.001364	0.046755	0.043014	GLYR1/RAD23A/HDAC6/BAZ2A/ANKRD13D/TAB3/UBQLN1/ING4/CXXC1/CHD8/ZFAND2B/USP15/ATRIP/BAG6/KMT2E/BPTF	16
MF	GO:0016829	lyase activity	17/730	213/20738	0.001484	0.048501	0.044621	UXS1/POLL/TSEN2/PDXDC1/PTGES2/HSD17B4/PTS/APMAP/OGG1/ADCY3/HMGB1/ME2/L3HYPDH/PCK2/ENO2/POLB/DGLUCY	17
MF	GO:0032182	ubiquitin-like protein binding	12/730	125/20738	0.001544	0.048501	0.044621	SERBP1/SIRT2/RAD23A/HDAC6/GGA3/TAB3/NUP62/RBCK1/USP16/ZFAND2B/USPL1/RNF111	12

Table 8-2. Gene Ontology analysis for aberrantly spliced genes in monocytic cell population in MDS with *SF3B1* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0018205	peptidyl-lysine modification	27/366	434/20973	1.36E-08	5.12E-05	4.46E-05	MBIP/PIH1D1/ARNT/CRTC2/ATXN7L3/KMT2C/EZH1/KDM3A/EPC1/KAT6B/KDM4C/PHF20/GLYR1/ARID4A/ASH2L/NFYB/LIAS/DNMT1/SETD5/RIF1/CAPN3/UHRF2/MAP3K7/KMT2E/ATF2/PRKAA1/SUPT6H	27
BP	GO:0032259	methylation	25/366	408/20973	6.37E-08	7.74E-05	6.74E-05	METTL5/PIH1D1/KMT2C/MTO1/EZH1/BUD23/KDM3A/METTL17/KDM4C/PHF20/MPHOSPH8/TARBP1/ARID4A/ASH2L/MTRR/NFYB/TRMO/DNMT1/METTL13/SETD5/RIF1/PRMT2/KMT2E/PAF1/SUPT6H	25
BP	GO:0031056	regulation of histone modification	17/366	198/20973	7.76E-08	7.74E-05	6.74E-05	MBIP/PIH1D1/KDM3A/PPP4C/PINK1/KDM4C/PHF20/LRRK2/GLYR1/NFYB/DNMT1/SAP30BP/SETD5/RIF1/KMT2E/PAF1/SUPT6H	17
BP	GO:0016571	histone methylation	16/366	178/20973	9.96E-08	7.74E-05	6.74E-05	PIH1D1/KMT2C/EZH1/KDM3A/KDM4C/PHF20/ARID4A/ASH2L/NFYB/DNMT1/SETD5/RIF1/PRMT2/KMT2E/PAF1/SUPT6H	16
BP	GO:0043414	macromolecule methylation	23/366	360/20973	1.03E-07	7.74E-05	6.74E-05	METTL5/PIH1D1/KMT2C/MTO1/EZH1/BUD23/KDM3A/KDM4C/PHF20/MPHOSPH8/TARBP1/ARID4A/ASH2L/MTRR/NFYB/TRMO/DNMT1/SETD5/RIF1/PRMT2/KMT2E/PAF1/SUPT6H	23
BP	GO:0008380	RNA splicing	27/366	497/20973	2.16E-07	0.000136	0.000118	SNRPN/THOC1/TIA1/CELF2/CCNL1/PPIE/NSRP1/FAM50A/EFTUD2/SUGP2/LUC7L/U2AF1/PABPC1/DDX46/HNRNPM/CLK1/CLK3/SRRM1/HNRNPA3/RBM5/PPIL3/ACIN1/DHX9/CLK4/RBM41/SUPT6H/WBP11	27
BP	GO:0034968	histone lysine methylation	14/366	145/20973	2.62E-07	0.000141	0.000123	PIH1D1/KMT2C/EZH1/KDM3A/KDM4C/PHF20/ARID4A/ASH2L/NFYB/DNMT1/SETD5/RIF1/KMT2E/SUPT6H	14
BP	GO:0006979	response to oxidative stress	26/366	487/20973	5.08E-07	0.000239	0.000208	APTX/STAU1/MAPK9/ARNT/THG1L/MPV17/APP/CAPN2/APEX1/CD36/PINK1/GSS/HNRNPM/AKT1/ALOX5/PPIF/ERCC6L2/MCTP1/LRRK2/LIAS/GCH1/ANXA1/ATF2/PRKAA1/STX4/SLC8A1	26
BP	GO:0018022	peptidyl-lysine methylation	14/366	161/20973	9.43E-07	0.000395	0.000344	PIH1D1/KMT2C/EZH1/KDM3A/KDM4C/PHF20/ARID4A/ASH2L/NFYB/DNMT1/SETD5/RIF1/KMT2E/SUPT6H	14
BP	GO:1903311	regulation of mRNA metabolic process	20/366	323/20973	1.15E-06	0.000434	0.000378	SERBP1/TIA1/CELF2/NSRP1/APEX1/SLC11A1/PABPC1/HNRNPM/AKT1/CNOT2/SRRM1/SECISBP2/RBM5/SAFB/EXOSC9/DHX9/HNRNPAB/PAF1/CPSF6/SUPT6H	20
BP	GO:0006479	protein methylation	16/366	220/20973	1.74E-06	0.000547	0.000477	PIH1D1/KMT2C/EZH1/KDM3A/KDM4C/PHF20/ARID4A/ASH2L/NFYB/DNMT1/SETD5/RIF1/PRMT2/KMT2E/PAF1/SUPT6H	16
BP	GO:0008213	protein alkylation	16/366	220/20973	1.74E-06	0.000547	0.000477	PIH1D1/KMT2C/EZH1/KDM3A/KDM4C/PHF20/ARID4A/ASH2L/NFYB/DNMT1/SETD5/RIF1/PRMT2/KMT2E/PAF1/SUPT6H	16
BP	GO:0034470	ncRNA processing	24/366	472/20973	3.27E-06	0.000948	0.000826	METTL5/PIH1D1/INTS8/INTS11/THG1L/INTS6L/ADAR/MRPS11/MTO1/BUD23/RPS16/POP5/WDR74/RPP21/RIOK3/RIOK1/ELP5/TARBP1/TRMO/EXOSC9/DTWD1/NCOR1/NOL8/WBP11	24
BP	GO:0000302	response to reactive oxygen species	15/366	213/20973	5.44E-06	0.001382	0.001204	APTX/MAPK9/MPV17/CAPN2/APEX1/CD36/PINK1/AKT1/PPIF/ERCC6L2/LRRK2/GCH1/ANXA1/PRKAA1/SLC8A1	15
BP	GO:0002713	negative regulation of B cell mediated immunity	5/366	16/20973	5.87E-06	0.001382	0.001204	THOC1/PTPN6/FCGR2B/CD55/CD46	5
BP	GO:0002890	negative regulation of immunoglobulin mediated immune response	5/366	16/20973	5.87E-06	0.001382	0.001204	THOC1/PTPN6/FCGR2B/CD55/CD46	5
BP	GO:0061647	histone H3-K9 modification	8/366	57/20973	6.25E-06	0.001384	0.001205	PIH1D1/CRTC2/KDM3A/KDM4C/ARID4A/DNMT1/SETD5/RIF1	8
BP	GO:0051567	histone H3-K9 methylation	7/366	46/20973	1.39E-05	0.002902	0.002527	PIH1D1/KDM3A/KDM4C/ARID4A/DNMT1/SETD5/RIF1	7
BP	GO:0034614	cellular response to reactive oxygen species	12/366	155/20973	1.85E-05	0.003662	0.003188	MAPK9/MPV17/APEX1/CD36/PINK1/AKT1/PPIF/ERCC6L2/LRRK2/GCH1/ANXA1/PRKAA1	12
BP	GO:0002924	negative regulation of humoral immune response mediated by circulating immunoglobulin	4/366	11/20973	2.73E-05	0.005147	0.004481	PTPN6/FCGR2B/CD55/CD46	4
BP	GO:0062197	cellular response to chemical stress	19/366	370/20973	3.04E-05	0.005449	0.004744	STAU1/MAPK9/ARNT/DDIT3/MPV17/APEX1/CD36/PINK1/AKT1/ALOX5/PPIF/ERCC6L2/LRRK2/GCH1/CAPN3/ANXA1/ATF2/PRKAA1/STX4	19
BP	GO:0031058	positive regulation of histone modification	10/366	116/20973	3.67E-05	0.006279	0.005467	PIH1D1/PINK1/PHF20/LRRK2/GLYR1/DNMT1/SAP30BP/RIF1/KMT2E/PAF1	10
BP	GO:0045069	regulation of viral genome replication	9/366	96/20973	4.64E-05	0.007599	0.006616	STAU1/TASOR/ADAR/PPIE/PABPC1/PLSCR1/MPHOSPH8/IFI16/OASL	9
BP	GO:0034599	cellular response to oxidative stress	17/366	320/20973	5.21E-05	0.007914	0.00689	STAU1/MAPK9/ARNT/MPV17/APEX1/CD36/PINK1/AKT1/ALOX5/PPIF/ERCC6L2/LRRK2/GCH1/ANXA1/ATF2/PRKAA1/STX4	17
BP	GO:0019079	viral genome replication	11/366	146/20973	5.27E-05	0.007914	0.00689	STAU1/TASOR/ADAR/PPIE/PABPC1/PLSCR1/MPHOSPH8/IFI16/OASL/LAMTOR5/ATG16L2	11
BP	GO:0031060	regulation of histone methylation	9/366	98/20973	5.46E-05	0.007914	0.00689	PIH1D1/KDM3A/KDM4C/PHF20/DNMT1/RIF1/KMT2E/PAF1/SUPT6H	9
BP	GO:0051573	negative regulation of histone H3-K9 methylation	4/366	13/20973	5.76E-05	0.008035	0.006996	PIH1D1/KDM3A/KDM4C/DNMT1	4
BP	GO:0016573	histone acetylation	12/366	176/20973	6.46E-05	0.008419	0.00733	MBIP/PIH1D1/CRTC2/ATXN7L3/EPC1/KAT6B/PHF20/GLYR1/NFYB/SETD5/MAP3K7/ATF2	12
BP	GO:0007254	JNK cascade	13/366	204/20973	6.48E-05	0.008419	0.00733	MBIP/MAP4K4/MAPK9/APP/GPS2/HDAC3/PINK1/FCGR2B/NAIP/LRRK2/NCOR1/MAP3K7/ATF2	13
BP	GO:0018394	peptidyl-lysine acetylation	13/366	205/20973	6.81E-05	0.008555	0.007449	MBIP/PIH1D1/CRTC2/ATXN7L3/EPC1/KAT6B/PHF20/GLYR1/NFYB/SETD5/MAP3K7/ATF2/PRKAA1	13
BP	GO:0016032	viral process	22/366	498/20973	7.10E-05	0.008623	0.007508	STAU1/TASOR/CD4/EIF3F/ADAR/PPIE/EIF3B/PABPC1/TPCN1/CD55/CD46/PLSCR1/CTSB/MPHOSPH8/IFI16/CHD1/TAF11/OASL/GSK3B/DHX9/LAMTOR5/ATG16L2	22

BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	18/366	364/20973	7.95E-05	0.009072	0.007899	SNRPN/TIA1/CELF2/PPIE/NSRP1/EFTUD2/LUC7L/U2AF1/PABPC1/DDX46/HNRNPM/SRRM1/HNRNPA3/RBM5/PPIL3/DHX9/RBM41/WBP11	18
BP	GO:0000398	mRNA splicing, via spliceosome	18/366	364/20973	7.95E-05	0.009072	0.007899	SNRPN/TIA1/CELF2/PPIE/NSRP1/EFTUD2/LUC7L/U2AF1/PABPC1/DDX46/HNRNPM/SRRM1/HNRNPA3/RBM5/PPIL3/DHX9/RBM41/WBP11	18
BP	GO:0006473	protein acetylation	14/366	239/20973	8.50E-05	0.009418	0.008201	MBIP/PIH1D1/CRTC2/ATXN7L3/EPC1/KAT6B/PHF20/GLYR1/NFYB/SETD5/GSK3B/MAP3K7/ATF2/PRKAA1	14
BP	GO:0000375	RNA splicing, via transesterification reactions	18/366	368/20973	9.13E-05	0.009696	0.008443	SNRPN/TIA1/CELF2/PPIE/NSRP1/EFTUD2/LUC7L/U2AF1/PABPC1/DDX46/HNRNPM/SRRM1/HNRNPA3/RBM5/PPIL3/DHX9/RBM41/WBP11	18
BP	GO:0031061	negative regulation of histone methylation	5/366	27/20973	9.27E-05	0.009696	0.008443	PIH1D1/KDM3A/KDM4C/DNMT1/SUPT6H	5
BP	GO:0031929	TOR signaling	10/366	131/20973	0.000103	0.010482	0.009127	SIK3/PIH1D1/PDCD6/HDAC3/NPRL2/PINK1/AKT1/LAMTOR5/PRKAA1/RPS6KB1	10
BP	GO:0051568	histone H3-K4 methylation	8/366	84/20973	0.000109	0.010741	0.009353	PIH1D1/KMT2C/PHF20/ARID4A/ASH2L/NFYB/DNMT1/KMT2E	8
BP	GO:0090200	positive regulation of release of cytochrome c from mitochondria	5/366	28/20973	0.000111	0.010741	0.009353	BAX/PINK1/PPIF/MFF/DNM1L	5
BP	GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	14/366	250/20973	0.000137	0.012878	0.011214	THOC1/PTPN6/CD4/LILRB1/SLC11A1/FCGR2B/CD55/NLRP3/CD46/RIF1/ANXA1/MAP3K7/FBXO38/SUPT6H	14
BP	GO:0002921	negative regulation of humoral immune response	4/366	16/20973	0.000141	0.012921	0.01125	PTPN6/FCGR2B/CD55/CD46	4
BP	GO:0018393	internal peptidyl-lysine acetylation	12/366	194/20973	0.000163	0.01461	0.012721	MBIP/PIH1D1/CRTC2/ATXN7L3/EPC1/KAT6B/PHF20/GLYR1/NFYB/SETD5/MAP3K7/ATF2	12
BP	GO:0006475	internal protein amino acid acetylation	12/366	198/20973	0.000197	0.01723	0.015003	MBIP/PIH1D1/CRTC2/ATXN7L3/EPC1/KAT6B/PHF20/GLYR1/NFYB/SETD5/MAP3K7/ATF2	12
BP	GO:0090199	regulation of release of cytochrome c from mitochondria	6/366	49/20973	0.000201	0.01723	0.015003	BAX/PINK1/AKT1/PPIF/MFF/DNM1L	6
BP	GO:0051570	regulation of histone H3-K9 methylation	5/366	32/20973	0.000215	0.017533	0.015266	PIH1D1/KDM3A/KDM4C/DNMT1/RIF1	5
BP	GO:0009451	RNA modification	11/366	171/20973	0.000216	0.017533	0.015266	METTL5/THG1L/ADAR/MTO1/BUD23/ELP5/ALKBH3/TARBP1/TRMO/DTWD1/HNRNPAB	11
BP	GO:0140694	non-membrane-bounded organelle assembly	19/366	430/20973	0.000219	0.017533	0.015266	MAPK9/TIA1/MRPS11/HDAC3/CSNK1D/BIN2/RNF213/CSRP1/CETN2/ZNF207/CNOT2/AKAP13/HAUS4/BSCL2/LSM14A/ATXN2/CAPN3/NCOR1/PRKAA1	19
BP	GO:0008637	apoptotic mitochondrial changes	9/366	121/20973	0.000276	0.021676	0.018874	BAX/PINK1/AKT1/PPIF/MFF/GSK3B/DNM1L/ATF2/RHOT1	9
BP	GO:0098586	cellular response to virus	8/366	98/20973	0.00032	0.024112	0.020995	ADAR/BAX/RIOK3/NLRP3/USP15/LSM14A/OASL/ATF2	8
BP	GO:1905897	regulation of response to endoplasmic reticulum stress	8/366	98/20973	0.00032	0.024112	0.020995	DDIT3/XBP1/BAX/FCGR2B/PTPN2/USP25/ALOX5/LRRK2	8
BP	GO:0002819	regulation of adaptive immune response	14/366	273/20973	0.000338	0.024964	0.021737	THOC1/PTPN6/CD4/LILRB1/SLC11A1/FCGR2B/CD55/NLRP3/CD46/RIF1/ANXA1/MAP3K7/FBXO38/SUPT6H	14
BP	GO:0050684	regulation of mRNA processing	10/366	153/20973	0.000365	0.026474	0.023051	TIA1/CELF2/NSRP1/SRRM1/RBM5/SAFB/DHX9/PAF1/CPSF6/SUPT6H	10
BP	GO:2001242	regulation of intrinsic apoptotic signaling pathway	11/366	184/20973	0.000405	0.028796	0.025073	DDIT3/XBP1/BAX/PINK1/AKT1/PTPN2/PPIF/BCLAF1/LRRK2/DNM1L/CD44	11
BP	GO:0031057	negative regulation of histone modification	6/366	56/20973	0.000422	0.029172	0.025401	PIH1D1/KDM3A/PPP4C/KDM4C/DNMT1/SUPT6H	6
BP	GO:0070584	mitochondrion morphogenesis	4/366	21/20973	0.000432	0.029172	0.025401	BAX/CERT1/MFF/DNM1L	4
BP	GO:0036336	dendritic cell migration	5/366	37/20973	0.000434	0.029172	0.025401	SLAMF8/DOCK8/NLRP12/ALOX5/ANO6	5
BP	GO:0032386	regulation of intracellular transport	16/366	349/20973	0.000455	0.029976	0.026101	ERGIC3/GCC2/ANXA2/ITGB1BP1/HDAC3/OS9/CD36/PINK1/MFF/LRRK2/GSK3B/DHX9/SCP2/CPSF6/PRKAA1/SUPT6H	16
BP	GO:0043543	protein acylation	14/366	282/20973	0.000468	0.029976	0.026101	MBIP/PIH1D1/CRTC2/ATXN7L3/EPC1/KAT6B/PHF20/GLYR1/NFYB/SETD5/GSK3B/MAP3K7/ATF2/PRKAA1	14
BP	GO:0031123	RNA 3'-end processing	9/366	130/20973	0.000469	0.029976	0.026101	INTS8/INTS6L/APP/RPRD1A/PABPC1/TENT2/EXOSC9/PAF1/CPSF6	9
BP	GO:0009615	response to virus	19/366	460/20973	0.000504	0.030964	0.026961	DDX60L/CARD8/ADAR/LILRB1/BAX/RIOK3/ABCF3/NLRP3/PLSCR1/IFI44/NLRP1/USP15/IFI16/LSM14A/OASL/SAP30BP/DHX9/LAMTOR5/ATF2	19
BP	GO:0061013	regulation of mRNA catabolic process	11/366	189/20973	0.000508	0.030964	0.026961	SERBP1/APEX1/SLC11A1/PABPC1/HNRNPM/AKT1/CNOT2/SECISBP2/EXOSC9/DHX9/HNRNPAB	11
BP	GO:0018023	peptidyl-lysine trimethylation	6/366	58/20973	0.000511	0.030964	0.026961	PIH1D1/KDM4C/ARID4A/NFYB/SETD5/KMT2E	6
BP	GO:0042542	response to hydrogen peroxide	9/366	132/20973	0.000525	0.030964	0.026961	APTX/CAPN2/APEX1/PINK1/PPIF/LRRK2/ANXA1/PRKAA1/SLC8A1	9
BP	GO:0045862	positive regulation of proteolysis	18/366	425/20973	0.000526	0.030964	0.026961	CTSH/PSME2/MAPK9/CARD8/PDCD6/ANXA2/APP/CSNK1D/BAX/PINK1/AKT1/NLRP3/NLRP12/NLRP1/LRRK2/IFI16/GSK3B/CAPN3	18
BP	GO:0050858	negative regulation of antigen receptor-mediated signaling pathway	5/366	39/20973	0.000557	0.031978	0.027844	PTPN6/GPS2/PHPT1/FCGR2B/PTPN2	5
BP	GO:0001836	release of cytochrome c from mitochondria	6/366	59/20973	0.00056	0.031978	0.027844	BAX/PINK1/AKT1/PPIF/MFF/DNM1L	6

BP	GO:0043653	mitochondrial fragmentation involved in apoptotic process	3/366	10/20973	0.000577	0.032466	0.028269	BAX/MFF/DNM1L	3
BP	GO:1901987	regulation of cell cycle phase transition	19/366	467/20973	0.000605	0.033349	0.029037	PSME2/THOC1/ANAPC5/PTPN6/APP/APEX1/ZFYVE19/CDK5RAP3/AKT1/ZNF207/RAD51C/METTL13/ANXA1/INIP/PRMT2/KMT2E/ZNF655/PAF1/ATF2	19
BP	GO:0002712	regulation of B cell mediated immunity	7/366	83/20973	0.00062	0.033349	0.029037	THOC1/PTPN6/FCGR2B/CD55/CD46/RIF1/SUPT6H	7
BP	GO:0002889	regulation of immunoglobulin mediated immune response	7/366	83/20973	0.00062	0.033349	0.029037	THOC1/PTPN6/FCGR2B/CD55/CD46/RIF1/SUPT6H	7
BP	GO:0034976	response to endoplasmic reticulum stress	14/366	291/20973	0.000638	0.033826	0.029453	DDIT3/DNAJB12/OS9/XBP1/BAX/EIF2B5/FCGR2B/CDK5RAP3/PTPN2/USP25/ALOX5/CERT1/LRRK2/GSK3B	14
BP	GO:0043484	regulation of RNA splicing	10/366	165/20973	0.000662	0.034624	0.030148	TIA1/CELF2/CCNL1/NSRP1/FAM50A/CLK1/CLK3/SRRM1/RBM5/CLK4	10
BP	GO:1905898	positive regulation of response to endoplasmic reticulum stress	5/366	41/20973	0.000704	0.036341	0.031643	DDIT3/XBP1/BAX/FCGR2B/PTPN2	5
BP	GO:0043281	regulation of cysteine-type endopeptidase activity involved in apoptotic process	12/366	230/20973	0.00076	0.038706	0.033702	CTSH/PIH1D1/CARD8/PDCD6/BAX/AKT1/NLRP3/NAIP/NLRP12/NLRP1/LAMTOR5/CD44	12
BP	GO:2000622	regulation of nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	3/366	11/20973	0.000784	0.039365	0.034276	PABPC1/SECISBP2/HNRNPAB	3
BP	GO:2000116	regulation of cysteine-type endopeptidase activity	13/366	264/20973	0.000796	0.039477	0.034373	CTSH/PIH1D1/CARD8/PDCD6/BAX/AKT1/NLRP3/NAIP/NLRP12/NLRP1/IFI16/LAMTOR5/CD44	13
BP	GO:0070585	protein localization to mitochondrion	9/366	141/20973	0.000844	0.041301	0.035962	DDIT3/OXA1L/BAX/PINK1/AKT1/MFF/LRRK2/DNM1L/PRKAA1	9
BP	GO:1901989	positive regulation of cell cycle phase transition	9/366	143/20973	0.000933	0.045068	0.039241	THOC1/ANAPC5/APP/APEX1/AKT1/RAD51C/ANXA1/KMT2E/PAF1	9
BP	GO:0009895	negative regulation of catabolic process	16/366	374/20973	0.000953	0.045427	0.039554	ANXA2/SLC11A1/PINK1/PABPC1/CDK5RAP3/AKT1/USP25/TENT2/SECISBP2/LRRK2/BSCL2/DHX9/NCOR1/MYCBP2/HNRNPAB/PRKAA1	16
BP	GO:2001233	regulation of apoptotic signaling pathway	17/366	412/20973	0.000993	0.046757	0.040712	CTSH/PIH1D1/MAPK9/DDIT3/XBP1/BAX/PINK1/AKT1/PTPN2/PPIF/BCLAF1/LRRK2/GSK3B/DNM1L/CD44/STX4/RPS6KB1	17
BP	GO:0046831	regulation of RNA export from nucleus	3/366	12/20973	0.001032	0.047387	0.041261	DHX9/CPSF6/SUPT6H	3
BP	GO:0090309	positive regulation of DNA methylation-dependent heterochromatin assembly	3/366	12/20973	0.001032	0.047387	0.041261	TASOR/MPHOSPH8/DNMT1	3
BP	GO:0000266	mitochondrial fission	5/366	45/20973	0.001085	0.049225	0.042861	COX10/PINK1/MFF/LRRK2/DNM1L	5
BP	GO:0048524	positive regulation of viral process	6/366	67/20973	0.001104	0.049524	0.043122	STAU1/CD4/ADAR/PPIE/PABPC1/DHX9	6
BP	GO:0050792	regulation of viral process	11/366	208/20973	0.001118	0.049566	0.043158	STAU1/TASOR/CD4/ADAR/PPIE/PABPC1/PLSCR1/MPHOSPH8/IFI16/OASL/DHX9	11
CC	GO:0005681	spliceosomal complex	14/381	229/21993	5.00E-05	0.011237	0.009158	SNRPN/ADAR/PPIE/EFTUD2/LUC7L/U2AF1/PABPC1/HNRNPM/SRRM1/HNRNPA3/RBM5/PPIL3/RBM41/WBP11	14
CC	GO:0071013	catalytic step 2 spliceosome	9/381	98/21993	5.18E-05	0.011237	0.009158	SNRPN/PPIE/EFTUD2/U2AF1/PABPC1/HNRNPM/SRRM1/HNRNPA3/PPIL3	9
CC	GO:0005852	eukaryotic translation initiation factor 3 complex	4/381	16/21993	0.000137	0.014222	0.01159	EIF3F/EIF3M/EIF3B/EIF3C	4
CC	GO:0033290	eukaryotic 48S preinitiation complex	4/381	16/21993	0.000137	0.014222	0.01159	EIF3F/EIF3M/EIF3B/EIF3C	4
CC	GO:0101002	ficolin-1-rich granule	13/381	232/21993	0.000219	0.014222	0.01159	CTSH/PPIE/DYNLL1/BIN2/SLC11A1/FCAR/PNP/CD55/CTSB/GUSB/ALOX5/SERPINA1/PSMD7	13
CC	GO:0016282	eukaryotic 43S preinitiation complex	4/381	18/21993	0.000224	0.014222	0.01159	EIF3F/EIF3M/EIF3B/EIF3C	4
CC	GO:0000123	histone acetyltransferase complex	8/381	94/21993	0.000229	0.014222	0.01159	MBIP/ATXN7L3/EPC1/KAT6B/SUPT20H/PHF20/MAP3K7/ATF2	8
CC	GO:0070993	translation preinitiation complex	4/381	19/21993	0.00028	0.01517	0.012363	EIF3F/EIF3M/EIF3B/EIF3C	4
CC	GO:0061702	inflammasome complex	4/381	20/21993	0.000345	0.016531	0.013471	CARD8/NLRP3/NAIP/NLRP1	4
CC	GO:0016607	nuclear speck	20/381	492/21993	0.000407	0.016531	0.013471	THOC1/FAM76B/HP1BP3/CCNL1/PPIE/NSRP1/EFTUD2/APEX1/U2AF1/DDX46/CLK3/SRRM1/BCLAF1/MBD1/IFI16/ASCC1/ACIN1/KMT2E/CPSF6/PRKAA1	20
CC	GO:0031248	protein acetyltransferase complex	8/381	104/21993	0.000457	0.016531	0.013471	MBIP/ATXN7L3/EPC1/KAT6B/SUPT20H/PHF20/MAP3K7/ATF2	8
CC	GO:1902493	acetyltransferase complex	8/381	104/21993	0.000457	0.016531	0.013471	MBIP/ATXN7L3/EPC1/KAT6B/SUPT20H/PHF20/MAP3K7/ATF2	8
CC	GO:0005819	spindle	18/381	434/21993	0.000621	0.020727	0.016891	SEPTIN6/SEPTIN2/ANAPC5/APP/DYNLL1/TAF1D/HDAC3/CSNK1D/NUP85/AKT1/ZNF207/HAUS4/IRAG2/LSM14A/RIF1/NCOR1/VRK1/CEP170	18
CC	GO:0031313	extrinsic component of endosome membrane	3/381	11/21993	0.000768	0.023793	0.01939	SNX5/ANXA1/USP8	3
CC	GO:1904813	ficolin-1-rich granule lumen	9/381	146/21993	0.00103	0.029787	0.024275	CTSH/PPIE/BIN2/PNP/CTSB/GUSB/ALOX5/SERPINA1/PSMD7	9

CC	GO:1990124	messenger ribonucleoprotein complex	3/381	14/21993	0.001629	0.044187	0.03601	HNRNPA3/LSM14A/HNRNPAB	3
CC	GO:0030027	lamellipodium	11/381	222/21993	0.00178	0.045453	0.037042	PSTPIP1/DOCK8/APP/ITGB1BP1/PABPC1/AKT1/KLHL2/CD44/FGD4/CAPZB/STX4	11
MF	GO:0140098	catalytic activity, acting on RNA	23/373	453/20738	8.78E-06	0.002526	0.002261	METTL5/DDX60L/INTS11/WARS1/MARS1/DDX24/BUD23/APEX1/POP5/EIF4A2/RPP21/POLR1C/DDX46/CNOT2/TENT2/ALKBH3/TARBP1/POLR1H/DDX50/TRMO/EXOSC9/DHX9/DTWD1	23
MF	GO:0061629	RNA polymerase II-specific DNA-binding transcription factor binding	20/373	375/20738	1.68E-05	0.002526	0.002261	ARNT/DDIT3/HDAC3/KDM3A/APEX1/MLX/KAT6B/CDK5RAP3/PTPN2/KDM4C/TAF11/OASL/EXOSC9/GSK3B/UHRF2/DHX9/NCOR1/PRMT2/LRIF1/ATF2	20
MF	GO:0016504	peptidase activator activity	7/373	47/20738	1.95E-05	0.002526	0.002261	CTSH/PSME2/CARD8/APP/PINK1/NLRP12/NLRP1	7
MF	GO:0031369	translation initiation factor binding	6/373	32/20738	1.98E-05	0.002526	0.002261	EIF3F/EIF3M/EIF3B/EIF2B5/EIF3C/GCH1	6
MF	GO:0106310	protein serine kinase activity	20/373	380/20738	2.03E-05	0.002526	0.002261	SIK3/MAP4K4/MAPK9/MKNK2/RPS6KC1/CDK16/CSNK1D/RIOK3/RIOK1/PINK1/AKT1/CLK1/CLK3/LRRK2/GSK3B/CLK4/VRK1/MAP3K7/PRKAA1/RPS6KB1	20
MF	GO:0042393	histone binding	16/373	270/20738	3.36E-05	0.003475	0.00311	TDRD3/PIH1D1/NAP1L4/KMT2C/PTMA/KAT6B/CHD2/ZMYND8/CHD8/MPHOSPH8/GLYR1/USP15/CHD1/UHRF2/KMT2E/SUPT6H	16
MF	GO:0003743	translation initiation factor activity	7/373	53/20738	4.35E-05	0.003863	0.003457	EIF3F/EIF4A2/EIF3M/EIF3B/EIF2B5/EIF3C/MTIF3	7
MF	GO:0003725	double-stranded RNA binding	8/373	77/20738	7.25E-05	0.005629	0.005038	DDX60L/APTX/STAU1/ADAR/NLRP1/LSM14A/OASL/DHX9	8
MF	GO:0004674	protein serine/threonine kinase activity	21/373	470/20738	0.000134	0.009256	0.008284	SIK3/MAP4K4/MAPK9/MKNK2/RPS6KC1/CDK16/CSNK1D/RIOK3/RIOK1/PINK1/AKT1/CLK1/CLK3/AKAP13/LRRK2/GSK3B/CLK4/VRK1/MAP3K7/PRKAA1/RPS6KB1	21
MF	GO:0008168	methyltransferase activity	13/373	226/20738	0.000243	0.014569	0.01304	METTL5/CIAPIN1/KMT2C/EZH1/BUD23/METTL17/TARBP1/ASH2L/TRMO/DNMT1/METTL13/SETD5/PRMT2	13
MF	GO:0008656	cysteine-type endopeptidase activator activity involved in apoptotic process	4/373	18/20738	0.000258	0.014569	0.01304	CTSH/CARD8/NLRP12/NLRP1	4
MF	GO:0004386	helicase activity	11/373	177/20738	0.000375	0.019172	0.017158	DDX60L/SHPRH/DDX24/EIF4A2/DDX46/CHD2/CHD8/ERCC6L2/DDX50/CHD1/DHX9	11
MF	GO:0004712	protein serine/threonine/tyrosine kinase activity	20/373	476/20738	0.000427	0.019172	0.017158	SIK3/MAP4K4/MAPK9/MKNK2/RPS6KC1/CDK16/CSNK1D/RIOK3/RIOK1/PINK1/AKT1/CLK1/CLK3/LRRK2/GSK3B/CLK4/VRK1/MAP3K7/PRKAA1/RPS6KB1	20
MF	GO:0016741	transferase activity, transferring one-carbon groups	13/373	240/20738	0.000432	0.019172	0.017158	METTL5/CIAPIN1/KMT2C/EZH1/BUD23/METTL17/TARBP1/ASH2L/TRMO/DNMT1/METTL13/SETD5/PRMT2	13
MF	GO:0016505	peptidase activator activity involved in apoptotic process	4/373	21/20738	0.000484	0.020025	0.017922	CTSH/CARD8/NLRP12/NLRP1	4
MF	GO:0050321	tau-protein kinase activity	4/373	22/20738	0.000583	0.022044	0.019729	SIK3/CSNK1D/GSK3B/PRKAA1	4
MF	GO:0035064	methylated histone binding	7/373	81/20738	0.000639	0.022044	0.019729	TDRD3/ZMYND8/CHD8/MPHOSPH8/GLYR1/CHD1/KMT2E	7
MF	GO:0140034	methylation-dependent protein binding	7/373	81/20738	0.000639	0.022044	0.019729	TDRD3/ZMYND8/CHD8/MPHOSPH8/GLYR1/CHD1/KMT2E	7
MF	GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	10/373	167/20738	0.000914	0.029863	0.026727	METTL5/KMT2C/EZH1/BUD23/TARBP1/ASH2L/TRMO/DNMT1/SETD5/PRMT2	10
MF	GO:0043028	cysteine-type endopeptidase regulator activity involved in apoptotic process	5/373	44/20738	0.001118	0.034728	0.031081	CTSH/CARD8/NAIP/NLRP12/NLRP1	5
MF	GO:0043522	leucine zipper domain binding	3/373	13/20738	0.001444	0.040751	0.036472	DDIT3/PMF1/ATF2	3
MF	GO:0140658	ATP-dependent chromatin remodeler activity	3/373	13/20738	0.001444	0.040751	0.036472	CHD2/CHD8/CHD1	3
MF	GO:0090079	translation regulator activity, nucleic acid binding	8/373	122/20738	0.001651	0.04458	0.039898	EIF3F/EIF4A2/EIF3M/EIF3B/EIF2B5/EIF3C/PABPC1/MTIF3	8
MF	GO:0008135	translation factor activity, RNA binding	7/373	97/20738	0.001849	0.047844	0.04282	EIF3F/EIF4A2/EIF3M/EIF3B/EIF2B5/EIF3C/MTIF3	7

Table 8-3. Gene Ontology analysis for aberrantly spliced genes in granulocytic cell population in MDS with *SF3B1* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0008380	RNA splicing	29/295	497/20973	1.11E-10	3.29E-07	3.17E-07	DDX39B/SRSF5/PPIE/SETX/RBM39/PPIL3/PRPF4B/SRSF7/SRRM1/SREK1/SNRNP25/DDX23/CLK1/HNRNPC/SRSF11/SON/RBM15/CLK4/JMJD6/RBM25/U2AF1L4/ZRANB2/SART3/LSM1/RBM6/TMBIM6/CWF19L1/PRPF39/SRSF3	29
BP	GO:0000375	RNA splicing, via transesterification reactions	23/295	368/20973	2.79E-09	4.13E-06	3.99E-06	DDX39B/SRSF5/PPIE/SETX/RBM39/PPIL3/PRPF4B/SRSF7/SRRM1/SNRNP25/DDX23/HNRNPC/SON/RBM15/JMJD6/RBM25/U2AF1L4/SART3/LSM1/RBM6/CWF19L1/PRPF39/SRSF3	23
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	22/295	364/20973	1.15E-08	8.50E-06	8.20E-06	DDX39B/SRSF5/PPIE/SETX/RBM39/PPIL3/PRPF4B/SRSF7/SRRM1/SNRNP25/DDX23/HNRNPC/SON/RBM15/JMJD6/RBM25/U2AF1L4/SART3/RBM6/CWF19L1/PRPF39/SRSF3	22
BP	GO:0000398	mRNA splicing, via spliceosome	22/295	364/20973	1.15E-08	8.50E-06	8.20E-06	DDX39B/SRSF5/PPIE/SETX/RBM39/PPIL3/PRPF4B/SRSF7/SRRM1/SNRNP25/DDX23/HNRNPC/SON/RBM15/JMJD6/RBM25/U2AF1L4/SART3/RBM6/CWF19L1/PRPF39/SRSF3	22
BP	GO:1903311	regulation of mRNA metabolic process	19/295	323/20973	1.79E-07	0.000106	0.000102	SERBP1/SECISBP2/CNOT1/MAPK14/RBM39/SRSF7/SRRM1/YBX3/CNOT2/HNRNPC/SAFB2/SON/EXOSC9/RBM15/JMJD6/RBM25/SLTM/LSM1/SRSF3	19
BP	GO:0043484	regulation of RNA splicing	13/295	165/20973	6.39E-07	0.000316	0.000305	SRSF5/SETX/RBM39/SRSF7/SRRM1/CLK1/SON/RBM15/CLK4/JMJD6/RBM25/TMBIM6/SRSF3	13
BP	GO:0050684	regulation of mRNA processing	10/295	153/20973	6.37E-05	0.026971	0.026026	RBM39/SRSF7/SRRM1/SAFB2/SON/RBM15/JMJD6/RBM25/SLTM/SRSF3	10
BP	GO:1905820	positive regulation of chromosome separation	4/295	19/20973	0.000126	0.046651	0.045017	CDC27/ANAPC5/PLSCR1/DLGAP5	4
BP	GO:0048024	regulation of mRNA splicing, via spliceosome	8/295	108/20973	0.000145	0.0479	0.046222	RBM39/SRSF7/SRRM1/SON/RBM15/JMJD6/RBM25/SRSF3	8
CC	GO:0005819	spindle	19/306	434/21993	1.21E-05	0.004482	0.003963	CDC27/SEPTIN6/HAUS4/CEP295/SEPTIN2/ANAPC5/MAPK14/DYNLL1/SEPTIN7/AAAS/NUP85/MMS19/DLGAP5/ZNF207/KATNA1/BRCC3/KIF23/UBXN2B/APP	19
CC	GO:0016607	nuclear speck	20/306	492/21993	2.08E-05	0.004482	0.003963	DDX39B/SRSF5/PPIE/MAPK14/RBM39/PNISR/PRPF4B/SRSF7/SRRM1/SREK1/BCLAF1/NCAPG2/SRSF11/SON/RBM15/RBM25/U2AF1L4/SART3/HECTD1/SRSF3	20
CC	GO:0000152	nuclear ubiquitin ligase complex	6/306	51/21993	7.35E-05	0.010118	0.008945	CDC27/BCOR/ANAPC5/BABAM2/BRCA2/BRCC3	6
CC	GO:0005681	spliceosomal complex	12/306	229/21993	9.39E-05	0.010118	0.008945	DDX39B/PPIE/PPIL3/PRPF4B/SRRM1/SREK1/SNRNP25/DDX23/HNRNPC/U2AF1L4/CWF19L1/PRPF39	12
CC	GO:0000779	condensed chromosome, centromeric region	10/306	167/21993	0.000121	0.010455	0.009244	SEPTIN6/SEPTIN2/PPP1R12A/DYNLL1/SEPTIN7/NUP85/NCAPD2/ORC2/ZNF207/CLIP1	10
CC	GO:0098687	chromosomal region	16/306	405/21993	0.000193	0.013858	0.012252	SEPTIN6/SEPTIN2/ORC3/PPP1R12A/SETX/RAD50/DYNLL1/SEPTIN7/NUP85/NCAPD2/ORC2/ATR/MIS18BP1/ZNF207/BRCA2/CLIP1	16
CC	GO:0000776	kinetochore	9/306	157/21993	0.000364	0.020199	0.017858	SEPTIN6/SEPTIN2/PPP1R12A/DYNLL1/SEPTIN7/NUP85/ORC2/ZNF207/CLIP1	9
CC	GO:0000793	condensed chromosome	12/306	266/21993	0.000375	0.020199	0.017858	SEPTIN6/SEPTIN2/PPP1R12A/RAD50/DYNLL1/SEPTIN7/NUP85/NCAPD2/ORC2/ZNF207/BRCA2/CLIP1	12
CC	GO:0000151	ubiquitin ligase complex	13/306	319/21993	0.000569	0.027232	0.024077	CDC27/DERL2/BCOR/ANAPC5/PDCD6/DMAC2/OS9/BABAM2/BRCA2/BRCC3/KLHL2/CRBN/DCAF6	13
CC	GO:0000775	chromosome, centromeric region	11/306	244/21993	0.000656	0.028268	0.024992	SEPTIN6/SEPTIN2/PPP1R12A/DYNLL1/SEPTIN7/NUP85/NCAPD2/ORC2/MIS18BP1/ZNF207/CLIP1	11
CC	GO:0005940	septin ring	3/306	15/21993	0.001072	0.033675	0.029773	SEPTIN6/SEPTIN2/SEPTIN7	3
CC	GO:0031105	septin complex	3/306	15/21993	0.001072	0.033675	0.029773	SEPTIN6/SEPTIN2/SEPTIN7	3
CC	GO:0000314	organellar small ribosomal subunit	4/306	33/21993	0.001094	0.033675	0.029773	MRPS22/MRPS18A/MRPS18C/MRPS11	4
CC	GO:0005763	mitochondrial small ribosomal subunit	4/306	33/21993	0.001094	0.033675	0.029773	MRPS22/MRPS18A/MRPS18C/MRPS11	4
CC	GO:0032156	septin cytoskeleton	3/306	16/21993	0.001306	0.037437	0.033099	SEPTIN6/SEPTIN2/SEPTIN7	3
CC	GO:0140534	endoplasmic reticulum protein-containing complex	8/306	153/21993	0.00139	0.037437	0.033099	DERL2/KRTCAP2/DNAJC10/EMC9/HLA-A/DPM2/OS9/GPAA1	8
CC	GO:0000313	organellar ribosome	6/306	90/21993	0.001616	0.038705	0.03422	MRPS22/MRPL48/MRPS18A/MRPS18C/MRPL35/MRPS11	6
CC	GO:0005761	mitochondrial ribosome	6/306	90/21993	0.001616	0.038705	0.03422	MRPS22/MRPL48/MRPS18A/MRPS18C/MRPL35/MRPS11	6
CC	GO:0032592	integral component of mitochondrial membrane	6/306	93/21993	0.001912	0.043365	0.03834	IMMT/MFF/OXA1L/TIMM17B/PINK1/RHOT1	6
CC	GO:0098573	intrinsic component of mitochondrial membrane	6/306	95/21993	0.00213	0.045902	0.040583	IMMT/MFF/OXA1L/TIMM17B/PINK1/RHOT1	6
CC	GO:0031965	nuclear membrane	12/306	329/21993	0.002335	0.047927	0.042373	TMEM120A/RNF13/P2RX4/AAAS/NUP85/PHF20/STAU2/UTP18/RBM15/SYNE1/NUDT9/TMEM18	12
CC	GO:0071013	catalytic step 2 spliceosome	6/306	98/21993	0.002492	0.048824	0.043166	PPIE/PPIL3/PRPF4B/SRRM1/DDX23/HNRNPC	6
MF	GO:0045296	cadherin binding	16/307	357/20738	9.18E-05	0.043783	0.040676	SEPTIN2/SERBP1/NOP56/ADD1/EIF4G2/GAPVD1/SEPTIN7/P2RX4/DLG1/EIF5/PICALM/MACF1/CAST/ASAP1/KTN1/YWHAB	16

Table 8-4. Gene Ontology analysis for aberrantly spliced genes in erythrocytic cell population in MDS with *SF3B1* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0034250	positive regulation of cellular amide metabolic process	9/137	188/20973	4.12E-06	0.004271	0.003967	PICALM/POLR2G/PKM/VIM/RPS27L/EIF4G2/RPS9/FXR1/HNRNPD	9
BP	GO:0008380	RNA splicing	14/137	497/20973	4.85E-06	0.004271	0.003967	SREK1/YTHDC1/CLK4/RBM39/PRPF39/SRRM1/FXR1/U2AF1/TMBIM6/NONO/USP39/DHX15/SNU13/GRSF1	14
BP	GO:0033108	mitochondrial respiratory chain complex assembly	7/137	110/20973	7.91E-06	0.004606	0.004279	NDUFS2/OXA1L/PET100/NDUFS4/COX14/NDUFB3/NDUFS8	7
BP	GO:0045727	positive regulation of translation	8/137	160/20973	1.05E-05	0.004606	0.004279	POLR2G/PKM/VIM/RPS27L/EIF4G2/RPS9/FXR1/HNRNPD	8
BP	GO:1903311	regulation of mRNA metabolic process	10/137	323/20973	5.33E-05	0.018752	0.017419	YTHDC1/ZC3HAV1/POLR2G/RBM39/VIM/SRRM1/FXR1/HNRNPD/PAN2/PAPOLA	10
BP	GO:0002181	cytoplasmic translation	7/137	165/20973	0.000107	0.018752	0.017419	PKM/RPS9/RPL13A/EIF4A2/RPLP0/HNRNPD/RPS24	7
BP	GO:0043484	regulation of RNA splicing	7/137	165/20973	0.000107	0.018752	0.017419	YTHDC1/CLK4/RBM39/SRRM1/FXR1/TMBIM6/GRSF1	7
BP	GO:0010257	NADH dehydrogenase complex assembly	5/137	73/20973	0.000116	0.018752	0.017419	NDUFS2/OXA1L/NDUFS4/NDUFB3/NDUFS8	5
BP	GO:0032981	mitochondrial respiratory chain complex I assembly	5/137	73/20973	0.000116	0.018752	0.017419	NDUFS2/OXA1L/NDUFS4/NDUFB3/NDUFS8	5
BP	GO:0045842	positive regulation of mitotic metaphase/anaphase transition	3/137	15/20973	0.000117	0.018752	0.017419	CDC16/ANAPC5/CDC27	3
BP	GO:1901970	positive regulation of mitotic sister chromatid separation	3/137	15/20973	0.000117	0.018752	0.017419	CDC16/ANAPC5/CDC27	3
BP	GO:1902101	positive regulation of metaphase/anaphase transition of cell cycle	3/137	17/20973	0.000173	0.022791	0.021171	CDC16/ANAPC5/CDC27	3
BP	GO:0071826	ribonucleoprotein complex subunit organization	8/137	240/20973	0.000184	0.022791	0.021171	YTHDC1/RPS27L/PRPF39/RPL13A/RPLP0/USP39/SNU13/ZFAND1	8
BP	GO:0061912	selective autophagy	5/137	82/20973	0.000201	0.022791	0.021171	AUP1/BECN1/UBQLN1/HTRA2/RB1CC1	5
BP	GO:0010506	regulation of autophagy	10/137	380/20973	0.000202	0.022791	0.021171	RMC1/STAT3/BECN1/EIF4G2/UBQLN1/HTRA2/WDR45/NPRL3/DRAM2/TMEM39A	10
BP	GO:0031145	anaphase-promoting complex-dependent catabolic process	3/137	18/20973	0.000207	0.022791	0.021171	CDC16/ANAPC5/CDC27	3
BP	GO:1905820	positive regulation of chromosome separation	3/137	19/20973	0.000245	0.025351	0.023549	CDC16/ANAPC5/CDC27	3
BP	GO:0045840	positive regulation of mitotic nuclear division	4/137	51/20973	0.000343	0.033575	0.031188	CDC16/PHIP/ANAPC5/CDC27	4
BP	GO:0033617	mitochondrial cytochrome c oxidase assembly	3/137	23/20973	0.000439	0.040627	0.037738	OXA1L/PET100/COX14	3
BP	GO:0019646	aerobic electron transport chain	5/137	98/20973	0.000461	0.040627	0.037738	NDUFS2/UQCRB/NDUFS4/NDUFB3/NDUFS8	5
CC	GO:0005680	anaphase-promoting complex	3/142	21/21993	0.000322	0.023547	0.020909	CDC16/ANAPC5/CDC27	3
CC	GO:0044391	ribosomal subunit	7/142	207/21993	0.000402	0.023547	0.020909	RPS27L/RPS9/RPL13A/RPLP0/RPS24/MRPS18C/MTERF4	7
CC	GO:0098803	respiratory chain complex	5/142	105/21993	0.000602	0.023547	0.020909	NDUFS2/UQCRB/NDUFS4/NDUFB3/NDUFS8	5
CC	GO:0005746	mitochondrial respirasome	5/142	109/21993	0.000713	0.023547	0.020909	NDUFS2/UQCRB/NDUFS4/NDUFB3/NDUFS8	5
CC	GO:0005681	spliceosomal complex	7/142	229/21993	0.000733	0.023547	0.020909	SREK1/PRPF39/SRRM1/U2AF1/USP39/DHX15/SNU13	7
CC	GO:0005747	mitochondrial respiratory chain complex I	4/142	64/21993	0.000784	0.023547	0.020909	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4
CC	GO:0030964	NADH dehydrogenase complex	4/142	64/21993	0.000784	0.023547	0.020909	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4
CC	GO:0045271	respiratory chain complex I	4/142	64/21993	0.000784	0.023547	0.020909	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4
CC	GO:0101002	ficolin-1-rich granule	7/142	232/21993	0.000791	0.023547	0.020909	COMMD3/PKM/DYNLL1/GMFG/CD55/VCL/CTSB	7
CC	GO:0070469	respirasome	5/142	116/21993	0.000944	0.023547	0.020909	NDUFS2/UQCRB/NDUFS4/NDUFB3/NDUFS8	5
CC	GO:0098800	inner mitochondrial membrane protein complex	6/142	176/21993	0.001005	0.023547	0.020909	NDUFS2/HSPA9/UQCRB/NDUFS4/NDUFB3/NDUFS8	6
CC	GO:0098798	mitochondrial protein-containing complex	8/142	315/21993	0.001035	0.023547	0.020909	NDUFS2/HSPA9/UQCRB/NDUFS4/MRPS18C/MTERF4/NDUFB3/NDUFS8	8
CC	GO:0005840	ribosome	7/142	245/21993	0.001087	0.023547	0.020909	RPS27L/RPS9/RPL13A/RPLP0/RPS24/MRPS18C/MTERF4	7
CC	GO:0000151	ubiquitin ligase complex	8/142	319/21993	0.001122	0.023547	0.020909	AUP1/WDR26/ELOB/CDC16/STUB1/ANAPC5/CDC27/UBE2V1	8
CC	GO:0005925	focal adhesion	10/142	480/21993	0.001142	0.023547	0.020909	GIT2/VIM/HSPA9/RPS9/RPL13A/CD46/ACTR3/RPLP0/VCL/PPP1R12A	10
CC	GO:0022626	cytosolic ribosome	5/142	122/21993	0.001183	0.023547	0.020909	RPS27L/RPS9/RPL13A/RPLP0/RPS24	5
CC	GO:0030055	cell-substrate junction	10/142	489/21993	0.001311	0.023547	0.020909	GIT2/VIM/HSPA9/RPS9/RPL13A/CD46/ACTR3/RPLP0/VCL/PPP1R12A	10
CC	GO:0000407	phagophore assembly site	3/142	34/21993	0.001362	0.023547	0.020909	BECN1/WDR45/RB1CC1	3
CC	GO:0016607	nuclear speck	10/142	492/21993	0.001372	0.023547	0.020909	SREK1/YTHDC1/SMC4/RBM39/SRRM1/U2AF1/HBP1/NONO/DHX15/RUFY1	10
CC	GO:1990204	oxidoreductase complex	5/142	143/21993	0.00238	0.038801	0.034454	NDUFS2/UQCRB/NDUFS4/NDUFB3/NDUFS8	5
CC	GO:1904813	ficolin-1-rich granule lumen	5/142	146/21993	0.002604	0.040431	0.035901	COMMD3/PKM/GMFG/VCL/CTSB	5
CC	GO:0032592	integral component of mitochondrial membrane	4/142	93/21993	0.003119	0.046213	0.041035	ABCB6/HSPA9/OXA1L/PET100	4
CC	GO:0098573	intrinsic component of mitochondrial membrane	4/142	95/21993	0.003367	0.047488	0.042167	ABCB6/HSPA9/OXA1L/PET100	4
CC	GO:0015935	small ribosomal subunit	4/142	96/21993	0.003496	0.047488	0.042167	RPS27L/RPS9/RPS24/MRPS18C	4
MF	GO:0031625	ubiquitin protein ligase binding	10/140	351/20738	0.000139	0.020482	0.017967	AUP1/BECN1/NDUFS2/HSPA9/ELOB/STUB1/TUBA1B/VCL/TMBIM6/TANK	10
MF	GO:0044389	ubiquitin-like protein ligase binding	10/140	369/20738	0.000208	0.020482	0.017967	AUP1/BECN1/NDUFS2/HSPA9/ELOB/STUB1/TUBA1B/VCL/TMBIM6/TANK	10
MF	GO:0019843	rRNA binding	5/140	80/20738	0.000209	0.020482	0.017967	RPS9/RPLP0/MRPS18C/MTERF4/GTF3A	5
MF	GO:0008137	NADH dehydrogenase (ubiquinone) activity	4/140	57/20738	0.000596	0.030398	0.026665	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4
MF	GO:0050136	NADH dehydrogenase (quinone) activity	4/140	58/20738	0.000636	0.030398	0.026665	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4
MF	GO:0003954	NADH dehydrogenase activity	4/140	60/20738	0.000724	0.030398	0.026665	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4
MF	GO:0003955	NAD(P)H dehydrogenase (quinone) activity	4/140	60/20738	0.000724	0.030398	0.026665	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4
MF	GO:0005200	structural constituent of cytoskeleton	5/140	118/20738	0.00124	0.04557	0.039974	TUBGCP4/VIM/TUBA1B/ACTR3/TPM1	5
MF	GO:0016655	oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	4/140	73/20738	0.00151	0.049323	0.043266	NDUFS2/NDUFS4/NDUFB3/NDUFS8	4

Table 8-5. Gene Ontology analysis for aberrantly spliced genes in $CD34^+$ cell population in MDS with *SRSF2* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0033157	regulation of intracellular protein transport	24/697	239/20973	1.58E-06	0.006966	0.00659	UBE2D3/MFF/DERL3/JUP/HPS4/EMD/ATP13A2/BAP1/LEPROT/HTRA2/NUP58/XPO4/PDCD10/ATG13/BARD1/ANGPT1/SP100/PKIG/HUWE1/BCAS3/PCM1/ECT2/RAB29/PARL	24
BP	GO:0006261	DNA-templated DNA replication	18/697	162/20973	7.81E-06	0.010321	0.009763	DDX11/FAM111A/POLA2/RTEL1/NOC3L/MUS81/METTL4/MCM3/RAD51/POLG/SLFN11/CCNE1/ATR/SRPK2/KAT7/POLQ/CENPX/EME1	18
BP	GO:0006260	DNA replication	26/697	302/20973	1.01E-05	0.010321	0.009763	DDX11/FAM111A/POLA2/POLL/RTEL1/NOC3L/TREX1/MUS81/METTL4/MCM3/ANKRD17/RAD51/POLG/SLFN11/UCHL5/NFIC/CCNE1/ATR/SRPK2/KAT7/POLQ/TBRG1/CENPX/EME1/CHEK1/NFRKB	26
BP	GO:1903311	regulation of mRNA metabolic process	27/697	323/20973	1.18E-05	0.010321	0.009763	FASTKD3/RBM23/SLTM/HNRNPC/SAFB/CNOT1/METTL3/FASTK/RNPS1/TIA1/PABPC4/PUM1/PUF60/TRA2A/SRSF2/CTR9/TUT7/GIGYF2/CELF1/SRPK2/BARD1/RBM10/MBNL2/RBM5/NSRP1/EXOSC3/SRRM1	27
BP	GO:0050684	regulation of mRNA processing	17/697	153/20973	1.39E-05	0.010321	0.009763	RBM23/SLTM/SAFB/RNPS1/TIA1/PUF60/TRA2A/SRSF2/CTR9/CELF1/SRPK2/BARD1/RBM10/MBNL2/RBM5/NSRP1/SRRM1	17
BP	GO:0006310	DNA recombination	28/697	345/20973	1.41E-05	0.010321	0.009763	AP5Z1/POLL/CYREN/RTEL1/CCNB1IP1/WRAP53/TREX1/MUS81/PSMC3IP/MCM3/RAD51/BRD8/SETX/UCHL5/RAD51AP1/INIP/ARID2/HROB/POLQ/APEX2/CENPX/EME1/FANCB/CHEK1/EXOSC3/RIF1/NFRKB/PTPRC	28
BP	GO:0043173	nucleotide salvage	6/697	19/20973	2.47E-05	0.014334	0.013559	NAPRT/UCK1/DGUOK/UPP1/UCKL1/AMPD2	6
BP	GO:0071108	protein K48-linked deubiquitination	8/697	38/20973	2.89E-05	0.014334	0.013559	USP34/OTUD5/USP20/BAP1/USP15/USP8/MINDY3/USP33	8
BP	GO:0045005	DNA-templated DNA replication maintenance of fidelity	9/697	49/20973	2.93E-05	0.014334	0.013559	DDX11/FAM111A/RTEL1/MUS81/RAD51/SLFN11/ATR/CENPX/EME1	9
BP	GO:0043484	regulation of RNA splicing	17/697	165/20973	3.71E-05	0.016335	0.015452	RBM23/HNRNPH3/FASTK/METTL4/RNPS1/TIA1/PUF60/SETX/TRA2A/SRSF2/CELF1/SRPK2/RBM10/MBNL2/RBM5/NSRP1/SRRM1	17
BP	GO:0031297	replication fork processing	8/697	41/20973	5.17E-05	0.020638	0.019523	DDX11/FAM111A/RTEL1/MUS81/RAD51/ATR/CENPX/EME1	8
BP	GO:0048024	regulation of mRNA splicing, via spliceosome	13/697	108/20973	6.10E-05	0.020638	0.019523	RBM23/RNPS1/TIA1/PUF60/TRA2A/SRSF2/CELF1/SRPK2/RBM10/MBNL2/RBM5/NSRP1/SRRM1	13
BP	GO:0008380	RNA splicing	34/697	497/20973	6.27E-05	0.020638	0.019523	SREK1/RBM23/DDX41/HNRNPC/PPIE/HNRNPH3/PPIL3/TSEN2/LUC7L/SUGP2/U2AF1L4/METTL3/FASTK/METTL4/RNPS1/THOC2/TIA1/DDX47/SNU13/PUF60/SETX/SMN1/TRA2A/SRSF2/GEMIN8/CELF1/SRPK2/RBM10/DDX39B/PRMT7/MBNL2/RBM5/NSRP1/SRRM1	34
BP	GO:0006221	pyrimidine nucleotide biosynthetic process	7/697	32/20973	7.08E-05	0.020638	0.019523	NME6/UCK1/NME4/SHMT1/UPP1/UCKL1/DCTD	7
BP	GO:0043543	protein acylation	23/697	282/20973	7.48E-05	0.020638	0.019523	PBXIP1/NOC2L/NFYC/GLYR1/KAT2A/OGDH/PHF20/POR/ZDHHC13/ATAT1/PER1/SETD5/PORCN/BRD8/ATF2/NAA20/KAT7/KAT6A/NFYB/ARRB1/MECP2/CHEK1/NAA16	23
BP	GO:0009129	pyrimidine nucleoside monophosphate metabolic process	6/697	23/20973	8.20E-05	0.020638	0.019523	NT5C/UCK1/SHMT1/UPP1/UCKL1/DCTD	6
BP	GO:0006302	double-strand break repair	23/697	285/20973	8.78E-05	0.020638	0.019523	DDX11/AP5Z1/POLL/CYREN/RTEL1/WRAP53/PAXX/MUS81/MCM3/RAD51/BRD8/SETX/PARP2/RAD51AP1/INIP/ARID2/ERCC8/POLQ/EME1/FANCB/CHEK1/TDP1/RIF1	23
BP	GO:0018393	internal peptidyl-lysine acetylation	18/697	194/20973	8.83E-05	0.020638	0.019523	PBXIP1/NOC2L/NFYC/GLYR1/KAT2A/PHF20/POR/ATAT1/PER1/SETD5/BRD8/ATF2/KAT7/KAT6A/NFYB/ARRB1/MECP2/CHEK1	18
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	27/697	364/20973	9.39E-05	0.020638	0.019523	RBM23/DDX41/HNRNPC/PPIE/HNRNPH3/PPIL3/LUC7L/U2AF1L4/METTL3/RNPS1/TIA1/SNU13/PUF60/SETX/SMN1/TRA2A/SRSF2/GEMIN8/CELF1/SRPK2/RBM10/DDX39B/PRMT7/MBNL2/RBM5/NSRP1/SRRM1	27
BP	GO:0000398	mRNA splicing, via spliceosome	27/697	364/20973	9.39E-05	0.020638	0.019523	RBM23/DDX41/HNRNPC/PPIE/HNRNPH3/PPIL3/LUC7L/U2AF1L4/METTL3/RNPS1/TIA1/SNU13/PUF60/SETX/SMN1/TRA2A/SRSF2/GEMIN8/CELF1/SRPK2/RBM10/DDX39B/PRMT7/MBNL2/RBM5/NSRP1/SRRM1	27
BP	GO:0000375	RNA splicing, via transesterification reactions	27/697	368/20973	0.000113	0.020638	0.019523	RBM23/DDX41/HNRNPC/PPIE/HNRNPH3/PPIL3/LUC7L/U2AF1L4/METTL3/RNPS1/TIA1/SNU13/PUF60/SETX/SMN1/TRA2A/SRSF2/GEMIN8/CELF1/SRPK2/RBM10/DDX39B/PRMT7/MBNL2/RBM5/NSRP1/SRRM1	27
BP	GO:0006475	internal protein amino acid acetylation	18/697	198/20973	0.000115	0.020638	0.019523	PBXIP1/NOC2L/NFYC/GLYR1/KAT2A/PHF20/POR/ATAT1/PER1/SETD5/BRD8/ATF2/KAT7/KAT6A/NFYB/ARRB1/MECP2/CHEK1	18
BP	GO:0031056	regulation of histone modification	18/697	198/20973	0.000115	0.020638	0.019523	BCOR/SAP30BP/NOC2L/NFYC/GLYR1/KAT2A/PHF20/NSD1/SETD5/MTF2/AKAP8/CTR9/KAT7/NFYB/ARRB1/MECP2/CHEK1/RIF1	18
BP	GO:0006399	tRNA metabolic process	19/697	216/20973	0.000116	0.020638	0.019523	CARS1/DUS3L/TRMU/RPP30/TRMT2B/TSEN2/NARS1/MARS1/CTU2/WDR6/KARS1/AARSD1/TRMO/EXOSC3/DTWD1/LARS2/WARS1/CDKAL1/CARS2	19
BP	GO:0032386	regulation of intracellular transport	26/697	349/20973	0.000117	0.020638	0.019523	UBE2D3/ANXA2/MFF/DERL3/MAP2K2/JUP/HPS4/EMD/ATP13A2/BAP1/LEPROT/HTRA2/NUP58/XPO4/PDCD10/ATG13/BARD1/ANGPT1/SP100/PKIG/HUWE1/BCAS3/PCM1/ECT2/RAB29/PARL	26
BP	GO:0060271	cilium assembly	27/697	370/20973	0.000123	0.02083	0.019704	TMEM107/OFD1/ODF2/TTC8/CFAP20/ATAT1/ARHGAP35/CFAP410/WDR11/SEPTIN2/IFT46/CDK10/ODF2L/TCTN1/IFT122/C2CD3/CDKL1/AHI1/RO60/IFT88/PCM1/CIBAR1/SCLT1/TBC1D7/CEP162/CEP41/SDCCAG8	27
BP	GO:0010569	regulation of double-strand break repair via homologous recombination	10/697	72/20973	0.000129	0.020965	0.019831	RTEL1/WRAP53/RAD51/BRD8/RAD51AP1/ARID2/POLQ/FANCB/CHEK1/RIF1	10
BP	GO:0090316	positive regulation of intracellular protein transport	16/697	167/20973	0.000148	0.023209	0.021954	UBE2D3/MFF/JUP/HPS4/EMD/BAP1/LEPROT/HTRA2/XPO4/PDCD10/ATG13/HUWE1/BCAS3/PCM1/ECT2/RAB29	16
BP	GO:0006473	protein acetylation	20/697	239/20973	0.000153	0.02322	0.021964	PBXIP1/NOC2L/NFYC/GLYR1/KAT2A/PHF20/POR/ATAT1/PER1/SETD5/BRD8/ATF2/NAA20/KAT7/KAT6A/NFYB/ARRB1/MECP2/CHEK1/NAA16	20
BP	GO:0035065	regulation of histone acetylation	9/697	61/20973	0.000174	0.023749	0.022465	NOC2L/NFYC/GLYR1/KAT2A/SETD5/KAT7/NFYB/ARRB1/CHEK1	9
BP	GO:0009130	pyrimidine nucleoside monophosphate biosynthetic process	5/697	17/20973	0.000177	0.023749	0.022465	UCK1/SHMT1/UPP1/UCKL1/DCTD	5

BP	GO:0018394	peptidyl-lysine acetylation	18/697	205/20973	0.000178	0.023749	0.022465	PBXIP1/NOC2L/NFYC/GLYR1/KAT2A/PHF20/POR/ATAT1/PER1/SETD5/BRD8/ATF2/KAT7/KAT6A/NFYB/ARRB1/MECP2/CHEK1	18
BP	GO:0044782	cilium organization	28/697	399/20973	0.000179	0.023749	0.022465	TMEM107/OFD1/ODF2/TTC8/CFAP20/ATAT1/ARHGAP35/CFAP410/WDR11/SLC9A3R1/SEPTIN2/IFT46/CDK10/ODF2L/TCTN1/IFT122/C2CD3/CDKL1/AHI1/RO60/IFT88/PCM1/CIBAR1/SCLT1/TBC1D7/CEP162/CEP41/SDCCAG8	28
BP	GO:0072528	pyrimidine-containing compound biosynthetic process	7/697	37/20973	0.000187	0.023749	0.022465	NME6/UCK1/NME4/SHMT1/UPP1/UCKL1/DCTD	7
BP	GO:0043414	macromolecule methylation	26/697	360/20973	0.000192	0.023749	0.022465	BCOR/NFYC/TRMT2B/PHF20/NSD1/NDUFAF7/METTL3/SETD4/NOP2/WDR6/SETD5/METTL4/EED/MTF2/NTMT1/EZH2/CTR9/FDXACB1/NFYB/MECP2/TRMO/PRMT7/MTRR/RIF1/NSUN4/PRDM2	26
BP	GO:0000725	recombinational repair	16/697	171/20973	0.000194	0.023749	0.022465	AP5Z1/POLL/RTEL1/WRAP53/MUS81/MCM3/RAD51/BRD8/RAD51AP1/INIP/ARID2/HROB/POLQ/FANCB/CHEK1/RIF1	16
BP	GO:0031023	microtubule organizing center organization	15/697	155/20973	0.000211	0.025092	0.023735	KAT2A/ODF2/NUP62/HAUS7/FES/TUBE1/NDE1/C2CD3/BCAS3/PCM1/CHORDC1/CHEK1/SGO1/SDCCAG8/USP33	15
BP	GO:0000381	regulation of alternative mRNA splicing, via spliceosome	9/697	64/20973	0.000252	0.028301	0.026771	RNPS1/TIA1/PUF60/SRSF2/CELF1/RBM10/MBNL2/RBM5/NSRP1	9
BP	GO:0007098	centrosome cycle	14/697	141/20973	0.00026	0.028301	0.026771	KAT2A/ODF2/NUP62/HAUS7/FES/TUBE1/NDE1/C2CD3/PCM1/CHORDC1/CHEK1/SGO1/SDCCAG8/USP33	14
BP	GO:0070585	protein localization to mitochondrion	14/697	141/20973	0.00026	0.028301	0.026771	PITRM1/UBE2D3/DDIT3/MFF/HPS4/BAP1/IMMP1L/LEPROT/HTRA2/ATG13/HUWE1/TOMM40L/PARL/DNM1L	14
BP	GO:1903955	positive regulation of protein targeting to mitochondrion	7/697	39/20973	0.000264	0.028301	0.026771	UBE2D3/HPS4/BAP1/LEPROT/HTRA2/ATG13/HUWE1	7
BP	GO:0016573	histone acetylation	16/697	176/20973	0.00027	0.028301	0.026771	PBXIP1/NOC2L/NFYC/GLYR1/KAT2A/PHF20/PER1/SETD5/BRD8/ATF2/KAT7/KAT6A/NFYB/ARRB1/MECP2/CHEK1	16
BP	GO:0006220	pyrimidine nucleotide metabolic process	8/697	52/20973	0.000295	0.029519	0.027923	NME6/NT5C/UCK1/NME4/SHMT1/UPP1/UCKL1/DCTD	8
BP	GO:1903214	regulation of protein targeting to mitochondrion	8/697	52/20973	0.000295	0.029519	0.027923	UBE2D3/HPS4/BAP1/LEPROT/HTRA2/ATG13/HUWE1/PARL	8
BP	GO:0016571	histone methylation	16/697	178/20973	0.000307	0.029696	0.02809	BCOR/NFYC/PHF20/NSD1/SETD4/SETD5/EED/MTF2/NTMT1/EZH2/CTR9/NFYB/MECP2/PRMT7/RIF1/PRDM2	16
BP	GO:0018205	peptidyl-lysine modification	29/697	434/20973	0.000315	0.029696	0.02809	BCOR/PBXIP1/NOC2L/NFYC/GLYR1/HDAC7/KAT2A/PHF20/POR/NSD1/SETD4/ATAT1/PER1/SETD5/BRD8/MTF2/ATF2/EZH2/SIRT5/CTR9/RWDD3/KAT7/KAT6A/NFYB/ARRB1/MECP2/CHEK1/RIF1/UHRF2	29
BP	GO:1903008	organelle disassembly	13/697	127/20973	0.000317	0.029696	0.02809	KLC1/ATP13A2/CDC37/VRK1/MRRF/WDR45/HTRA2/CAMKK2/MTIF3/ATG13/HUWE1/PARL/DNM1L	13
BP	GO:0009218	pyrimidine ribonucleotide metabolic process	6/697	29/20973	0.000326	0.02985	0.028236	NME6/NT5C/UCK1/NME4/UPP1/UCKL1	6
BP	GO:0006626	protein targeting to mitochondrion	12/697	113/20973	0.000381	0.033882	0.03205	PITRM1/UBE2D3/MFF/HPS4/BAP1/IMMP1L/LEPROT/HTRA2/ATG13/HUWE1/TOMM40L/PARL	12
BP	GO:0006418	tRNA aminoacylation for protein translation	8/697	54/20973	0.000385	0.033882	0.03205	CARS1/NARS1/MARS1/KARS1/AARSD1/LARS2/WARS1/CARS2	8
BP	GO:2000779	regulation of double-strand break repair	12/697	114/20973	0.000413	0.035425	0.03351	DDX11/CYREN/RTEL1/WRAP53/RAD51/BRD8/RAD51AP1/ARID2/POLQ/FANCB/CHEK1/RIF1	12
BP	GO:0072594	establishment of protein localization to organelle	30/697	463/20973	0.000419	0.035425	0.03351	PITRM1/ZFAND6/UBE2D3/NUP107/DDIT3/AP4M1/MFF/JUP/HPS4/NUP62/WRAP53/GGA3/BAP1/IMMP1L/LEPROT/ZFAND2B/ATF2/HACL1/LONP2/HTRA2/NUP58/ATG13/ANGPT1/PKIG/HUWE1/TOMM40L/ECT2/ZFYVE16/PARL/STK4	30
BP	GO:0045916	negative regulation of complement activation	4/697	12/20973	0.000484	0.039233	0.037111	A2M/CD55/CD46/CD59	4
BP	GO:1903749	positive regulation of establishment of protein localization to mitochondrion	7/697	43/20973	0.000494	0.039233	0.037111	UBE2D3/HPS4/BAP1/LEPROT/HTRA2/ATG13/HUWE1	7
BP	GO:0002090	regulation of receptor internalization	9/697	70/20973	0.000499	0.039233	0.037111	ANXA2/FLOT1/ANKRD13D/ATXN2/ANGPT1/ARRB2/AHI1/ARRB1/NUMB	9
BP	GO:2000756	regulation of peptidyl-lysine acetylation	9/697	70/20973	0.000499	0.039233	0.037111	NOC2L/NFYC/GLYR1/KAT2A/SETD5/KAT7/NFYB/ARRB1/CHEK1	9
BP	GO:0002275	myeloid cell activation involved in immune response	11/697	101/20973	0.000532	0.040597	0.038402	LAT/GRN/S100A13/STX4/DNASE1/PIK3CD/TREX1/FES/KARS1/LILRA2/ADGRE2	11
BP	GO:0000724	double-strand break repair via homologous recombination	15/697	169/20973	0.000535	0.040597	0.038402	AP5Z1/POLL/RTEL1/WRAP53/MUS81/MCM3/RAD51/BRD8/RAD51AP1/INIP/ARID2/POLQ/FANCB/CHEK1/RIF1	15
BP	GO:0043039	tRNA aminoacylation	8/697	57/20973	0.00056	0.041168	0.038942	CARS1/NARS1/MARS1/KARS1/AARSD1/LARS2/WARS1/CARS2	8
BP	GO:0043094	cellular metabolic compound salvage	6/697	32/20973	0.000571	0.041168	0.038942	NAPRT/UCK1/DGUOK/UPP1/UCKL1/AMPD2	6
BP	GO:0072655	establishment of protein localization to mitochondrion	13/697	135/20973	0.000573	0.041168	0.038942	PITRM1/UBE2D3/DDIT3/MFF/HPS4/BAP1/IMMP1L/LEPROT/HTRA2/ATG13/HUWE1/TOMM40L/PARL	13
BP	GO:0032259	methylation	27/697	408/20973	0.00058	0.041168	0.038942	BCOR/METTL17/NFYC/TRMT2B/PHF20/NSD1/NDUFAF7/METTL3/SETD4/NOP2/WDR6/SETD5/METTL4/EED/MTF2/NTMT1/EZH2/CTR9/FDXACB1/NFYB/MECP2/TRMO/PRMT7/MTRR/RIF1/NSUN4/PRDM2	27
BP	GO:0043038	amino acid activation	8/697	58/20973	0.000631	0.044021	0.041641	CARS1/NARS1/MARS1/KARS1/AARSD1/LARS2/WARS1/CARS2	8
BP	GO:0009220	pyrimidine ribonucleotide biosynthetic process	5/697	22/20973	0.000657	0.044021	0.041641	NME6/UCK1/NME4/UPP1/UCKL1	5
BP	GO:0035456	response to interferon-beta	6/697	33/20973	0.000678	0.044021	0.041641	XAF1/PLSCR1/CAPN2/TREX1/HTRA2/IRF1	6
BP	GO:0036297	interstrand cross-link repair	6/697	33/20973	0.000678	0.044021	0.041641	RAD51/RAD51AP1/ATR/HROB/CENPX/ERCC6L2	6

BP	GO:0032388	positive regulation of intracellular transport	17/697	210/20973	0.00068	0.044021	0.041641	UBE2D3/ANXA2/MFF/JUP/HPS4/EMD/BAP1/LEPROT/HTRA2/XPO4/PDCD10/ATG13/HUWE1/BCAS3/PCM1/ECT2/RAB29	17
BP	GO:0006241	CTP biosynthetic process	4/697	13/20973	0.00068	0.044021	0.041641	NME6/UCK1/NME4/UCKL1	4
BP	GO:0006282	regulation of DNA repair	14/697	156/20973	0.000727	0.046328	0.043823	DDX11/CYREN/RTEL1/WRAP53/RAD51/BRD8/RAD51AP1/ARID2/ERCC8/KAT7/POLQ/FANCB/CHEK1/RIF1	14
BP	GO:0009124	nucleoside monophosphate biosynthetic process	7/697	46/20973	0.000752	0.047284	0.044728	UCK1/SHMT1/DGUOK/UPP1/UCKL1/AMPD2/DCTD	7
BP	GO:1903747	regulation of establishment of protein localization to mitochondrion	8/697	60/20973	0.000795	0.049278	0.046614	UBE2D3/HPS4/BAP1/LEPROT/HTRA2/ATG13/HUWE1/PARL	8
CC	GO:0036064	ciliary basal body	20/716	168/21993	6.10E-07	0.000226	0.000205	OFD1/TTC8/CFAP20/ARHGAP35/CFAP410/WDR11/EXOC7/IFT46/CDK10/ODF2L/IFT122/SPATA7/C2CD3/AHI1/IFT88/PCM1/CIBAR1/TBC1D7/CEP41/SDCCAG8	20
CC	GO:0090734	site of DNA damage	15/716	100/21993	8.22E-07	0.000226	0.000205	CBX3/POLL/CYREN/WRAP53/PAXX/RAD51/ATF2/SLFN11/CUL5/PARP2/ININP/KAT7/HROB/OARD1/RIF1	15
CC	GO:0005819	spindle	33/716	434/21993	6.47E-06	0.001186	0.001079	CBX3/DDX11/PTPN7/ODF2/NUBP2/TUBGCP4/APP/NUP62/HAUS7/TUBGCP6/EMD/WDR73/ATAT1/CIAO1/MAEA/VRK1/RMDN1/YPEL5/ZNF207/SPDL1/CDC27/STAG2/NDE1/SEPTIN2/CSPP1/TFDP2/ECT2/MAD1L1/CEP162/KIF23/SGO1/RIF1/IRAG2	33
CC	GO:0035869	ciliary transition zone	11/716	69/21993	1.32E-05	0.001812	0.001647	TMEM107/TTC8/CFAP410/SEPTIN2/TCTN1/IFT122/SPATA7/CDKL1/AHI1/PCM1/CIBAR1	11
CC	GO:0000922	spindle pole	17/716	177/21993	7.03E-05	0.007733	0.00703	DDX11/ODF2/NUBP2/TUBGCP4/NUP62/TUBGCP6/EMD/WDR73/RMDN1/YPEL5/SPDL1/STAG2/NDE1/CSPP1/MAD1L1/SGO1/IRAG2	17
CC	GO:0098687	chromosomal region	28/716	405/21993	0.000165	0.015103	0.01373	CBX3/NUP107/RTEL1/WRAP53/THOC2/MCM3/RAD51/SETX/ZNF207/EZH2/SPDL1/STAG2/PPP2R5C/NDE1/SEPTIN2/NCAPD3/RAD51AP1/ATR/TINF2/KAT7/SP100/CENPX/MAD1L1/CHEK1/SGO1/ITGB3BP/RIF1/UHRF2	28
CC	GO:0000793	condensed chromosome	20/716	266/21993	0.000479	0.032107	0.029188	STAG3L4/CBX3/NUP107/CCNB1IP1/PSMC3IP/RAD51/RAD1/AKAP8/ZNF207/SPDL1/STAG2/NDE1/SEPTIN2/NCAPD3/CENPX/MAD1L1/CHEK1/SGO1/ITGB3BP/RIF1	20
CC	GO:0005814	centriole	15/716	171/21993	0.000491	0.032107	0.029188	OFD1/ODF2/NUBP2/CFAP20/SFI1/ODF2L/C2CD3/AHI1/IFT88/PCM1/CIBAR1/SCLT1/CEP162/CEP41/SDCCAG8	15
CC	GO:0031965	nuclear membrane	23/716	329/21993	0.000525	0.032107	0.029188	CBX3/NUP107/TOR1AIP1/MLX/NPIPA1/ERBIN/PHF20/SMOX/P2RX4/CREB3L4/NUP62/EMD/TMC6/SYNE1/MX1/ANKRD17/TMEM120A/CPNE1/NUP58/MINDY3/DPY19L4/PCM1/RIF1	23
CC	GO:0000151	ubiquitin ligase complex	22/716	319/21993	0.000823	0.04527	0.041155	BCOR/DERL3/DCAF13/ZSWIM8/SYVN1/UBR1/MAEA/AUP1/FBXL12/RAD51/YPEL5/CDC27/CUL5/CRBN/ERCC8/BARD1/PHC3/KLHL24/DCAF17/PCGF3/USP33/KLHL20	22
MF	GO:0140098	catalytic activity, acting on RNA	36/717	453/20738	3.46E-06	0.002589	0.002372	CARS1/DDX41/DDX55/DUS3L/TRMU/CNOT1/RPP30/TRMT2B/TSEN2/NARS1/MARS1/METTL3/RPAP1/NOP2/METTL4/DDX24/DDX47/REXO2/DDX58/TOE1/EIF4A2/KARS1/TUT7/FDXACB1/DDX50/POLR3A/AARSD1/DDX39B/TRMO/EXOSC3/DTWD1/LARS2/NSUN4/WARS1/CDKAL1/CARS2	36
MF	GO:0004386	helicase activity	19/717	177/20738	1.27E-05	0.004753	0.004355	DDX11/DDX41/DDX55/RTEL1/RAD54L2/DDX24/MCM3/DDX47/RAD51/DDX58/SETX/EIF4A2/SLFN11/DDX50/SMARCA2/POLQ/ERCC6L2/DDX39B/CHD2	19
MF	GO:0140097	catalytic activity, acting on DNA	20/717	223/20738	0.000101	0.016434	0.015056	DDX11/POLL/RTEL1/DNASE1/RAD54L2/TREX1/MUS81/METTL4/MCM3/RAD51/REXO2/SETX/POLG/BPTF/RAD1/SMARCA2/POLQ/APEX2/CHD2/TDP1	20
MF	GO:0140101	catalytic activity, acting on a tRNA	16/717	156/20738	0.000104	0.016434	0.015056	CARS1/DUS3L/TRMU/RPP30/TRMT2B/TSEN2/NARS1/MARS1/KARS1/AARSD1/TRMO/DTWD1/LARS2/WARS1/CDKAL1/CARS2	16
MF	GO:0004843	cysteine-type deubiquitinase activity	13/717	110/20738	0.00011	0.016434	0.015056	USP11/USP34/OTUD5/USP20/BAP1/ALG13/UCHL5/USP15/USP10/USP8/MINDY3/TANK/USP33	13
MF	GO:0101005	deubiquitinase activity	13/717	117/20738	0.000206	0.025641	0.023491	USP11/USP34/OTUD5/USP20/BAP1/ALG13/UCHL5/USP15/USP10/USP8/MINDY3/TANK/USP33	13
MF	GO:0008234	cysteine-type peptidase activity	17/717	185/20738	0.000246	0.026251	0.024049	USP11/CAPN2/USP34/OTUD5/USP20/BAP1/CTSB/CASP8/ALG13/CASP10/UCHL5/USP15/USP10/USP8/MINDY3/TANK/USP33	17
MF	GO:0042393	histone binding	21/717	270/20738	0.000487	0.034034	0.031179	CBX3/NOC2L/GLYR1/MPND/MLLT10/WRAP53/SMARCC2/BPTF/MTF2/NCAPD3/SCML2/USP15/SMARCA2/KAT7/KAT6A/BCAS3/PHC3/SCMH1/PRMT7/CHD2/UHRF2	21
MF	GO:0019783	ubiquitin-like protein peptidase activity	13/717	128/20738	0.000498	0.034034	0.031179	USP11/USP34/OTUD5/USP20/BAP1/ALG13/UCHL5/USP15/USP10/USP8/MINDY3/TANK/USP33	13
MF	GO:0004812	aminoacyl-tRNA ligase activity	8/717	54/20738	0.000501	0.034034	0.031179	CARS1/NARS1/MARS1/KARS1/AARSD1/LARS2/WARS1/CARS2	8
MF	GO:0016875	ligase activity, forming carbon-oxygen bonds	8/717	54/20738	0.000501	0.034034	0.031179	CARS1/NARS1/MARS1/KARS1/AARSD1/LARS2/WARS1/CARS2	8
MF	GO:0016741	transferase activity, transferring one-carbon groups	19/717	240/20738	0.000717	0.040204	0.036832	METTL17/AMT/TRMT2B/NSD1/NDUFAF7/METTL3/SETD4/NOP2/SETD5/METTL4/SHMT1/EED/NTMT1/EZH2/FDXACB1/TRMO/PRMT7/NSUN4/PRDM2	19
MF	GO:0008408	3'-5' exonuclease activity	8/717	57/20738	0.000726	0.040204	0.036832	CNOT1/TREX1/REXO2/POLG/TOE1/RAD1/APEX2/EXOSC3	8
MF	GO:0050660	flavin adenine dinucleotide binding	10/717	86/20738	0.00076	0.040204	0.036832	DUS3L/ACAD10/POR/COQ6/ACAD8/MICAL1/PPOX/ACADM/ACOX3/MTRR	10
MF	GO:0031491	nucleosome binding	9/717	72/20738	0.000816	0.040204	0.036832	NOC2L/GLYR1/HNRNPC/MLLT10/HMGN2/SMARCC2/EED/PARP2/HMGN1	9
MF	GO:0004527	exonuclease activity	10/717	88/20738	0.000911	0.040204	0.036832	REXO5/CNOT1/TREX1/REXO2/POLG/TOE1/RAD1/APEX2/TDP1/EXOSC3	10
MF	GO:0008168	methyltransferase activity	18/717	226/20738	0.000914	0.040204	0.036832	METTL17/AMT/TRMT2B/NSD1/NDUFAF7/METTL3/SETD4/NOP2/SETD5/METTL4/EED/NTMT1/EZH2/FDXACB1/TRMO/PRMT7/NSUN4/PRDM2	18

Table 8-6. Gene Ontology analysis for aberrantly spliced genes in monocytic cell population in MDS with *SRSF2* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0008380	RNA splicing	24/352	497/20973	4.04E-06	0.008455	0.007869	SRRM1/TIA1/HNRNPH1/HNRNPH3/SETX/PPIL3/SUGP2/PRPF4B/RBM5/QKI/U2AF1/SNRPD2/RBM23/IVNS1ABP/VIRMA/ISY1/LSM1/HNRNPA2B1/CCNL1/SNRPA1/MAGOH/WBP11/NSRP1/ZCCHC8	24
BP	GO:0000375	RNA splicing, via transesterification reactions	20/352	368/20973	4.63E-06	0.008455	0.007869	SRRM1/TIA1/HNRNPH1/HNRNPH3/SETX/PPIL3/PRPF4B/RBM5/QKI/U2AF1/SNRPD2/RBM23/ISY1/LSM1/HNRNPA2B1/SNRPA1/MAGOH/WBP11/NSRP1/ZCCHC8	20
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	19/352	364/20973	1.42E-05	0.01292	0.012025	SRRM1/TIA1/HNRNPH1/HNRNPH3/SETX/PPIL3/PRPF4B/RBM5/QKI/U2AF1/SNRPD2/RBM23/ISY1/HNRNPA2B1/SNRPA1/MAGOH/WBP11/NSRP1/ZCCHC8	19
BP	GO:0000398	mRNA splicing, via spliceosome	19/352	364/20973	1.42E-05	0.01292	0.012025	SRRM1/TIA1/HNRNPH1/HNRNPH3/SETX/PPIL3/PRPF4B/RBM5/QKI/U2AF1/SNRPD2/RBM23/ISY1/HNRNPA2B1/SNRPA1/MAGOH/WBP11/NSRP1/ZCCHC8	19
BP	GO:0043484	regulation of RNA splicing	12/352	165/20973	2.35E-05	0.017166	0.015977	SRRM1/TIA1/HNRNPH1/HNRNPH3/SETX/RBM5/QKI/RBM23/HNRNPA2B1/CCNL1/MAGOH/NSRP1	12
BP	GO:0031056	regulation of histone modification	13/352	198/20973	3.19E-05	0.019445	0.018098	PIH1D1/NELFE/ZZZ3/PPP4C/SETD5/PINK1/SAP30BP/KANSL2/MTF2/NFYB/DNMT1/GLYR1/KDM4C	13
BP	GO:0043029	T cell homeostasis	6/352	41/20973	5.87E-05	0.030602	0.028481	CORO1A/BAX/P2RX7/LGALS9/PPP2R3C/FAS	6
BP	GO:0031098	stress-activated protein kinase signaling cascade	15/352	282/20973	9.15E-05	0.041698	0.038808	MAP2K6/QARS1/APP/PINK1/FCGR2B/STK3/NAIP/LGALS9/COPS5/RNF13/DUSP3/MAP3K4/ARHGEF6/CCDC88C/FAS	15
BP	GO:0051567	histone H3-K9 methylation	6/352	46/20973	0.000114	0.041698	0.038808	PIH1D1/SETD5/SETDB1/ARID4A/DNMT1/KDM4C	6
BP	GO:0045862	positive regulation of proteolysis	19/352	425/20973	0.000114	0.041698	0.038808	PLK3/CASP8/CAPN3/RNF14/APP/ANXA2/RAD23A/BAX/UBQLN1/PINK1/PSMC5/LGALS9/APH1B/NLRP1/PDCD2/PSEN1/NLRP12/CARD8/FAS	19
CC	GO:0071013	catalytic step 2 spliceosome	11/359	98/21993	5.94E-07	0.000258	0.000224	SRRM1/HNRNPH1/PPIL3/PRPF4B/U2AF1/SNRPD2/ISY1/HNRNPA2B1/SNRPA1/MAGOH/ZCCHC8	11
CC	GO:0005681	spliceosomal complex	15/359	229/21993	5.92E-06	0.001286	0.001119	SRRM1/HNRNPH1/HNRNPH3/PPIL3/PRPF4B/RBM5/U2AF1/SNRPD2/IVNS1ABP/ISY1/HNRNPA2B1/SNRPA1/MAGOH/WBP11/ZCCHC8	15
CC	GO:0036464	cytoplasmic ribonucleoprotein granule	14/359	281/21993	0.000232	0.033534	0.029198	TIA1/PUM1/ATXN2/POLDIP3/G3BP2/UBAP2/CNOT7/LSM1/PSMA6/SERPINB1/PABPC4/PSMA4/BTBD1/EIF4E2	14
CC	GO:0035770	ribonucleoprotein granule	14/359	301/21993	0.000464	0.044542	0.038784	TIA1/PUM1/ATXN2/POLDIP3/G3BP2/UBAP2/CNOT7/LSM1/PSMA6/SERPINB1/PABPC4/PSMA4/BTBD1/EIF4E2	14
CC	GO:0000932	P-body	8/359	117/21993	0.000679	0.044542	0.038784	PUM1/UBAP2/CNOT7/LSM1/PSMA6/PSMA4/BTBD1/EIF4E2	8
CC	GO:0031968	organelle outer membrane	12/359	245/21993	0.000742	0.044542	0.038784	CASP8/DNM1L/MFN1/BAX/PINK1/ARMCX6/CPT1B/MICOS10/RHOT1/LETMD1/PSEN1/VDAC3	12
CC	GO:0019867	outer membrane	12/359	247/21993	0.000797	0.044542	0.038784	CASP8/DNM1L/MFN1/BAX/PINK1/ARMCX6/CPT1B/MICOS10/RHOT1/LETMD1/PSEN1/VDAC3	12
CC	GO:0030062	mitochondrial tricarboxylic acid cycle enzyme complex	3/359	12/21993	0.000851	0.044542	0.038784	IDH3B/BCKDK/IDH3G	3
CC	GO:0005741	mitochondrial outer membrane	11/359	217/21993	0.000924	0.044542	0.038784	CASP8/DNM1L/MFN1/BAX/PINK1/ARMCX6/CPT1B/MICOS10/RHOT1/LETMD1/VDAC3	11

Table 8-7. Gene Ontology analysis for aberrantly spliced genes in granulocytic cell population in MDS with *SRSF2* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0008380	RNA splicing	38/486	497/20973	1.32E-10	4.77E-07	4.57E-07	THOC1/RBM7/SNRPB/HNRNPK/HNRNPH1/LARP7/NSRP1/SRRM1/CLK1/HNRNPM/ARL6IP4/NCL/DHX8/SRSF5/PRPF4B/U2AF1L4/ZNF638/SRSF3/SREK1/TRA2A/RBM5/CWF19L1/LSM1/GEMIN7/PRPF40A/FXR1/STRAP/SRSF2/RBM6/SUPT6H/SRSF6/TRA2B/METTL3/RBM25/CCNL1/ZRANB2/SNRPD2/ZBTB8OS	38
BP	GO:0000375	RNA splicing, via transesterification reactions	29/486	368/20973	1.11E-08	2.00E-05	1.91E-05	RBM7/SNRPB/HNRNPK/HNRNPH1/LARP7/NSRP1/SRRM1/HNRNPM/NCL/DHX8/SRSF5/PRPF4B/U2AF1L4/SRSF3/TRA2A/RBM5/CWF19L1/LSM1/GEMIN7/PRPF40A/FXR1/STRAP/SRSF2/RBM6/SRSF6/TRA2B/METTL3/RBM25/SNRPD2	29
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	28/486	364/20973	3.35E-08	3.02E-05	2.89E-05	RBM7/SNRPB/HNRNPK/HNRNPH1/LARP7/NSRP1/SRRM1/HNRNPM/NCL/DHX8/SRSF5/PRPF4B/U2AF1L4/SRSF3/TRA2A/RBM5/CWF19L1/GEMIN7/PRPF40A/FXR1/STRAP/SRSF2/RBM6/SRSF6/TRA2B/METTL3/RBM25/SNRPD2	28
BP	GO:0000398	mRNA splicing, via spliceosome	28/486	364/20973	3.35E-08	3.02E-05	2.89E-05	RBM7/SNRPB/HNRNPK/HNRNPH1/LARP7/NSRP1/SRRM1/HNRNPM/NCL/DHX8/SRSF5/PRPF4B/U2AF1L4/SRSF3/TRA2A/RBM5/CWF19L1/GEMIN7/PRPF40A/FXR1/STRAP/SRSF2/RBM6/SRSF6/TRA2B/METTL3/RBM25/SNRPD2	28
BP	GO:0043484	regulation of RNA splicing	18/486	165/20973	5.78E-08	4.17E-05	3.99E-05	RBM7/HNRNPK/HNRNPH1/LARP7/NSRP1/SRRM1/CLK1/NCL/SRSF5/SRSF3/TRA2A/RBM5/FXR1/SRSF2/SRSF6/TRA2B/RBM25/CCNL1	18
BP	GO:0048024	regulation of mRNA splicing, via spliceosome	14/486	108/20973	2.04E-07	0.000123	0.000117	RBM7/HNRNPK/LARP7/NSRP1/SRRM1/NCL/SRSF3/TRA2A/RBM5/FXR1/SRSF2/SRSF6/TRA2B/RBM25	14
BP	GO:0010452	histone H3-K36 methylation	6/486	15/20973	6.29E-07	0.000324	0.00031	NSD3/PAXIP1/SMYD2/SETD5/NSD2/BCOR	6
BP	GO:0043414	macromolecule methylation	25/486	360/20973	1.23E-06	0.000554	0.000529	SNRPB/MTO1/MYB/LARP7/NSD3/PAXIP1/MACROH2A1/SMYD2/KDM3A/KMT2C/DALRD3/TRMT13/BAZ2A/WDR6/SETD5/CREBBP/ARID4A/DNMT3A/PIWIL4/SUPT6H/NSD2/METTL3/PCMTD1/KMT2E/BCOR	25
BP	GO:0050684	regulation of mRNA processing	15/486	153/20973	2.86E-06	0.001145	0.001095	RBM7/HNRNPK/LARP7/NSRP1/SRRM1/NCL/SRSF3/TRA2A/RBM5/FXR1/SRSF2/SUPT6H/SRSF6/TRA2B/RBM25	15
BP	GO:0032259	methylation	26/486	408/20973	3.69E-06	0.00133	0.001271	METTL17/SNRPB/MTO1/MYB/LARP7/NSD3/PAXIP1/MACROH2A1/SMYD2/KDM3A/KMT2C/DALRD3/TRMT13/BAZ2A/WDR6/SETD5/CREBBP/ARID4A/DNMT3A/PIWIL4/SUPT6H/NSD2/METTL3/PCMTD1/KMT2E/BCOR	26
BP	GO:1903311	regulation of mRNA metabolic process	22/486	323/20973	7.27E-06	0.002383	0.002279	RBM7/HNRNPK/PABPC4/LARP7/NSRP1/CNOT2/GIGYF2/SRRM1/HNRNPM/NCL/SRSF3/TRA2A/RBM5/LSM1/FXR1/SRSF2/SUPT6H/SRSF6/TRA2B/METTL3/PUM2/RBM25	22
BP	GO:0006479	protein methylation	17/486	220/20973	1.59E-05	0.004418	0.004225	SNRPB/MYB/NSD3/PAXIP1/MACROH2A1/SMYD2/KDM3A/KMT2C/SETD5/CREBBP/ARID4A/DNMT3A/SUPT6H/NSD2/PCMTD1/KMT2E/BCOR	17
BP	GO:0008213	protein alkylation	17/486	220/20973	1.59E-05	0.004418	0.004225	SNRPB/MYB/NSD3/PAXIP1/MACROH2A1/SMYD2/KDM3A/KMT2C/SETD5/CREBBP/ARID4A/DNMT3A/SUPT6H/NSD2/PCMTD1/KMT2E/BCOR	17
BP	GO:0000380	alternative mRNA splicing, via spliceosome	10/486	81/20973	1.75E-05	0.004519	0.004321	RBM7/NSRP1/HNRNPM/RBM5/FXR1/STRAP/SRSF2/SRSF6/TRA2B/RBM25	10
BP	GO:0070646	protein modification by small protein removal	15/486	181/20973	2.20E-05	0.0053	0.005068	SUPT20H/COPS3/SHMT2/USP48/USP8/GPS1/OTUD5/UBXN1/BABAM2/TNIP1/ATXN3/TAF2/USP3/USPL1/TANK	15
BP	GO:0018205	peptidyl-lysine modification	25/486	434/20973	3.13E-05	0.007058	0.00675	PHF20L1/MYB/NSD3/PAXIP1/MACROH2A1/MDM2/SMYD2/KDM3A/KMT2C/ING4/KAT6B/SETD5/CAPN3/CREBBP/BRD8/ARID4A/UHRF2/SUPT6H/TAF2/NSD2/KMT2E/BCOR/ATF2/USPL1/MORF4L2	25
BP	GO:0034968	histone lysine methylation	13/486	145/20973	3.40E-05	0.007208	0.006893	MYB/NSD3/PAXIP1/MACROH2A1/SMYD2/KDM3A/KMT2C/SETD5/ARID4A/SUPT6H/NSD2/KMT2E/BCOR	13
BP	GO:0016571	histone methylation	14/486	178/20973	7.28E-05	0.014591	0.013953	MYB/NSD3/PAXIP1/MACROH2A1/SMYD2/KDM3A/KMT2C/SETD5/CREBBP/ARID4A/SUPT6H/NSD2/KMT2E/BCOR	14
BP	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	25/486	460/20973	8.01E-05	0.015209	0.014544	RAD23A/HECTD1/FBXO38/MDM2/ZSWIM8/UBE2D3/FBXO9/OS9/DNAJB12/ANAPC5/COP1/UBXN1/NEMF/CRBN/NSFL1C/PSMD7/TLK2/ERLEC1/ATXN3/LTN1/PPP2R5C/UBXN2B/GSK3B/TRIP4/DNAJB14	25
BP	GO:0018022	peptidyl-lysine methylation	13/486	161/20973	0.0001	0.018089	0.017299	MYB/NSD3/PAXIP1/MACROH2A1/SMYD2/KDM3A/KMT2C/SETD5/ARID4A/SUPT6H/NSD2/KMT2E/BCOR	13
BP	GO:0000381	regulation of alternative mRNA splicing, via spliceosome	8/486	64/20973	0.000111	0.019113	0.018278	RBM7/NSRP1/RBM5/FXR1/SRSF2/SRSF6/TRA2B/RBM25	8
BP	GO:0030330	DNA damage response, signal transduction by p53 class mediator	8/486	74/20973	0.00031	0.048182	0.046077	SP100/COPS3/PAXIP1/MDM2/SMYD2/ING4/CASP2/MDM4	8
BP	GO:0000910	cytokinesis	13/486	181/20973	0.00032	0.048182	0.046077	USP8/KIF23/STMN1/TTC19/CETN2/JTB/CALM1/PRC1/PIK3C3/PRPF40A/UVRAG/CSPP1/SEPTIN6	13
BP	GO:0042770	signal transduction in response to DNA damage	14/486	205/20973	0.000321	0.048182	0.046077	SP100/THOC1/MRE11/COPS3/PAXIP1/GIGYF2/MDM2/SMYD2/ING4/MUS81/CASP2/BABAM2/MDM4/ATF2	14
CC	GO:0010494	cytoplasmic stress granule	12/505	83/21993	4.16E-07	0.00021	0.000184	HNRNPK/PABPC4/STAU2/GIGYF2/NXF1/DDX19A/ATXN2/G3BP2/PABPC1L/UBAP2L/PUM2/PRRC2C	12
CC	GO:0036464	cytoplasmic ribonucleoprotein granule	21/505	281/21993	2.44E-06	0.000615	0.000538	HNRNPK/PABPC4/STAU2/SYNE1/CNOT2/GIGYF2/NXF1/NCL/DDX19A/ATXN2/G3BP2/LSM1/AGO3/FXR1/PABPC1L/PIWIL4/UBAP2L/PUM2/USP3/PRRC2C/SERPINB1	21
CC	GO:0030117	membrane coat	11/505	88/21993	5.46E-06	0.000687	0.000601	CLINT1/PICALM/AP1G2/NECAP1/SEC24C/PDCD6/SCLT1/AP2M1/AP3S2/ARCN1/SAR1A	11
CC	GO:0048475	coated membrane	11/505	88/21993	5.46E-06	0.000687	0.000601	CLINT1/PICALM/AP1G2/NECAP1/SEC24C/PDCD6/SCLT1/AP2M1/AP3S2/ARCN1/SAR1A	11
CC	GO:0035770	ribonucleoprotein granule	21/505	301/21993	7.13E-06	0.000719	0.000629	HNRNPK/PABPC4/STAU2/SYNE1/CNOT2/GIGYF2/NXF1/NCL/DDX19A/ATXN2/G3BP2/LSM1/AGO3/FXR1/PABPC1L/PIWIL4/UBAP2L/PUM2/USP3/PRRC2C/SERPINB1	21
CC	GO:0000775	chromosome, centromeric region	18/505	244/21993	1.51E-05	0.00127	0.001111	CENPK/MACROH2A1/PPP1R12A/DCTN1/DCTN2/UHRF2/CENPM/ZNF207/MIS18BP1/DNMT3A/UVRAG/CFDP1/ITGB3BP/SEPTIN6/PPP2R5C/DCTN6/CENPC/PMF1	18
CC	GO:0005681	spliceosomal complex	17/505	229/21993	2.39E-05	0.001724	0.001509	SNRPB/HNRNPK/HNRNPH1/SRRM1/HNRNPM/NCL/DHX8/PRPF4B/U2AF1L4/SREK1/TRA2A/RBM5/CWF19L1/PRPF40A/SRSF2/TRA2B/SNRPD2	17
CC	GO:0030120	vesicle coat	8/505	55/21993	3.45E-05	0.002173	0.001902	CLINT1/AP1G2/NECAP1/SEC24C/PDCD6/AP2M1/ARCN1/SAR1A	8

CC	GO:0016363	nuclear matrix	11/505	113/21993	5.92E-05	0.003173	0.002777	THOC1/ZNF350/MYB/XPOT/PAXIP1/SRRM1/HNRNPM/PRPF40A/DNMT3A/ATXN3/KIN	11
CC	GO:0034399	nuclear periphery	12/505	134/21993	6.30E-05	0.003173	0.002777	THOC1/ZNF350/MYB/XPOT/PAXIP1/SRRM1/HNRNPM/PRPF40A/DNMT3A/ATXN3/KIN/NARF	12
CC	GO:0016607	nuclear speck	26/505	492/21993	7.98E-05	0.003657	0.0032	THOC1/HECTD1/NSRP1/PNISR/SRRM1/NXF1/ARL6IP4/BAZ2A/SRSF5/PRPF4B/COP1/U2AF1L4/ZNF638/BCLAF1/SRSF3/SREK1/TCF12/PRPF40A/SRSF2/SRSF6/FNBP4/METTL3/KMT2E/RBM25/CCNL1/RUFY1	26
CC	GO:0098687	chromosomal region	22/505	405/21993	0.000189	0.007958	0.006964	SP100/THOC1/CENPK/MRE11/MACROH2A1/PPP1R12A/DCTN1/DCTN2/UHRF2/CENPM/PTGES3/ZNF207/MIS18BP1/DNMT3A/UVRAG/CFDP1/ITGB3BP/SEPTIN6/PPP2R5C/DCTN6/CENPC/PMF1	22
CC	GO:0000776	kinetochore	12/505	157/21993	0.000283	0.010974	0.009603	CENPK/PPP1R12A/DCTN1/DCTN2/CENPM/ZNF207/CFDP1/ITGB3BP/SEPTIN6/DCTN6/CENPC/PMF1	12
CC	GO:0000123	histone acetyltransferase complex	9/505	94/21993	0.000312	0.011233	0.00983	SUPT20H/PHF20L1/ING4/KAT6B/CREBBP/BRD8/TAF2/ATF2/MORF4L2	9
CC	GO:0098573	intrinsic component of mitochondrial membrane	9/505	95/21993	0.000338	0.011351	0.009933	IMMT/HSPA9/MTX1/PINK1/SLC8B1/RHOT1/FIS1/PPOX/MFF	9
CC	GO:0030118	clathrin coat	6/505	42/21993	0.00037	0.011655	0.010199	CLINT1/PICALM/AP1G2/NECAP1/SCLT1/AP2M1	6
CC	GO:0000779	condensed chromosome, centromeric region	12/505	167/21993	0.000496	0.014589	0.012767	CENPK/PPP1R12A/DCTN1/DCTN2/CENPM/ZNF207/CFDP1/ITGB3BP/SEPTIN6/DCTN6/CENPC/PMF1	12
CC	GO:0005741	mitochondrial outer membrane	14/505	217/21993	0.000521	0.014589	0.012767	IMMT/LETMD1/TAFAZZIN/HSPA9/TRAF3IP3/CPT1B/MTX1/PINK1/RHOT1/OPA1/FIS1/UBA52/ATF2/MFF	14
CC	GO:0031968	organelle outer membrane	15/505	245/21993	0.000578	0.015096	0.013211	IMMT/LETMD1/TAFAZZIN/SYNE1/HSPA9/TRAF3IP3/CPT1B/MTX1/PINK1/RHOT1/OPA1/FIS1/UBA52/ATF2/MFF	15
CC	GO:0019867	outer membrane	15/505	247/21993	0.000628	0.015096	0.013211	IMMT/LETMD1/TAFAZZIN/SYNE1/HSPA9/TRAF3IP3/CPT1B/MTX1/PINK1/RHOT1/OPA1/FIS1/UBA52/ATF2/MFF	15
CC	GO:0031248	protein acetyltransferase complex	9/505	104/21993	0.000659	0.015096	0.013211	SUPT20H/PHF20L1/ING4/KAT6B/CREBBP/BRD8/TAF2/ATF2/MORF4L2	9
CC	GO:1902493	acetyltransferase complex	9/505	104/21993	0.000659	0.015096	0.013211	SUPT20H/PHF20L1/ING4/KAT6B/CREBBP/BRD8/TAF2/ATF2/MORF4L2	9
CC	GO:0034708	methyltransferase complex	9/505	106/21993	0.000757	0.016579	0.014508	SNRPB/PAXIP1/KMT2C/RIOK1/C17orf49/MAX/METTL3/WDR77/SNRPD2	9
CC	GO:0098800	inner mitochondrial membrane protein complex	12/505	176/21993	0.000791	0.016619	0.014543	IMMT/PAM16/NDUFB10/HSPA9/MICU2/MTX1/NDUFV3/ATP5MK/IMMP1L/NDUFV1/NDUFS8/SDHD	12
CC	GO:0034719	SMN-Sm protein complex	4/505	20/21993	0.000994	0.020029	0.017528	SNRPB/GEMIN7/STRAP/SNRPD2	4
CC	GO:0098798	mitochondrial protein-containing complex	17/505	315/21993	0.001059	0.020527	0.017963	IMMT/PAM16/MRPL48/NDUFB10/HSPA9/IDH3G/MICU2/PDHB/MTX1/MRPS18B/NDUFV3/ATP5MK/IMMP1L/NDUFV1/NDUFS8/MRPS18C/SDHD	17
CC	GO:0030496	midbody	13/505	210/21993	0.00119	0.022221	0.019446	CCDC66/USP8/KIF23/TTC19/JTB/RALB/PRC1/PIK3C3/UVRAG/SEPTIN6/EXOC7/USP3/CENPC	13
CC	GO:0000793	condensed chromosome	15/505	266/21993	0.001329	0.023481	0.020548	CENPK/MACROH2A1/PPP1R12A/DCTN1/STAG3L1/DCTN2/AGO3/CENPM/ZNF207/CFDP1/ITGB3BP/SEPTIN6/DCTN6/CENPC/PMF1	15
CC	GO:0032592	integral component of mitochondrial membrane	8/505	93/21993	0.001351	0.023481	0.020548	IMMT/HSPA9/MTX1/PINK1/SLC8B1/RHOT1/FIS1/MFF	8
CC	GO:0071013	catalytic step 2 spliceosome	8/505	98/21993	0.001891	0.031762	0.027795	SNRPB/HNRNPK/HNRNPH1/SRRM1/HNRNPM/DHX8/PRPF4B/SNRPD2	8
CC	GO:0031307	integral component of mitochondrial outer membrane	4/505	24/21993	0.002026	0.032945	0.02883	PINK1/RHOT1/FIS1/MFF	4
CC	GO:0034709	methylosome	3/505	12/21993	0.002268	0.035728	0.031266	SNRPB/WDR77/SNRPD2	3
CC	GO:0031306	intrinsic component of mitochondrial outer membrane	4/505	25/21993	0.002369	0.036182	0.031663	PINK1/RHOT1/FIS1/MFF	4
CC	GO:0001401	SAM complex	3/505	13/21993	0.002899	0.040586	0.035517	IMMT/HSPA9/MTX1	3
CC	GO:0031010	ISWI-type complex	3/505	13/21993	0.002899	0.040586	0.035517	BPTF/C17orf49/BAZ1A	3
CC	GO:0140275	MIB complex	3/505	13/21993	0.002899	0.040586	0.035517	IMMT/HSPA9/MTX1	3
CC	GO:0030125	clathrin vesicle coat	4/505	28/21993	0.003632	0.049474	0.043295	CLINT1/AP1G2/NECAP1/AP2M1	4

Table 8-8. Gene Ontology analysis for aberrantly spliced genes in erythrocytic cell population in MDS with *SRSF2* mutations

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0008380	RNA splicing	27/238	497/20973	1.86E-11	5.39E-08	5.05E-08	ZC3H13/PRPF3/HNRNPC/LUC7L3/RBM39/DDX39A/PRPF39/HNRNPH1/PQBP1/NONO/SRRM1/RTRAF/STRAP/THOC2/MAGOHB/SNRPA1/ZRANB2/SNU13/TXNL4A/ZBTB8OS/SRSF11/LSM1/MAGOH/HNRNPA3/RBM4/U2AF1/TRA2B	27
BP	GO:0000375	RNA splicing, via transesterification reactions	20/238	368/20973	8.67E-09	1.26E-05	1.18E-05	PRPF3/HNRNPC/LUC7L3/RBM39/DDX39A/PRPF39/HNRNPH1/PQBP1/SRRM1/STRAP/MAGOHB/SNRPA1/SNU13/TXNL4A/LSM1/MAGOH/HNRNPA3/RBM4/U2AF1/TRA2B	20
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	19/238	364/20973	3.94E-08	2.86E-05	2.68E-05	PRPF3/HNRNPC/LUC7L3/RBM39/DDX39A/PRPF39/HNRNPH1/PQBP1/SRRM1/STRAP/MAGOHB/SNRPA1/SNU13/TXNL4A/MAGOH/HNRNPA3/RBM4/U2AF1/TRA2B	19
BP	GO:0000398	mRNA splicing, via spliceosome	19/238	364/20973	3.94E-08	2.86E-05	2.68E-05	PRPF3/HNRNPC/LUC7L3/RBM39/DDX39A/PRPF39/HNRNPH1/PQBP1/SRRM1/STRAP/MAGOHB/SNRPA1/SNU13/TXNL4A/MAGOH/HNRNPA3/RBM4/U2AF1/TRA2B	19
BP	GO:0015980	energy derivation by oxidation of organic compounds	15/238	349/20973	1.17E-05	0.006813	0.006386	PRELID1/PFKM/SOD2/NDUFS2/ARL2/ACADVL/ATP5F1D/ISCU/NDUFV3/PTGES3/SNCA/RB1CC1/HIF1A/PDHA1/NDUFB3	15
BP	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	16/238	460/20973	7.81E-05	0.037783	0.035414	PCBP2/WAC/PSMD2/XPO1/ANKZF1/UBQLN1/UBXN1/ELOB/CDC27/RNF19A/DCAF11/SDCBP/CRBN/DERL2/TRIP12/ARMC8	16
CC	GO:0005681	spliceosomal complex	15/248	229/21993	5.58E-08	2.24E-05	1.98E-05	PRPF3/HNRNPC/LUC7L3/WAC/PRPF39/HNRNPH1/SRRM1/MAGOHB/SNRPA1/SNU13/TXNL4A/MAGOH/HNRNPA3/U2AF1/TRA2B	15
CC	GO:0016607	nuclear speck	21/248	492/21993	2.09E-07	4.20E-05	3.70E-05	ZC3H13/PRPF3/APEX1/HP1BP3/LUC7L3/RBM39/WAC/DDX39A/PQBP1/NONO/SRRM1/THOC2/SNRPA1/COPS4/SRSF11/MAGOH/HIF1A/EAF2/TRIP12/RBM4/U2AF1	21
CC	GO:0071013	catalytic step 2 spliceosome	8/248	98/21993	1.53E-05	0.002055	0.001813	HNRNPC/HNRNPH1/SRRM1/MAGOHB/SNRPA1/MAGOH/HNRNPA3/U2AF1	8
CC	GO:0098687	chromosomal region	15/248	405/21993	6.13E-05	0.005028	0.004437	APEX1/RAD51AP1/EZH2/MACROH2A1/XPO1/DCLRE1B/TINF2/PTGES3/THOC2/NSMCE2/ZNF207/SKA2/SGO1/RAD51/CENPE	15
CC	GO:0000781	chromosome, telomeric region	10/248	190/21993	6.25E-05	0.005028	0.004437	APEX1/RAD51AP1/EZH2/MACROH2A1/DCLRE1B/TINF2/PTGES3/THOC2/NSMCE2/RAD51	10
CC	GO:0120114	Sm-like protein family complex	8/248	136/21993	0.000159	0.010647	0.009395	PRPF3/LUC7L3/PRPF39/STRAP/SNRPA1/SNU13/TXNL4A/LSM1	8
CC	GO:0098798	mitochondrial protein-containing complex	12/248	315/21993	0.000258	0.014057	0.012405	MRPS11/MRPS18C/NDUFS2/HSPA9/ATP5F1D/NDUFV3/IMMP1L/MRPL22/MICOS10/ATP5MK/PDHA1/NDUFB3	12
CC	GO:0000151	ubiquitin ligase complex	12/248	319/21993	0.00029	0.014057	0.012405	ELOB/CDC27/RNF19A/TMEM183A/RNF7/DCAF11/CRBN/DERL2/MKLN1/USP33/RAD51/ARMC8	12
CC	GO:0005684	U2-type spliceosomal complex	7/248	114/21993	0.000315	0.014057	0.012405	PRPF3/LUC7L3/PRPF39/MAGOHB/SNRPA1/SNU13/TXNL4A	7
CC	GO:0000153	cytoplasmic ubiquitin ligase complex	3/248	15/21993	0.000583	0.023438	0.020683	ELOB/DERL2/USP33	3
CC	GO:0008180	COP9 signalosome	3/248	16/21993	0.000712	0.02384	0.021037	COPS2/COPS5/COPS4	3
CC	GO:1990023	mitotic spindle midzone	3/248	16/21993	0.000712	0.02384	0.021037	KIF20B/PRC1/CENPE	3
CC	GO:0098800	inner mitochondrial membrane protein complex	8/248	176/21993	0.000893	0.027597	0.024353	NDUFS2/HSPA9/ATP5F1D/NDUFV3/IMMP1L/MICOS10/ATP5MK/NDUFB3	8
CC	GO:0072686	mitotic spindle	8/248	178/21993	0.000961	0.027597	0.024353	KIF20B/HAUS1/TAF1D/CDC27/RTRAF/PRC1/CENPE/TFDP2	8
CC	GO:0071005	U2-type precatalytic spliceosome	5/248	72/21993	0.001325	0.035508	0.031333	PRPF3/MAGOHB/SNRPA1/SNU13/TXNL4A	5
CC	GO:0071011	precatalytic spliceosome	5/248	74/21993	0.001498	0.037635	0.03321	PRPF3/MAGOHB/SNRPA1/SNU13/TXNL4A	5
CC	GO:0097525	spliceosomal snRNP complex	6/248	111/21993	0.001635	0.038653	0.034109	PRPF3/LUC7L3/PRPF39/SNRPA1/SNU13/TXNL4A	6

Table 8-9. Aberrantly spliced genes in the category of "RNA splicing (GO:0008380)" of GO term in MDS patients harbouring *SF3B1* mutations. Results obtained by rMATS bioinformatic tool (FDR < 0.05 and |ΔPSI| > 0.1). *Assigning multiple Event IDs in a gene means different splice variants in the genes.

Gene Symbol	Chromosome	strand	P Value	FDR	|ΔPSI|	Event ID*	Event Type	BM subpopulations			
								CD34	MON	GRA	ERY
ACIN1	14	-	4.31E-12	4.31E-10	0.158	19157	SE	1			
ACIN1	14	-	0.000596	0.008791	0.106	19159	SE	1			
ACIN1	14	-	6.91E-05	0.002068	0.188	19157	SE		1		
C14orf166	14	+	0.000208	0.00351	0.125	31813	SE	1			
CCDC130	19	+	0.00227	0.009774	0.144	2584	RI	1			
CCNL1	3	-	0.000177	0.005804	0.157	54	A5SS		1		
CELF1	11	-	0.003554	0.040055	0.18	37235	SE	1			
CELF2	10	+	0.000151	0.005168	0.173	692	A5SS		1		
CIRBP	19	+	0.004317	0.049335	0.336	4686	A3SS	1			
CIRBP	19	+	2.21E-11	3.21E-10	0.175	3413	RI	1			
CIRBP	19	+	1.67E-15	2.85E-13	0.15	23189	SE	1			
CIRBP	19	+	5.67E-07	1.85E-05	0.137	23188	SE	1			
CLK1	2	-	0	0	0.235	1506	SE		1		
CLK1	2	-	3.10E-11	7.79E-09	0.25	1506	SE			1	
CLK3	15	+	2.27E-07	8.12E-06	0.107	21759	SE	1			
CLK3	15	+	5.13E-11	1.04E-08	0.243	21755	SE		1		
CLK3	15	+	2.49E-10	3.88E-08	0.284	21761	SE		1		
CLK4	5	-	2.11E-05	0.000595	0.218	4242	A5SS	1			
CLK4	5	-	2.65E-05	0.000163	0.215	4793	RI	1			
CLK4	5	-	9.05E-08	3.61E-06	0.2	32953	SE	1			
CLK4	5	-	6.35E-05	0.001247	0.217	32952	SE	1			
CLK4	5	-	0.000366	0.007845	0.174	32953	SE		1		
CLK4	5	-	2.01E-05	0.000875	0.286	32952	SE			1	
CLK4	5	-	0.000229	0.006611	0.168	32953	SE			1	
CLK4	5	-	7.99E-05	0.002962	0.524	4242	A5SS				1
CLK4	5	-	0.000357	0.009749	0.472	32952	SE				1
CWF19L1	10	-	0.001092	0.023567	0.117	25124	SE			1	
CWF19L2	11	-	0.000666	0.009694	0.127	34167	SE	1			
DDX23	12	-	0.001834	0.018935	0.119	4045	RI			1	
DDX39A	19	-	0	0	0.114	4585	RI	1			
DDX39B	6	-	2.57E-06	1.86E-05	0.128	3965	RI	1			
DDX39B	6	-	1.67E-06	9.36E-05	0.201	5510	A3SS			1	
DDX41	5	-	0.000525	0.008301	0.101	3446	A3SS	1			
DDX41	5	-	1.11E-16	2.75E-15	0.112	2526	RI	1			
DDX46	5	+	0.005087	0.035695	0.113	5271	RI		1		
DDX47	12	+	1.31E-09	1.49E-08	0.203	3654	RI	1			
DHX15	4	-	0.00023	0.006628	0.165	33364	SE				1
DHX9	1	+	0.000262	0.006054	0.128	25021	SE		1		
EFTUD2	17	-	2.76E-06	5.32E-05	0.152	1209	RI		1		
FAM50A	X	+	6.81E-07	1.56E-05	0.161	4361	RI		1		

FASTK	7	-	1.14E-05	7.46E-05	0.123	1066	RI	1			
FASTK	7	-	0.000125	0.000683	0.12	1064	RI	1			
FASTK	7	-	0.000919	0.004251	0.211	1063	RI	1			
FXR1	3	+	0.001419	0.020605	0.24	5312	RI				1
GEMIN2	14	+	2.42E-05	0.000529	0.108	4146	SE	1			
GEMIN8	X	-	0.000197	0.003357	0.13	34560	SE	1			
GRSF1	4	-	0.001741	0.039577	0.103	2582	SE				1
HNRNPA3	2	+	5.22E-08	2.24E-06	0.228	33196	SE	1			
HNRNPA3	2	+	9.09E-07	4.95E-05	0.189	33196	SE		1		
HNRNPC	14	-	3.43E-07	2.84E-05	0.246	20948	SE			1	
HNRNPH1	5	-	0	0	0.112	4512	RI	1			
HNRNPH1	5	-	2.69E-06	7.72E-05	0.112	30997	SE	1			
HNRNPH3	10	+	1.47E-09	1.22E-07	0.135	637	A5SS	1			
HNRNPH3	10	+	0.002203	0.031796	0.121	638	A5SS	1			
HNRNPH3	10	+	7.55E-08	3.10E-06	0.168	4610	SE	1			
HNRNPH3	10	+	9.87E-07	3.09E-05	0.184	4609	SE	1			
HNRNPM	19	+	0.007204	0.047669	0.276	2091	RI		1		
IVNS1ABP	1	-	0.000242	0.00401	0.194	12092	SE	1			
JMJD6	17	-	6.45E-05	0.002396	0.17	30268	SE			1	
JMJD6	17	-	0.000682	0.01606	0.213	30270	SE			1	
KIAA1967	8	+	6.86E-10	8.15E-09	0.155	873	RI	1			
LSM1	8	-	0.000712	0.016691	0.127	30216	SE			1	
LUC7L	16	-	0	0	0.245	1434	RI	1			
LUC7L	16	-	1.14E-05	0.000275	0.19	9036	SE	1			
LUC7L	16	-	0.000378	0.003866	0.366	1433	RI		1		
MBNL2	13	+	0.001055	0.014401	0.326	11589	SE	1			
METTL3	14	-	1.44E-09	1.63E-08	0.137	4179	RI	1			
NHP2L1	22	-	0	0	0.128	17539	SE	1			
NHP2L1	22	-	2.43E-10	1.71E-08	0.106	17542	SE	1			
NHP2L1	22	-	0.000625	0.01596	0.45	17542	SE				1
NONO	X	+	2.73E-05	0.001001	0.101	27465	SE				1
NSRP1	17	+	0.001832	0.047196	0.125	1655	MXE		1		
PABPC1	8	-	0.002542	0.01985	0.153	357	RI		1		
PAXBP1	21	-	0.002045	0.029657	0.15	1325	A5SS	1			
PPIE	1	+	3.86E-13	5.37E-11	0.251	2151	A5SS	1			
PPIE	1	+	0.000664	0.009679	0.169	16466	SE	1			
PPIE	1	+	0.002357	0.042692	0.189	2151	A5SS		1		
PPIE	1	+	0.000416	0.004194	0.113	2407	RI		1		
PPIE	1	+	0	0	0.42	2151	A5SS			1	
PPIL3	2	-	3.47E-05	0.000735	0.23	12649	SE	1			
PPIL3	2	-	4.34E-05	0.001436	0.129	12654	SE		1		
PPIL3	2	-	0.001402	0.04777	0.367	742	MXE			1	
PRPF3	1	+	1.20E-11	1.09E-09	0.136	7207	SE	1			
PRPF39	14	+	1.27E-05	8.22E-05	0.102	773	RI	1			
PRPF39	14	+	0.001211	0.025705	0.313	4904	SE			1	
PRPF39	14	+	2.21E-06	5.95E-05	0.224	774	RI				1
PRPF39	14	+	1.32E-05	0.000533	0.244	4903	SE				1

PRPF40A	2	-	5.55E-16	1.03E-13	0.127	31113	SE	1			
PRPF4B	6	+	3.61E-05	0.000709	0.241	878	RI			1	
QKI	6	+	6.96E-05	0.001351	0.15	3839	A3SS	1			
QKI	6	+	9.03E-06	5.99E-05	0.104	2777	RI	1			
RBM10	X	+	0.000555	0.008263	0.144	2409	SE	1			
RBM15	1	+	1.12E-05	0.000525	0.346	16697	SE			1	
RBM23	14	-	1.56E-11	8.60E-10	0.13	6455	A3SS	1			
RBM23	14	-	0.00027	0.008974	0.133	1956	MXE	1			
RBM23	14	-	0.000144	0.002553	0.167	31757	SE	1			
RBM23	14	-	0.000162	0.002836	0.168	31755	SE	1			
RBM25	14	+	0	0	0.107	2511	RI	1			
RBM25	14	+	9.27E-05	0.003255	0.131	17127	SE			1	
RBM39	20	-	0.000332	0.015365	0.207	1126	MXE			1	
RBM39	20	-	7.80E-12	1.74E-09	0.295	1126	MXE				1
RBM39	20	-	3.34E-09	3.62E-07	0.404	19229	SE				1
RBM41	X	-	0.002272	0.027737	0.109	6370	SE	1			
RBM41	X	-	0.001632	0.026873	0.444	6371	SE		1		
RBM4B	11	-	8.83E-08	3.55E-06	0.271	36144	SE	1			
RBM5	3	+	4.77E-11	6.65E-10	0.134	4137	RI	1			
RBM5	3	+	1.16E-05	0.000463	0.185	28360	SE		1		
RBM6	3	+	0	0	0.152	3849	RI	1			
RBM6	3	+	6.34E-05	0.001245	0.145	26094	SE	1			
RBM6	3	+	0.000859	0.019215	0.184	26094	SE			1	
SART3	12	-	0.000426	0.011056	0.132	19249	SE			1	
SETX	9	-	8.42E-06	0.000209	0.178	15598	SE	1			
SETX	9	-	0.000109	0.005815	0.201	899	MXE			1	
SMN1	5	+	8.74E-06	0.000217	0.142	23848	SE	1			
SNRNP25	16	+	0.000744	0.009284	0.121	3727	RI			1	
SNRPN	15	+	0	0	0.242	2359	A3SS	1			
SNRPN	15	+	0	0	0.31	2362	A3SS	1			
SNRPN	15	+	9.75E-11	5.15E-09	0.307	2361	A3SS	1			
SNRPN	15	+	0	0	0.245	1520	A5SS	1			
SNRPN	15	+	0	0	0.21	11475	SE	1			
SNRPN	15	+	0	0	0.209	11476	SE	1			
SNRPN	15	+	6.67E-09	4.81E-07	0.29	2359	A3SS		1		
SNRPN	15	+	2.25E-07	1.03E-05	0.339	2362	A3SS		1		
SNRPN	15	+	1.74E-13	5.62E-11	0.29	11475	SE		1		
SNRPN	15	+	2.84E-05	0.001013	0.236	11476	SE		1		
SON	21	+	2.72E-06	0.000158	0.176	17849	SE			1	
SREK1	5	+	3.42E-09	1.50E-07	0.158	5065	A3SS	1			
SREK1	5	+	8.87E-11	1.21E-09	0.125	3688	RI	1			
SREK1	5	+	8.72E-10	5.40E-08	0.2	25002	SE	1			
SREK1	5	+	0.000384	0.006015	0.194	25003	SE	1			
SREK1	5	+	0.000661	0.008646	0.12	3688	RI			1	
SREK1	5	+	1.76E-10	3.32E-08	0.222	5065	A3SS				1
SRRM1	1	+	1.65E-06	4.92E-05	0.151	2548	SE	1			
SRRM1	1	+	7.39E-08	5.54E-06	0.193	2548	SE		1		

SRRM1	1	+	0.000123	0.00201	0.182	418	RI			1	
SRRM1	1	+	0.001311	0.019341	0.224	418	RI				1
SRRM1	1	+	1.33E-14	5.16E-12	0.127	2551	SE				1
SRSF11	1	+	2.81E-07	9.90E-06	0.115	23921	SE	1			
SRSF11	1	+	1.44E-06	9.44E-05	0.109	23917	SE			1	
SRSF11	1	+	3.95E-05	0.001578	0.277	23921	SE			1	
SRSF3	6	+	0.00148	0.030466	0.165	16619	SE			1	
SRSF5	14	+	0.000571	0.012763	0.176	6350	A3SS			1	
SRSF5	14	+	0.00063	0.01561	0.169	4030	A5SS			1	
SRSF5	14	+	0.00065	0.008549	0.178	4542	RI			1	
SRSF5	14	+	0.00552	0.044158	0.204	4543	RI			1	
SRSF7	2	-	7.44E-05	0.001324	0.153	2738	RI			1	
SRSF7	2	-	0.000126	0.004134	0.141	18628	SE			1	
SRSF7	2	-	0.000376	0.010023	0.162	18626	SE			1	
SUGP1	19	-	0	0	0.119	6131	A3SS	1			
SUGP2	19	-	5.55E-16	1.27E-14	0.167	2931	RI	1			
SUGP2	19	-	2.65E-14	3.64E-12	0.205	19997	SE	1			
SUGP2	19	-	3.31E-13	4.07E-11	0.237	19995	SE	1			
SUGP2	19	-	0.000254	0.002781	0.234	2931	RI		1		
SUPT6H	17	+	0.002977	0.04327	0.524	8649	SE		1		
THOC1	18	-	0	0	0.294	4127	A3SS	1			
THOC1	18	-	0.000318	0.001604	0.202	2983	RI	1			
THOC1	18	-	1.53E-07	7.11E-06	0.407	4127	A3SS		1		
TIA1	2	-	4.96E-05	0.002016	0.118	3148	A5SS		1		
TMBIM6	12	+	6.53E-11	5.07E-09	0.137	25213	SE	1			
TMBIM6	12	+	0.000975	0.021387	0.155	25214	SE			1	
TMBIM6	12	+	1.69E-05	0.000661	0.177	25214	SE				1
TSEN15	1	+	0.002287	0.027906	0.149	2702	SE	1			
TSEN2	3	+	0.002322	0.047665	0.108	676	MXE	1			
TSEN2	3	+	0.00774	0.02989	0.164	1734	RI	1			
TSEN2	3	+	4.73E-05	0.000961	0.121	11280	SE	1			
TXNL4A	18	-	0.004068	0.044715	0.108	23010	SE	1			
U2AF1	21	-	0.001415	0.011824	0.122	1529	RI		1		
U2AF1	21	-	0.001731	0.023626	0.256	1528	RI				1
U2AF1L4	19	-	5.22E-15	1.08E-13	0.106	665	RI	1			
U2AF1L4	19	-	0.000124	0.000679	0.164	668	RI	1			
U2AF1L4	19	-	0.000108	0.003667	0.235	4005	SE			1	
USP39	2	+	7.50E-05	0.002481	0.322	31131	SE				1
WBP11	12	-	0.003491	0.049309	0.118	4757	SE		1		
YTHDC1	4	-	0.000624	0.019307	0.282	3326	A3SS				1
ZBTB8OS	1	-	1.18E-11	6.65E-10	0.169	4113	A3SS	1			
ZRANB2	1	-	0.000274	0.007719	0.241	37046	SE			1	

Table 8-10. Aberrantly spliced genes in the category of "RNA splicing (GO:0008380)" of GO term in MDS patients harbouring *SRSF2* mutations. Results obtained by rMATS bioinformatic tool (FDR < 0.05 and |ΔPSI| > 0.1). *Assigning multiple Event IDs in a gene means different splice variants in the genes.

Gene Symbol	Chromosome	strand	P Value	FDR	|ΔPSI|	Event ID*	Event Type	BM subpopulations			
								CD34	MON	GRA	ERY
ARL6IP4	12	+	0.000155627	0.001815342	0.19	5306	RI			1	
C14orf166	14	+	7.73E-14	1.32E-11	0.161	31814	SE				1
CCNL1	3	-	0.000601252	0.010449002	0.111	332	SE		1		
CCNL1	3	-	0.001830385	0.023393795	0.168	329	SE			1	
CCNL1	3	-	0.002008634	0.024941551	0.17	331	SE			1	
CELF1	11	-	3.22E-05	0.000719355	0.352	37235	SE	1			
CLK1	2	-	1.32E-06	2.68E-05	0.302	248	RI			1	
CLK1	2	-	0.004206828	0.028957244	0.172	249	RI			1	
CLK1	2	-	4.84E-08	2.62E-06	0.382	1506	SE			1	
CWF19L1	10	-	1.79E-07	8.04E-06	0.134	25124	SE			1	
CWF19L1	10	-	0.00160257	0.020993437	0.15	25122	SE			1	
DDX39A	19	-	6.33E-09	2.63E-07	0.106	4586	RI				1
DDX39B	6	-	0.000648835	0.009153335	0.113	26936	SE	1			
DDX41	5	-	5.07E-05	0.001587419	0.121	3445	A3SS	1			
DDX41	5	-	7.36E-05	0.002059344	0.124	3446	A3SS	1			
DDX41	5	-	5.55E-16	4.59E-14	0.12	2526	RI	1			
DDX47	12	+	6.66E-16	1.34E-13	0.119	24795	SE	1			
DHX8	17	+	0.000545271	0.005106668	0.132	2950	RI			1	
FASTK	7	-	0.000593866	0.00441028	0.236	1063	RI	1			
FXR1	3	+	7.00E-05	0.001519794	0.116	36095	SE			1	
GEMIN7	19	+	1.05E-05	0.000293832	0.117	13136	SE			1	
GEMIN8	X	-	1.00E-05	0.000264826	0.204	34560	SE	1			
HNRNPA2B1	7	-	7.62E-05	0.001974557	0.109	6457	SE		1		
HNRNPA3	2	+	0.000975526	0.015938518	0.152	33196	SE				1
HNRNPC	14	-	0.001603002	0.029220689	0.117	4221	A3SS	1			
HNRNPC	14	-	1.48E-08	9.71E-07	0.325	2757	A5SS				1
HNRNPC	14	-	7.77E-16	2.19E-13	0.106	20955	SE				1
HNRNPH1	5	-	1.74E-07	1.40E-05	0.119	4014	A5SS		1		
HNRNPH1	5	-	0.001291798	0.011616645	0.34	4512	RI		1		
HNRNPH1	5	-	0.003780722	0.04538673	0.139	30994	SE		1		
HNRNPH1	5	-	1.07E-08	1.07E-06	0.116	4013	A5SS			1	
HNRNPH1	5	-	5.33E-07	1.22E-05	0.141	4518	RI			1	
HNRNPH1	5	-	0.000211231	0.002337947	0.485	4512	RI			1	
HNRNPH1	5	-	0.001870303	0.023733961	0.149	30990	SE			1	
HNRNPH1	5	-	0.001675638	0.016843999	0.333	4513	RI				1
HNRNPH1	5	-	1.81E-12	2.40E-10	0.193	30988	SE				1
HNRNPH1	5	-	0.002207343	0.032093273	0.379	30994	SE				1
HNRNPH3	10	+	0.00151003	0.026120711	0.105	637	A5SS	1			
HNRNPH3	10	+	0.002322337	0.035931713	0.13	638	A5SS	1			
HNRNPH3	10	+	0.000859126	0.011670615	0.135	4608	SE	1			

HNRNPH3	10	+	7.25E-06	0.000360992	0.126	637	A5SS		1		
HNRNPH3	10	+	0.001147763	0.017601565	0.175	4610	SE		1		
HNRNPH3	10	+	0.001312234	0.019809786	0.224	4609	SE		1		
HNRNPK	9	-	0.001037868	0.018206937	0.121	5560	A3SS			1	
HNRNPM	19	+	6.42E-05	0.000861989	0.462	2091	RI			1	
ISY1	3	-	2.38E-05	0.000769202	0.11	17501	SE		1		
IVNS1ABP	1	-	8.38E-06	0.000323285	0.348	12092	SE		1		
KIAA1429	8	-	1.08E-05	0.000404443	0.272	17777	SE		1		
LARP7	4	+	0.000365677	0.006716272	0.107	3889	A5SS			1	
LSM1	8	-	7.20E-05	0.00189007	0.208	30216	SE		1		
LSM1	8	-	1.30E-06	4.83E-05	0.341	30216	SE			1	
LSM1	8	-	0.00012686	0.002791786	0.511	30216	SE				1
LUC7L	16	-	1.44E-15	1.10E-13	0.239	1434	RI	1			
LUC7L	16	-	2.36E-06	7.48E-05	0.124	9035	SE	1			
LUC7L	16	-	0.001797606	0.02122566	0.182	9036	SE	1			
LUC7L3	17	+	2.31E-06	0.000165649	0.22	799	MXE				1
LUC7L3	17	+	9.75E-08	5.53E-06	0.145	13752	SE				1
MAGOH	1	-	0.001898621	0.026270674	0.106	14745	SE		1		
MAGOH	1	-	0.000364525	0.006936691	0.162	14745	SE				1
MAGOHB	12	-	2.02E-09	1.55E-07	0.122	4812	SE				1
MAGOHB	12	-	1.79E-05	0.000520714	0.43	4811	SE				1
MBNL2	13	+	0.001408897	0.017570562	0.362	11589	SE	1			
METTL3	14	-	5.06E-06	7.00E-05	0.129	4179	RI	1			
METTL3	14	-	0.000470198	0.007424141	0.105	28684	SE			1	
METTL4	18	-	0.009700473	0.047391997	0.151	365	RI	1			
NCL	2	-	0.000472298	0.004623295	0.103	2106	RI			1	
NHP2L1	22	-	2.05E-13	2.66E-11	0.131	17539	SE	1			
NHP2L1	22	-	3.24E-06	0.000118381	0.442	17539	SE				1
NHP2L1	22	-	0.000538489	0.009651027	0.199	17541	SE				1
NONO	X	+	0.001852565	0.018033193	0.148	4033	RI				1
NSRP1	17	+	0.002217558	0.025207474	0.51	27232	SE	1			
NSRP1	17	+	0.003867768	0.046194361	0.112	27234	SE		1		
NSRP1	17	+	1.05E-05	0.00046605	0.187	1655	MXE			1	
PPIE	1	+	3.31E-05	0.001182985	0.175	2151	A5SS	1			
PPIL3	2	-	0	0	0.148	743	MXE	1			
PPIL3	2	-	0.000211177	0.008478771	0.261	741	MXE	1			
PPIL3	2	-	6.43E-09	5.76E-07	0.185	743	MXE		1		
PPIL3	2	-	0.000148073	0.006301994	0.507	741	MXE		1		
PQBP1	X	+	0.001806403	0.017809285	0.46	576	RI				1
PRMT7	16	+	0.001096331	0.014216878	0.113	7815	SE	1			
PRPF3	1	+	0.001234142	0.022770579	0.379	1434	A3SS				1
PRPF39	14	+	0.000432755	0.005159509	0.238	774	RI				1
PRPF39	14	+	0.000895611	0.014851372	0.262	4903	SE				1
PRPF40A	2	-	3.10E-05	0.000736384	0.347	31113	SE			1	
PRPF4B	6	+	0.000453589	0.004723917	0.248	878	RI		1		
PRPF4B	6	+	1.87E-06	8.91E-05	0.301	5347	SE		1		
PRPF4B	6	+	4.76E-06	0.00019748	0.31	5349	SE		1		

PRPF4B	6	+	0.001410525	0.011307418	0.248	878	RI			1	
PUF60	8	-	3.93E-09	2.68E-07	0.148	37214	SE	1			
QKI	6	+	0.000954018	0.008976613	0.244	2777	RI		1		
QKI	6	+	4.98E-05	0.001411341	0.147	19004	SE		1		
RBM10	X	+	0.000604402	0.008613012	0.156	2409	SE	1			
RBM23	14	-	5.22E-07	3.23E-05	0.122	6455	A3SS	1			
RBM23	14	-	0.001051514	0.030155912	0.13	1956	MXE	1			
RBM23	14	-	0.000479884	0.007053265	0.171	31757	SE	1			
RBM23	14	-	0.000511904	0.007474993	0.172	31755	SE	1			
RBM23	14	-	3.20E-06	0.000141167	0.139	31758	SE		1		
RBM23	14	-	9.22E-05	0.002330487	0.182	31762	SE		1		
RBM25	14	+	0.001472697	0.019608854	0.213	17127	SE			1	
RBM39	20	-	0.000546028	0.021780466	0.19	1126	MXE				1
RBM39	20	-	4.72E-07	2.18E-05	0.203	19229	SE				1
RBM4	11	+	0.001893493	0.028405739	0.219	11664	SE				1
RBM5	3	+	0.001664698	0.019980334	0.114	28359	SE	1			
RBM5	3	+	0.000879058	0.008401039	0.109	4136	RI		1		
RBM5	3	+	9.24E-09	6.02E-07	0.214	28360	SE			1	
RBM6	3	+	0.000176286	0.003317817	0.245	26094	SE			1	
RBM7	11	+	2.03E-05	0.00067807	0.235	2075	A3SS			1	
RNPS1	16	-	0	0	0.15	16635	SE	1			
SETX	9	-	5.42E-09	3.55E-07	0.244	15598	SE	1			
SETX	9	-	2.42E-09	2.87E-07	0.111	899	MXE		1		
SNRPA1	15	-	0.001137118	0.017484508	0.189	5014	SE		1		
SNRPA1	15	-	2.71E-08	1.72E-06	0.105	5012	SE				1
SNRPB	20	-	2.94E-05	0.000871474	0.244	472	A3SS			1	
SNRPD2	19	-	2.74E-08	2.25E-06	0.126	34278	SE		1		
SNRPD2	19	-	0.003418283	0.038200622	0.105	34278	SE			1	
SREK1	5	+	3.19E-08	2.80E-06	0.16	5065	A3SS	1			
SREK1	5	+	7.23E-06	9.57E-05	0.107	3688	RI	1			
SREK1	5	+	1.72E-06	5.74E-05	0.149	25002	SE	1			
SREK1	5	+	7.26E-12	8.87E-10	0.457	25003	SE			1	
SREK1	5	+	0.000591501	0.009035584	0.39	25002	SE			1	
SRPK2	7	-	4.10E-05	0.000873564	0.135	442	SE	1			
SRRM1	1	+	0.002937695	0.031680966	0.102	2548	SE	1			
SRRM1	1	+	8.03E-05	0.002349269	0.12	488	A3SS		1		
SRRM1	1	+	0.003625307	0.043898366	0.229	2548	SE		1		
SRRM1	1	+	1.98E-07	4.77E-06	0.205	418	RI			1	
SRRM1	1	+	0.000453562	0.007186907	0.176	2548	SE			1	
SRRM1	1	+	0	0	0.102	2551	SE				1
SRSF11	1	+	8.39E-05	0.001953777	0.169	23920	SE				1
SRSF2	17	-	1.10E-08	6.57E-07	0.133	21644	SE	1			
SRSF2	17	-	0.000141019	0.00273484	0.494	21644	SE			1	
SRSF3	6	+	9.44E-15	1.83E-12	0.284	16619	SE			1	
SRSF5	14	+	0.000763432	0.006790708	0.108	4543	RI			1	
SRSF6	20	+	0.000264431	0.004580562	0.537	11883	SE			1	
STRAP	12	+	7.97E-05	0.001673425	0.204	24374	SE			1	

STRAP	12	+	3.79E-12	4.90E-10	0.409	24374	SE				1
SUGP2	19	-	2.26E-09	6.36E-08	0.18	2931	RI	1			
SUGP2	19	-	1.74E-06	5.80E-05	0.172	19997	SE	1			
SUGP2	19	-	2.26E-06	7.18E-05	0.205	19995	SE	1			
SUGP2	19	-	0.00018435	0.002180882	0.23	2931	RI		1		
SUPT6H	17	+	0.000179757	0.003368927	0.393	8649	SE			1	
THOC1	18	-	5.14E-09	4.72E-07	0.565	4127	A3SS			1	
THOC2	X	-	0	0	0.122	23325	SE	1			
THOC2	X	-	1.69E-10	1.56E-08	0.106	23325	SE				1
TIA1	2	-	0	0	0.222	23987	SE	1			
TIA1	2	-	0.000283362	0.006866526	0.158	4848	A3SS		1		
TRA2A	7	-	8.94E-09	5.50E-07	0.186	22810	SE	1			
TRA2A	7	-	1.17E-09	9.15E-08	0.273	22810	SE			1	
TRA2B	3	-	0.000390486	0.00636708	0.111	24463	SE			1	
TRA2B	3	-	0.003446272	0.045754048	0.177	24463	SE				1
TSEN2	3	+	0.000264032	0.009637179	0.188	676	MXE	1			
TSEN2	3	+	1.33E-13	1.79E-11	0.357	11281	SE	1			
TSEN2	3	+	1.33E-07	5.94E-06	0.236	11280	SE	1			
TXNL4A	18	-	3.49E-05	0.000925319	0.669	23013	SE				1
U2AF1	21	-	0.001901983	0.016263626	0.112	1529	RI		1		
U2AF1	21	-	0.002626533	0.036703381	0.234	9821	SE				1
U2AF1L4	19	-	3.70E-09	1.02E-07	0.112	665	RI	1			
U2AF1L4	19	-	1.34E-07	5.94E-06	0.158	4005	SE	1			
U2AF1L4	19	-	0.001602881	0.019387264	0.127	4006	SE	1			
U2AF1L4	19	-	0.005906224	0.038869339	0.211	666	RI			1	
WBP11	12	-	0.002859687	0.036124162	0.142	4757	SE		1		
ZBTB8OS	1	-	0.003709018	0.040849565	0.118	20493	SE			1	
ZBTB8OS	1	-	7.73E-05	0.001824866	0.395	20494	SE				1
ZC3H13	13	-	4.21E-13	7.24E-11	0.747	1240	A3SS				1
ZCCHC8	12	-	0.004084944	0.047660422	0.176	34448	SE		1		
ZNF638	2	+	3.33E-16	7.62E-14	0.321	13043	SE			1	
ZRANB2	1	-	0.002314388	0.027825677	0.274	37046	SE			1	
ZRANB2	1	-	1.35E-06	5.52E-05	0.348	37046	SE				1